Springer Handbook of Ocean Engineering

Dhanak
Xiros
Editors

Springer

Springer Handbook of Ocean Engineering

Springer Handbooks provide a concise compilation of approved key information on methods of research, general principles, and functional relationships in physical and applied sciences. The world's leading experts in the fields of physics and engineering will be assigned by one or several renowned editors to write the chapters comprising each volume. The content is selected by these experts from Springer sources (books, journals, online content) and other systematic and approved recent publications of scientific and technical information.

The volumes are designed to be useful as readable desk reference book to give a fast and comprehensive overview and easy retrieval of essential reliable key information, including tables, graphs, and bibliographies. References to extensive sources are provided.

Springer Handbook

# of Ocean Engineering

Manhar R. Dhanak, Nikolaos I. Xiros (Eds.)

With 1001 Figures and 150 Tables

Springer

Editors
Manhar R. Dhanak
Florida Atlantic University
The Inst. for Ocean and Systems Engineering – SeaTech
Dania Beach, Florida, USA
dhanak@fau.edu

Nikolaos I. Xiros
University of New Orleans
School of Naval Architecture & Marine Engineering
New Orleans, Louisiana, USA
nxiros@uno.edu

ISBN: 978-3-319-16648-3 e-ISBN: 978-3-319-16649-0
DOI 10.1007/978-3-319-16649-0
Library of Congress Control Number: 2016937503

© Springer Dordrecht Heidelberg London New York 2016

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herin or for any errors or omissions that may have been made.

Production and typesetting: le-tex publishing services GmbH, Leipzig
Typography and layout: schreiberVIS, Seeheim
Illustrations: Hippmann GbR, Schwarzenbruck
Cover design: eStudio Calamar Steinen, Barcelona
Cover production: WMXDesign GmbH, Heidelberg
Printing and binding: Printer Trento s.r.l., Trento

Printed on acid free paper

This Springer imprint is published by Springer Nature
The registered company is Springer-Verlag GmbH Berlin Heidelberg

Preface

It is our honor and pleasure to be part of this Springer Handbook that is intended to serve as a ready reference for ocean engineers, including practitioners in the maritime industry and government, as well as for researchers, educators, and students, providing a review of important fundamental and applied materials as well as updates of the contemporary topics in ocean engineering and ocean technologies. The handbook should be of interest to those involved in the many aspects of ocean engineering, including the design, development and operations of marine vehicles, coastal systems, and offshore technologies, as well as in the extraction of renewable ocean energy. At the same time, it will serve as a guide for anyone interested in the ocean and human activities in the ocean and the coastal environment. The handbook is divided into five parts comprising 49 chapters that cover ocean engineering basics and four important areas of application: *autonomous ocean vehicles, subsystems and control*; *coastal design*; *offshore technologies*; and *renewable ocean energy*. The coverage includes fundamental concepts, basic theories, methodologies, tools, and techniques involved in various aspects of the subject. The authors of the chapters of the handbook are specialists in their respective fields worldwide, and include well-accomplished individuals from academia, industry, and government. Each chapter has been peer reviewed. The participation of these select authors and peer reviewers has helped ensure that the handbook is an excellent and timely reference. The Springer editorial team has beautifully produced each chapter, including numerous custom drawings and figures. Keywords on each page are properly indexed for the convenience of the reader browsing through the handbook, enabling relatively easy location of material of interest in the handbook.

Our sincere gratitude are due, first of all, to the part editors for their efforts in scoping out the five parts, identifying and enlisting subject-matter experts as chapter authors, guiding the arrangement of the chapters in the parts, following up with authors, and finally seeking peer reviews of the chapters, thereby determining the scope and quality of the handbook. Second, our sincere thanks to all the authors for taking the time from their busy schedules to participate in this project, and devoting much time to carefully preparing their chapters. Further, we very much appreciate the selfless efforts of the peer reviewers for providing critical reviews of the chapters. Finally, we are significantly indebted to Springer's entire publishing team, including Werner Skolaut, Leontina Di Cecco, Veronika Hamm, Judith Hinterberg, and Constanze Ober for their knowledgeable advice, guidance, dedication, significant patience, and effective and efficient editing, which have been instrumental in ensuring timely and quality production of this handbook in support of the ocean engineering community. Part C on coastal design is dedicated to the late Professor Robert Dean for his many important contributions to coastal engineering.

Manhar R. Dhanak and Nikolaos I. Xiros

About the Editors

Manhar R. Dhanak, PhD, is Professor of Ocean Engineering and Director of the Institute for Ocean and Systems Engineering (SeaTech) at Florida Atlantic University (FAU). He is the past chair (2003–2009) of the Department of Ocean Engineering at FAU and a graduate of Imperial College, University of London. He served as a Research Associate at Imperial College, as a Research Scientist at Topexpress Ltd., Cambridge, UK, and as a Senior Research Associate at University of Cambridge before joining Florida Atlantic University. Dr. Dhanak has research interests in hydrodynamics, physical oceanography, autonomous underwater vehicles (AUV), and ocean energy. His sponsored research activities have included development of advanced energy-efficient autonomous surface vehicles and tools for assessment of advanced-hull ships, characterization of electromagnetic fields associated with oceanographic features in a coastal environment, and assessment of electromagnetic field emissions associated with live submarine cables.

Nikolaos I. Xiros is an Associate Professor of Naval Architecture and Marine Engineering at the University of New Orleans. His career spans more than 15 years in both industry and academia and his expertise lies within the fields of marine, electrical and ocean engineering. He holds an Electrical Engineer's degree and a Marine Engineering doctorate. His research interests are process modeling and simulation, system dynamics, identification and control, reliability, signal and data analysis. He is author of several technical papers and a Springer monograph. His current research encompasses projects within nonlinear process dynamics, applied mathematics, energy engineering and watercraft systems.

About the Part Editors

Manhar R. Dhanak Part A For biographical profile, please see the section "About the Editors".

Cortis K. Cooper

Chevron Energy Technology Corp.
San Ramon, CA 94583, USA
cortcooper@chevron.com

Part A

Cortis Cooper is a Chevron Fellow. His primary job is pollutant modeling and providing wind, wave, and current criteria for Chevron's worldwide operations. He received a BSc and an MSc in Engineering at MIT, followed by a PhD from the University of Maine. Outside the industry, he has served on seven National Academy of Sciences committees and boards, testified before Congress, and been a frequent advisor to government agencies.

Thomas B. Curtin

University of Washington
Applied Physics Laboratory
Arlington, VA 22205, USA
tbcurtin@verizon.net

Part B

Thomas Curtin has served as Chief Scientist at NURC in Italy, Chief Knowledge Officer at AUVSI, Program Manager at ONR, Assistant Professor at NCSU, and oceanographer at the Fisheries Research Institute in Malaysia. He received the BS degree in Physics from Boston College, MS and PhD degrees in Physical Oceanography from Oregon State University and the University of Miami, and the MBA degree from Massachusetts Institute of Technology.

Jennifer L. Irish

Virginia Tech
Dep. Civil and Environmental Engineering
Blacksburg, VA 24060, USA
jirish@vt.edu

Part C

Dr Jennifer L. Irish is an Associate Professor of Coastal Engineering at Virginia Tech. She is an expert in storm dynamics at the coast, vegetative effects, and coastal hazard risk assessment. She received the Department of the Army Superior Civilian Service Award and Texas A&M University's Civil Engineering Excellence in Research Award, and has served as Secretary of the Coasts, Oceans, Ports, and Rivers Institute Board of Governors of the ASCE.

James M. Kaihatu

Texas A&M University
Zachry Dep. of Civil Engineering
College Station, TX 77843-3136, USA
jkaihatu@civil.tamu.edu

Part C

James M. Kaihatu is an Associate Professor of Civil Engineering and of Ocean Engineering at Texas A&M University, where he has been since 2006. He has also previously held positions at the Naval Research Laboratory (1995–2006) and the Coastal Engineering Research Center, US Army Corps of Engineers (1987–1989). He received his PhD in Civil Engineering from the University of Delaware in 1994. His research interests include nearshore wave modeling, nearshore nonlinear wave dynamics, nearshore circulation, and wave propagation over cohesive sediments and vegetation.

R. Cengiz Ertekin

University of Hawaii at Manoa
Dep. Ocean & Resources Engineering
Honolulu, HI 96822, USA
Ertekin@hawaii.edu

Part D

R. Cengiz Ertekin has been Professor of Hydromechanics at the University of Hawaii's Department of Ocean and Resources Engineering since 1986. Before that he worked for Chevron and Shell oil companies. He holds a PhD degree from UC Berkeley. He has been teaching and doing research on marine renewable energy, hydroelasticity, nonlinear waves, wave loads on offshore structures and motions, among others. He is a fellow of SNAME and ASME.

Dominique Roddier

Principle Power
Berkeley, CA 94710, USA
droddier@principlepowerinc.com

Part D

Dr Dominique Roddier is a Naval Architect specializing in complex hydrodynamic problems. He serves as Principle Power CTO and is responsible for the WindFloat technology development and manages the company engineering team. He is actively involved in ASME OMAE conference management and in SNAME. He also serves as Associate Editor of multiple journals in the field of offshore engineering.

Nikolaos I. Xiros Part E For biographical profile, please see the section "About the Editors".

List of Authors

Raju Abraham
Naval Research Board
Pallickaranai Campus
Velachery-Tambram Road, Pallickaranai
600 100 Chennai, India
abraham@niot.res.in

Gabriel M. Alsenas
Florida Atlantic University
Southeast National Marine Renewable Energy Center
777 Glades Road
Boca Raton, FL 33431, USA
galsenas@fau.edu

Pak-Cheung Edgar An
Florida Atlantic University
Dep. Ocean and Mechanical Engineering
777 Glades Road
Boca Raton, FL 33431, USA
pan@fau.edu

Palaniswamy Ananthakrishnan
Florida Atlantic University
Dep. Ocean and Mechanical Engineering
777 Glades Road
Boca Raton, FL 33431, USA
panantha@fau.edu

Alexia Aubault
Principle Power Inc.
2321 4th Street
Berkeley, CA 94710, USA
aaubault@principlepowerinc.com

Alexander Bahr
Ecole Polytechnique Fédérale de Lausanne
ENAC IIE DISAL
GR A2 454, Station 2
1015 Lausanne, Switzerland
alexander.bahr@epfl.ch

Promode R. Bandyopadhyay
Naval Undersea Warfare Center
Undersea Warfare Weapons, Vehicles, and Defensive Systems Dep.
1176 Howell Street
Newport, RI 02841, USA
promode.bandyopadhya@navy.mil

David R. Basco
Old Dominion University
Dep. Civil and Environmental Engineering
135 Kaufman Hall
Norfolk, VA 23529, USA
dbasco@odu.edu

Pierre-Philippe J. Beaujean
Florida Atlantic University
Dep. Ocean and Mechanical Engineering
777 Glades Road
Boca Raton, FL 33431, USA
pbeaujea@fau.edu

James G. Bellingham
Woods Hole Oceanographic Institution
266 Woods Hole Road
Woods Hole, MA02543-1050, USA
jbellingham@whoi.edu

Michael R. Benjamin
Massachusetts Inst. of Technology
Laboratory for Autonomous Marine Sensing Systems
77 Massachusetts Avenue
Cambridge, MA 02139, USA
mikerb@mit.edu

Michael M. Bernitsas
University of Michigan
Dep. Naval Architecture & Marine Engineering
2600 Draper Road
Ann Arbor, MI 48109-2145, USA
michaelb@umich.edu

Rémi Bourguet
University of Toulouse, CNRS
Inst. Fluid Mechanics
Allée du Pr. C. Soula
31400 Toulouse, France
remi.bourguet@imft.fr

Mario P. Brito
University of Southampton
Centre for Risk Research
Southampton, S017 1BJ, UK
m.p.brito@soton.ac.uk

Robert A. Brizzolara
Office of Naval Research
Sea Platforms and Weapons Division
875 N. Randolph Street
Arlington, VA 22203, USA
robert.brizzolara@navy.mil

Stefano Brizzolara
Massachusetts Institute of Technology
MIT Innovative Ship Design Lab
77 Massachusetts Avenue
Cambridge, MA 02139-4307, USA
stebriz@mit.edu

Wendell S. Brown
University of Massachusetts – Dartmouth
School for Marine Science and Technology
706 S Rodney French Boulevard
New Bedford, MA 02744-1221, USA
wbrown@umassd.edu

Andrew Cairns
AECOM
125 Broad Street
New York, NY 10004, USA
Andrew.cairns@aecom.com

Thomas J. Campbell
The Shaw Group
Coastal Planning and Engineering Inc.
2481 NW Boca Raton Blvd
Boca Raton, FL 33431, USA
sanddiver@icloud.com

John M. Carel
13658 Lucera Court
Estero, FL 33928, USA
johnmcarel@gmail.com

Bruce Cathers
University of New South Wales
Water Research Laboratory
110 King Street
Manly Vale, NSW 2093, Australia
b.cathers@unsw.edu.au

Ioannis K. Chatjigeorgiou
National Technical University of Athens
School of Naval Architecture and Marine Engineering
9 Heroon Polytechniou Ave
15780 Athens, Greece
chatzi@naval.ntua.gr

Cortis K. Cooper
Chevron Energy Technology Corp.
6001 Bollinger Canyon Road
San Ramon, CA 94583, USA
cortcooper@chevron.com

Joseph A. Curcio
P.O. Box 1259
Gray, ME 04039, USA
jacurcio@mit.edu

Jason Dahl
University of Rhode Island
217 Sheets Laboratory
215 South Ferry Road
Narragansett, RI 02882, USA
dahl@egr.uri.edu

Michael S. Dean
Naval Sea Systems Command
1333 Isaac Hull Avenue, SE Stop 1075
Washington D.C., DC 20376, USA
michael.s.dean@navy.mil

Robert G. Dean (deceased)

Manhar R. Dhanak
Florida Atlantic University
The Inst. for Ocean and Systems Engineering – SeaTech
101 North Beach Road
Dania Beach, FL 33004, USA
dhanak@fau.edu

Alana E.S. Duerr
New West Technologies LLC
10333 East Dry Creek Road, Suite 200
Englewood, CO 80112, USA
a.e.smentek@gmail.com

Gerald D'Spain
University of California San Diego
Scripps Inst. of Oceanography
291 Rosecrans Street
San Diego, CA 92106, USA
gdspain@ucsd.edu

Frank Ehlers
Bundeswehr Technical Center for Ships and Naval Weapons (WTD 71)
Klausdorfer Weg 2-24
24148 Kiel, Germany
drfrank.ehlers@gmail.com

Naomi Ehrich Leonard
Princeton University
Mechanical and Aerospace Engineering
D-234 Engineering Quadrangle
Princeton, NJ 08544, USA
naomi@princeton.edu

Karl Dietrich von Ellenrieder
Florida Atlantic University
SeaTech Inst. for Ocean Systems Engineering
101 North Beach Road
Dania Beach, FL 33004-3023, USA
ellenrie@fau.edu

R. Cengiz Ertekin
University of Hawaii at Manoa
Dep. Ocean & Resources Engineering
2540 Dole Street, Holmes Hall 402
Honolulu, HI 96822, USA
Ertekin@hawaii.edu

Merv Fingas
Spill Science
1717 Rutherford Point S.W.
Edmonton, T6W 1J6, Canada
fingasmerv@shaw.ca

George Z. Forristall
Forristall Ocean Engineering, Inc.
101 Chestnut Street
Camden, ME 04843, USA
george@forocean.com

David Fries
University of South Florida
Inst. for Research in Arts
4202 E. Fowler Avenue
Tampa, FL 33620, USA
dfries@usf.edu

Gwyn Griffiths
Autonomous Analytics
14 Bassett Close
Southampton, SO16 7PE, UK
gwyn@autonomousanalytics.com

Patrick J. Haley Jr.
Massachusetts Institute of Technology
Dep. Mechanical Engineering
77 Massachusetts Avenue
Cambridge, MA 02139-43071, USA
phaley@mit.edu

Andrew Hamilton
Monterey Bay Aquarium Research Institute
7700 Sandholdt Road
Moss Landing, CA 95039, USA
hamilton@mbari.org

Howard P. Hanson
Florida Atlantic University
Southeast National Marine Renewable Energy Center
777 Glades Road
Boca Raton, FL 33431, USA
hphanson@fau.edu

John J. Holmes
Naval Surface Warfare Center
Underwater Electromagnetic Signatures and Technology Division
9500 MacArthur Boulevard
West Bethesda, MD 20817-5700, USA
john.j.holmes@navy.mil

Jennifer L. Irish
Virginia Tech
Dep. Civil and Environmental Engineering
Patton Hall 221E, 750 Drillfield Drive
Blacksburg, VA 24060, USA
jirish@vt.edu

Dong-Sheng Jeng
Griffith University Gold Coast Campus
Griffith School of Engineering
Griffith, QLD 4222, Australia
d.jeng@griffith.edu.au

Scott A. Jenkins
University of California San Diego
Scripps Inst. of Oceanography
291 Rosecrans Street
San Diego, CA 92106, USA
sjenkins@ucsd.edu

Jason Jonkman
National Renewable Energy Laboratory
National Wind Technology Center
15013 Denver West Parkway
Golden, CO 80401, USA
jason.jonkman@nrel.gov

James M. Kaihatu
Texas A&M University
Zachry Dep. of Civil Engineering
3136 TAMU
College Station, TX 77843-3136, USA
jkaihatu@civil.tamu.edu

Tae Won Kim
FMC Technologies Schilling Robotics
260 Cousteau Place Suite 200
Davis, CA 95618, USA
twon.kim@gmail.com

William Kirkwood
Monterey Bay Aquarium Research Institute
7700 Sandholdt Road
Moss Landing, CA 95039, USA
kiwi@mbari.org

John J. Leonard
Massachusetts Institute of Technology
Dep. Mechanical Engineering
77 Massachusetts Avenue
Cambridge, MA 02139-4307, USA
jleonard@mit.edu

Pierre F.J. Lermusiaux
Massachusetts Institute of Technology
Dep. Mechanical Engineering
77 Massachusetts Avenue
Cambridge, MA 02139-43071, USA
pierrel@mit.edu

Wayne G. Leslie
American Meteorological Society
45 Beacon Street
Boston, MA 02108-3693, USA
wgleslie@mit.edu

Xiao Li
4 Macopin Drive
Livingston, NJ 07039, USA
Steven.li@aecom.com

Tapovan Lolla
Massachusetts Institute of Technology
Dep. Mechanical Engineering
77 Massachusetts Avenue
Cambridge, MA 02139-43071, USA
ltapovan@mit.edu

Raymond Lum
DSO National Laboratories
14 Science Park Drive
118226 Singapore, Singapore
raylum@alum.mit.edu

Patrick J. Lynett
University of Southern California
Sonny Astani Dep. of Civil and Environmental Engineering
3620 South Vermont Avenue
Los Angeles, CA 90089, USA
plynett@usc.edu

Giacomo Marani
West Virginia University
WV Robotic Technology Center
1000 Galliher Drive
Fairmont, WV 26554, USA
gmarani@gmail.com

Spyros A. Mavrakos
National Technical University of Athens
School of Naval Architecture and Marine Engineering
9 Heroon Polytechniou Ave
15780 Athens, Greece
mavrakos@naval.ntua.gr

Robert E. Melchers
The University of Newcastle
Centre for Infrastructure Performance and Reliability
University Drive
Callaghan, NSW 2308, Australia
Rob.Melchers@newcastle.edu.au

Yahya Modarres-Sadeghi
University of Massachusetts
Mechanical and Industrial Engineering
160 Governors Drive
Amherst, MA 01003, USA
modarres@engin.umass.edu

Stephanie M. Petillo
Woods Hole Oceanographic Institution
266 Woods Hole Road
Woods Hole, MA 02543, USA
spetillo@whoi.edu

Roshanka Ranasinghe
UNESCO-IHE Inst. for Water Education
Westvest 7
2611 AX, Delft, The Netherlands
r.ranasinghe@unesco-ihe.org

Muthukamatchi Ravindran
Naval Research Board
Pallickaranai Campus
Velachery-Tambram Road, Pallickaranai
600 100 Chennai, India
ravinmv@gmail.com

Dronnadula V. Reddy
Florida Atlantic University
Dep. Civil, Environmental and Geomatics Engineering
777 Glades Road EW201
Boca Raton, FL 33431, USA
dreddy@fau.edu

Donald T. Resio
University of North Florida
Dep. Civil Engineering
1 UNF Drive
Jacksonville, FL 32224, USA
don.resio@unf.edu

H. Ronald Riggs
University of Hawaii at Manoa
Dep. Civil & Environmental Engineering
2540 Dole Street, Holmes Hall 402
Honolulu, HI 96822, USA
riggs@hawaii.edu

George Rodenbusch
Rodenbusch Consulting
213 Heather Hills Drive
Dripping Springs, TX 78620, USA
george@rodenbusch.net

J.A. Dano Roelvink
UNESCO-IHE Inst. for Water Education
Westvest 7
2611 AX, Delft, The Netherlands
d.roelvink@unesco-ihe.org

Salvatore Scaturro
Expert Water Modelling
Level 26, 44 Market Street
Sydney, NSW 2000, Australia
sal.scaturro@ewmworld.com

Henrik Schmidt
Massachusetts Institute of Technology
Center for Ocean Engineering
77 Massachusetts Avenue
Cambridge, MA 02139, USA
henrik@mit.edu

Robert Seah
Chevron Energy Technology Company
Facilities Engineering Dep.
1400 Smith Street
Houston, TX 77002, USA
RSeah@Chevron.com

Thomas Sondergaard
TrackMan A/S
Stubbeled 2
2950 Vedbæk, Denmark
tsondergaard@gmail.com

Matthias Stammler
Fraunhofer Institute for Wind Energy and Energy System Technology IWES
Appelstraße 9A
30167 Hannover, Germany
matthias.stammler@iwes.fraunhofer.de

Milica Stojanovic
Northeastern University
ECE Department
Boston, MA 02115, USA
millitsa@ece.neu.edu

Mareike Strach-Sonsalla
Fraunhofer Institute for Wind Energy and Energy System Technology IWES
Am Seedeich 45
27572 Bremerhaven, Germany
mareike.strach-sonsalla@outlook.com

Arisi S.J. Swamidas
Memorial University of Newfoundland
Faculty of Engineering and Applied Science
230 Elizabeth Avenue
St. John's, A1B 2C6 NL, Canada
aswamidas@mun.ca

Peter M. Tate
Sydney Water Corporation
PO Box 399
Parramatta, NSW 2124, Australia
peter.tate@sydneywater.com.au

Krish P. Thiagarajan
University of Maine
Dep. Mechanical Engineering
5711 Boardman Hall
Orono, ME 04469, USA
Krish.Thiagarajan@maine.edu

Michael S. Triantafyllou
Massachusetts Institute of Technology
Dep. Mechanical Engineering
77 Massachusetts Avenue
Cambridge, MA 02139, USA
mistetri@mit.edu

Mark A. Tumeo
University of North Florida
College of Computing, Engineering & Construction
1 UNF Drive
Jacksonville, FL 32224, USA
m.tumeo@unf.edu

Mattheus P. Ueckermann
Creare
16 Great Hollow Road
Hanover, NH 30755, USA
mpu@creare.com

James H. VanZwieten
Florida Atlantic University
Southeast National Marine Renewable Energy Center
777 Glades Road
Boca Raton, FL 33431, USA
jvanzwi@fau.edu

Fabian Vorpahl
Leading Expert Offshore Engineering Tower & Substructure Senvion GmbH
Franz-Lenz-Straße 1
49084 Osnabrück, Germany
fabian.vorpahl@senvion.com

Dirk-Jan R. Walstra
Deltares
Unit Hydraulic Engineering
Rotterdamseweg 185
2629 HD, Delft, The Netherlands
Dirkjan.walstra@deltares.nl

Mick van der Wegen
UNESCO-IHE Inst. for Water Education
Westvest 7
2611 AX, Delft, The Netherlands
m.vanderwegen@unesco-ihe.org

Robert Weiss
Virginia Tech
Dep. Geosciences
4044 Derring Hall
Blacksburg, VA 24060, USA
weiszr@vt.edu

Jan Wenske
Fraunhofer Institute for Wind Energy and Energy System Technology IWES
Am Seedeich 45
27572 Bremerhaven, Germany
jan.wenske@iwes.fraunhofer.de

Nikolaos I. Xiros
University of New Orleans
School of Naval Architecture & Marine Engineering
2000 Lakeshore Drive
New Orleans, LA 70148, USA
nxiros@uno.edu

Konuralp Yigit
University of Illinois at Urbana Champaign
Dep. Aerospace Engineering
104S Wright Street
Urbana, IL 61801, USA
yigit2@illinois.edu

Solomon Yim
Oregon State University
School of Civial and Construction Engineering
Kearney Hall 101
Corvallis, OR 97331, USA
solomon.yim@oregonstate.edu

Junku Yuh
Korea Institute of Science and Technology
National Agenda Research Division
Hwarangno 14-gil 5, Seongbuk-gu
Seoul, 136-791, Korea
yuh.junku@gmail.com

Contents

List of Abbreviations XXIX

1 **Introduction**
Manhar R. Dhanak, Nikolaos I. Xiros 1
1.1 Enabling Maritime Design and Development 1
1.2 History 2
1.3 Basics 2
1.4 Applications 4
1.5 Future Trends 7
References 8

Part A Fundamentals

2 **Elements of Physical Oceanography**
Wendell S. Brown 15
2.1 Atmospheric Processes 16
2.2 Ocean Structure 17
2.3 Oceanic Processes 18
2.4 Surface Gravity Waves 19
2.5 Wind-Forced Ocean Processes 28
2.6 Deep Ocean Currents 38
2.7 Coastal Ocean Currents 39
2.8 Ocean Surface Tides 40
2.9 Oceanic Internal Waves and Tides 43
References 45

3 **Metocean Extreme and Operating Conditions**
George Z. Forristall, Cortis K. Cooper 47
3.1 Quantifying the Metocean Environment 48
3.2 Overview of WWC Processes 49
3.3 Measurements 55
3.4 Modeling 58
3.5 Joint Events 61
3.6 Operational Criteria 64
3.7 Extreme Criteria 66
3.8 Conclusions 72
References 72

4 **Mechanics of Ocean Waves**
James M. Kaihatu, Palaniswamy Ananthakrishnan 77
4.1 Ocean Surface Waves 77
4.2 Wave Theories 78
4.3 Properties of Small Amplitude Gravity Waves 80
4.4 Weakly Nonlinear Deep Water Wave Theories 83
4.5 Shallow Water Wave Theories 87
4.6 Transformation of Waves Approaching Land 90

4.7 Computational Method for Fully Nonlinear Waves 93
4.8 Wave Forces on Fixed and Floating Structures 94
4.9 Concluding Remarks 97
References 98

5 **Physical Properties of Seawater**
Wendell S. Brown 101
5.1 Hydrostatic Pressure 101
5.2 Temperature 101
5.3 Salinity 103
5.4 Density 104
5.5 Temperature–Salinity Relationships 105
5.6 Specific Heat 106
5.7 Freezing of Sea Water and Sea Ice 106
5.8 Coefficient of Thermal Expansion 106
5.9 Sound Velocity 107
5.10 Acoustic Ambient Noise 107
5.11 Light Transmission 108
References 109

6 **Principles of Marine Corrosion**
Robert E. Melchers 111
6.1 Chemical and Physical Composition of Seawater 111
6.2 Materials Used in Marine Environments 113
6.3 Marine Corrosion of Steel 113
6.4 Modeling Longer Term Corrosion of Steel 116
6.5 Other Influences on Steel Corrosion 118
6.6 Pitting Corrosion of Steel 119
6.7 Some Other Important Materials 121
6.8 Conclusion 123
References 123

7 **Hydromechanics**
Karl Dietrich von Ellenrieder, Manhar R. Dhanak 127
7.1 Dimensional Analysis, Basic Estimation, and Model Testing 127
7.2 Fluid Statics 153
7.3 Hydrodynamics 155
References 174

8 **Ocean Electromagnetics**
John J. Holmes 177
8.1 Electromagnetism in an Ocean Environment 177
8.2 Electromagnetic Field Theory 178
8.3 Plane Wave Propagation 180
8.4 Reflection and Transmission of a Plane Wave at the Surface of Fresh Water 182
8.5 Plane Wave Incident on Seawater 184
8.6 Magnetic and Electric Dipoles in an Unbounded Ocean 186
8.7 Magnetic and Electric Dipoles in a Bounded Ocean 188
8.8 Electromagnetic Propagation in the Ocean at Optical Wavelengths 193
References 195

9 **Digital Signal Processing**
Nikolaos I. Xiros 197
9.1 Discrete-Time Systems 197
9.2 Digital Filters 206
9.3 The Fast Fourier Transform (FFT) 211
9.4 Waveform Analysis 216
9.5 Optimal Signal Estimation 220
9.6 Concluding Remarks 225
References 225

10 **Control Theory and Applications**
Nikolaos I. Xiros, Pak-Cheung Edgar An 227
10.1 System Theory 227
10.2 Analysis of LTI Systems 237
10.3 SISO System Controls 247
10.4 Pole Placement of LTI Systems 261
10.5 Course-Keeping Autopilots 267
References 275

Part B Autonomous Ocean Vehicles, Subsystems and Control

11 **Highly Maneuverable Biorobotic Underwater Vehicles**
Promode R. Bandyopadhyay 281
11.1 Biorobotics 281
11.2 Theoretical Foundation of Animal-Inspired Hydrodynamics and Control 286
11.3 Description of Biology-Inspired Vehicles of Emergent Maturity 289
11.4 Reliability, Low Power Consumption, and Disturbance Rejection of Bio-Inspired Propulsion 296
11.5 Demonstrated Maneuverings of NUWC Bio-Inspired Vehicles 296
11.6 Discussion 297
11.7 Concluding Remarks 298
11.8 Nomenclature 298
References 299

12 **Autonomous Underwater Gliders**
Scott A. Jenkins, Gerald D'Spain 301
12.1 Concept 301
12.2 Hydrodynamics of Wings Versus Propellers 305
12.3 Underwater Glider Attributes and Limitations 306
12.4 Optimal Size and Shape for Horizontal Transport Efficiency 308
12.5 Thermal Glider 318
12.6 Discussion and Conclusions 319
References 320

13 **Autonomous Sea Surface Vehicles**
Stefano Brizzolara, Robert A. Brizzolara 323
13.1 Platforms 324
13.2 Autonomous Maneuvering and Navigation 326
13.3 Naval Architecture of AUSV Design 327

13.4 Optimized Class of Autonomous Unmanned Surface Vehicles 330
13.5 Conclusions 337
References 339

14 **Autonomous Underwater Vehicle Navigation**
John J. Leonard, Alexander Bahr 341
14.1 Sensors 343
14.2 Algorithms 346
14.3 Summary 352
14.4 Conclusion 353
References 354

15 **Acoustic Communication**
Milica Stojanovic, Pierre-Philippe J. Beaujean 359
15.1 A Brief History 360
15.2 Current and Emerging Modem Applications 360
15.3 Existing Technology 361
15.4 Propagation Channel 364
15.5 Point-to-Point Links: Signal Processing 374
15.6 Future Trends 383
References 383

16 **Autonomous Underwater Vehicle Docking**
James G. Bellingham 387
16.1 Technical Elements of Docking 388
16.2 AUV Characteristics 390
16.3 Sensors For Homing 391
16.4 Capture and Connection Mechanisms 395
16.5 Coupling Power and Communications 399
16.6 AUV Control Considerations 400
16.7 Conclusions and Future Prospects 403
References 404

17 **Underwater Vehicle Manipulators**
Tae Won Kim, Giacomo Marani, Junku Yuh 407
17.1 Underwater Vehicles for Intervention Missions 407
17.2 Dynamics of Underwater Vehicle Manipulators 409
17.3 Teleoperation of Underwater Vehicle Manipulators 410
17.4 Sensor-Based Manipulator Control 411
17.5 Coordinated Motion Control of Underwater Vehicle-Manipulator Systems 413
17.6 Underwater Autonomous Manipulation 414
17.7 Conclusions 419
References 419

18 **Non-Acoustic Sensors**
David Fries, William Kirkwood 423
18.1 Non-Acoustic Ocean Sensors: Sourcing and Classification 423
18.2 Classical Non-Acoustic Ocean Sensors 424
18.3 Chemical Sensor Systems 426
18.4 Biological Sensor Systems 429

18.5 Physical Sensor Systems 432
18.6 AUV-Based Physical Sensors – Horizons 435
18.7 AUV-Chemistry Sensors – Horizons 435
18.8 AUV-Based Biological Sensors – Horizons 436
18.9 Autonomous Sampling Systems – Extending Real-Time AUV Sensors 436
18.10 Non-Acoustic Sensor Packaging 436
18.11 The Essential Need for Sensors 437
References 437

19 **Cooperative Vehicle Environmental Monitoring**
Naomi Ehrich Leonard 441
19.1 Motivation 441
19.2 Background and History 443
19.3 Advances in Cooperative Vehicle Ocean Monitoring 445
19.4 Recent Developments and Future Directions 452
References 454

20 **Nested Autonomy for Distributed Ocean Sensing**
Henrik Schmidt, Michael R. Benjamin, Stephanie M. Petillo, Raymond Lum 459
20.1 Nested Autonomy 460
20.2 Concept of Operations (CONOPS) 461
20.3 Autonomy 463
20.4 Acoustic Communication Infrastructure 466
20.5 On-Board, Real-Time Signal Processing 467
20.6 Application Examples 467
20.7 Conclusion 478
References 479

21 **Science of Autonomy: Time-Optimal Path Planning and Adaptive Sampling for Swarms of Ocean Vehicles**
Pierre F.J. Lermusiaux, Tapovan Lolla, Patrick J. Haley Jr., Konuralp Yigit, Mattheus P. Ueckermann, Thomas Sondergaard, Wayne G. Leslie 481
21.1 Time-Optimal Path Planning for Swarms of Ocean Vehicles 482
21.2 Adaptive Sampling for Swarms of Ocean Vehicles 489
21.3 Conclusions and Outlook 494
References 496

22 **Cooperative Vehicle Target Tracking**
Frank Ehlers 499
22.1 General Theoretical Framework 499
22.2 Distributed Sensing, Control, and Decisions 503
22.3 Multistatic Sonar 505
22.4 Maritime Surveillance 510
22.5 Effective Coordination Schemes 511
22.6 Conclusions and Recommendations 512
References 513

23 **Rules of the Road for Unmanned Marine Vehicles**
Joseph A. Curcio 517

23.1 COLREGS 517
23.2 Sensing the World 518
23.3 Proper Behaviors Yield Compliance 519
23.4 Integrating UMVs into Public Water Space 523
23.5 Developing Standards 524
23.6 The Road Ahead 525
23.7 Conclusion 525
References 526

24 **Autonomy: Risk Assessment**
Mario P. Brito, Gwyn Griffiths 527
24.1 Risk Management Process for Autonomous Ocean Vehicles 528
24.2 Risk of Failure 529
24.3 Risk of Collision 532
24.4 Risk of Unavailability 534
24.5 Risk of Loss 535
24.6 Legal Risks 541
References 542

Part C Coastal Design

25 **Physical Characteristics of Coastal Hazards**
Jennifer L. Irish, Robert Weiss, Donald T. Resio 549
25.1 Types of Coastal Hazards 549
25.2 Coastal Impacts 555
25.3 Summary 561
25.4 Nomenclature 561
References 562

26 **Statistical Characterization of Hazards and Risk in Coastal Areas**
Donald T. Resio, Mark A. Tumeo, Jennifer L. Irish 567
26.1 Overview of Risk and Uncertainty 567
26.2 Quantifying Coastal Hazards/Risks 570
26.3 Historical Perspective 574
26.4 Summary 587
26.5 Nomenclature 587
26.A Appendix: Glossary of Probability and Risk Terms 587
References 593

27 **Modeling of Coastal Waves and Hydrodynamics**
Patrick J. Lynett, James M. Kaihatu 597
27.1 Wind Wave Modeling 598
27.2 Modeling Long Waves 604
27.3 Coupled and Nested Techniques 605
27.4 Summary of Model Properties 606
27.5 Conclusions 608
27.6 Nomenclature 608
References 608

28 **Modeling of Coastal Morphological Processes**
J.A. Dano Roelvink, Dirk-Jan R. Walstra, Mick van der Wegen, Roshanka Ranasinghe 611
28.1 Types of Coastal Models 611
28.2 Principles of Process-Based Morphodynamic Modeling 613
28.3 Modeling Approaches 619
28.4 Future Directions 629
28.5 Nomenclature 631
References 631

29 **Beach Nourishment**
Robert G. Dean, Thomas J. Campbell 635
29.1 Advantages of Beach Nourishment Over Other Approaches 636
29.2 Methods of Delivery of Sand for Beach Nourishment 636
29.3 Role of Structures in Beach Nourishment 639
29.4 Design and Prediction Approaches and Methods 640
29.5 Additional Design Considerations 643
29.6 Legacy Beach Nourishment Projects 645
29.7 Other Beach Nourishment Projects 648
29.8 Summary and Conclusions 649
29.9 Nomenclature 650
References 650

30 **Storm Hazard Mitigation Structures**
David R. Basco 653
30.1 Design Criteria, Philosophy, and Constraints 655
30.2 Coastal Armoring Structures 658
30.3 Shoreline Stabilization Structures 674
30.4 Websites and Sea Level Rise Trends 681
References 681

31 **Port and Harbor Design**
Andrew Cairns, John M. Carel, Xiao Li 685
31.1 Port and Harbor Layout and Design 685
31.2 Structure Types 692
31.3 Loads on Structures due to Vessel Mooring and Berthing 703
31.4 Suggested Reading 707
31.5 Notation 709
References 709

32 **Marine Outfalls**
Peter M. Tate, Salvatore Scaturro, Bruce Cathers 711
32.1 Terminology 712
32.2 Governance 713
32.3 Predicting Near-Field Dilutions 716
32.4 Hydraulic Analysis and Design 723
32.5 Outfall Construction 734
32.6 Environmental Monitoring 736
References 739

Part D Offshore Technologies

33 **Offshore Platforms**
Arisi S.J. Swamidas, Dronnadula V. Reddy 745
33.1 Relevance 745
33.2 Types of Offshore Platforms 747
33.3 Future Trends and Developments in Offshore Platforms 751
References 752

34 **Stability of Offshore Systems**
Alexia Aubault, R. Cengiz Ertekin 755
34.1 Stability Criteria 756
34.2 Fundamentals 757
34.3 Hydrostatic Forces and Moments 761
34.4 Stability 763
34.5 Loads 776
34.6 Lightship Parameters 779
34.7 Subdivision 781
34.8 Analysis 784
References 785

35 **Wave, Current and Wind Loads**
R. Cengiz Ertekin, George Rodenbusch 787
35.1 Wave Loads 787
35.2 Current Loads 802
35.3 Wind Loads 806
35.4 Model Tests 810
35.5 CFD Tools 814
35.6 Extreme Response Estimation 814
References 816

36 **Vortex-Induced Vibrations**
Michael S. Triantafyllou, Rémi Bourguet, Jason Dahl, Yahya Modarres-Sadeghi 819
36.1 VIV Prediction of Wide-Span Rigid Cylinders in Uniform Cross-Flow 820
36.2 VIV Prediction of Flexible Structures in Nonuniform Flow 831
36.3 Experimental Studies and Fatigue Analysis 838
36.4 Effectiveness of Vortex Canceling Devices 843
36.5 Multiple Interfering Bluff Bodies 845
36.6 Effect of Reynolds Number 845
References 846

37 **Structural Dynamics**
H. Ronald Riggs, Solomon Yim 851
37.1 Single Degree-of-Freedom System 852
37.2 Multi-Degree of Freedom Systems 861
37.3 Linear Hydroelasticity for Inviscid Fluid Flow 864
37.4 Linear Response to Random Seas 867
37.5 Nonlinear Hydroelasticity (Nonlinear FSI) 870
References 873

38 **Cable Dynamics for Marine Applications**
Ioannis K. Chatjigeorgiou, Spyros A. Mavrakos 875
38.1 Mathematical Formulation 876
38.2 The Eigenvalue Problem of a Catenary Mooring Cable 882
38.3 High Tension Cables; Snap-Slack Conditions 885
38.4 Dynamics of Catenary Moorings 888
38.5 Second-Order Nonlinear Dynamics of Cables Formulated as Euler–Bernoulli Beams 894
References 903

39 **Offshore Geotechnics**
Dong-Sheng Jeng 907
39.1 Basic Models 907
39.2 Mechanisms of Seabed Dynamics 916
39.3 Wave(Current)-Induced Soil Response in Marine Sediments 922
39.4 Seabed Stability Around Caisson Breakwaters 925
39.5 Remarks 931
References 932

40 **Buoy Technology**
Andrew Hamilton 937
40.1 Buoy and Mooring Types and Design Considerations 937
40.2 Buoy and Mooring System Components 942
40.3 Analysis Techniques 949
40.4 Example Designs 956
References 961

41 **Liquefied Natural Gas Carriers**
Krish P. Thiagarajan, Robert Seah 963
41.1 Types of LNG Carriers 963
41.2 Thermodynamics of LNG 968
41.3 Environmental Challenges 970
41.4 Fluid Structure Interaction of LNG Systems 971
41.5 Design Methodologies of LNG Containment Systems 978
References 983

42 **Salvage Operations**
Michael S. Dean 985
42.1 The Casualty and Response 986
42.2 Introduction to Salvage Engineering 988
42.3 Data, Surveys, and Planning 991
42.4 Types of Operations 996
42.A Appendix: Dewatering 1003
42.B Appendix: Common Formulas, Calculations, and References 1008
42.C Appendix: De-Beaching 1028
42.D Appendix: Refloating 1037
42.E Appendix: Wrecking in Place 1055
References 1065

43 **Oil Spills and Response**
Merv Fingas 1067
43.1 Frequency of Oil Spills 1068
43.2 Response to Oil Spills 1070
43.3 Typical Oils and Their Properties 1070
43.4 Behavior of Oil in the Environment 1071
43.5 Analysis, Detection, and Remote Sensing of Oil Spills 1075
43.6 Containment on Water 1078
43.7 Oil Recovery on Water 1080
43.8 Separation, Pumping, Decontamination, and Disposal 1084
43.9 Spill-Treating Agents 1086
43.10 In-Situ Burning 1087
43.11 Shoreline Cleanup and Restoration 1089
References 1092

Part E Ocean Renewable Energy

44 **Marine Hydrokinetic Energy Resource Assessment**
Manhar R. Dhanak, Alana E.S. Duerr, James H. VanZwieten 1099
44.1 Wave Energy Resource 1100
44.2 Tidal and Ocean Current Energy Resource 1101
44.3 Assessment of Global Ocean Current Resources 1106
44.4 Other Considerations 1113
References 1114

45 **Ocean Wave Energy Conversion Concepts**
Nikolaos I. Xiros, Manhar R. Dhanak 1117
45.1 Basic Concepts in Primary Energy Capture 1118
45.2 Power Takeoff Systems 1127
45.A Appendix: Practical Applications of Wave Energy Conversion Technologies 1133
References 1143

46 **Ocean Current Energy Conversion**
Howard P. Hanson, James H. VanZwieten, Gabriel M. Alsenas 1147
46.1 Fundamentals 1148
46.2 The Betz Limit 1149
46.3 Conversion Systems 1151
46.4 Supporting Infrastructure 1156
46.5 Beyond Engineering 1160
46.6 Summary 1160
References 1160

47 **Harvesting Energy by Flow Included Motions**
Michael M. Bernitsas 1163
47.1 Hydrokinetic Energy in Horizontal Flow 1166
47.2 Alternating-Lift Technologies: The VIVACE Converter as a Case Study 1178
47.3 Methodology and Tools in Support of Development 1202
47.4 Nomenclature 1237
References 1238

48 **Ocean Thermal Energy Conversion**
Muthukamatchi Ravindran, Raju Abraham 1245
48.1 OTEC Principles and Systems 1245
48.2 History of OTEC Installations Worldwide 1250
48.3 Current Status of OTEC Technologies 1254
48.4 Design Considerations for Future OTEC Plants 1257
48.5 Conclusion 1265
References 1265

49 **Offshore Wind Energy**
Mareike Strach-Sonsalla, Matthias Stammler, Jan Wenske, Jason Jonkman, Fabian Vorpahl 1267
49.1 Current Offshore Wind Turbine Technology 1268
49.2 Fundamentals of Turbine Dynamics and the Offshore Environment 1276
49.3 Outlook on Future Technology 1283
References 1284

Acknowledgements 1287
About the Authors 1289
Detailed Contents 1303
Index 1327

List of Abbreviations

Symbols

1-D	one-dimensional
2-D	two-dimensional
2-D-URANS	two-dimensional unsteady, Reynolds-Averaged, Navier–Stokes
2.5-D	two-and-a-half-dimensional
3-D	three-dimensional
4-D	four-dimensional

A

A/D	analog-to-digital
A/R	asphaltenes and resins
ABE	autonomous benthic explorer
ACDP	acoustic current Doppler profiler
ACF	actual cubic feet
ACTUV	anti-submarine warfare unmanned surface vehicle
ACV	air cushion vehicle
ADCP	acoustic Doppler current profiler
ADV	acoustic Doppler velocimeter
AGV	automated guided vehicle
AHRS	attitude-heading reference system
AIS	automatic identification system
ALIVE	autonomous light intervention vehicle
ALT	alternating-lift technology
AM	analog modulation
AMOC	Atlantic meridional overturning circulation
AOP	apparent optical properties
AOSN	autonomous ocean sampling network
AOV	autonomous ocean vehicle
ARMA	Auto-Regressive Moving Average
ARQ	automatic repeat request
ASAP	adaptive sampling and prediction
ASC	autonomous surface craft
ASL	autonomous systems laboratory
ATA	absolute pressure in atmospheres
ATB	articulated tug barge
AUSV	autonomous USV
AUV	autonomous underwater vehicle
AWACS	autonomous wide aperture cluster for surveillance
AWM	added-weight method
Awp	areas of waterplane
AWS	Archimedes wave swing

B

BAUV	biorobotic autonomous underwater vehicle
Bbl	billion barrels
BCH	Bose-Chaudhuri-Hocquenghem
BENS	Broadband Environmental Networked Sensor
BHP	brake horsepower
BM	metacentric radius
BML	longitudinal metacentric radius
BOD	biochemical oxygen demand
BOG	boil-off gas
BOP	blow-out-preventer
BTEX	benzene, toluene, ethyl-benzene and xylenes
BTR	beam-time record

C

C3MR	air products propane precooled mixed refrigerant
CA	cellular automata
CAB	captured air bubble
CAD	computer-aided design
CALM	Catenary Anchor Leg Mooring
CARACaS	control architecture for robotic agent command and sensing
CBM	constant-buoyancy method
CCD	charge-coupled device
CCL	compact control language
CDF	cumulative distribution function
CDMA	code-division multiple access
CE	control and estimation
CEP	circular error probability
CF	cross-flow
CFD	computational fluid dynamics
CFL	Courant–Friedrichs–Lévy
CNG	compressed natural gas
COB	center of buoyancy
COBRAS	Cornell breaking wave and structure
COG	center of gravity
COM	center of mass
COMCOT	Cornell multi-grid coupled tsunami model
CONOPS	concept of operations
COSI	coastal storm impulse
COULWAVE	Cornell University long and intermediate wave model

CP	cathodic protection
CPC	characteristic current profile
CSM	computational structural mechanics
CSMA	carrier sensing multiple access
CT	inorganic carbon
CTD	conductivity, temperature and depth
CTS	clear-to-send
CWP	cold water pipe

D

D/A	digital-to-analog
DAB	digital audio broadcast
DAC	digital-to-analog conversion
DATS	digital acoustic telemetry system
DCC	direct contact condenser
DCCL	dynamic compact control language
DEC-POMDP	decentralized partially observable Markov decision process
DEMUS	deployable experimental multistatic undersea surveillance
DEPTHX	deep phreatic thermal explorer
DFE	decision-feedback equalizer
DFIG	doubly-fed generator
DFT	discrete Fourier transform
DGPS	differential global positioning system
DIDSON	dual-frequency identification sonar
DIF	decimation-in-frequency
DIT	decimation-in-time
DOC	dissolved organic compound
DOF	degree of freedom
DON	organic nitrogen compound
DP	dynamic positioning
DR	dead reckoning
DS-SS	direct-sequence spread-spectrum
DSL	demanded static load
DSP	Digital Signal Processor
DTFT	discrete-time Fourier transform
DTG	dynamically tuned gyroscope
DTN	delay-tolerant network
DVB	digital video broadcast
DVL	Doppler velocity log
DWT	deadweight tonnage

E

ECC	equatorial countercurrent
EDS	energy density spectrum
EESG	electrically-excited-synchronous generator
EHP	indicative effective horse power
EHS	erosional hot spot
EKF	extended Kalman filter
ELF	extremely low frequency
EMF	electromotive force
ENSO	El Nino-Southern Oscillation
EO	electro-optic
EOF	empirical orthogonal function
ER	echo-repeater
ERS1	Earth Resources Satellite
ESDU	Engineering Sciences Data Unit
ESP	environmental sampler processor
ESSE	error subspace statistical estimation
EST	empirical simulation technique
ETM	empirical track method
EUC	equatorial undercurrent

F

FDMA	frequency-division multiple access
FEC	forward error coding
FEM	finite element method
FFT	fast Fourier transform
FID	flame ionization detection
FIM	flow-induced motion
FIR	finite-duration impulse response
FLNG	floating liquefied natural gas
FMEA	failure model end effect analysis
FOG	fiber optic gyroscope
FOH	first-order hold
FORM	first-order reliability method
FOV	field of view
FPSO	floating production, storage and offloading
FR	Froude number
FRP	fiber reinforced plastic
FSI	flow–structure interaction
FSK	frequency shift keying
FSRU	floating regasification and storage unit
FSW	feet of seawater
FT	force–torque
FUNWAVE	fully nonlinear Boussinesq wave model
FVM	finite-volume method

G

GA	genetic algorithm
GBS	gravity-based substructure
GC	gas chromatograph
GC-MS	gas chromatography with mass spectrometric detection
GCCS	glider coordinated control system
GCT	glider coordinated trajectory
GDP	global drifter program
GDW	generalized dynamic wake
GFM	Green function method
GIS	geographic information system
GM	metacentric height

GML	longitudinal metacentric height
GMM	Gaussian Mixture Model
GMT	Greenwich mean time
GOOS	global ocean observing system
GPD	generalized Pareto distribution
GPS	global positioning system
GRP	glass-reinforced plastic
GTECCA	global tropical and extratropical cyclone climatic atlas
GWP	global warming potential

H

HAWT	horizontal-axis wind turbine
HDPE	high-density polyethylene
HED	horizontal electric dipole
HF	high frequency
HFR	high frequency radar
HMD	horizontal magnetic dipole
HMI	human–machine interface
HOPS	Harvard Ocean Prediction System
HOV	high-occupancy vehicle
HP	high pressure
HSLA	high strength low alloy
HTF	high tow force
HWM	high water mark
HYCOM	hybrid coordinate ocean model
HYSWAS	hydrofoil small water area ship

I

I-AUV	intervention-AUV
I/O	input–output
ICI	inter-carrier interference
ICON	Innovative Coastal Ocean Observing Network
ID	internal diameter
IFORM	inverse first-order reliability methods
IG	induction generator
IGBT	insulated-gate bipolar transistor
IHI	Ishikawajima–Harima heavy industries
IIR	infinite-duration impulse response
IMU	inertial measurement unit
INS	inertial navigation system
IO	inferior-olive
IOOS	Integrated Ocean Observing System
IPTS	international practical temperature scale
IR	infrared
IRM	inspection, repair, and maintenance
IRPA	individual risk per annum
ISI	inter-symbol interference
IT	inertia of the waterplane
ITB	integrated tug barge
IvP	interval programming

J

JPM	joint probability method

K

KC	Keulegan–Carpenter
KdV	Korteweg–de Vries

L

L/D	lift-to-drag ratio
LBM	lost-buoyancy method
LC-MS	liquid chromatography mass spectrometry
LCB	longitudinal center of buoyancy
LCF	longitudinal center of flotation
LCG	longitudinal center of gravity
LCO	limit cycle oscillation
LCOE	levelized cost of energy
LDPC	low-density parity-check
LDV	laser-Doppler velocimetry
LED	light-emitting diode
LEL	lower explosive limit
LES	large eddy simulation
LEV	leading-edge vortex
LFCTT	low-frequency cooperative target tracking
LGPL	GNU Lesser General Public License
LIBS	laser-induced breakdown spectroscopy
LIMPET	land-installed marine power energy transmitter
LISST	laser in-situ scattering and transmissometry
LNG	liquefied natural gas
LNGC	liquefied natural gas carrier
LOA	length overall
LOPC	laser optical particle counter
LPD	low probability of detection
LPG	liquefied petroleum gas
LTFSW	low-turbulence free-surface water
LTI	linear time invariant
LWL	length on waterline

M

MAC	medium access control
MACA	multiple-access collision avoidance
MAS	multiagent system
MBM	multi-buoy mooring
MBS	multibody simulation
MCTC	moment to change trim by one centimeter
MDOF	multiple degrees-of-freedom

MEG3	mooring equipment guideline
MEL	mixed Eulerian–Lagrangian
MEMS	micro-electro-mechanical system
MF	momentum flux
MFC	multi fluid cascade
MGSVA	Mariano Global Surface Velocity Analysis
MHK	marine hydrokinetic
MHW	mean high water
MIC	microbially induced corrosion
MILP	mixed integer linear programming
MIMO	multi-input multi-output
MIMS	membrane inlet mass spectrometry
MIP	mixed integer programming
MODU	mobile offshore drilling unit
MOOS	mission oriented operating suite
MOST	method of splitting tsunami
MPI	maximum possible intensity
MRE	marine renewable energy
MS	mass spectrometry
MSE	mean squared error
MSEAS	multidisciplinary simulation, estimation and assimilation system
MSL	mean sea level
MSP	marine spatial planning
MTBF	mean time between failure
MVC	main vehicle computer
MWL	mean-water level

N

NA	navigation aid
NCOM	Navy Coastal Ocean Model
NCV	nearly constant velocity
NEC	north equatorial current
NG	natural gas
NIP	net inlet pressure
NP	neutral loading point
NPSH	net positive suction head
NRV	NATO research vessel
NSE	Navier–Stokes equation
NSW	nonlinear shallow water
NTE	net transport economy

O

OA	objective analysis
OCT	ocean current turbine
OD	outer diameter
ODP	ozone depletion potential
OEX	Ocean Explorer
OFDM	orthogonal frequency division multiplexing
OHMSETT	oil and hazardous materials simulated environmental test tank
OHS	oil and hazardous substance
OM	operations and maintenance
ORQ	oil rig quality
OS	optimal sampling
OTEC	ocean thermal energy conversion
OTFT	organic thin-film transistor
OTM	overturning moment
OWC	oscillating water column
OWSC	oscillating wave surge converter
OWT	offshore wind turbine
OWTT	one-way travel-time

P

P–M	Pierson–Moskowitz
P-C	Pelnard Considère
PBL	planetary boundary layer
PCR	polymerase chain reaction
PDE	partial differential equation
PDF	probability density function
PDMS	polydimethylsiloxane
PGM	probabilistic graphical model
PHINS	photonic inertial navigation system
PISO	pressure implicit with splitting of operators
PIV	particle image velocimetry
PLL	phase-locked loop
PMH	probable maximum hurricane
PMSG	permanent-magnet-excited-synchronous generator
PMT	photomultiplier tube
POL	petroleum, oil and lubricants
POM	Princeton Ocean Model
POMDP	partially observable Markov decision process
POSG	partially observable stochastic games
PPP	point-to-point protocol
PROVERBS	probabilistic design tools for vertical breakwaters
PSD	power spectral density
PSK	phase shift keying
PSS	practical salinity scale
PTC	passive turbulence control
PU	polyurethane
PVC	polyvinyl chloride

Q

QAM	quadrature amplitude modulation
QPSK	quadrature amplitude modulation
QTF	quadratic transfer function

R

RANS	Reynolds-averaged Navier–Stokes equation
RAO	response amplitude operator
RC	Resistor and Capacitor
REF/DIF	refraction/diffraction model
REFLICS	real-time flow imaging and classification
REMUS	remote environmental monitoring units
RF	radio frequency
RHIB	rigid hull inflatable boat
RIB	rigid inflatable boat
RLG	ring laser gyroscope
RLS	recursive least square
RM	red muscle
RMG	rail mounted gantrie
RMMV	remote multi-mission vehicle
RMS	root mean square
RNA	rotor-nacelle assembly
ROMS	Regional Oceanic Modeling System
ROV	remotely operated vehicle
RPS	renewable portfolio standard
RS	Reed–Solomon
RSLR	relative sea level rise
RTG	rubber tire gantrie
RTS	request-to-send

S

SALM	single anchor leg mooring
SAR	synthetic aperture radar
SARUMS	Safety and Regulations for European Unmanned Maritime System
SAT	satellite communication
SAUVIM	semiautonomous underwater vehicle for intervention mission
SBL	short baseline
SCADA	supervisory control and data acquisition system
SCF	standard cubic feet
SCFM	standard cubic feet per minute
SCOUT	surface craft for oceanographic and unmanned testing
SDOF	single-degree-of-freedom
SDV	swimmer delivery vehicle
SEAS	spectrophotometric elemental analysis system
SEC	south equatorial current
SHP	shaft horsepower
SISO	single-input single-output
SLAM	simultaneous localization and mapping
SLR	sea-level rise
SLT	steady-lift technology
SNR	signal-to-noise ratio
SODAR	sonic detection and ranging
SOFAR	sound fixing and ranging
SOM	self-organizing map
SPH	Standard Project Hurricane
SPLINE	self-propelled line array
SPM	single point mooring
SPR	self-referential phase reset
SS	strong suppression
SSC	suspended sediment concentration
SSF	Saybolt furol second
SSP	Sea Solar Power
SST	sea-surface-temperature
SSU	Saybolt universal second
St	Strouhal number
STS	ship-to-ship
STWAVE	steady state spectral wave model
SWA	small waterplane area
SWAN	simulating waves nearshore
SWATH	small waterplane area twin hull

T

TA	total alkalinity
TAT	turn around time
TCB	transverse center of buoyancy
Tcf	trillion cubic feet
TCG	transverse center of gravity
TCXO	temperature compensated crystal oscillator
TDD	time division duplexing
TDHA	total dynamic head available
TDHR	total dynamic head required
TDMA	time division multiple access
TDP	touch down point
TDZ	touch down zone
TEF	transportation efficiency factor
TEU	twenty-foot equivalent unit
TLP	tension leg platform
TOF	time-of-flight
TPH	total petroleum hydrocarbon
TPI	tons per inch immersion
TRL	technology readiness level
TSL	tunnel sea lion
TSR	tip-speed ratio
TWh	terawatt-hours

U

UAV	unmanned aerial vehicle
UBAT	underwater bioluminescence assessment tool
UEL	upper explosive limit
UFC	Unified Facilities Criteria
UGV	unmanned ground vehicles

UHF	ultra high frequency
ULCC	ultra large crude carrier
ULCS	ultra large container ship
ULF	ultralow frequency
UMV	unmanned marine vehicle
UMVS	unmanned maritime vehicle system
UPS	uninterruptable power supply
USBL	ultrashort baseline
USSV	unmanned sea surface vehicle
UUV	unmanned underwater vehicle
UV	ultraviolet

V

VARANS	volume-averaged Reynolds-averaged Navier–Stokes equations
VBAP	virtual body and artificial potential methodology
VCB	vertical center of buoyancy
VCG	vertical center of gravity
VCR	virtual control room
VED	vertical electric dipole
VFD	variable frequency drive
VHE	Vortex Hydro Energy
VHF	very high frequency
vi	virtual instrument
VIM	vortex-induced motion
VIV	vortex-induced vibration
VIVACE	vortex-induced vibrations for aquatic clean energy
VLCC	very large crude carrier
VLFS	very large floating structure
VLSI	Very Large Scale Integration
VMD	vertical magnetic dipole
VO	velocity obstacle
VOC	volatile organic compound
VOF	volume of fluid
VS	Varley-Seymour
VSF	volume scattering function

W

WEC	wave energy conversion
WECA	wave energy conversion actuator
WET	whole effluent toxicity
WIG	wing in ground-effect
WIS	Wave Information Studies
WSS	wide sense stationary
WSSI	wave–seabed–structure interactions
WT	water tracing
WWC	winds, waves, and currents

Z

ZOH	zero order hold

1. Introduction

Manhar R. Dhanak, Nikolaos I. Xiros

This brief introduction is aimed to make the reader aware of the many interesting topics involved in ocean engineering, from basic engineering science to engineering applications, in support of enabling a host of commercial, naval, and recreational activities in the ocean and in the coastal zones. It is not possible to cover all the topics in detail in a single book. Instead, in this Handbook we aim to provide an updated account of key topics in ocean engineering and ocean technologies, including a review of important fundamental and applied subject matter as well as contemporary developments in leading-edge ocean technologies.

1.1 **Enabling Maritime Design and Development** 1
1.2 **History** 2
1.3 **Basics** 2
1.4 **Applications** 4
1.4.1 Automated Unmanned Systems 4
1.4.2 Coastal Design 5
1.4.3 Offshore Systems 6
1.4.4 Ocean Energy 6
1.5 **Future Trends** 7
References 8

Ocean engineering is an engineering discipline that includes elements of multiple traditional engineering disciplines and is specialized to provide the engineer with the required background to effectively undertake engineering projects in the marine environment. It is a study of science, technology, engineering, and mathematics, including ocean sciences, mechanical, electrical, civil and computer engineering, and naval architecture. Professional ocean engineers are involved in the design, development, and operation of ocean systems, technologies, and structures as well as in enabling a host of endeavors in the marine environment. New demands in commerce, national defence, climate change, and renewable energy, coupled with advances in electronics and marine materials drive innovations.

1.1 Enabling Maritime Design and Development

Covering over 70% of the surface area of the earth, the oceans are integral to life on the planet, to its atmosphere, weather and climate, and to carbon and other biogeochemical cycles. Maritime commerce and transportation are vital to the economy and security of many nations and the coasts support significant professional and recreational maritime activities while serving as living interfaces between land and the oceans. Further, the oceans represent rich sources of food, renewable energy, oil, and minerals. Engineering expertise and requisite ocean technologies enable responsible maintenance of the health, productivity and resiliency of the oceans and the coasts, safe navigation and exploration of the oceans, and environmentally sustainable tapping of their resources. Smart and innovative solutions are critical in developing robust and resilient coastal and offshore structures and platforms, in mitigating the impact of storms and global climate change, in managing the coasts and beaches, and in meeting commercial, energy, defense, security, and other needs of nations. With emerging advances in both land and ocean based technologies, and expanding anthropogenic activities in the oceans and coastal waters, the need to understand the oceans and to exercise care and wisdom in conducting these activities has never been greater. A US Commission on Ocean Policy [1.1] suggests that this expansion of activities coupled with the advances in technologies is threatening the quality of the oceans and their marine life. At the same time, the Commission regards the technological advances as also offering hope for us to

change course to achieve a vision of a future in which the world's oceans and coasts are clean, safe, and sustainably managed.

Ocean engineering is a multidisciplinary engineering field aimed at wisely and innovatively solving problems associated with working in the ocean environment, exploring the oceans, and harnessing its resources. It is a study of civil, computer, electrical, and mechanical aspects of engineering in combination with fundamentals of mathematics, physics, material science, naval architecture, and the ocean sciences. Today's professional ocean engineers are involved in the design, development, and operation of commercial and naval ships, submarines, and autonomous marine vehicles and systems, offshore and coastal structures, ocean energy converters, underwater sensor systems and technologies, and in meeting a host of other challenges associated with our endeavors in the ocean.

This Handbook aims to provide an updated account of key topics in ocean engineering and ocean technologies, including a review of important fundamental and applied subject matter as well as contemporary developments in ocean technologies. The intended readers include practitioners in a range of maritime industries as well as academic researchers and students in ocean engineering and areas related to ocean technologies. At the same time, the Handbook aims to serve as an important guide for anyone interested in the ocean and human activities in the maritime environment. The chapters in the Handbook are grouped into five parts. In Part A, the focus is on fundamental topics of ocean engineering; in Part B, it is on topics in marine transport and automation; in Part C the focus is on topics in coastal design; in Part D it is on topics in offshore systems; and in Part E the focus is on renewable ocean energy.

1.2 History

Although engineers have been involved in working in the ocean environment and in design and development of ships and submarines over centuries, ocean engineering as a recognized formal discipline is relatively new, dating back to 1960s. In the United States, concerns were expressed that engineers involved with design and construction of ships, submarines and other ocean systems lacked the necessary experience of the ocean environment. In 1963, the loss of the submarine USS Thresher [1.2] heightened these concerns and led to the development of this new engineering discipline at the undergraduate level; the Merriam-Webster dictionary notes the first known use of the term *ocean engineering* to be in 1964. Whereas *naval architecture* traditionally deals with design and construction of ships and offshore structures, and *marine engineering* deals with the construction and operation of power plants on ships, at docks, and harbors, *ocean engineering* is associated with operations in the ocean environment and deals with the application of engineering principles and techniques of design, construction, and maintenance to working in the ocean environment and to development and use of requisite ocean technologies. Today's ocean systems and operations are complex and require multidisciplinary approaches. Therefore, increasingly these disciplines overlap. *Naval engineering* and *offshore engineering* draw from all of these disciplines in design, construction, and maintenance of naval ships and submarines, and offshore platforms, respectively.

1.3 Basics

Besides mathematics, physics, chemistry, and oceanography, a typical core ocean engineering curriculum at an undergraduate level covers the topics of statics, dynamics, thermodynamics, heat transfer, graphics, electronics, programming, probability and statistics, data analysis, hydromechanics, wave mechanics, materials, structures, underwater acoustics, and control and dynamics systems. The curriculum culminates in design of ocean systems. The topics covered aim to develop an understanding of the ocean environment and its resources as well as fundamentals of engineering, and provide the knowledge, the principles, and techniques in:

1. Design, construction, maintenance, and survivability of ships, submarines and ocean structures
2. Understanding natural and anthropogenic signatures in the ocean
3. Performance enhancement of ocean systems
4. Control and automation and
5. System integration.

The coverage of the ocean environment encompasses characteristics of seawater and geochemical properties, and such physical processes as surface and internal waves, ocean and tidal currents, thermal gradients, and air–sea interaction. The engineering topics

aim to provide skills for determining stability and hydrodynamics of ships and submarines; for evaluating strength and integrity of marine materials and structures; for protecting ocean systems against corrosion and biofouling; for characterizing generation, propagation, and reception of underwater sound; for processing and analyzing ocean data; for developing algorithms for control and automation of marine vehicles; and for designing, developing, testing, and demonstrating ocean systems.

An understanding of the physics of oceanographic processes enables modeling and prediction of at-sea conditions and their variability. It develops an awareness of various phenomena in the water column, including such processes as the thermocline, and wind-driven and geostrophic currents. Winds, waves, and currents with random characteristics lead to dynamic *sea states* that range from states corresponding to routine operational conditions to those associated with extreme events. The forces associated with such meteorological and oceanic (metocean) conditions are correspondingly dynamic, random, and range from routine to extreme. They have to be accounted for in designing robust ocean and coastal structures and in planning and conducting offshore operations. Oceanographic and metocean processes are described in Chaps. 2–4.

Subsurface pressure, temperature, and salinity are physical properties of some importance in the deep ocean environment. Pressure ordinarily increases linearly with water depth, but is also a function of temperature and salinity. Submarines and housings for electronics of deep-water systems have to be designed so that they can withstand such pressures and require specially designed gaskets to prevent high-pressure water leaking into compartments of the systems. It is believed that Woods Hole Oceanographic Institution's unmanned deep-sea research submarine Nereus was lost in 2014 off New Zealand at a depth of nearly 10 km, where the pressure exceeds 1000 times the atmospheric pressure, due to an implosion of one of its components. While the temperature in the ocean below a mixed layer drops rapidly in a thermocline, it typically does not drop below approximately 2 °C. Properties of seawater are described in detail in Chap. 5.

Damage due to marine corrosion that arises through contact with seawater or through exposure to the atmosphere in coastal areas continues to be a major problem; this includes corrosion in engines operating at sea or exposed to salt-laden air. It typically accounts for 30% of failures on ships and other marine equipment. A World Corrosion Organization report [1.3] estimates that the annual cost of damage due to marine corrosion worldwide is over $1.8 trillion. Corrosion destroys materials through chemical reaction with its environment and the rate at which it impacts structures depends on the type of metal or metal alloys, and the design of the structures as well as the environmental conditions; presence of microbial organisms; and processes such as cavitation that can damage material surfaces [1.4]. Corrosion is typically controlled through use of coatings, cathodic protection, use of inhibitors, chemical dosing of the local environment, or use of less corrosive materials, including composites, in structures. Properties of marine corrosion are discussed in Chap. 6.

Maximizing the operational efficiency of ships and submarines continues to be an important goal in marine transportation. In hydromechanics, fundamentals of drag, lift, and propulsion as well as such features as turbulence, boundary layers, jets, shear layers and wake resistance, and vortex-induced vibrations and galloping carry over from aerodynamics with the difference that the density of seawater is over 800 times greater than that of air. However, phenomena such as cavitation and bubble generation, and wave-induced forces on vehicles and offshore structures require special consideration in hydromechanics. In addition, for vehicles operating on or near the free surface, an additional contribution to vehicle drag or resistance arises that is associated with waves generated by the vehicle. Correspondingly, for vehicles operating on or near the free surface, in addition to Reynolds number, a dimensionless parameter, the Froude number, $\mathrm{Fr} = U/\sqrt{gl}$, where U and l are, respectively, the characteristic speed and length, and g is the acceleration due to gravity, governs the flow characteristics. The wave-making resistance of ships is typically characterized by the Froude number. Hydromechanics is discussed in Chap. 7.

Acoustics, electromagnetics, and optics are important sensing mechanisms in underwater operations and have led to significant efforts in development of associated sensor and sensor platform technologies and operations. Underwater sound is generated through vibrational activities as a pressure pulse in the water that propagates at speeds in the range of 1400–1600 m/s at frequencies typically in the band 1 Hz–1 MHz with corresponding wavelengths in the range of 1.5 km–1.5 mm. The sound undergoes transmission losses through absorption and scattering through refraction, reflection, and destructive interference of sound waves. The speed of sound decreases with decrease in temperature and increases with depth. This gives rise to existence of a minimum at mid-depth and the variability in sound speed with depth leads to bending or refraction of sound waves toward region of lower speed. A sound wave generated in the thermocline bends downward at first toward region of lower speed and then upward as the speed increases with depth. This leads to a particularly interesting phenomenon called the SOFAR (sound

fixing and ranging) channel, existing at 600–1200 m depth in mid-latitudes, through which sound travels long distances without attenuation [1.5]. Distances traveled by sound waves satisfy the inverse-square law and decrease with increase in frequency – low-frequency waves may travel tens of thousands of kilometers without significant attenuation, while very high frequency waves typically penetrate distances of order 1 m. Hydrophones measure pressure fluctuations induced by sound in μPa (micro Pascal), and the amplitude or the loudness of the underwater sound is measured in decibel with reference to a standard pressure level at a standard distance, written dB re 1 μPa at 1 m. As a reference, blue whales vocalize at 10–40 Hz at source sound levels of 155–188 dB re 1 μPa at 1 m [1.6, 7], whereas large ships and fast-moving small boats can produce broadband (20–1000 Hz) sound levels with source levels of 150–200 dB re 1 μPa at 1 m. An excellent description of underwater acoustics is provided in a sister handbook [1.5].

Since seawater is a conducting fluid, with conductivity in the range 2.5–6 S/m, electric fields in the ocean maybe generated through induction by time varying external fields or by motion of the seawater through the Earth's magnetic field. Strong temporal variations in the magnetic fields in the ionosphere and the magnetosphere induce electric fields in the ocean and generate secondary magnetic fields. Further, the dynamo effect of currents in the ocean, involving motion of conducting seawater through the Earth's magnetic field, induce electric fields, and, in turn, give rise to secondary magnetic fields in the water column [1.8]. Whereas the electromagnetic waves in air propagate at the speed of light (3×10^8 m/s), in the ocean the speed depends on the wave frequency. At 1 Hz, it is approximately 1600 m/s, several orders of magnitude lower than in air, while at 10 kHz, it is 100 times faster. The waves undergo transmission losses, the rate of attenuation increasing with frequency; a 10 kHz electromagnetic wave is attenuated at a rate over 80 times faster than a 1 Hz wave. Finally, at frequencies of order 10^{14} s^{-1}, with corresponding wavelengths of 400–700 nm (in air), is the visible part of the electromagnetic spectrum. Light waves propagate at the speed of 2.24×10^8 m/s in seawater compared with 3×10^8 m/s in air. The attenuation of light is wavelength-selective, leading underwater objects having bluish or greenish tints [1.9]. Ocean electromagnetics are discussed in Chap. 8.

Ocean signals are typically noisy and have to be processed to discern the underlying signatures that enable detection, identification, and location of objects underwater. Typical processing is in terms of spectral analysis of time series of signals, using for example Fast Fourier transforms. Sophisticated algorithms are required to extract underlying signatures from excessively noisy signals. The science of signal processing enables development of the algorithms that facilitate such extraction. It is discussed in Chap. 9.

In the past decade, significant effort has gone into shipboard automation and development of unmanned ocean systems. Underlying this effort are major advances in microelectronics and computer technologies. Unmanned systems typically involve sensors and actuators that may be linked via a computer that acquires and stores data from the sensors, formulates a response using onboard control algorithms and sends out an actuation control signal, in support of achieving a desired state. The error between the outcome of the actuation, determined via an appropriate sensor measurement, and the desired state is continuously monitored in a feedback loop and used to improve the control signal in an iterative process. A simple proportional-integral-derivative (PID) controller is based on using the present values of the error, accumulated past values of the error, and predicted future values of the error in developing an improved actuation control signal. The control theories and case studies illustrating the theories are discussed in Chap. 10.

1.4 Applications

Human activities offshore and in coastal regions span across a range of sectors, including shipping and maritime transport, offshore energy, security and defense, development of ports, harbors and other coastal structures, fisheries and aquaculture, recreational activities and activities related to mitigating environmental impacts of these activities. Planning, design, and conduct of these activities as well as development of enabling technologies involve applications of ocean engineering at various levels through the maritime industry. In Parts B through E, we cover four major areas of applications of ocean engineering in the maritime domain: automated unmanned systems, coastal design and structures, offshore platforms, and offshore renewable energy.

1.4.1 Automated Unmanned Systems

Automated unmanned systems are systems that are pre-programmed to carry out desired tasks. Taking advantage of the advances in electronic and computer technologies, significant strides are being made in the

level of autonomy as well as in the diversity of applications in commercial activities in the maritime domain. The applications range from automated port container terminals with automated guided vehicles (AGVs), stacking cranes, gantry systems, and other automated systems that are significantly revolutionizing port and container shipment operations, to autonomous underwater and surface vehicles (AUVs and USVs) for commercial and military applications, including for hydrographic surveys, underwater pipeline inspections, surveillance, asset protection, and mine-counter measures. Typically several subsystems integrated together make up an unmanned automated system. Design and development of such systems utilize many of the basic elements of ocean engineering. In the case of an AUV, while the size of the vehicle may be determined by the payload requirements, the shape of its hull and the designs of its propulsion and control surfaces, for stable and efficient motion and maneuverability, are determined through application of the principles of hydromechanics. Selection of materials and structures for operations at given water depths and for durability and reduced maintenance is based on principles of materials engineering and on corrosion science, as well as on expected hydrodynamic loads on the structures. Batteries and powering systems are selected for long endurance and together with other electronic elements require appropriately designed cooling systems. Designs of vehicle navigation, obstacle avoidance, and underwater communication systems are based on application of the principles of underwater acoustics as well as optics. Typically, an AUV may act as a mobile sensor platform carrying a range of mission-based acoustic and nonacoustic sensors. The acoustic sensors operate over a range of mission-specific frequencies that encompass human hearing range (20–20 000 Hz) and beyond. They include the well-known sidescan sonar for detection, classification and location of targets in the water column, and high-frequency (kHz–MHz range) sonar for detection of buried objects, and for imaging subsurface objects. Nonacoustic sensors include electromagnetic sensors, with important applications in the areas of geophysical surveys and searches of the seafloor and sub-bottom, communication across the sea–air boundary, and high data transfer rate at short ranges; optical sensors, including flashing light-emitting diode (LED) and laser, for sensing, detection and communication; electrochemical sensors for environmental monitoring [1.10]; and other bio-geo-chemical sensors with a host of applications in marine science and environmental monitoring. States of the art in key areas of autonomous marine vehicles are presented in Part B.

1.4.2 Coastal Design

Coastlines are interfaces between the challenging ocean environment and densely populated coastal areas and associated human activities. It is estimated that over 1.2 billion people worldwide live within 100 km of the coast [1.11]. Significant effort is therefore involved in taking measures to protect the coastline from storm surges, winds, waves, flooding, and erosion as well as in accommodating, sometime conflicting, needs of the coastal population and associated human activities, while maintaining a healthy coastal environment. Shore protection against beach erosion and storm damage, which fall under coastal management, involves construction of hard structures such as sea walls, breakwaters and revetments as coastal armor, and groins as sand-trapping devices, as well as methods for periodic beach nourishment and stabilization as measures for mitigation of beach erosion. Structures that accommodate human needs include ports and harbors, piers, and marine outfalls. Designing robust coastal structures and measures for maintaining a healthy shoreline, as well as assessing risks of damage due to coastal hazards, involve estimating and predicting the forces associated with coastal waves and extreme storms through modeling and simulation. Typically, the structures are designed to withstand a 100-year or 50-year wave, which are statistical projections of wave heights that on average would be exceeded once in 100 or 50 years, respectively, and are based on past observations. A harbor is defined as a protected place that offers safety to ships, whereas a port is defined as harbor with terminal facilities that accommodate intermodal transportation and storage of cargo, in support of commerce. The amount of cargo throughput and its cost-effective handling at a port are based on the number and size of the berths, storage capacity, cargo handling equipment and the size of ships that it can accommodate, as well as the impact of local environmental conditions [1.12]. Demands posed by new super-large cargo ships and move toward automation, in terms of channel depth and berth space and cargo handling, have provided new worldwide impetus to the process of planning and design of ports and harbors. Safety and efficiency of operations at a port are significantly based on the design of its approach channel, its berthing, anchoring and mooring systems, turning basins, and its terminal facilities. These and other considerations involved in coastal design and development, including, in quantification of the physical coastal environment, in practice of beach nourishment, in design of structures for protection against coastal hazards, and in design of ports, harbors, and marine outfalls are discussed in Part C, which is dedicated to Dr. Robert G. Dean in

recognition of his significant contributions to coastal engineering.

1.4.3 Offshore Systems

Offshore platforms have traditionally been critical in extracting, processing, and temporarily storing offshore oil and gas. The platforms can be fixed, bottom-supported structures that extend to the bottom or floating structures that are moored to the bottom. Design and construction of a robust offshore platform, buoy or other offshore system, as well as offshore operations in open water pose significant challenges, depending on the water depth, the local hydrodynamic conditions, and payload requirements. Major design considerations for an offshore platform includes the unsteady hydrodynamic loads that the platform would likely experience due to waves, winds, currents, and turbulent eddies, and the interaction of the platform with the flow, choice of materials and the design life of the system. The type of platform and its mooring and/or other support systems, as well as the condition of its foundation, characterize the fluid–structure interactions and the dynamic responses of the structure and its floating subsystems [1.13]. Consideration is also given to the damage and degradation resulting from corrosion, fatigue, biofouling, and wear and tear that play a significant role in determining the frequency of requirement for duty-cycle maintenance and the design life of the platform. Important elements of offshore systems and operations are cables or tethers for mooring platforms, for towing systems through the water, for securing instrument packages, and for meeting a host of other requirements. The cables or tethers in energetic ocean environments are typically subjected to random dynamic tension forces and the failure of a cable or tether can have costly consequences, possibly resulting in loss of a platform, instrument, or other system. Good estimates of these forces are required in designing and in selecting material for a marine cable or tether.

Large mobile offshore systems include floating production, storage, and offloading (FPSO) vessels used for production and processing of hydrocarbons, high-capacity crane vessels for offshore construction, commercial and naval ships for transport of goods and people across the oceans, and other purpose-built vessels for transporting large specialized subsystems. Among the latter category are container ships, bulk carriers, super tankers, liquefied natural gas (LNG) carriers, ocean liners, cruise ships, aircraft carriers and battleships. These vessels, which can be over 300 m long, call for special engineering requirements in terms of maritime technology and operations. For example, a LNG carrier, which is specifically discussed in Chap. 41 in Part D, is used for transporting flammable liquefied natural gas at cryogenic temperature of $-163\,^{\circ}\mathrm{C}$ and is designed to provide the necessary thermal isolation and operational safety.

Major offshore operations, such as oil-spill remediation and salvage of ships and submarines, require significant planning and preparation and merit special considerations. Major oil spills, such as in the Gulf of Mexico in 2010, pose significant environmental hazards and typically call for rapid responses. Complex physical, biological, and chemical weathering processes govern the fate of an oil spill in water, including interaction with suspended sediments, and the processes being dependent on the type of oil and prevailing local environmental conditions. As a result, following a major spill, there is large uncertainty in the portion of the oil that is at the water surface, where it may be subjected to significant evaporation, and the portions that remain suspended within the water column, at what depths, and in what form, as well as how much gets deposited on the seabed [1.14]. The state of the art in engineering expertise and technology critically governs the effectiveness and timeliness of the response to an oil spill, from detection of the spill, to identification of source location(s), to characterization of the extent of the spill, to surveillance and monitoring, to intervention, mitigation, and remediation. Salvage of ships and submarines lost at sea also entails major broad-based, interdisciplinary engineering operations. Salvage engineers apply the principles of naval architecture and ocean engineering in assessing the strength and stability of damaged vessels and in recovering them from the sea. Working knowledge of basic ocean engineering topics, maritime safety engineering, and the theory and practice associated with rigging offshore systems are essential for a practicing salvage engineer. Oil-spill remediation and salvage operations are considered in two major chapters in Part D.

1.4.4 Ocean Energy

Worldwide theoretical potential of ocean energy (in the forms of offshore wind, wave, ocean and tidal currents, and thermal and salinity gradients) as a renewable source for electricity is striking. It is estimated that this potential ranges from 20 000 to 80000 terawatt-hours (TWh) of electricity annually, which is 100–400% of current global demand [1.15]. Actual resources that can be harvested cost-effectively are currently much smaller. Ocean energy development is significantly behind in technical maturity, compared with other renewables such as onshore wind and photovoltaic solar. This is due to the technical, socioeconomic, environ-

mental, regulatory, and infrastructural challenges posed by the harsh, energetic ocean environment. The levelized cost of energy (LCOE) for ocean energy is currently uncertain or at best much higher than that of fossil fuels and other renewables [1.15] because capacity factor and design life of offshore systems, which are key drivers of LCOE, are generally not well understood. However, as has been demonstrated in the case of development of onshore wind and solar energy [1.16], the rate at which the challenges to ocean energy development are overcome and costs reduced will be driven by investment as well as by a concerted effort in learning, innovation, and deployment of prototypes that enables informed decision-making. The good news is that a number of studies, including proof of concepts, and computational and laboratory investigations, have been conducted [1.17, 18] and various breakthrough energy conversion devices have been designed and built, with over 150 global patents in ocean energy technologies filed annually between 2009 and 2013 [1.15]. Industry investment in ocean energy, having lagged in view of the risks involved, is being spurred in Europe and elsewhere with the help of government subsidies. Government regulators in Europe and the United States are increasingly looking to develop ecosystem-based coastal and marine spatial planning (MSP [1.19]), a process for making informed and coordinated decisions for multiple maritime activities while conserving biodiversity in the coastal environment, in support of reducing conflicts between commercial-scale development of ocean energy and other competing maritime activities. MSP aids in identifying a policy framework for siting, permitting, and developing ocean energy [1.20, 21]. MSP aims to site ocean energy development at ocean energy *hotspots* that have the lowest potential conflict with other maritime activities, such as fishing, shipping, and whale watching, for example, as well as with lowest impact on biodiversity. Topics in various forms of ocean energy, in terms of resource characterization, and requisite technologies for harvesting the energy are covered in six chapters in Part E.

1.5 Future Trends

Emerging new demands in commerce, national defense, and energy requirements on the one hand, and advances in electronics, computer chip industry, and marine materials on the other represent significant new challenges and great opportunities for practicing ocean engineers and educators in developing and implementing new ocean technologies and educating and training the next generation engineer workforce. Ocean engineering topics of interest in the 21st Century include:

- *Ocean exploration*: Vast regions of the oceans remain unexplored and a new era of ocean exploration, particularly in the Arctic is expected. Based on current progress, it will result in development and use of custom smart ships, submersibles, remotely operated vehicles, AUVs, and gliders that will carry smart sensor systems for navigation and ocean observation. AUVs and USVs will be used routinely with unmanned aerial vehicles (UAVs), and underwater robotics will become increasingly sophisticated.
- *Shipboard automation*: The drive to reduce manning onboard ships through automation has been in progress for several years and advances are being made utilizing developments in power electronics. It will result in highly automated, all electric ships that are highly reconfigurable for robust operation, and that will use new, efficient, low emission engines and fuels [1.22].
- *Coastline security technologies*: The new threats due to terrorism from the seas call for greater surveillance of our harbors, ports, and coastline and call for development of new sensor systems and small, automated sensor platforms.
- *Coastal structures*: Coastal hazards, compounded by the new significant threats associated with sea-level rise as well as rising populations in low-elevation coastal zones [1.23], will continue to require innovative engineering solutions for coastal structures and shoreline management, in support of sustainable management of coastal zones.
- *Offshore platforms*: Oil and gas exploration and extraction is being extended to deep waters. This will present new challenges in construction, implementation, and operation and maintenance of platforms far from shore. New concepts in multiuse offshore platforms that combine, for example, energy extraction, aquaculture, and platform-related transport are being explored [1.24]. Such platforms facilitate effective marine/ocean spatial planning as well as consolidate various offshore activities, and can be green platforms, benefiting from local renewable ocean energy.
- *Renewable ocean energy*: As discussed above and in Part E, reducing the cost of harnessing ocean energy in the forms of offshore wind, waves, ocean and tidal currents, and thermal gradients in a sustainable manner is a key factor in commercial-scale de-

velopment of ocean energy. Innovative engineering methods and ocean technologies will be required to achieve the required solutions.

- *Marine materials*: Lighter, stronger ship hulls, and ocean structures made of composites that exploit developments in nanotechnology show promise. At the same time, protection against corrosion and biofouling using safe, durable coatings and materials continue to drive new research. These materials and coatings will reduce duty cycle maintenance of maritime systems. Further, there is renewed interest in utilizing surface chemistry of nano- or microtextured polymer coatings for skin-friction drag reduction and improved hydrodynamic performance of ships and submarines; for streamlined vehicles, such as ships and submarines, skin-friction typically accounts for over 50% of the total drag so that its reduction provides a significant boost to the propulsive efficiency of the vehicle and can result in fuel saving or increased range.
- *At-sea operations*: The necessary science and technology for operating in high seas continues to present both significant ocean engineering challenges and opportunities for innovation in ship design, cargo transfer, sea-keeping, and hydrodynamic performance in high sea-states.

Educators are challenged to respond to future trends and improve ocean-related science and engineering education, in support of developing an educated engineering workforce with requisite knowledge of the ocean environment. Ocean engineering will continue to offer good career opportunities to young people, however challenging.

References

1.1 US Commission on Ocean Policy: *An Ocean Blueprint for the 21st Century. Final Report* (COP, Washington 2004)

1.2 J. Bentley: *The Thresher Disaster; The Most Tragic Dive in Submarine History* (Doubleday, Garden City 1975)

1.3 G. Schmitt: *Global Needs for Knowledge Dissemination, Research, and Development in Materials Deterioration and Corrosion Control* (WCO, New York 2009)

1.4 K.A. Chandler: *Marine and Offshore Corrosion* (Butterworths, London 1985)

1.5 W. Kuperman, P. Roux: Underwater acoustics. In: *Springer Handbook of Acoustics*, ed. by T.D. Rossing (Springer, Berlin, Heidelberg 2007) pp. 149–201

1.6 W.C. Cummings, P.O. Thompson: Underwater sounds from the blue whale Balaenoptera musculus, J. Acoust. Soc. Am. **50**(4), 1193–1198 (1971)

1.7 W.J. Richardson, C.R. Greene, C.I. Malme, D.H. Thomson: *Marine Mammals and Noise* (Academic Press, San Diego 1995)

1.8 S. Maus: Ocean, Electromagnetic Effects. In: *Encyclopedia of Geomagnetism and Paleomagnetism*, ed. by D. Gubbins, E. Herrero-Bervera (Springer, Dordrecht 2007) pp. 740–742

1.9 J.R. Apel: *Principles of Ocean Physics* (Academic Press, San Diego 1987)

1.10 T.S. Moore, K.M. Mullaugh, R.R. Holyoke, A.S. Madison, M. Yücel, G.W. Luther: Marine chemical technology and sensors for marine waters: Potentials and limits, Ann. Rev. Mar. Sci. **1**, 91–115 (2009)

1.11 C. Small, R.J. Nicholls: A Global analysis of human settlement in coastal zones, J. Coast. Res. **19**(3), 584–599 (2003)

1.12 G.P. Tsinker: Port (Harbor) elements: Design principles and considerations. In: *Handbook of Port and Harbor Engineering*, ed. by G.P. Tsinker (Springer, Berlin, Heidelberg 1997) pp. 69–241

1.13 N. Haritos: Introduction to the analysis and design of offshore structures – An overview, Electron. J. Struct. Eng. **7**, 55–65 (2007)

1.14 National Research Council: *Oil in the Sea III: Inputs, Fates, and Effect* (National Acad. Press, Washington 2003)

1.15 I.R.E. Agency: *Ocean Energy – Technology Readiness, Patents, Deployment Status and Outlook* (IRENA, Paris 2014), http://www.irena.org/DocumentDownloads/Publications/IRENA_Ocean_Energy_report_2014.pdf

1.16 International Energy Agency: *Technology Roadmap for Wind energy* https://www.iea.org/publications/freepublications/publication/Wind_2013_Roadmap.pdf (2013)

1.17 Ocean Energy Systems: IES-OES Annual Report http://www.ocean-energy-systems.org/documents/82577_oes_annual_report_2013.pdf/ (2013)

1.18 NREL (National Renewable Energy Laboratory): Ocean energy technology overview, http://www.nrel.gov/docs/fy09osti/44200.pdf (2009)

1.19 K. McLeod, J. Lubchenco, S. Palumbi, A.A. Rosenberg: *Scientific Consensus Statement on Marine Ecosystem-Based Management* (Communication Partnership for Science and the Sea, Washington 2005)

1.20 A. Howell, C. Drake: *Scoping Study on Socio-Economic Impacts of Tidal Energy Development in Nova Scotia: A Research Synthesis and Priorities for Future Action*, Technical Report #2012-01 (Fundy Energy Research Network, Wolfville 2012)

1.21 M. Portman: Involving the public in the impact assessment of offshore renewable energy facilities, Mar. Policy **33**(2), 332–338 (2009)

1.22 M. Triantafyllou: *Science and Technology Challenges and Potential Game-Changing Opportunities*, Trans-

portation Research Board Special Report 306: Naval Engineering in the 21st Century (The Science and Technology Foundation for Future Naval Fleets, Washington 2011)

1.23 B. Neumann, A.T. Vafeidis, J. Zimmermann, R.J. Nicholls: Future coastal population growth and exposure to sea-level rise and coastal flooding-a global assessment, PLOS One **10**(3), e0118571 (2015)

1.24 European Commission: Summary Report: MERMAID (Innovative Multi-purpose off-shore platforms: Planning, Design and operation) Project, Project 288710 (2014) http://cordis.europa.eu/result/rcn/59211_en.html

Part A Fundamentals

Ed. by Manhar R. Dhanak, Cortis K. Cooper

2 **Elements of Physical Oceanography**
Wendell S. Brown, New Bedford, USA

3 **Metocean Extreme and Operating Conditions**
George Z. Forristall, Camden, USA
Cortis K. Cooper, San Ramon, USA

4 **Mechanics of Ocean Waves**
James M. Kaihatu, College Station, USA
Palaniswamy Ananthakrishnan, Boca Raton, USA

5 **Physical Properties of Seawater**
Wendell S. Brown, New Bedford, USA

6 **Principles of Marine Corrosion**
Robert E. Melchers, Callaghan, Australia

7 **Hydromechanics**
Karl Dietrich von Ellenrieder, Dania Beach, USA
Manhar R. Dhanak, Dania Beach, USA

8 **Ocean Electromagnetics**
John J. Holmes, West Bethesda, USA

9 **Digital Signal Processing**
Nikolaos I. Xiros, New Orleans, USA

10 **Control Theory and Applications**
Nikolaos I. Xiros, New Orleans, USA
Pak-Cheung Edgar An, Boca Raton, USA

In Part A, a range of fundamental topics in ocean engineering that underlie practical engineering applications in the marine environment are covered in nine chapters, encompassing physical aspects of oceanography and marine meteorology, seawater properties, hydrodynamics, marine corrosion, marine electromagnetics, signal processing and control theory. These topics form part of a core ocean engineering curriculum aimed at providing an understanding of the marine environment together with basic concepts of engineering. These include design, development, and operation of ocean and coastal structures, marine vehicles, offshore platforms, ocean renewable energy in its various forms, and a host of other applications in support of commercial, naval, and recreational activities in the ocean and in the coastal zones, and in accommodating sustainable coastal communities and coastal management.

In Chap. 2, elements of physical oceanography, including surface and internal gravity waves, inertial motion and hurricane-driven storm surge are described. An outline of the wind generated surface waves, from generation to dissipation is provided. The roles of turbulence-derived oceanic friction in Ekman dynamics/upwelling-downwelling are described. The tides and geostrophic flow in wind- and thermohaline-driven circulations are discussed.

In Chap. 3 the ocean and atmosphere (metocean) conditions that govern the design and operation of man-made structures, facilities, and vessels are described. Metocean environments that allow routine operations and extreme *storm-related* environments are considered. Methods commonly used in the offshore oil and gas industry to quantify the most important metocean variables, associated with winds, waves, and currents that impact offshore facilities are outlined. Applications of the methods include the design and operation of vessels, coastal structures, offshore wind farms, navigational aids, coastal geomorphology, and pollution studies.

In Chap. 4, the mechanics of water waves and wave-body interactions pertaining to ocean and coastal engineering based on linear and weakly-nonlinear wave theories are reviewed. Numerical methods based on Green's theorem and mixed Eulerian–Lagrangian formulation for fully nonlinear wave and wave-body interaction problems are discussed, together with methods to determine the wave forces on fixed and floating structures, including the viscous drag force.

In Chap. 5, the principal physical properties of seawater, including pressure, temperature, salinity, density, density anomaly, specific heat, and thermal expansion are described. The typical global ocean profiles and distribution maps of surface temperature, salinity, and density anomaly are provided. Other oceanic properties, including stability of oceanic water columns, use of temperature-salinity diagrams in identifying water masses, seawater freezing, as well as oceanic sound transmission and light are described.

In Chap. 6, the corrosion processes due to exposure to marine environments over extended periods are reviewed. Mathematical models for the prediction of corrosion and pitting of structural steels are described. The effects of various influencing factors are considered, including that of seawater quality and the effect of microorganisms on corrosion. Corrosion of various materials, including stainless steels, aluminum, copper-nickels, and steel reinforcing bars in reinforced concrete is considered.

In Chap. 7, a reference source for concepts in hydromechanics is provided with the aim of providing the working knowledge for solving practical ocean engineering problems. A range of topics is covered, including dimensional analysis, static and dynamic flows, potential and viscous flows, laminar and turbulent flows, boundary layers, wakes, jets and shear layers, and drag and lift forces.

In Chap. 8, properties of electric and magnetic fields within the electrically conducting ocean are described. The mathematical descriptions, based on Maxwell's equations, of electromagnetic fields and dipole sources within a conducting media are developed. The differences between plane wave reflection and transmission at the surface of freshwater and seawater are used to highlight how electromagnetic propagation within the electrically conducting ocean is so very different to the more familiar radio frequency transmissions in air. A brief discussion of ocean electromagnetics at optical wavelengths is also provided.

In Chap. 9, an overview of digital signal processing is provided. The concept of discrete-time signals and sampled-data systems implemented on digital hardware vs. those of practical continuous-time signals driving analog systems and processes is discussed. The discrete Fourier transform and the Z-transform are introduced as tools for analysis of time series of data. A brief account of digital filter structures and types is provided. Random signals and stochastic processes are considered, together with concepts of optimal signal estimation.

In Chap. 10, the basics of control theory are described and its application to ocean engineering is illustrated by specific examples. Fundamentals of systems science and theory are discussed. Stability and controllability of linear, time invariant systems are presented.

Bode plots and the root locus technique are discussed. For single-input, single-output systems, proportional-integral-derivative (PID) control is introduced and pole placement design is discussed. A brief presentation of digital controller implementations is given. Applications from ocean engineering include control of autonomous underwater vehicles and autopilots for surface vessels.

2. Elements of Physical Oceanography

Wendell S. Brown

This chapter discusses the most important elements of physical oceanography in terms of the various oceanic responses to atmospheric and astronomical forcing – as mediated by earth rotation, including surface, and internal gravity waves; inertial motion and hurricane-driven storm surge. We include an outline of the history of wind-generated surface waves: generation to dissipation on the beaches. The roles of turbulence-derived oceanic friction in Ekman dynamics/upwelling-downwelling are outlined. There is a discussion of the elements of practical tidal sea level prediction. There is a review of the central role of geostrophic flow in wind- and thermohaline-driven circulations on ocean basin as well as continental shelf scales.

2.1 **Atmospheric Processes** 16
2.2 **Ocean Structure** 17
2.3 **Oceanic Processes** 18
2.4 **Surface Gravity Waves** 19
2.4.1 Surface Gravity Waves 19
2.4.2 Definitions 20
2.4.3 Wave Generation and Evolution 25
2.5 **Wind-Forced Ocean Processes** 28
2.5.1 Frictional Effects 28
2.5.2 Earth Rotation Effects 29
2.5.3 Hurricane Wind-Forced Ocean Response 29
2.5.4 Wind-Forced Upper Ocean Response with Earth Rotation Effects 33
2.5.5 Wind-Driven Currents: Ocean Basin Scale 34
2.5.6 Gulf Stream Ring Flows 37
2.6 **Deep Ocean Currents** 38
2.7 **Coastal Ocean Currents** 39
2.8 **Ocean Surface Tides** 40
2.9 **Oceanic Internal Waves and Tides** 43
References 45

The ocean environment consists not only of the water column, but also the atmosphere above the water's surface and the seafloor and sediment and rocks of the subseafloor. A comprehensive study of the marine environment is enormous in scope and is thus often subdivided into disciplinary components, including biology, chemical, geology, physics, and meteorology. This chapter will focus on basic introductory information concerning aspects of the marine environment; however, it will also refer oceanographic literature for those seeking more comprehensive treatments.

Different aspects of ocean engineering design are influenced to varying degrees by one or more environmental factors. These factors comprise one of the several groups of the so-called external *design* constraints; or constraints that are independent of or external to the purpose on which the design is based. For example, submersibles encounter temperature, salinity, pressure, and density gradients as they move through the water column, irrespective of their individual missions. These *external constraints* may be subdivided into those which are fixed by mission requirements and those which are variable. For example, the design *maximum operating depth* of a submersible sets the maximum pressure that the submersible can encounter. However, this operating constraint is not usually linked to specific values of sea water temperature, salinity, and density which can be encountered; but rather only ranges of these variables.

While discussing different oceanic environmental factors, it is convenient to consider them in terms of the three basic regimes in a vertical profile that includes the atmosphere and the water column.

2.1 Atmospheric Processes

An instantaneous snapshot of the winds of the entire atmosphere would present an extremely chaotic view of the flow. Atmospheric circulations can be characterized as an enormously complex patterns of *eddies within eddies*. Thus, it is convenient to categorize the different features of atmospheric circulations according to their different sizes – or scales of motion.

The very small-scale eddies or whirls with time scales of only a few seconds can be eliminated by averaging the measured winds over periods of several minutes. Furthermore, observations are so widely spaced in time and area that most must be considered averages over horizontal distances of tens of kilometers, and vertical distances of tens of meters. Thus, even such relatively large circulation phenomena as thunderstorms and tornadoes often cannot be resolved by the usual weather station network of observations. Through this kind of averaging the range of atmospheric motions can be classified in terms of three general scales; microscale, mesoscale, and synoptic scales (Fig. 2.1).

On the microscale, we find small, short-lived eddies (or turbulence), that are strongly affected by local conditions of both surface roughness and temperature. These eddies are very significant as dispersers of pollutants. At the large end of the microscale are tornadoes and waterspouts. Because microscale motions have such short time scales, Earth rotation is not a significant factor.

Mesoscale phenomena of intermediate horizontal size include such processes as land–sea breezes, thunderstorms, and squall lines. Earth rotation can be an important factor affecting the larger mesoscale phenomena such as sea breezes.

Synoptic scale phenomena most notably represented by large-scale, low-pressure cyclones and high-pressure anticyclones that control our day-to-day weather changes – particularly at mid-latitudes. Historically these synoptic-scale features have been defined every 3 h by the large-scale weather network of simultaneous observations. Earth rotation-related forces are very important factors in the physics of synoptic-scale systems, which persist for days to weeks. Synoptic-scale features that persist for weeks to months tend to have planetary space scales and like jet stream loops play an important role in determining the seasonal characteristics of the weather.

The vertical displacements distinguish these different scales of atmospheric motion. For example, synoptic scale phenomena are dominated by horizontal motions; with vertical velocities generally only 1 to 2 cm/s. Even the great cyclonic storms in the middle and high latitudes, have vertical velocities of the order of 50 cm/s. In the smaller, more intense mesoscale circulations, horizontal and vertical velocities are more comparable; with thunderstorm vertical velocities are often of the order of 10 m/s. Microscale or turbulent phenomena, with nearly equal horizontal and vertical velocities, occur principally in a rather shallow layer adjacent to the Earth's surface and thus are important factors in air–sea interactions. A detailed description of the maritime atmosphere is provided in Chap. 3.

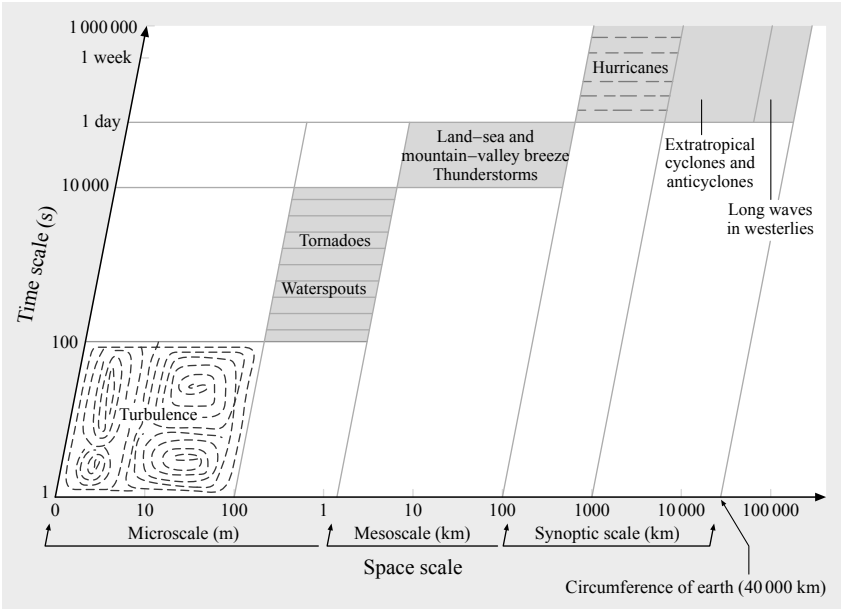


Fig. 2.1 Horizontal space and time scales of atmospheric circulations (after [2.1])

2.2 Ocean Structure

The global atmosphere and the ocean interact in ways that structure the ocean into fairly distinct *upper ocean layer*, which extends down to about 1000 m (3284 ft), and *deep ocean layer*, which extends to the bottom. On one hand, air–sea interactions create vast amounts of sinking cold, dense water in the polar oceans north and south of 40° latitude. On the other hand, in the subtropical and tropical regions between 40°S and 40°N, air–sea interactions warms a buoyant upper layer of the ocean that overlies a deep layer of cold, dense water throughout the world's oceans. The contoured-density section the Atlantic Ocean in Fig. 2.2 highlights the 2 layers. In particular, the region of closely packed of the isopycnals – the pycnocline – in the upper few 100 m represents a dynamical barrier to vertical mixing between upper and deep layer of the ocean.

The trio of oceanic water column density profiles in Fig. 2.3 demonstrates how the strength of the pynocline in the equatorial/tropical water column is much stronger than that in the high latitude polar regions. Density or specific weight of sea water is a function of temperature, salinity, and pressure. Variations in sea water temperature and salinity (which determine density) vary substantially in the upper layer water column forming their respective thermoclines and haloclines. These variations, in general, are due to the influences of heat exchange and precipitation/evaporation across the air–sea interface at a particular latitude and season. In the deep ocean layer, below 1000 m, these characteristics tend to be relatively constant. The upper ocean layer may be subdivided further into regions of seasonal and main pycnocline (and corresponding thermoclines and haloclines); particularly in the subtropics between 20° and 40° latitudes where density/temperature/salinity gradients are largest extending, respectively, from the surface to about 60 m and from this depth to about 1000 m (Fig. 2.4).

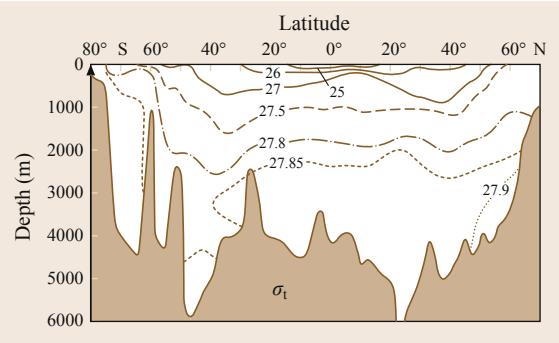


Fig. 2.2 A south–north section of contours of density anomaly (σ_t = water density − 1000) in the western Atlantic; showing how the upper and lower layers of the ocean between 50°S and 50°N are defined by closely packed isopycnals – a strong pycnocline or density stratification – at a depth of about 800 m (after [2.2])

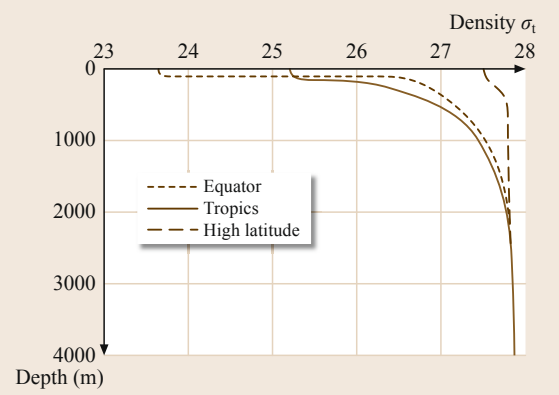


Fig. 2.3 Typical density anomaly versus depth profiles at low and high latitudes (after [2.2])

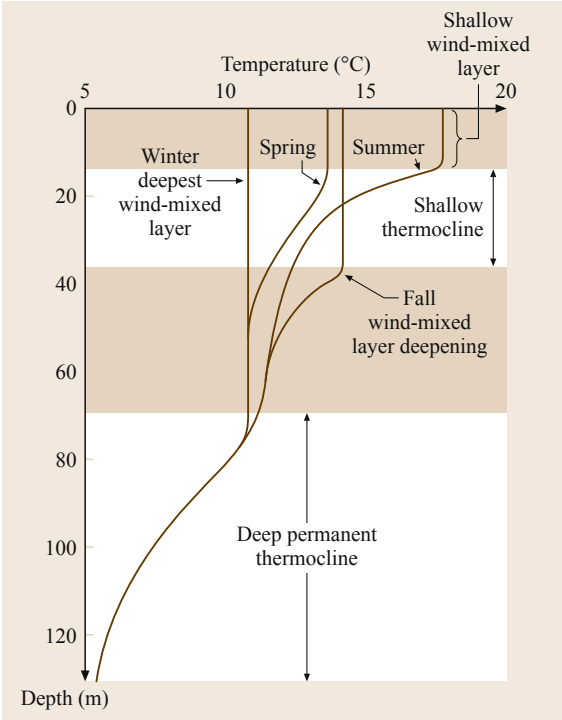


Fig. 2.4 The seasonal variability of a typical subtropical temperature profile (after [2.3])

2.3 Oceanic Processes

Like the atmospheric winds, ocean currents at a single place/time is a superposition of contributions from a complex set of processes many with different space-time scales. The following profiles many of the most important ocean processes in terms of their basic physical elements including forcing and restoring forces and relevance in the overall ocean system. The corresponding ranges of time and space scales of these processes are presented in Table 2.1 and in the dispersion diagram (Fig. 2.5) – a diagram that is particularly relevant to waves:

- Capillary waves. *Physics: surface tension, inertia, wind-driving* – First effect of winds in wave generation
- Surface gravity waves (short). *Physics: gravity, inertia, wind-driving* – Mid-ocean wind waves/surface chop, shipping, off-shore construction.
- Surface gravity waves (long). *Physics: gravity, inertia, bottom topography, wind-driving or earthquake* – Surf and breakers, storm surges, tsunamis, dangerous flooding.
- Internal gravity waves. *Physics: gravity or buoyancy, inertia, topography, indirect wind- driving* – Energy dissipation, mooring stresses, *dead water*.
- Turbulence, mixing. *Physics: inertia, buoyancy, nonlinear interaction, friction* – Dissipation of energy, mixing of momentum and dissolved chemicals including O_2.
- Upwelling (and downwelling). *Physics: buoyancy, local winds and currents* – Brings up nutrients/formation of water masses.
- Mid-Ocean mean flows. *Physics: buoyancy, Coriolis force, wind and heating driving, friction* – Maintaining climate, distributing chemicals, and heat for biology.
- Continental shelf currents. *Physics: buoyancy, Coriolis force, nonlinear inertia* – Transfer of heat and vorticity, climate and local weather, biology, chemistry.
- Tides. *Physics: gravity, inertia, sun and moon gravitational pull, Coriolis force* – Piloting and shipping, coastal construction.
- Internal tides. *Physics: gravity or buoyancy, inertia, topography, indirect wind-driving* – Internal waves at tidal periods with large vertical excursions.
- Estuary flows. *Physics: gravity-buoyancy, Coriolis, tidal, and wind forcing* – Pollutant dispersal, biological nutrient renewal, inflow of fresh water.

The energy distribution of these different oceanic processes tends to sort by time scale and can be conveniently presented in terms of frequency spectra in different forms. The energy density (i. e., energy per unit

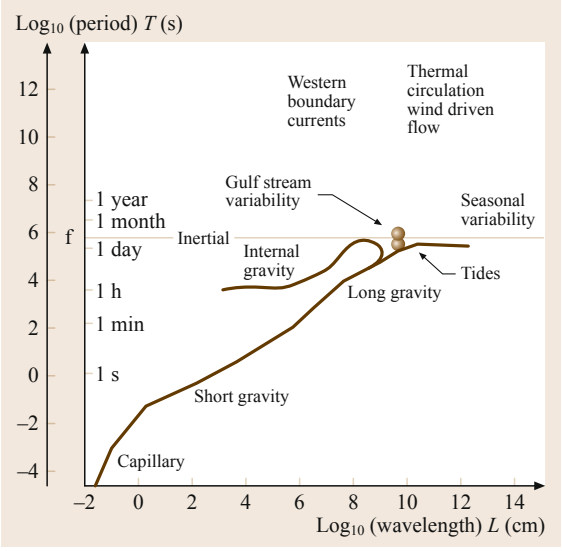


Fig. 2.5 Dispersion diagram for the principal oceanic processes

Table 2.1 Characteristic time and length scales associated with different ocean processes and currents

Type	Time Scales T	Length Scales L	Amplitude H
Capillary waves	< 0.1 s	< 1 cm	< 1 cm
Short surface gravity waves	0.1–60 s	cm–km	m
Long surface gravity waves	s/h	m–10's km	m
Internal gravity waves	min–h	m–10's km	m
Turbulence (Mixing)	s	cm–m	mm–cm/s
Upwelling/downwelling	Days–seasonal	100 km < 1	mm/day
Mid-ocean *mean* flow	1000km	1000 km	cm/s
Western boundary currents	Seasonal–decadal	100 km	≈ 200 cm/s
Continental shelf currents	Days–months	100 km	cm/s
Tides	Day	1000 km	m
Internal tides	Day	10–100 km	10 m
Estuarine currents	Day–seasonal	100 km	cm/sec

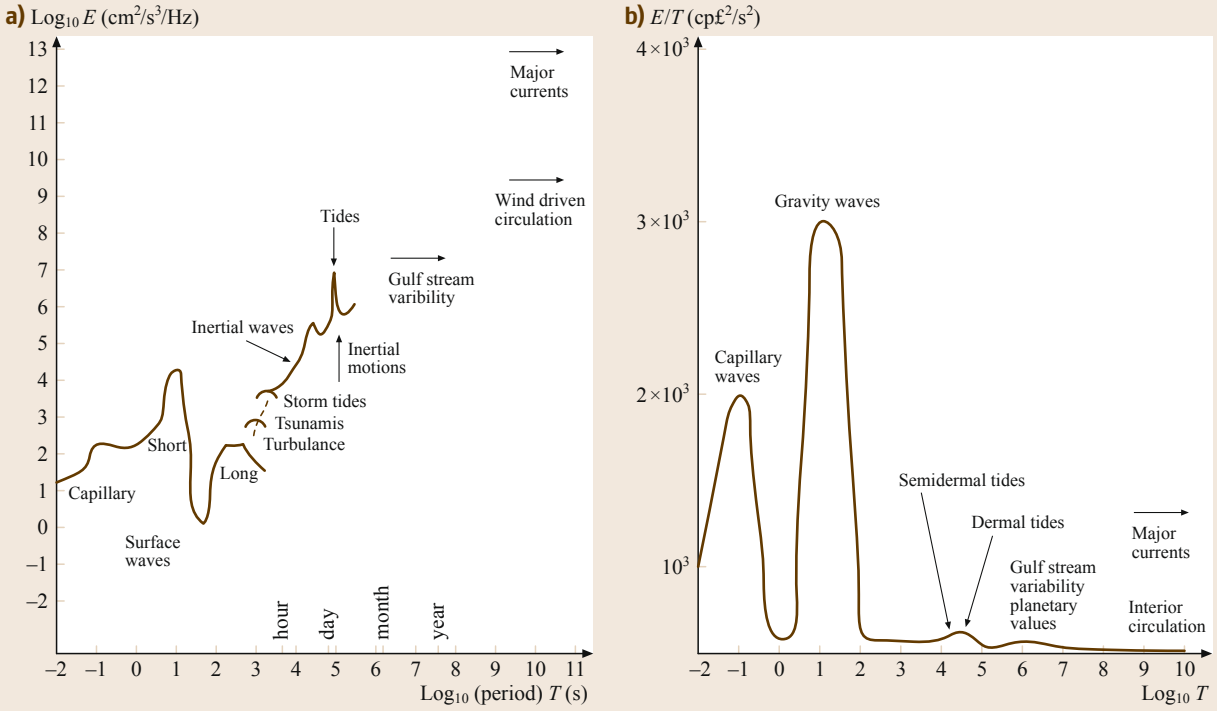


Fig. 2.6a,b **(a)** An *energy density spectrum* (EDS) of ocean current variability highlights the distinctive periodicities associated with many important oceanic processes. **(b)** The variance-preserving form of the corresponding EDS displays the period distribution of kinetic energy per unit ocean volume of these oceanic processes

frequency bandwidth) spectrum highlighting a range of principal ocean processes is presented in Fig. 2.6a. However, the corresponding *variance-preserving form* of the energy spectrum (Fig. 2.6b) is more useful for comparing the kinetic energy per unit volume of these different processes. Clearly energy concentration in physical space is greatest for surface gravity waves with capillary waves a close second. The next section treats the important class of surface gravity waves in more depth.

2.4 Surface Gravity Waves

Very often winds disturb the sea surface; leading to the generation a surface gravity waves with a broad range of wave periods. The gravitational interactions of the moon (and sun) with the Earth also disturb the sea surface in generating surface tidal waves. Tidal current interactions with bathymetry generate internal tides, which, if very nonlinear may spawn packets of internal solitons.

In all of these cases, as gravity acts to restore the sea surface distortion toward its equilibrium state, the potential energy of the original distortion is converted to kinetic energy in the form of a velocity field. The lowering sea surface then overshoots the equilibrium state as the gravity-related buoyancy forces reverse the process. As these surface gravity waves are generated, they with their energy propagate away from their generation site.

The wave periodogram in Fig. 2.7 shows how surface wave energy is distributed among many wave frequencies in the ocean. Wind-driven surface gravity waves of wind periods in the 1−30 s range are the most energetic, followed by tides. Note that short period (or high frequency) capillary wave motion is controlled by surface tension, while longer period (or low frequency) gravity waves, like storm surges and tides, are also influenced by earth rotation.

2.4.1 Surface Gravity Waves

The currents and pressures associated with surface gravity waves are attenuated with depth in accordance with their wavelength. The current and pressure amplitudes of surface waves become negligible (less than 5% of their surface values) at a depth of $L/2$, where L is the

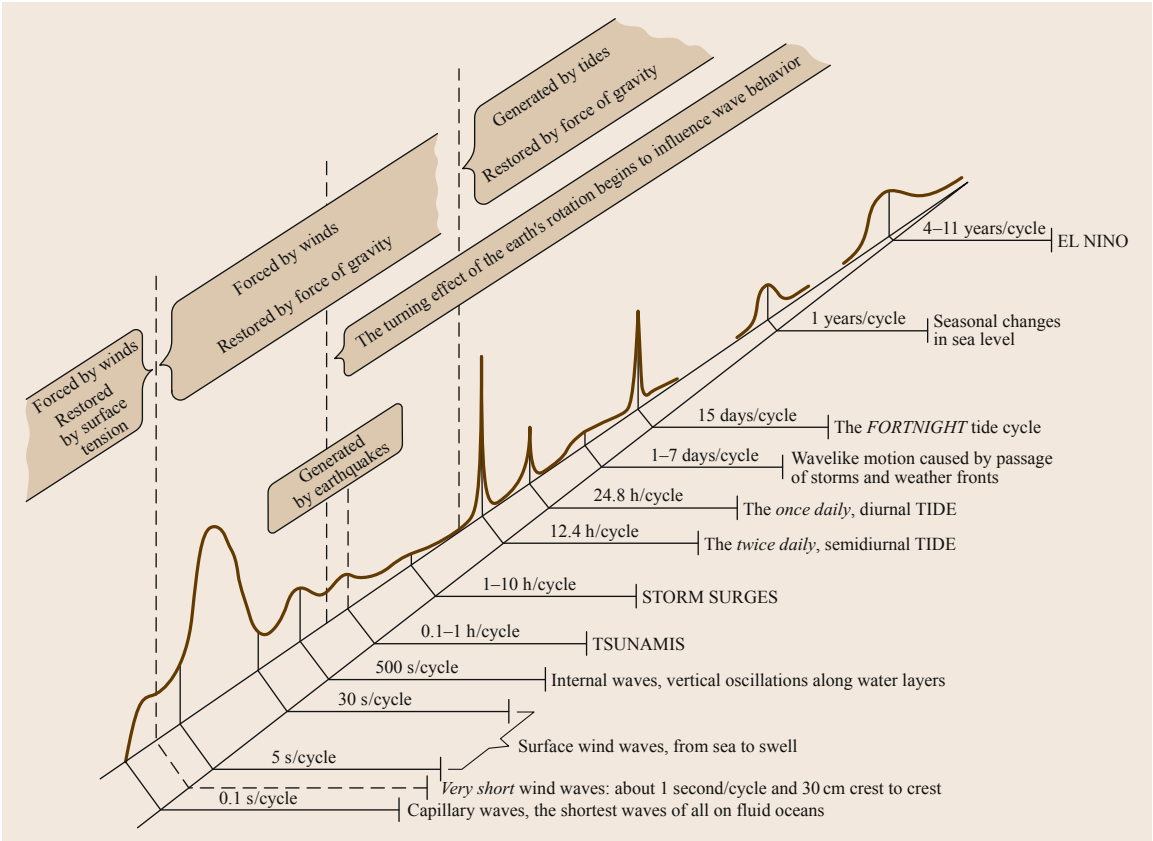


Fig. 2.7 This periodogram of sea level variability diagrams how energy of many different types of ocean gravity waves is distributed according to frequency. Waves are identified according to wave period and to the disturbing and restoring forces which act at different time scales (after [2.4])

wavelength. The currents associated with typical deep water wind waves (period of 20 s) are effectively attenuated at depths of about 300 m (984 ft) and are almost never felt at depths of 1000 m (3284 ft).

However, the longer period tides and tsunamis have such large wavelengths that their effects are virtually unattenuated throughout most of the world's oceans. However, because their periods are so much longer (minutes to a day) than typical wind waves, the water velocities associated with them are generally small. Tidal currents in coastal oceans settings are an important exception.

2.4.2 Definitions

Here we consider a simplified model for surface gravity waves. As we all have observed the sea surface at a particular location is a complicated superposition of waves with different characteristics propagating from different places where they were usually generated by the wind at different times. However, first we want to focus on the basic principles of wave motion, so we will explore a simple monochromatic (i. e., single frequency) wave model. Later, we will apply this understanding of the *essential physics* of wave motion to more realistic observed wave fields.

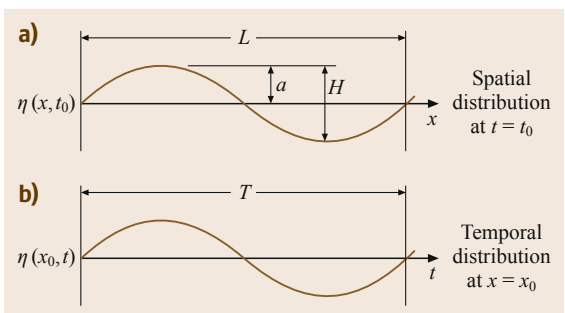


Fig. 2.8a,b The characteristics of a monochromatic sine wave $\eta(x, t)$ whose form is propagating with phase speed c in the x-direction is **(a)** *frozen* in space at time $t = t_0$; and **(b)** measured at a fixed station $x = x_0$

To start, we define the characteristics (with units indicated in [. . .]) of the monochromatic sine wave shown in Fig. 2.8:

- Elevation: $\eta(x, t)$ is the instantaneous vertical departure of the sea level from the undisturbed sea level.
- Wave height: H is the distance between the wave crest and trough. $[H] = (\text{unit length})^{-1}$
- Wave amplitude: $a = 1/2H$
- Wavelength: L is the distance between points of equal phase. $[L] = \text{length}$
- Wave number: $k = 2\pi/L$ is the number of times a wave can fit on the circumference of circle with a dimensionless radius = 1. $[k] = (\text{length})^{-1}$
- Wave period: T is the time between points of equal phase. $[T] = \text{unit time}$.
- Wave frequency: $f = 1/T$ is the number of complete cycles per unit time. $[f] = \text{cycles/time}$
- Angular frequency: $\omega = 2\pi/T = 2\pi f$ is the number of times a wave of period T can fit on the circumference of a circle with a dimensionless radius = 1. $[\omega] = (\text{time})^{-1}$
- Wave phase speed: $c \equiv L/T$ or $c \equiv \omega/k$ is the speed of a particular point on the waveform (or phase). $[c] = \text{length/time}$

The mathematical description of our monochromatic *right-traveling* waveform is

$$\eta(x,t) = a\cos\left(\frac{2\pi x}{L} - \frac{2\pi t}{T}\right)$$

or

$$= a\cos(kx - \omega t)\,. \tag{2.1}$$

Equation (2.1) can also be written in terms of phase speed as $\eta(x,t) = a\cos[k(x - ct)]$. The dynamical relation between wave period T and wavelength L can be defined by the solution to the two-dimensional (2-D) *Airy wave model* of a right-traveling wave at the surface of an ocean (mean depth $z = 0$) with a flat bottom at a depth of $-z = h$.

The solution to the approximate linearized Airy wave equations of motion and boundary conditions for small amplitude ($a/L \ll 1$) waves [2.5] yields the *dispersion relation*

$$\omega^2 = gk\tanh(kh)\,, \tag{2.2}$$

which describes the relation between wave frequency and wave number, and the dynamical variables of pressure p, and the x- and upward ($+z$)-directed velocity components u and w, respectively:

$$u = a\omega\frac{\cosh k(z+h)}{\sinh(kh)}\cos(kx - \omega t)\,, \tag{2.3a}$$

$$w = a\omega\frac{\sinh k(z+h)}{\sinh(kh)}\sin(kx - \omega t)\,, \tag{2.3b}$$

$$p = -\rho gz + \rho ga\frac{\cosh k(z+h)}{\cosh kh}\cos(kw - \omega t)\,. \tag{2.3c}$$

The dispersion relation given above can be rewritten as

$$c^2 = \frac{L^2}{T^2} = \frac{\omega^2}{k^2} = \frac{g}{k}\tanh(kh)\,,$$

which shows that, in general, Airy waves have phase speeds which depend upon wave number k (or wavelength L) (Fig. 2.9).

Two important approximations to the (2.2) and (2.3) wave solutions yield the classical *short* and *long* waves which are defined in terms of the wavelength to water depth ratio L/h.

For *short waves*, the water depth is greater than a half wavelength $h > L/2$ so that they are referred to as *deep water waves*. Mathematically, this means that kh is large (or $kh > \pi$) so that the *short wave* (deep water wave) dispersion relation (2.2) becomes

$$c^2 = \frac{\omega^2}{k^2} = \frac{g}{k} = g\frac{L}{2\pi}\,,$$

which reduces to

$$c = \sqrt{\frac{1}{2\pi}(gL)}\,,$$

so that short waves are dispersive, that is, the wave speed depends on wavelength. Since $L = cT$, the above

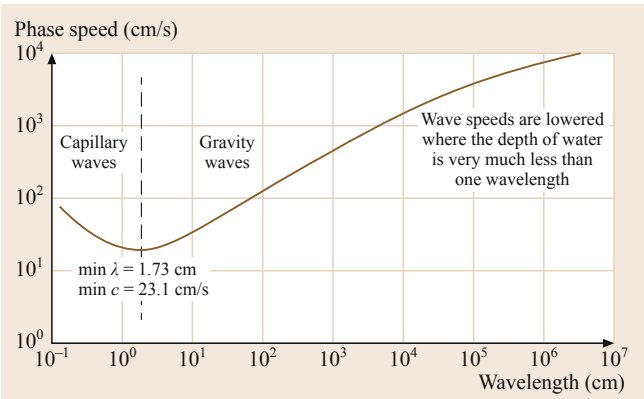


Fig. 2.9 The phase speed of gravity waves and capillary waves as a function of their wavelength (after [2.6])

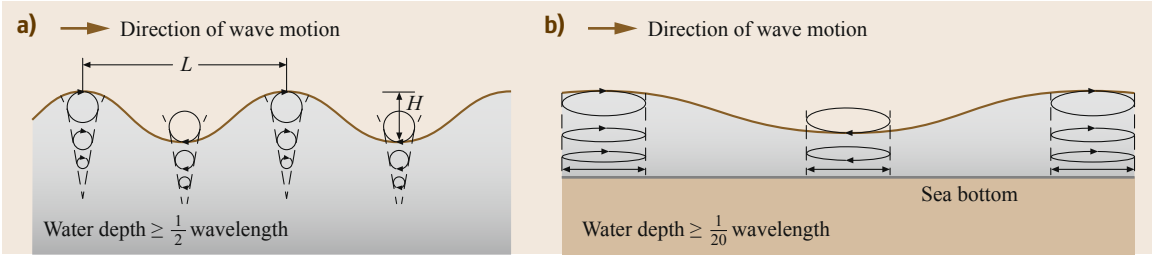


Fig. 2.10a,b **(a)** Progressive deep-water wave motion consists of near-circular water parcel orbits that become very small at water depths of $1/2$ wavelength. **(b)** Progressive shallow water wave motion ($h/L > 0.05$) is distinguished by water parcels that move in highly elliptical orbits, with widths that are constant with depth (after [2.7])

can also be written as

$$c = \frac{1}{2\pi}(\mathrm{g}T) \, .$$

For large kh, the *short wave solutions* to (2.3) for $\eta = a\cos(kx - \omega t)$ become

$$u = a\omega \mathrm{e}^{kz}\cos(kx - \omega t) \, ,$$

$$w = a\omega \mathrm{e}^{kz}\sin(kx - \omega t) \, ,$$

$$p = \rho g a \mathrm{e}^{kz}\cos(kx - \omega t) - \rho \mathrm{g} z \, .$$

The motions associated with short or deep water waves decreases with depth, such that amplitudes at a depth of $z = -L/2$ are $\mathrm{e}^{-\pi}$, or 4% of surface values. This is shown in the schematic of the water parcel trajectories as the wave passes (to the right in this case, Fig. 2.10a).

For *long waves*, the water depth is less than the wavelength or $h < L/20$ so that long waves *feel the bottom* and are called *shallow water waves* (Fig. 2.10b). Mathematically, this means that kh is small (or $kh > \pi/10$) so that the *long* (shallow water) *wave* dispersion relation (2.2) becomes

$$\tanh(kh) \approx kh \, ,$$

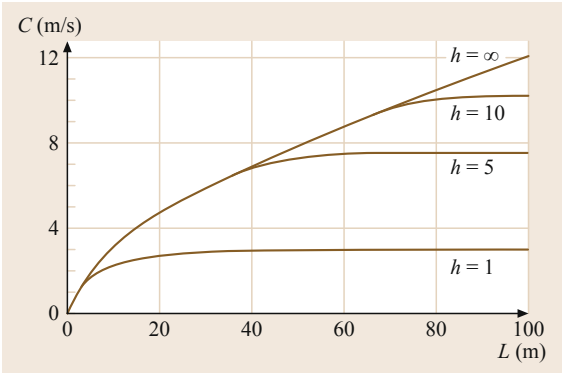


Fig. 2.11 Wave phase speed c versus wavelength L (m) – a dispersion diagram for Airy waves in different water depths h (m)

and the *long wave* (shallow water) dispersion relation becomes

$$c^2 = \frac{\omega^2}{k^2} = \frac{\mathrm{g}}{k}kh$$

or

$$c = \sqrt{\mathrm{g}h} \, .$$

Thus, long (or shallow water) waves are *nondispersive*, that is that their speed is independent of wavelength. This is demonstrated in Fig. 2.11 which depicts the relation of wave phase speed and wavelength for different water depths.

For small kh, the *long wave* (shallow water) pressure and horizontal velocity *solutions* to (2.3) for $\eta = a\cos(kx - \omega t)$ are undiminished with depth. The water parcel orbits during the passage of a *long water wave* in very shallow water are very elliptical as shown in Fig. 2.10b. Notice that most of the motion is horizontal, much like what scuba divers feel in shallow water with a long swell. Airy wave kinematics are shown schematically in Fig. 2.12 for different water depths. Note the relative phases of η, u, and w fields.

As Airy waves propagate from deep water through intermediate depth water into shallow water (Fig. 2.13), (2.2) and (2.3) describe their dynamics. As deep water waves propagate into shallow water, their orbital motions and pressure fields begin to interact with the bottom – they *feel the bottom*. Nearly circular water parcel orbits become more elliptical and wave velocities induce bottom stresses; which can have an effect on movable sediments.

However, while monochromatic Airy wave theory is very helpful in describing the basic dynamic characteristics of waves, the real ocean wave field is composed of contributions from waves with many wavelengths, wave periods and amplitudes. What are the consequences of the superposition of more than one Airy wave?

To answer the question, we increase the complexity of our model slightly by superposing two Airy waves η_1 and η_2 with the same amplitude but slightly different

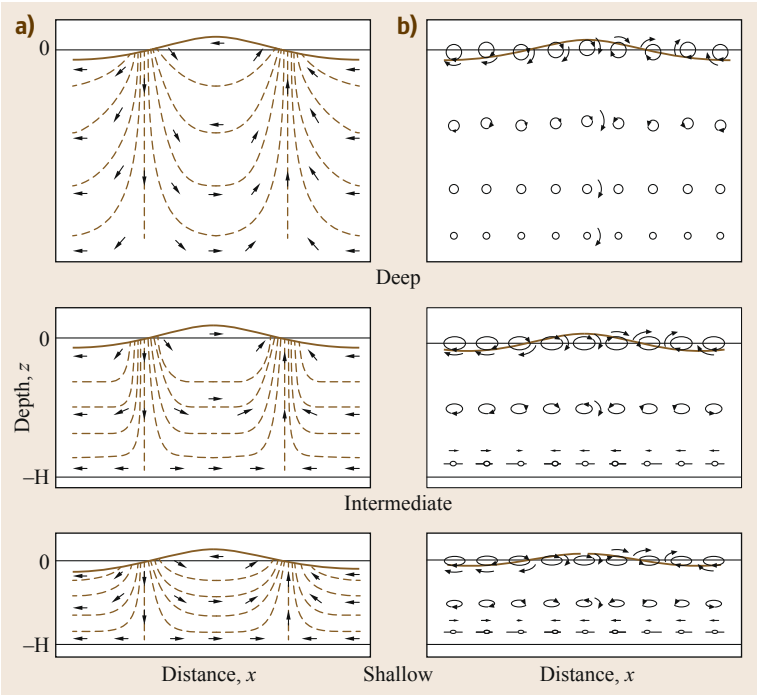


Fig. 2.12a,b Kinematics of waves propagating left to right in (top) deep; (middle) intermediate depth; and (bottom) shallow water – **(a)** instantaneous snapshots of the water parcel streamlines and **(b)** trajectories of selected water parcels over a full wave cycle (after [2.8])

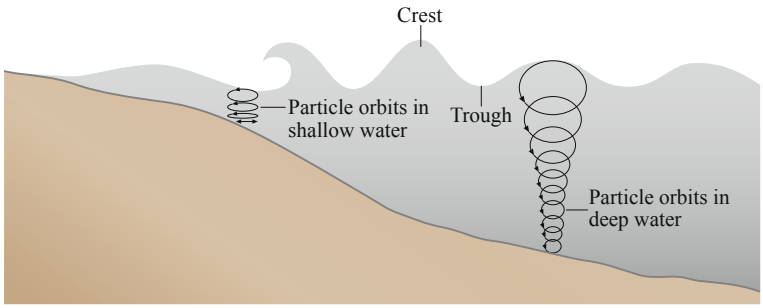


Fig. 2.13 Deep water waves propagating toward the left into *shallow water*. Note how the waves *feel the bottom* and water parcel orbits are squashed as the wave propagates into even shallower water

frequencies and wave numbers such that $\omega_2 > \omega_1$ and $k_2 > k_1$.

The amplitude of the resultant wave (shown in Fig. 2.14) is modulated according to

$$\eta = a\cos\left[\frac{\Delta k}{2}(x - c_g t)\right]\cos\left[\overline{k}(x - \overline{c}t)\right] , \qquad (2.4a)$$

where

$$\overline{k} = \frac{k_1 + k_2}{2}, \quad \overline{c} = \frac{\overline{w}}{\overline{k}} \quad \text{and} \quad \overline{\omega} = \frac{\omega_1 + \omega_2}{2} . \qquad (2.4b)$$

Note that on one hand, the cos[ ] term on the right in (2.4) represents a wave with intermediate frequency and wavelength and therefore is nearly indistinguishable from the parent waves. On the other hand, the cos[ ]

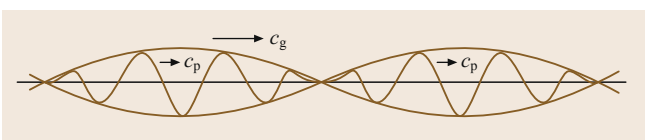


Fig. 2.14 The wave envelope – resulting from the superposition of a pair of surface gravity waves (with nearly identical phase speeds c_p) – travels at the wave group velocity c_g

term on the left represents a wave with much greater L (smaller k) that is traveling at the speed c_g given by

$$\frac{\Delta\omega}{\Delta k} = \frac{\omega_2 - \omega_1}{k_2 - k_1} = c_g .$$

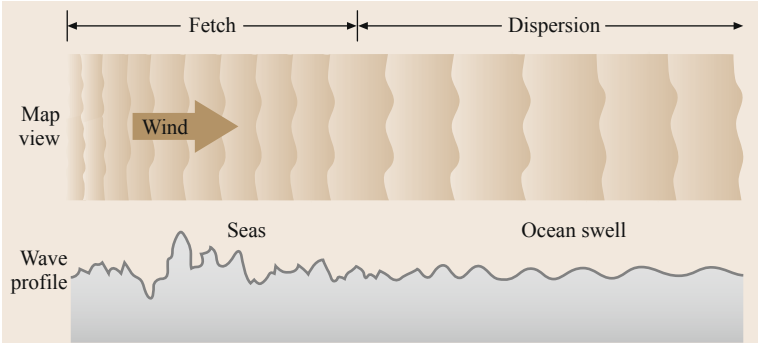


Fig. 2.15 As the wind increases from zero over a *glassy* ocean surface, first capillary waves are generated; which very rapidly evolve into ripples, chop, wind waves, and eventually a fully developed sea (after [2.7])

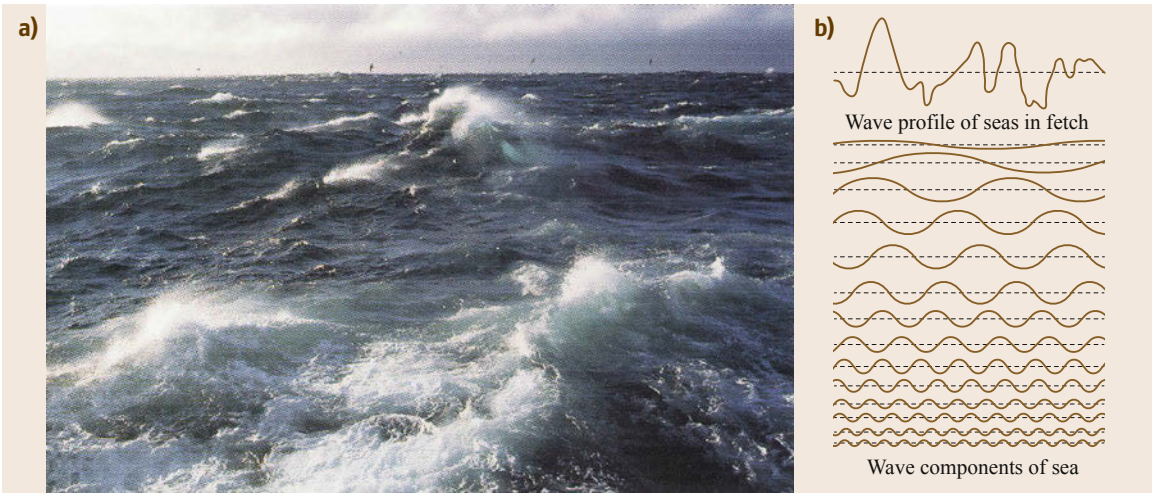


Fig. 2.16 **(a)** The open ocean sea surface is generally chaotic because it is **(b)** a superposition of many different waves, with different wavelengths and heights (after [2.7])

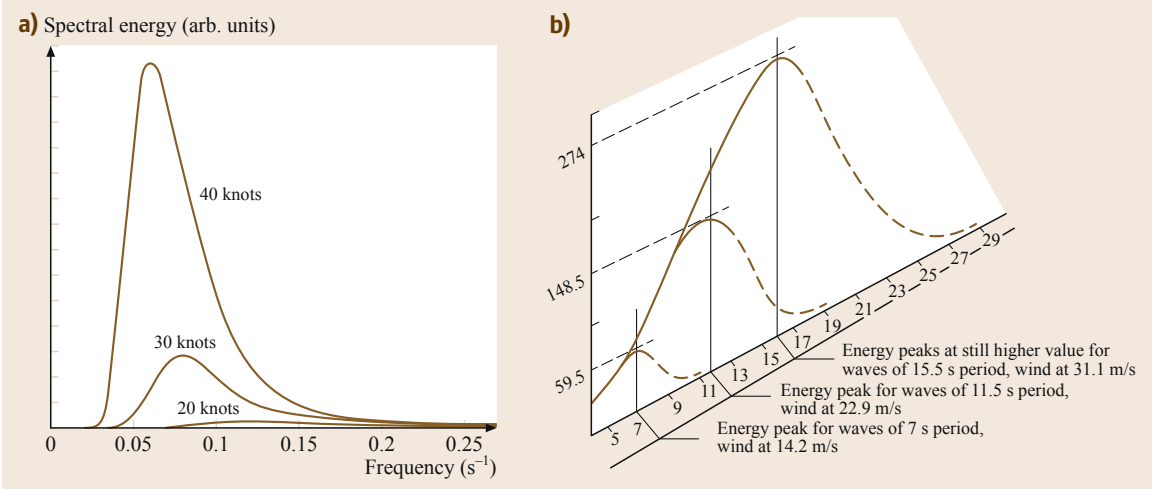


Fig. 2.17 **(a)** A frequency spectra of the wave *energy density*(variance per unit frequency bandwidth) for a fully developed sea under different wind speed conditions. Note how the larger wind speeds generate waves for which the energy density peak is at lower frequencies (after [2.8]). **(b)** A *periodogram* of *wave energy* (kJ/m^2) for fully developed wave fields generated by different wind speeds – derived from Table 2.2 data (after [2.4])

Table 2.2 Examples of the characteristics of wind waves observed for fully developed seas under different wind conditions (after [2.9])

Wind speed [m/s]	Wave phase speed [m/s]	Wave period [s]	Wavelength L [m]	Wave height H [m]	H/L	Energy [kJ/m^2]
14.2	11.5	7.0	78.0	6.9	0.088	59.5
16.0	12.8	8.0	103.0	7.7	0.074	74.1
19.2	16.3	9.5	147.0	9.2	0.063	106.0
22.9	18.3	11.5	209.0	10.9	0.052	148.5
27.0	21.5	13.5	290.0	13.0	0.045	211.0
31.1	25.0	15.5	384.0	14.8	0.039	274.0

Table 2.3 Conditions necessary for fully developed sea at given wind speed and the parameters of the resulting waves (after [2.9])

Wind speed [knots]	Fetch [nautical miles]	Duration [h]	Average height [ft]	[m]	Average length [ft]	[m]	Average period [s]
10	10	2	0.9	0.27	28	8.5	3.0
12	18	4	1.4	0.43	40	12.2	3.4
14	28	5	2.0	0.61	55	16.8	4.0
16	40	7	2.8	0.85	71	21.6	4.6
18	55	8	3.8	1.2	90	27.4	5.0
20	75	10	4.9	1.5	111	33.8	5.7
22	100	12	6.3	1.9	135	41.2	6.3
24	130	14	7.8	2.4	160	48.8	7.0
26	180	17	9.5	2.9	188	57.3	7.4
28	230	20	11.4	3.5	218	66.4	8.0
30	280	23	13.6	4.1	251	76.5	8.6
32	340	27	16.0	4.9	285	86.9	9.0
34	420	30	18.6	5.7	322	98.2	9.7
36	500	34	21.4	6.5	361	110.1	10.3
38	600	38	34.5	7.5	402	122.6	10.9
40	710	42	27.9	8.5	446	136.0	11.4
42	830	47	31.5	9.6	491	149.7	12.0
44	960	52	35.4	10.8	540	164.6	12.6

In general, the superposition of multiple surface gravity waves form packets called *wave groups* whose energy travels at the *group speed* given by $c_g = \mathrm{d}\omega/\mathrm{d}k$. Thus the group velocity for general Airy waves is

$$c_g = \frac{1}{2}c\left(1 + \frac{2kh}{\sin h2kh}\right) ,$$

which for short waves (large kh) becomes

$$c_g = \frac{1}{2}c\,; \quad \text{deep water waves}$$

and for long waves (small kh) becomes

$$c_g = c; \quad \text{shallow water waves.}$$

Wave energy is composed of both the kinetic and potential energy components which can be computed explicitly for Airy waves [2.5]. In particular for shallow water waves, which are nondispersive, the wave energy travels at the same speed for all waves at $c = \sqrt{gh}$. The topic of wave generation and wave energy propagation is considered next.

2.4.3 Wave Generation and Evolution

Most ocean surface waves are produced by the wind. As the wind increases from near zero, it generates capillary waves, with short wavelengths (< 1.7 cm) and short periods (< 0.1 s). As the wind continues to blow, the

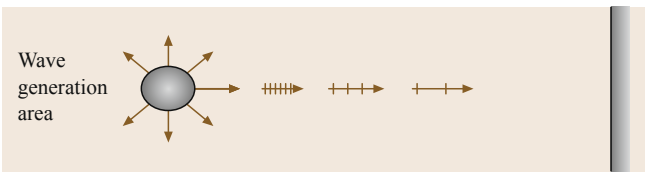


Fig. 2.18 Wave dispersion separates waves generated in the wave generation region

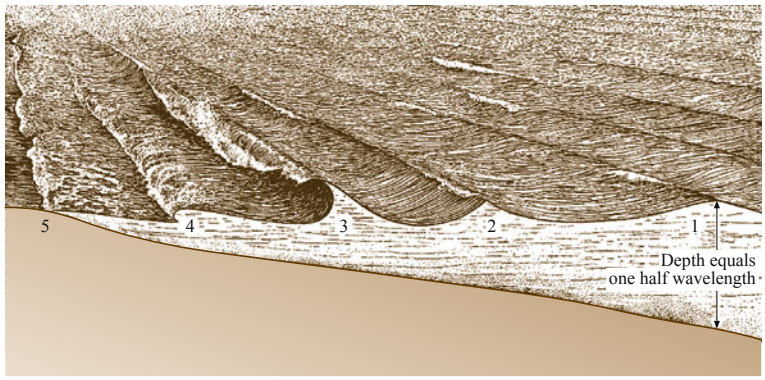


Fig. 2.19 (1) Deep water waves beginning to *feel the bottom*; (2) intermediate to shallow water wave transition; (3) shallow water waves begins to break; (4) fuel wave breaking; and (5) runup (after [2.10])

capillary waves can grow to become ripples and short choppy waves (Fig. 2.15). With increased wind velocity, duration, and fetch (length of the area over which the wind blows), more energetic (higher), longer waves are generated. The superposition of all of these waves in the wave generation region creates a fully developed sea with white capping. A fully developed sea appears chaotic because it is a superposition of waves with different wavelengths and corresponding wave periods (Fig. 2.16).

The spectral energy content of sea surface variability at the site can be analyzed in terms of their frequency content and displayed as energy density spectra (Fig. 2.17a) or a wave energy periodogram (Fig. 2.17b). These examples illustrate typical energy distribution as a function of wave frequency (or wave period) for a *fully developed sea* under different wind speeds. Both examples show generally that higher winds generate more energetic lower frequency waves. The wind strength, duration, and fetch determine the resulting range of wavelengths and wave amplitudes in the waves constituting the corresponding *fully developed* sea (Table 2.3).

Since weather systems move much faster than ocean wind waves, the storm winds will diminish over the immediate area of wave generation. Then this chaotic superposition of waves of different wave periods and wavelengths will propagate from the region in all directions at different speeds. Along a particular propagation trajectory, like that depicted in Fig. 2.18, groups of these deep ocean waves with the longer wavelengths (and thus the larger group velocities) will outrun the shorter wave groups. Because of deep water wave dispersion, wave energy spectra measured some distance from their generation site (e.g., a distant shore) will differ from those that are measured at a later time; as the slower, shorter wave groups arrive.

Eventually all of the deep water waves will propagate into water depths where they begin to *feel the bottom* (Fig. 2.19) and undergo significant changes. Once frictional contact begins, the wave changes in several ways simultaneously, namely the wave speed drops, its height increases, and its direction shoreward becomes more and more perpendicular to the beach line [2.4].

As these shallow water waves propagate shoreward both their phase and group speeds

$$c = c_g = \frac{L}{T} = \sqrt{gh}\,, \tag{2.5}$$

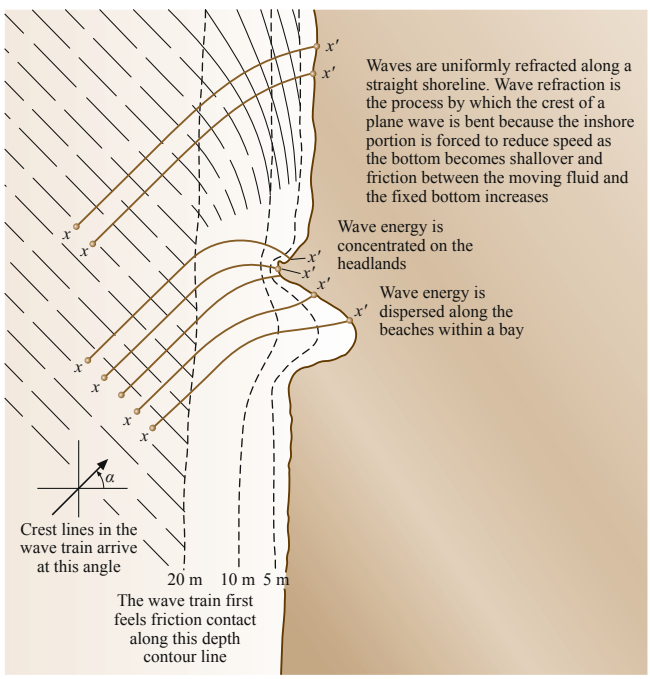


Fig. 2.20 Wave refraction or wavefront bending is due to the slowing of the part of the wavefront that shoals earliest. The pair of rays (or orthogonals that are perpendicular to the wavefront locally) mark the trajectory of equal amounts of wave energy (after [2.4]) ◄

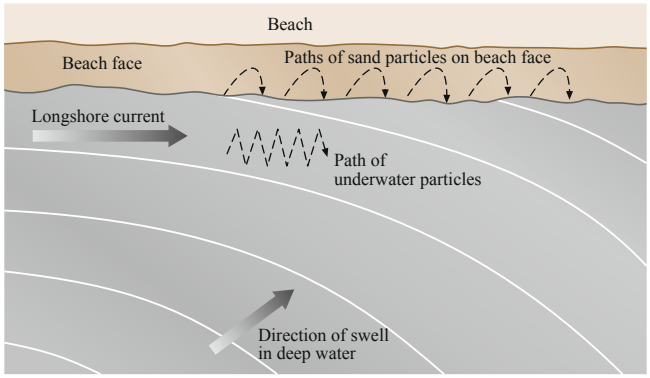


Fig. 2.21 Waves that approach the beach obliquely approach generates a longshore current in the surf zone. The wave-breaking suspends beach material (usually sand) that is transported in the surf zone by the longshore current; while the repeated wave run-up cycles also transport sand in the same longshore direction (after [2.10]) ►

decrease. Since the number of waves in the shoreward propagating wave train are conserved, the wave period T *remains constant*. Thus from a rearranged (2.5)

$$T = \frac{L}{\sqrt{gh}}\,, \tag{2.6}$$

the wavelength L must also decrease. For an idealized beach with straight coastline and parallel isobaths, the *wave power* per unit length of the shoreward propagating wavefront – given by

$$\overline{P} = \frac{1}{8}\rho g H^2 c_g = \frac{1}{8}\rho g H^2 \sqrt{gh} \tag{2.7}$$

is constant. Thus from a rearranged (2.7)

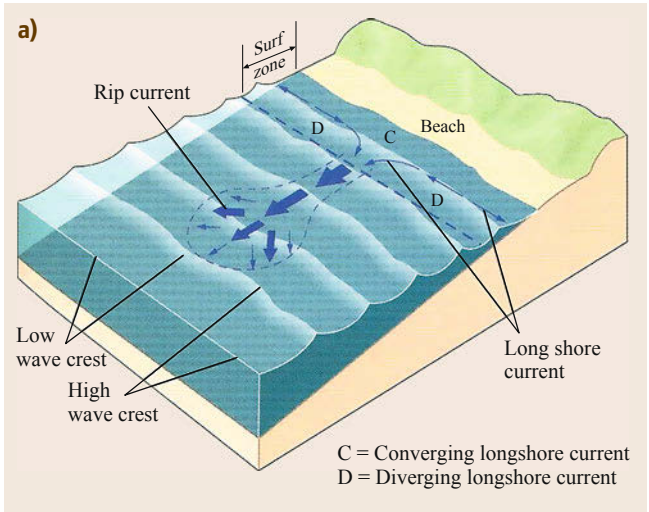
$$H^2 = \left(\frac{8\overline{P}}{\rho g^{3/2}}\right) h^{-1/2}\,, \tag{2.8}$$

we conclude that for shoreward propagating waves H must also increase! As a result, the wave slope which is proportional to H/L will increase to the point where Airy wave theory no longer applies.

At this stage, wave crests or fronts are generally propagating obliquely to the orientation of the beach (Fig. 2.20). The wavefronts are bent or refracted as the inshore end of the wavefront encounters even shallower water first and slows relative to the rest of the wave – bringing the wavefront into better alignment with the beach.

In the example of waves impinging on a straight beach (Fig. 2.20 – upper), by the refraction distorts the waves (and corresponding wave rays or orthogonals) so that the wave energy flux (energy/length/unit time) is reduced from its deep water value. Wave refraction near irregular coastlines (Fig. 2.20 – center) leads to distortion of the wavefronts. Thus the ray paths, or orthogonals indicating the direction of wave energy propagation, are also distorted. In the case of a *point of land* the equidistant rays in deep water are seen to converge on the point thereby concentrating wave energy there relative to other parts of the coastline. The opposite is true for a bay.

As the wave propagates shoreward, wave slopes generally increase and eventually lead to significant



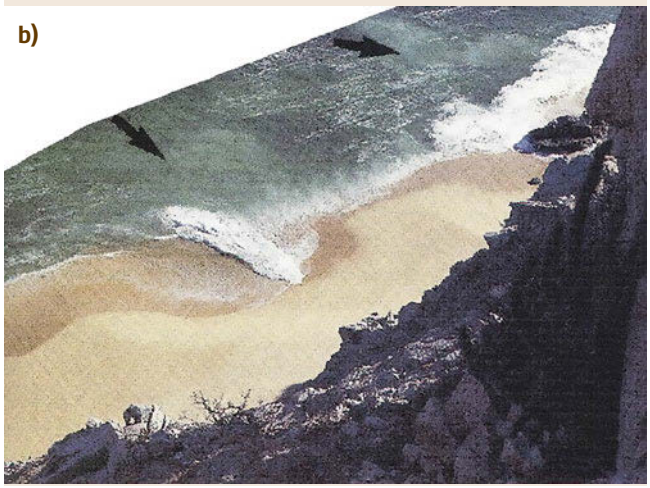


Fig. 2.22 **(a)** In this illustration, waves with crests parallel to the coast break and produce run-up that returns to the ocean via a system of longshore currents that feed narrow offshore rip currents. **(b)** narrow rapid rip current systems are observed frequently especially with large amplitude waves (after [2.7])

nonlinearities and subsequent *wave-breaking* as indicated at position 4 and *wave run-up* onto the beach front in position 5 in Fig. 2.19. It is noted that wave trains that approach the beach obliquely never fully align with the beach front. This asymmetry leads to the generation of an alongshore current in the direction indicated in Fig. 2.21. This alongshore current plus the zig–zag motion of the wave run-up transport beach material (sand in many situations) along shore as indicated. This is a dynamic longshore sand transport process under which a steady-state beach persists as long as the up-coast sand supply exists.

Of course, the wave run-up must return to the ocean. This return flow is concentrated in narrow rapidly moving offshore moving currents called rip currents (Fig. 2.22).

2.5 Wind-Forced Ocean Processes

2.5.1 Frictional Effects

Surface winds apply stress to the ocean surface to generate upper ocean currents and produce turbulent mixing. This process of transferring momentum to the ocean currents occurs through a complex process involving the generation of surface waves. In contrast, less complex stress-related processes remove momentum from near-bottom ocean flows and dissipate it as heat. Thus, ocean stress in the ocean mediates the vertical transport of horizontal momentum in the ocean.

The normal expression for fluid stress, in which horizontal stress τ is proportional to the local vertical gradient of horizontal velocity, is given by

$$\tau = \mu \frac{\partial u}{\partial z} , \tag{2.9}$$

in which μ is the coefficient of molecular viscosity. However, μ is too small for this relation to accurately define momentum transport (and thus stress and dissipation) in a turbulent ocean. Thus, we must incorporate turbulent eddy viscous effects into our consideration of momentum transfer in the ocean.

The nature of eddy viscous effects can be defined quantitatively by partitioning a variable eastward ocean flow into its temporal mean value $\overline{u}$ and eddy-induced fluctuating part u' (Fig. 2.23). A similar treatment defines the fluctuating *northward* and *upward* velocity components v' and w', respectively.

The time-average of the many random turbulent eddy events involving correlated fluctuations $\overline{u'w'}$ leads to an effective eastward stress S_x on a horizontal face of the fluid. The result is a turbulence-induced stress – called a Reynolds stress – that can be related to the shear of the mean flow via a coefficient of *eddy dynamic viscosity* A^e according to

$$S_x = A_z^e \frac{\partial \overline{u}}{\partial z} = -\rho \overline{u'w'} , \tag{2.10}$$

where $\nu_z^e = A_z^e/\rho$ is the *eddy kinematic viscosity*.

Similar relationships can be derived for stresses on other *faces* of a water parcel. Typical vertical eddy kinematic viscosities in the ocean range from 2 to $10^4\,\mathrm{cm^2\,s^{-1}}$. Lateral kinematic eddy viscosities are generally larger ranging from 10 to $10^8\,\mathrm{cm^2\,s^{-1}}$. The large difference between the vertical and lateral eddy coefficients relates to the fact that vertical turbulent momentum transport is considerably reduced relative to lateral momentum transport. This asymmetry is due to both the large geometrical aspect ratio or *thinness* of oceans; as well as the stabilizing effects of vertical stratification of the ocean.

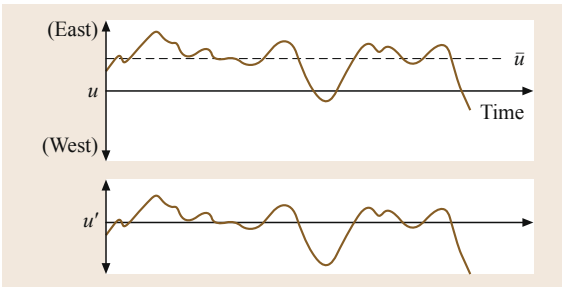


Fig. 2.23 The total time-varying of say an eastward flow is $u(t)$, which can be partitioned into its time-averaged component $\overline{u}$ and fluctuating component u'

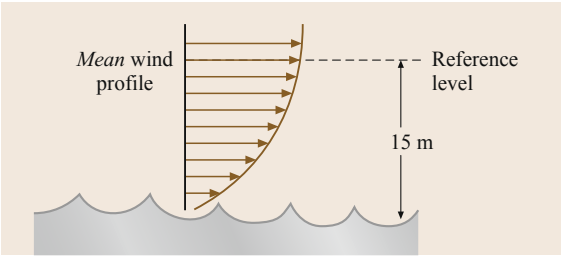


Fig. 2.24 Typical boundary layer air flow near a solid (or watery) horizontal boundary shows the effects of momentum extraction from the flow. The reference level of the wind at 15 m above the sea surface is used for estimating horizontal wind stress at the air-ocean interface

These considerations apply directly to oceanic stresses at the air–ocean and ocean–sediment interfaces. For example, dimensional considerations allows us to estimate sea surface stress according to is estimated as a function of the

$$\tau_w = \text{constant} \cdot \rho_a W_{15}^2 \,, \tag{2.11}$$

where W_{15} is the wind velocity at 15 m above the sea surface – the elevation of many ship and buoy anemometers.

To determine the constant in (2.11) we first define a friction velocity u^* such that $\tau = \rho_a u^{*2}$. Then consider the surface layer of turbulent winds near the sea surface – or *boundary layer* – with a time average wind profile that decreases from a maximum at the top of the boundary layer to near zero at the sea surface (Fig. 2.24).

It can be shown that the mean velocity profile $\overline{u(z)}$ in a constant stress turbulent boundary layer can be described in terms of a friction velocity u^* and Von Karman's constant $k_o = 0.4$ according to

$$\overline{u}(z) = \frac{u^*}{k_o} \ln \frac{z + z_o}{z_o} \,, \tag{2.12}$$

where z_o is a parameter related to ocean surface roughness. This form of the boundary layer mean velocity is the so-called Log Profile. This formulation permits us to compute u^* and z_o from the observations of $\overline{u}(z)$ so that the unknown constant in an expression for wind stress can be determined. Typically

$$\tau_w = 2.6 \times 10^{-3} \cdot \rho_a W_{15}^2 \tag{2.13}$$

for density units g cm^{-3} and wind units m s^{-1}.

The wind stress exerted on the sea surface can be expressed in terms of the direction of the wind vector $\boldsymbol{W}$ and is proportional to the square of the wind speed according to the relation

$$\boldsymbol{\tau}_w = \rho_a C_D |\boldsymbol{W}| \boldsymbol{W} \,, \tag{2.14}$$

where ρ_a is the air density and C_D is a drag coefficient.

2.5.2 Earth Rotation Effects

Earth rotation plays a central role in the physics (as viewed in our earthly rotating – hence accelerating frame of reference) of many of the most important atmospheric and ocean flows, (which have variability time scales exceeding 1/2 day). Earth rotation (rate $= \boldsymbol{\Omega}$) influences ocean flow through the effects of the Coriolis force which always acts to the right of the direction moving water (Fig. 2.25) according to

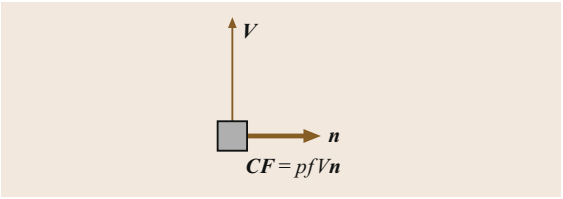


Fig. 2.25 An ocean water parcel is moving in an arbitrary direction with steady velocity $\boldsymbol{V}$ in the Northern Hemisphere. The Coriolis force $\boldsymbol{CF}$ acts on a moving water parcel to the right of the direction of $\boldsymbol{V}$ – or in the $\boldsymbol{n}$ direction

$$\boldsymbol{CF} = \rho f \boldsymbol{V} \boldsymbol{n} \,,$$

where ρ is the water density, $f = 2\Omega \sin(\text{latitude})$ is the Coriolis parameter, $\boldsymbol{n}$ is the unit vector in the direction perpendicular and to the right of the water parcel velocity $\boldsymbol{V}$ composed of an eastward u and northward v components.

In the Northern Hemisphere Cartesian coordinate system, the Coriolis force $\boldsymbol{CF}$ is given by $\boldsymbol{CF} = \rho f v \boldsymbol{i} - \rho f u \boldsymbol{j}$; where $i \neq j$ are the unit vectors in the eastward and northward directions, respectively.

2.5.3 Hurricane Wind-Forced Ocean Response

Tropical cyclones have a significant impact on coastal areas of the world. In the Atlantic Ocean they are called hurricanes, whereas similar storms in the Pacific Ocean are called typhoons. They are a significant part of global weather systems because they transfer large quantities of water and warm, moist air from equatorial regions to higher latitudes. In the Caribbean Sea, the Gulf of Mexico, and the southeastern United States, hurricanes have shaped low-lying coastal areas and, at times, have caused dramatic losses of life and property.

Historically, the hurricane season in the North Atlantic Ocean has extended from June 1 to November 30; when seawater temperatures in the equatorial oceans are warmest, usually above 25 °C (about 80 °F). Hurricanes begin as tropical waves, areas of organized clouds 200–500 km in diameter in the eastern equatorial Atlantic. Occasionally, the wave becomes a tropical depression, which when winds become 65–118 km/h (39–74 mph) becomes a tropical storm, which propagates generally westward. If winds exceed 118 km/h, the storm is called a hurricane (in the Atlantic and eastern Pacific oceans) or a typhoon (in the western Pacific Ocean) – powerful, self-sustaining atmospheric/oceanic heat engines.

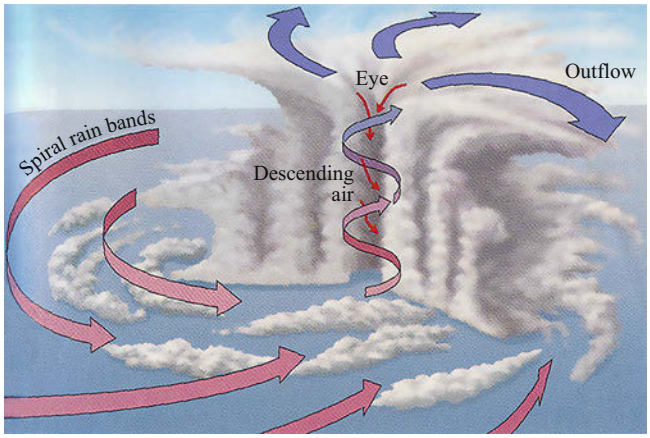


Fig. 2.26 Typical hurricane wind/cloud structure (after [2.11])

Tropical storms and hurricanes have cyclonic or counterclockwise circulation in the Northern Hemisphere because the Coriolis effect deflects the inward air flow toward to the right (Fig. 2.26). (In the Southern Hemisphere, the Coriolis deflection is to the left so that cyclones in the southern tropics have clockwise rotation). In a hurricane, inward flowing air spirals in an anticlockwise sense over the warm, wet ocean toward the eye, where it begins to spiral upward around the *eyewall* that surrounds the eye of the hurricane. As the air rises it cools and its water vapor condenses forming clouds and rain and releasing heat that propels the air even higher. At the top of the eyewall, most of this very cool, dry air begins to descend into the eye and the various spiral bands, warming and becoming even less humid (i. e., dryer) as it encounters the higher pressures at lower levels.Upon reaching the sea surface this very warm, dry air is able to absorb both heat and moisture from the ocean enabling it to rise again, thus re-energizing the hurricane which continues to propagate under its own power.

Table 2.4 The Saffir–Simpson hurricane scale is based on the minimum sustained wind speed criteria for each of the official categories. Typical minimum surface atmospheric pressures and storm surges are given along with more subjective damage descriptions (after [2.10])

Hurricane category	Maximum sustained wind speed		Storm surge [ft]	Minimum Surface pressure [mbar]	Damage
	kilometers/hour	miles/hour			
1	119–153	74–96	4–5	> 980	Minimal
2	154–178	97–111	6–8	979–965	Moderate
3	179–210	112–131	9–12	964–945	Extensive
4	211–250	132–155	13–18	944–920	Extreme
5	> 250	156+	18+	< 920	Catastrophic

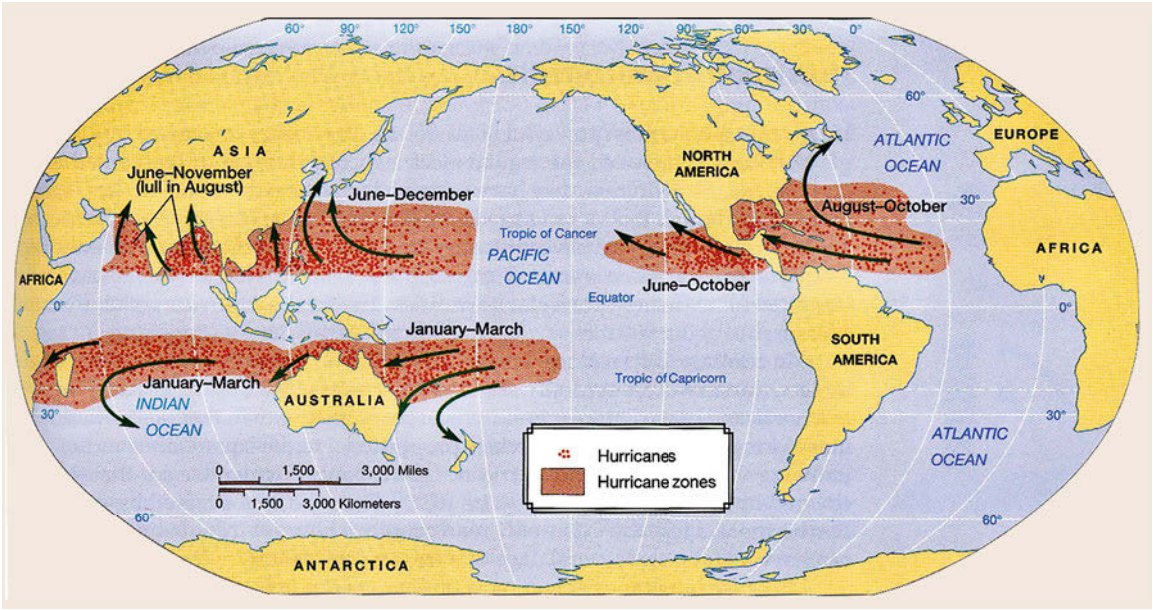


Fig. 2.27 Hurricane and typhoon locations around the globe (after [2.11])

The strength of hurricanes is measured by the Saffir–Simpson scale (Table 2.4), which uses minimum sustained wind speeds to categorize storms from Category 1, a minimal hurricane, to Category 5, the most intense hurricane. A list of the most intense US hurricanes appears in Table 2.5.

Hurricane trajectories are also influenced by larger scale meteorological forces – hence the uncertainty. Under certain conditions, the Atlantic hurricane trajectories tend toward the Caribbean, Florida and Gulf of Mexico; like hurricane Katrina in 2005 (Figs. 2.27 and 2.28). Under some conditions, the trajectories can take more northward paths along the eastern US seaboard; like hurricane Carol of 1954.

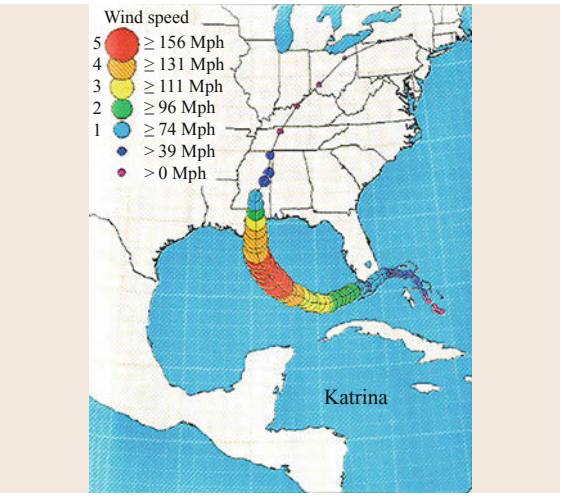


Fig. 2.28 The trajectory of hurricane Katrina showing the evolution in minimum sustained wind intensity and category (after [2.12])

The hurricane–ocean interaction also produces ocean surface waves, currents, and sea level changes or so-called storm surge. The group speeds of the most energetic (i. e., long) hurricane-generated surface waves are faster than typical hurricane propagation speeds and thus they speed out ahead of the hurricane. The local hurricane-generated ocean currents in the near-surface layer are intensified in the direction of the generating winds as the hurricane passes overhead. In the Northern Hemisphere, the highest hurricane winds are on the right hand side of the leading edge of the hurricane (the upper right quadrant in Fig. 2.29). This is because the forward speed of the hurricane itself adds to the speed of the wind around the storm on the right side of the storm.

As a hurricane approaches land, these ocean currents impinge on the land and thus can contribute to increased the sea level rise or storm surge. The storm surge is composed of contributions from the hurricane-generated current field and the elevated sea level that is induced by the very low sea level atmospheric pressure at the center of a hurricane. Thus the storm surge for a particular hurricane generally evolves in varying

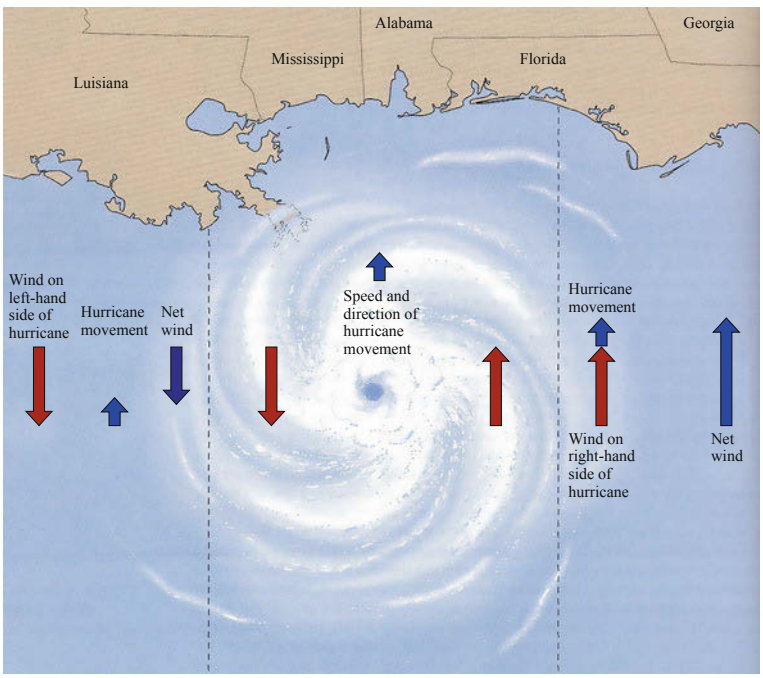


Fig. 2.29 Hurricane horizontal wind structure (after [2.7])

Table 2.5 Characteristics of the most intense hurricanes that made landfall in the US between 1900 and 1996 in terms of their categorization and central air pressure (after [2.10])

Hurricane	Year	Category	Pressure[a] [mb]	[in Hg]	[mm Hg]
Gilbert	1988	5	888	26.2	666.1
Florida Keys	1935	5	892	26.35	669.2
Camille	1969	5	909	26.84	681.7
Andrew	1992	4	922	27.23	691.6
Florida Keys/S. Texas	1919	4	927	27.37	695.2
Lake Okeechobee, Fl.	1928	4	929	27.43	696.7
Donna	1960	4	930	27.46	697.5
Galveston, TX	1900	4	931	27.49	698.2
Grand Isle, LA	1909	4	931	27.49	698.2
New Orleans, LA	1915	4	931	27.49	698.2
Carla	1961	4	931	27.49	698.2
Hugo	1989	4	934	27.58	700.5
Miami, FL/Pensacola, FL	1926	4	935	27.61	701.3
Hazel	1954	4[b]	938	27.70	703.6
SE FL/SE MS/AL	1947	4	940	27.76	705.1
North Texas	1932	4	941	27.79	705.9
Opal	1995	3[b]	942	27.82	706.6
Frederic	1979	3	946	27.94	709.7
Betsy	1965	3	948	27.99	710.9
Fran	1996	3	954	28.17	714.8

[a] Pressure exerted by the atmosphere at sea level is equal to 1 atmosphere, which is equal to 760 mm of mercury (760 mmHg), 1.01325 bar, 1013.25 mar, 29.92 inches of mercury (29.92 in Hg), and 14.7 pounds per square inch (14.7 psi)
[b] Hurricanses traveling more than 50 km/h.

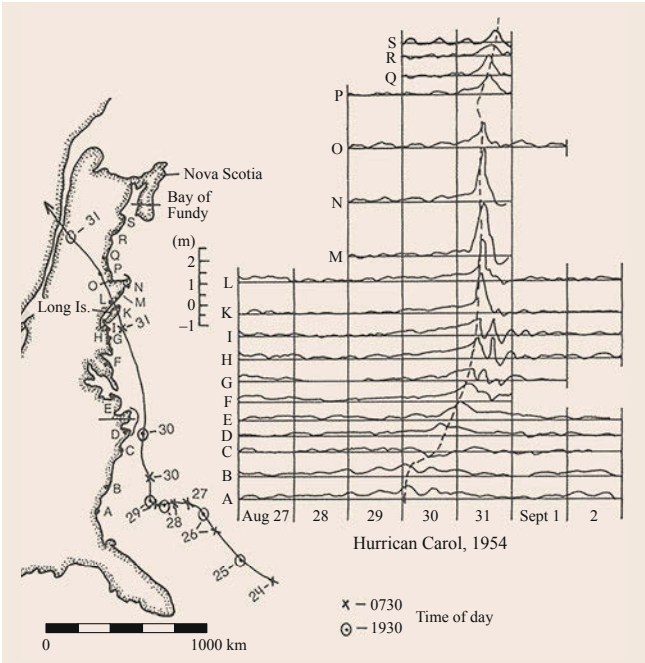


Fig. 2.30 The propagation of hurricane Carol along the US East Coast induced storm surge at coastal locations as hurricane Carol (after [2.13])

ways in tandem with the evolving hurricane (Hurricane Carol 1954, Fig. 2.30). Hurricane damage is caused by the combination of high winds, flooding, and storm surge. Surface winds can cause an abnormal (i. e., non-tidal) rise in sea level, reaching 1–6 m above normal sea level. Along the Gulf Coast, the estimated 330 km/h winds of Hurricane Camille in 1969 caused a maximum storm surge of 8 m, with 3-m-high wind-driven waves atop the surge. The end result was a l0-m-high wall of water approaching the Mississippi coastline. Failure to evacuate caused significant loss of life.

Hurricane-forced ocean currents are transient and intense in proportion to the wind intensity. The Integrated Ocean Observing System (IOOS) high frequency radar network along the US East Coast measured surface currents induced by 29 October 2012 *Superstorm* Sandy – the meteorological combination of a tropical hurricane and an extra-tropical storm that made landfall in New Jersey at about 2000 EDT 29 October 2012. The maps in Fig. 2.31 show that Sandy's winds overwhelmed the normal Mid-Atlantic 10–40 cm/s primarily wind-driven current patterns (e.g., 2000 EDT 27 October 2012) a full 24 h before landfall (2000 EDT 28 October 2012). The Sandy-induced intensifying southward currents that exceeded 1 kt (≈ 50 cm/s).

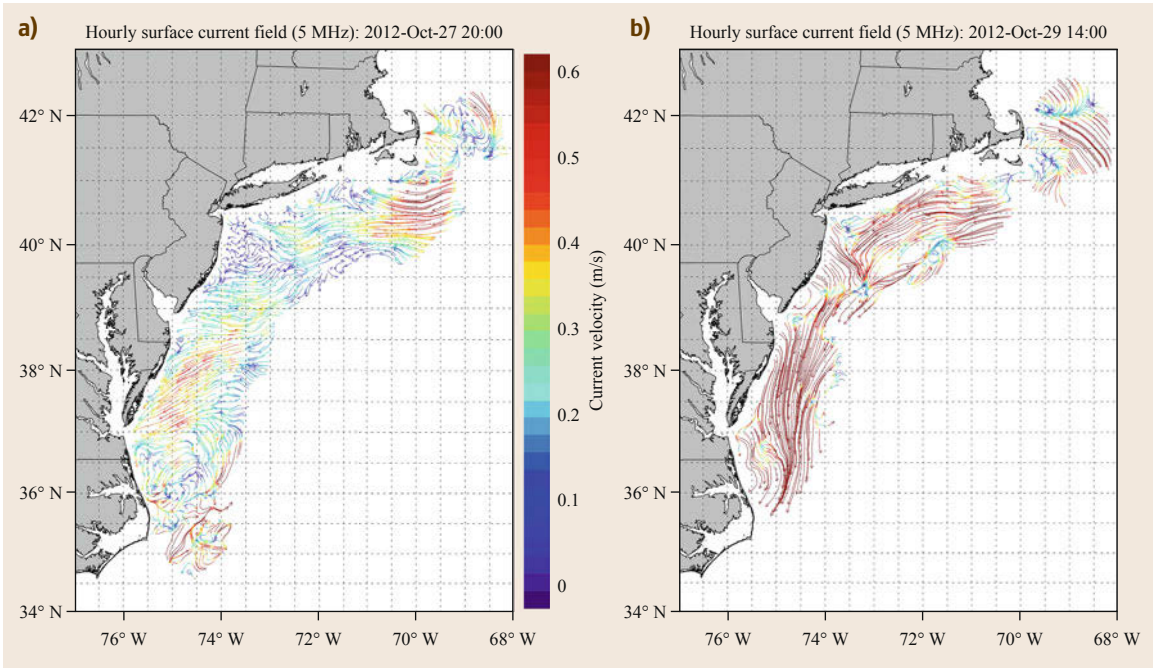


Fig. 2.31a,b Surface currents maps measured by a network of high frequency radar (HFR) sites in the coastal mid-Atlantic at: **(a)** 2000 EDT 27 October 2012 ... + 48 h before Superstorm Sandy landfall in New Jersey; and **(b)** 1400 EDT 29 October 2012 ... + 4 h – the last complete map before physical damage and power failures shut down the HFR network (after [2.14])

2.5.4 Wind-Forced Upper Ocean Response with Earth Rotation Effects

At time scale exceeding about one-half day, Earth rotation effects can become important. For example, in the aftermath of impulsive wind forcing, such as that associated with a weather front, the ocean often exhibits circular motion called inertial flow. More steady wind forcing of a frictional upper ocean often leads to Ekman flow. With larger scale wind shear, that forcing can induce Ekman divergence/convergence and associated upwelling and downwelling.

Inertial Flow

If there are no other significant horizontal forces present, then the Coriolis force will turn a moving water parcel into a clockwise (anticlockwise) circular inertial motion. The dynamics of *inertial motion* can be thought of as a dynamic balance between two pseudo-forces, namely the Coriolis force $\boldsymbol{CF}$ and the centrifugal force $\boldsymbol{f}_c$ that is associated with the acceleration of cir-

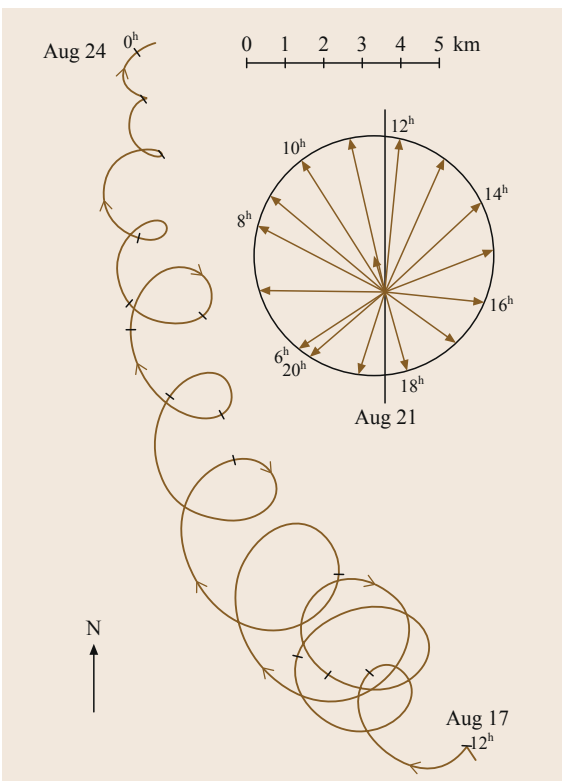


Fig. 2.32 A 7-day progressive vector diagram derived from moored current measurements time series, indicating clockwise circular inertial motion superposed on a northwestward flow (after [2.15]) ►

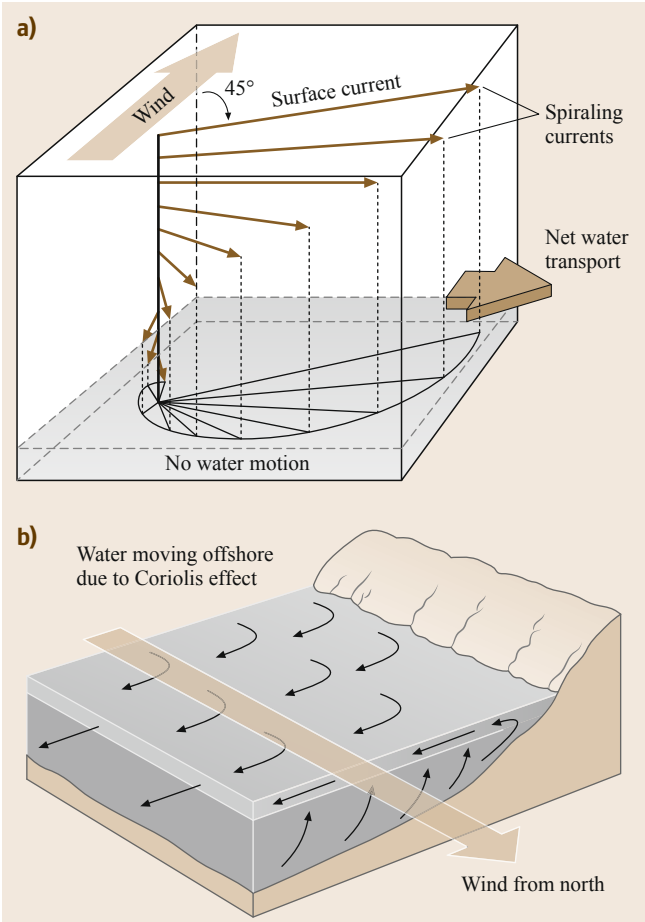


Fig. 2.33a,b Schematics of a Northern Hemisphere (NH) ocean being forced by a steady winds; showing **(a)** the surface velocity at 45° to the right of the surface wind stress and the downward spiral current structure – the vertical average of which yields a net Ekman transport to the right throughout the Ekman layer (after [2.7]). **(b)** Persistent winds along a NH coast produce a persistent Ekman transport *to the right* and an associated upwelling (after [2.10])

cular motion. Thus the force balance (per unit volume) for inertial motion is

$$\boldsymbol{CF} = \rho f \boldsymbol{V} = \rho \frac{\boldsymbol{V}^2}{R} = \boldsymbol{f}_c$$

so

$$f\boldsymbol{V} = \frac{\boldsymbol{V}^2}{R}$$

and the radius of the circle R

$$R = \frac{\boldsymbol{V}}{f} = \frac{\boldsymbol{V}}{2\Omega \sin\phi} .$$

The circular water parcel trajectory has an oscillation period called the inertial *period* T

$$T = \frac{2\pi R}{v} = \frac{2\pi}{f} = \frac{\pi}{\Omega \sin\phi}$$

which is independent of R! The inertial period is one half of a *pendulum day* which is defined as $2\pi/\Omega \sin\phi$. (A pendulum day is the time it takes for the vertical plane, in which pendulum swings, to rotate a full 360° (or 2π radians) relative to the Earth – a measure of the overhead Earth rotation rate at particular latitude). In practice, particularly after storm, it is not unusual to detect inertial motion (Fig. 2.32).

Ekman Flow

For wind stress-forced flow situations with times scales exceeding a half day, the effects of wind stress can be explored in terms a simplified form of the horizontal equations of motion in which Coriolis and frictional forces are balanced. The solution to these equations, assuming the application of a northward surface wind stress $\boldsymbol{\tau}_s = \tau_s \boldsymbol{j}$, is called *Ekman flow* whose respective eastward and northward components are

$$u^E = V_o e^{\pi z/D} \cos\left(\frac{\pi}{4} + \frac{\pi z}{D}\right) , \qquad (2.15a)$$

$$v^E = V_o e^{\pi z/D} \sin\left(\frac{\pi}{4} + \frac{\pi z}{D}\right) , \qquad (2.15b)$$

where $V_o = (\tau_s/\rho)/\sqrt{f\nu_z}$ is the surface velocity, $D = \pi\sqrt{2\nu_z/f}$ is the Ekman depth, and $\nu_z = A_z^e/\rho$ is the eddy kinematic viscosity.

While the detailed form of Ekman flow (Fig. 2.33) has only been recently verified experimentally, the existence of a vertically averaged flow to the right of the wind stress (Northern Hemisphere) called the Ekman transport has been observed for a long time. On ocean basin scales, as discussed below, Ekman transport convergence associated with the basin-scale wind fields create the conditions for large-scale ocean gyre currents. Near coasts, persistence along coast wind forcing generates offshore or onshore Ekman transports which lead to coastal upwelling/downwelling (Fig. 2.33). This process is particularly prominent along the US west coast during the summer when equatorward coastal winds persist.

2.5.5 Wind-Driven Currents: Ocean Basin Scale

Global-scale winds drive similar basin-scale gyre flows distinguished by intensified poleward flowing western boundary currents and less distinct equatorward flowing eastern boundary currents (Fig. 2.34). These mid-latitude gyre flows are connected by both tropical ocean current systems that straddle the equator in

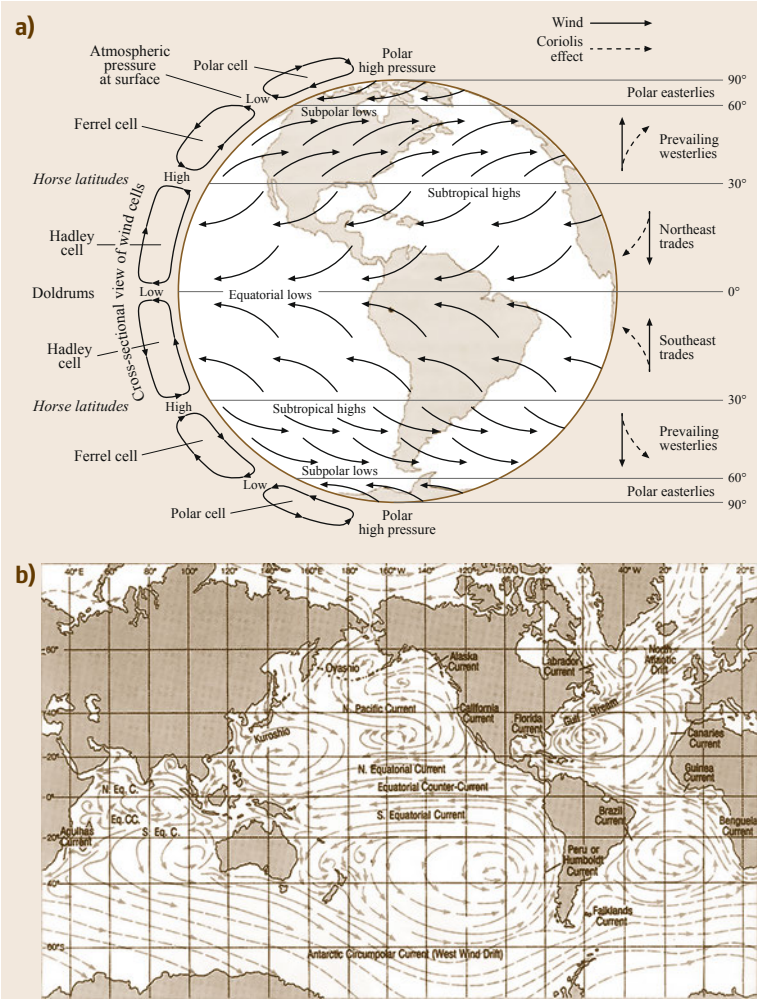


Fig. 2.34 **(a)** Average surface global winds divide into zonal bands due to Earth rotation effects. In both hemispheres, there are zones for the tropical easterly trade winds (equator–30°N and S), prevailing westerlies (30–60°N and S), and polar easterlies (60–90°N and S). **(b)** The major ocean basins feature gyre current systems that are interact through east–west equatorial current systems in the Atlantic and Pacific and the Antarctic Circumpolar Current in the Southern Ocean (after [2.7])

the Atlantic, Pacific, and Indian oceans and the Southern Ocean that circles Antarctica. The transports of the principal currents are presented in Table 2.6.

The dynamics of much of the open ocean flows is governed by one of the geostrophic force balances in which the Coriolis force balances the horizontal pressure gradient force for an arbitrary flow direction that is perpendicular to the force direction (Fig. 2.35).

The scalar form of the force balance is

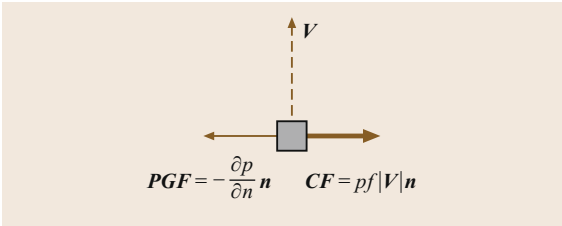
$$\rho f|\boldsymbol{V}| = \frac{\partial p}{\partial n}$$

So that the geostrophic flow magnitude is

$$V_{\mathrm{g}} = |\boldsymbol{V}| = \frac{1}{\rho f}\frac{\partial p}{\partial n}\,.$$

Table 2.6 Transport of major ocean surface currents

Location	Name	Maximum current [cm/s]	Volume transport [10^6 m^3/s]
Western Boundary Currents	Gulf Stream	200–300	400
	Kuroshio Current	> 200	50
	Brazil Current	50–100	10
Southern Ocean	Antarctic Circumpolar Current	15	150
Equatorial Currents	North Pacific Equatorial Current	20	45
	Equatorial Undercurrent	100–150	40
Eastern Ocean	Peru Current	10–50	20



The Cartesian representation of the velocity, Coriolis force, and pressure gradient force are

$$\boldsymbol{V} = u\boldsymbol{i} + v\boldsymbol{j}\,,$$
$$\boldsymbol{CF} = \rho f v\boldsymbol{i} - \rho f u\boldsymbol{j}\,,$$
$$-\partial p/\partial n = -\partial p/\partial x\boldsymbol{i} + -\partial p/\partial y\boldsymbol{j}\,.$$

Strictly speaking geostrophic flow is a steady-state, rectilinear flow that is perpendicular to the Coriolis/pressure gradient force balance which is to the right (left) in the Northern (Southern) Hemisphere as illustrated in Fig. 2.35. However, geostrophic flow is a practical model because much of the open ocean flows vary slowly compared to a pendulum day with very little curvature. The pair of diagrams in Fig. 2.36 show how the ocean has adjusted its internal mass structure (through seaward flows of warmer water and landward flows of cooler water) to generate the appropriate sea surface slope-induced pressure gradients that balance the Coriolis forces associated with what is a primarily wind-driven Gulf Stream flow.

Fig. 2.35 Generalized geostrophic flow force balance between the Coriolis force $\boldsymbol{CF} = \rho f|\boldsymbol{V}|\boldsymbol{n}$ and negative pressure gradient force $\boldsymbol{PGF} = -(\partial p/\partial n)\boldsymbol{n}$, which are acting on a water parcel with a velocity $\boldsymbol{V}$ in the orthogonal direction. The Cartesian components of different vectors are given ◄

The combined prevailing westerlies and easterly trade winds in the subtropical latitude band between 10° and 50° set up a poleward sheared surface wind stress field (Fig. 2.34a) that drives the basin-scale geostrophic gyre flows in the principal ocean basins. The Ekman transport generated by the westerlies and the trade winds converges in the vicinity of 30° in both the northern and southern Atlantic and Pacific Ocean basins. This situation creates the tendency for east–west zonal geostrophic flow perpendicular to Ekman transport-induced north/south pressure gradients and Coriolis force. However, in most ocean basins (the exception being the Southern Ocean around Antarctica), continents block the zonal ocean flow. So this wind-driven oceanic pressure field is a steady-state high pressure cell around which the geostrophic flow circulates. The combined effects of the geostrophic flow around the pressure cells and the earth's rotation lead to clockwise (or counterclockwise) basin scale *wind-driven* gyres in the Northern (or Southern) Hemisphere ocean basins (Fig. 2.34b). The gyroscopic interaction between the gyre motion and earth rotation leads to intensification of gyre currents along the western bound-

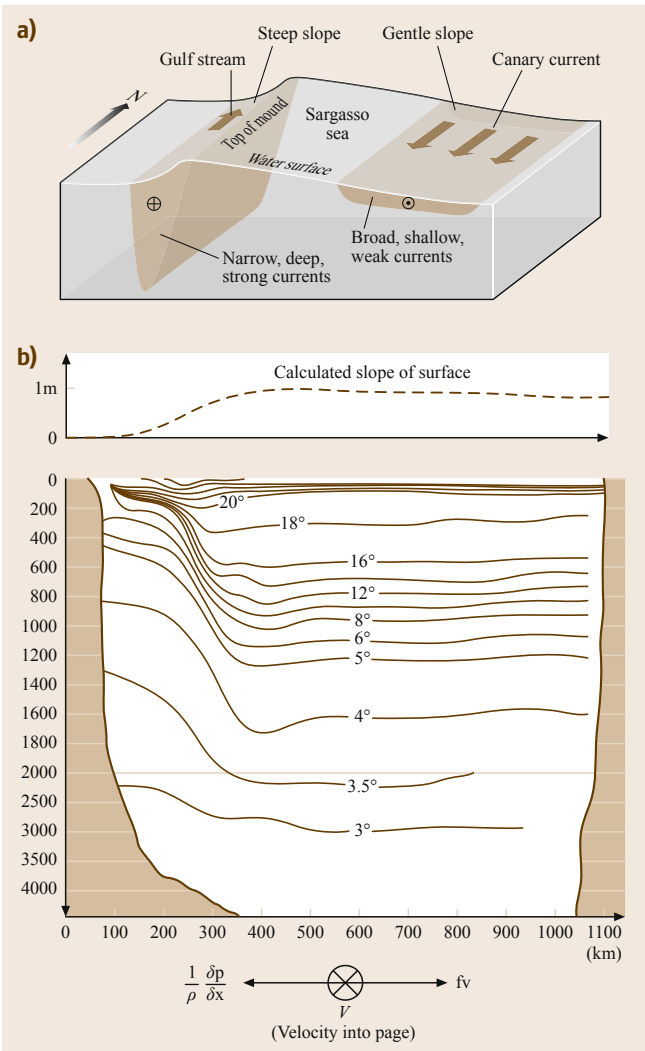


Fig. 2.36 **(a)** Schematic of the principal geostrophic flows in the North Atlantic basin. The exaggerated sea level distortions induce the pressure gradients that balance the Coriolis forces associated with the respective poleward Gulf Stream and equatorward Canary Current flows (after [2.7]). **(b)** A contoured temperature transect left-to-right across the Gulf Stream between Florida and Bermuda showing how the warm open ocean water raises sea level relative that associated with the cooler near-coast water and the relevant geostrophic flow force balance (after [2.6]) ◄

ary of all the ocean basins, exceeding 200 cm/s in the Gulf Stream of North Atlantic and in the Kuroshio of the North Pacific. The total transport of the Gulf Stream has been estimated to be in the range of 50–90 $\times 10^6$ m^3/s; with that of the Kuroshio being somewhat less (Table 2.6).

Despite the greater intensity of winds in the Southern Hemisphere, the intensity of the surface western boundary currents is less than that for either the Gulf Stream or Kuroshio. The Brazil current in the western South Atlantic has typical speeds of 50–100 cm/s and penetrates to depths of only about 400 m. in contrast to 800 m for the Gulf Stream. The reduced transport of the Brazil Current may be related to the presence of deeper western boundary undercurrents and their role in transporting heat poleward. The reduced intensity of the Australian current in the western Pacific and the Agulhas Current in the western Indian Ocean may be related to the lack of complete continental blockage of zonal flow.

The equatorward flow along the eastern boundaries of the wind-driven gyre as represented by the Canary and California Currents is generally less intense and less organized than the western boundary currents. The exception is the Benguela Current (Fig. 2.34b), which is an apparent extension of the Agulhas Current in the South Atlantic.

At tropical latitudes straddling the equator, significant trade winds can drive surface currents in the zonal circulation patterns shown in Figs. 2.34 and 2.37, patterns that exist in all major oceans. The wind-driven westward flow (Coriolis effects are weak near the equator) in a zone bracketing the equator is blocked by the presence of land masses. A resulting eastward pressure gradient force drives a return equatorial undercurrent jet along the equator and a seasonal countercurrent in the doldrum region. These countercurrents, while primarily zonal, are also very important connection between the major ocean basins.

The circulation at poleward latitudes greater than 50° differs greatly in the different oceans because of differences in the geography of the Northern and Southern Hemisphere ocean basins. The North Atlantic Ocean effectively ends at the sills at about 60°N and is weakly associated with the shallower Arctic Ocean, which has a circulation of its own. In the South Atlantic, there is a zonal ocean which circles the entire Antarctic continent. The winds are important in driving the broad Circumpolar Current (Fig. 2.34). Density adjustments in the interior produce pressure gradient forces which balance the Coriolis forces to produce a current profile which diminishes from about 15 cm/s at the surface to near zero at 3 km depth. So although Circumpolar Current speeds are considerably less than Gulf Stream speeds, the transports associated with the Circumpolar

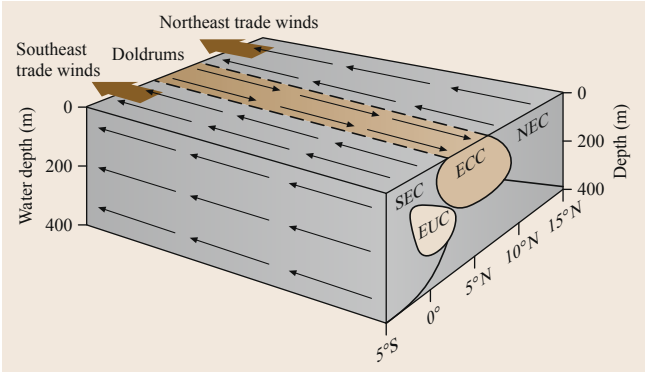


Fig. 2.37 Generalized Scheme of equatorial currents. Schematic of the westward trade wind-driven south (SEC) and north equatorial currents (NEC) bracketing the eastward flowing equatorial countercurrent (ECC) and equatorial undercurrent (EUC) (after [2.7])

Current are 130–190 $\times 10^6$ m^3/s and larger than Gulf Stream transports. As such, this current has the largest transport of any in the world's oceans and plays a crucial role in controlling the deep ocean circulation.

2.5.6 Gulf Stream Ring Flows

Before leaving the discussion of the wind-driven surface currents, it is important to note that western boundary currents are unstable leading to considerable changes on monthly time scales. For example, the Gulf Stream meanders and on occasion sheds pairs of eddies; warm core rings containing warm Sargasso Sea water or cold core rings (Fig. 2.38) containing colder waters with origins in the north. These rings are characterized by high currents – Gulf Stream-like currents of 50–100 cm/s. This eddy-shedding process enables effective poleward heat transport. Once these rings are shed, with the large Gulf Stream-like ocean currents, propagate affecting their local environments. Cold-core rings south of the Gulf Stream also propagate southwestward to be eventually entrained by the Gulf Stream in the region of Cape Hatteras. Warm-core rings with clockwise circulation are observed to propagate slowly southwestward along the continental margin on the northern side of the Gulf Stream. Satellite infrared imagery *is very* helpful in tracking these anomalously warm and cold eddies.

In reality, all of the upper oceans of the world are filled with eddy-like current structures embedded in the large-scale, wind-driven gyres described above. Typical currents in these eddies are about 10–20 cm/s. in contrast to the 1 cm/s gyre mean flow, and have lateral extents of about 100 km. Such eddies have been observed to propagate through a region in a period of months. The role of these eddies in the overall transport

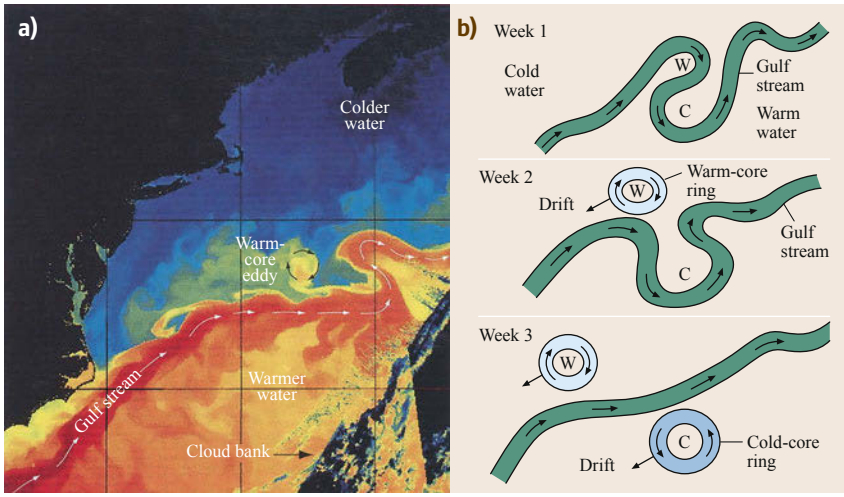


Fig. 2.38 **(a)** The warm Gulf Stream becomes unstable after leaving coast at Cape Hatteras, NC, and spawns both cold and warm core rings. **(b)** Gulf Stream ring formation begins with a meander that semienclosesses pockets of the warmer Sargasso sea water. and colder Slope Sea water. Every few months first a clockwise warm core ring and then an anticlockwise cold core ring will snap off. These rings will then drift equatorward on their respective side of the Gulf Stream – many times being reabsorbed by the Gulf Streams after several months of drifting (after [2.7])

of heat and other constituents is only just beginning to be understood.

The circulation in the tropical regions is likewise unsteady. A very important tropical ocean process is one in which warm equatorial waters in the western Pacific flow from west to east in response to moderating trade winds – with an uneven periodicity of 2 to 7 years. This anomalous flow of warm equatorial water has devastating effects on the fisheries along the Western Coast of South America. This ocean-centric process – called *El Nino* – is now known to be part of the larger scale atmospheric process called the Southern Oscillation. thus, the whole El Nino-Southern Oscillation process is referred to as ENSO.

2.6 Deep Ocean Currents

Superimposed on the wind-driven upper ocean circulation, which is mainly confined by the main pycnocline (or density gradient; Fig. 2.2) is a less intense thermohaline circulation. This global ocean scale overturning cell is driven primarily by atmospheric cooling-driven sinking of waters in the polar regions and closed through a poorly understood systems of vertical advections. The net effect of these vertical return flows (average vertical velocities of centimeters/day) is to maintain the depth of the main pycnocline against the downward heat diffusion in the upper ocean (Fig. 2.39).

Early ideas of deep ocean circulation were based on the ideas that water at polar latitudes – delivered primarily by western boundary currents – becomes relatively dense as it was cooled by the atmospheric winds and sank to depths in accordance with its density – with the coldest water sinking the deepest. As the water slowly warmed over hundreds if not thousands of years, it became less dense and rose to the surface distributed throughout the global oceans (Fig. 2.39).

The Stommel–Aron [2.16] dynamic theory of the thermohaline circulation of the world's oceans consisted of (a) polar region sinking – primarily North Atlantic and Antarctica – that (b) fed a network of deep western boundary currents (Fig. 2.40) that distributed the water to (c) a selected set upwelling zones from where (d) it returned to the upper ocean western boundary currents. Subsequent observations of the Atlantic deep western boundary current have verified one of the principal predictions of the Stommel–Aron theory.

Thus the overall picture of global ocean circulation is one that combines wind-driven and thermohaline flows in ways that transport enough heat poleward to

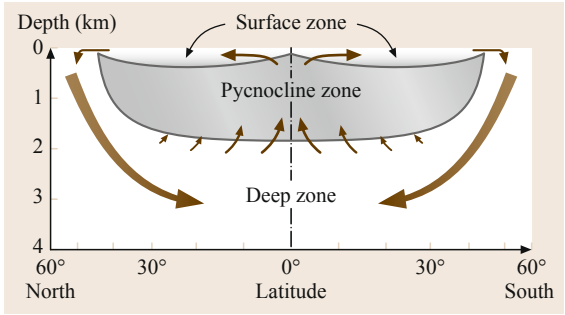


Fig. 2.39 Schematic of a three-zone global ocean in which thermohaline or meridional overturning circulation consists of poleward flow in the surface layers, sinking in the polar regions and upwelling through the very stable density gradient or pycnocline zone

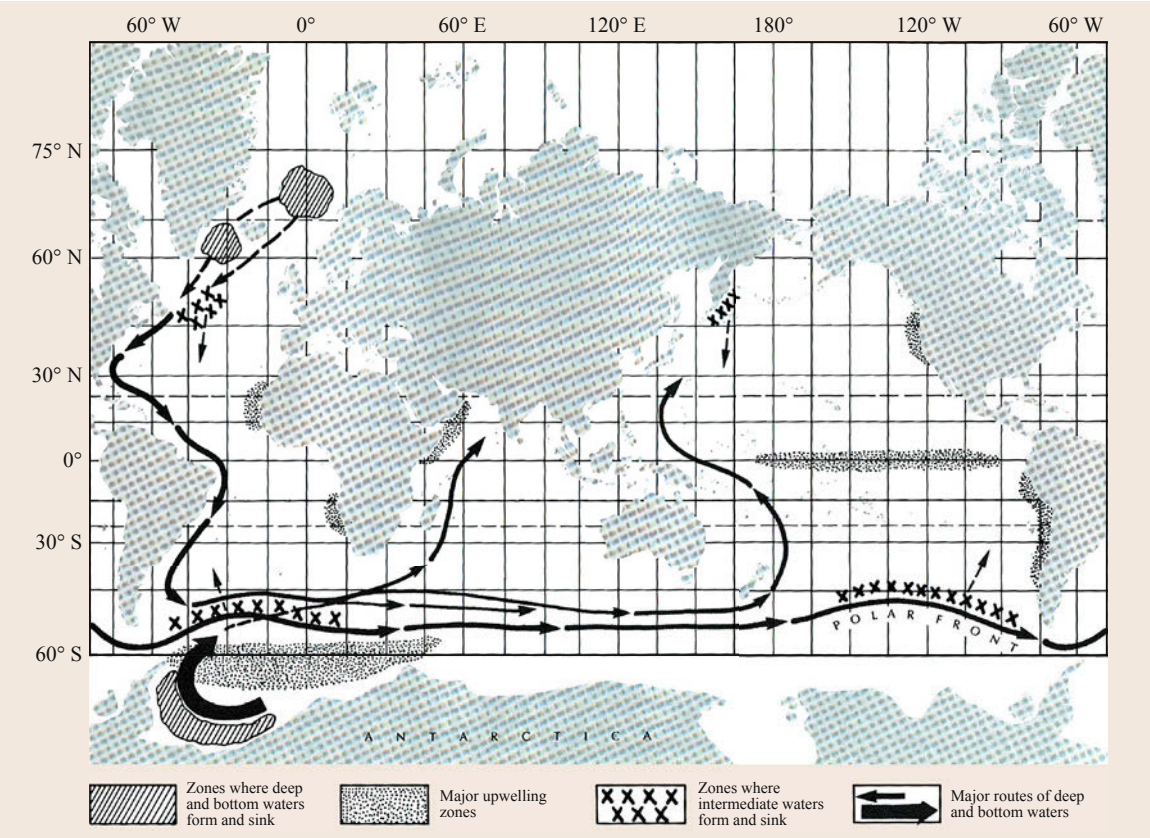


Fig. 2.40 The Stommel–Arons theory of deep ocean circulation consisting of a distributed set of deep western boundary currents that were connected via the Southern Ocean and to upwelling in the Southern, equatorial, and selected coastal oceans (after [2.4])

balance a polar hear deficit. The North Atlantic composite ocean circulation consists of an intense Gulf Stream that is opposed by the southward flowing deep western boundary currents composed of North Atlantic polar waters. The South Atlantic composite ocean circulation consists of a relatively weak Brazil current that is aided by a southward deep western boundary currents of North Atlantic polar waters. The medium intensity Kuroshio Current in the western Pacific is aided by deep underwater currents in transporting heat poleward.

2.7 Coastal Ocean Currents

The relative shallowness of the continental shelf and coastal regions of the world's oceans leads to a distinct current regime in which short term fluctuations are much larger than longer term mean flows. In general, the shallower coastal ocean waters respond more readily to wind forcing and exhibit important variability on the 2–10-day time scales of the mid-latitude synoptic weather systems which pass through the region. Density fronts caused in part by the proximity of the coastal ocean to significant fresh water river sources are associated with significant intensification of shelf and coastal currents. Of course, there are always tidal currents, which in many cases are as relatively weak (5 cm/s) as in the deep ocean. However, other places in the coastal ocean exhibit significantly stronger tidal currents that play a more important role in the mixing of shelf and coastal water masses (see next section for more on that).

Wind-forced shelf currents vary over longer time scales than tidal currents and thus are usually more effective in transporting nutrients, biota, pollutants, and other constituents on the shelf. Wind-forced

shelf fluctuations are usually most intense (typically 10–50 cm/s) in the along-shelf direction, with the maximum currents occurring at mid-shelf.

The presence of significant across-shelf density changes, such as those found along the East Coast of North America, leads to a modified current structure in which maximum along-shelf currents are concentrated in the density front along the outer shelf. These currents change over longer time scales in response to the more slowly changing density structures. Occasionally, deeper ocean currents in the form of Gulf Stream warm core rings penetrate onto the shelf and disturb the *normal* distribution of currents found there.

On many continental shelves, including the East Coast of North America, there is a more or less steady along-shelf flow superimposed on the fluctuating tidal, buoyancy-induced and wind-driven current field. Typical velocities of these steady flows are usually about 5–10 cm/s. Thus, while astronomical wind forcing plays an important role in shelf currents, their superimposition leads to complex and, as yet, unpredictable patterns crudely described by the characteristics discussed above.

2.8 Ocean Surface Tides

The astronomically forced surface gravitational tide has wavelengths that are long compared to the ocean depth and thus is a shallow water wave. Tidal waves are strongly forced by the motion of the earth relative to the moon and the sun and thus differ from *freely* propagating wind-generated surface gravity waves. Unlike most short period wind-driven waves, tidal waves have periods on the order of a day – enough to be influenced by the earth's rotation (Fig. 2.2).

The tidal response of the global oceans is complex because of the combined astronomical *tidal forcing* of the earth, moon and sun. The observed tides differ significantly from location to location because of the shallowness and geographical complexity of the global ocean basins. Observed tides throughout the world's ocean are composed of differing mixtures of both twice-a-day (semidiurnal) and once-a-day (diurnal) tidal variability. For example, the tides in Pakhoi, China (Fig. 2.41) are primarily once-a-day (or diurnal), while the tides in Port Adelaide, Australia are primarily twice-a-day or semidiurnal. Although the tides at both locations include both semidiurnal and diurnal tides, the minor tidal constituent is too small to detect except when the tidal range is small. Note that about every 14

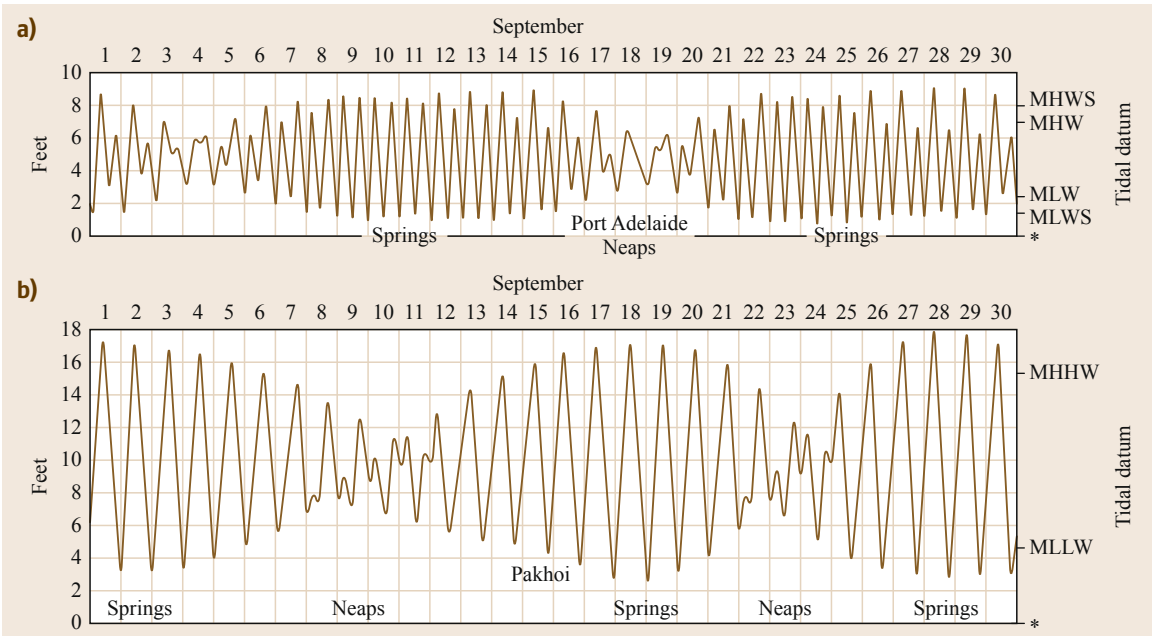


Fig. 2.41a,b Observed tides in **(a)** Port Adelaide are semidiurnal (twice-a-day) and in **(b)** Pakhoi diurnal (once-a-day). Tidal ranges in both locations exhibit a spring-neap cycles that are out of phase (after [2.15])

Table 2.7 The six basic astronomical tide-producing force frequencies are given in radians per unit time and degrees per solar day

Source	Angular frequency
Earth rotation	$\omega_\tau = \frac{2\pi}{\text{lunar day}} = 360° - 12.2°/d$
Lunar revolution	$\omega_s = \frac{2\pi}{\text{sidereal month}} = 13.176°/d$
Solar declination	$\omega_h = \frac{2\pi}{\text{solar year}} = 0.985/d$
Revolution of lunar perigee	$\omega_p = \frac{2\pi}{8.87\,\text{years}} = 0.111°/d$
Precession of lunar node	$\omega_N = \frac{2\pi}{18.6\,\text{years}} = 0.0529°/d$
Revolution of solar perigee	$\omega_{p1} = \frac{2\pi}{20\,000\,\text{years}} = 0.00005°/d$

days the range of the tidal sea level excursions and associated currents cycle through a maximum called spring tides and a minimum called neap tides (Fig. 2.41). This phenomenon occurs because the tide-producing forces associated with the moon and the sun reinforce (times of full and new moon) and oppose (times of half moons) each other twice a month.

As it turns out, there are five basic frequencies (Table 2.7) associated with astronomical tide- producing forces. Because the interaction of the astronomical tidal forcing and the oceanic response is highly nonlinear, the observed ocean tide anywhere on the earth can be decomposed into 400 *partial tides* (or *tidal species*). Each partial tides has a unique frequency ω_i that is determined by a set of integer weights a_j (for $j = 1, 2, 3, 4, 5,$ and 6) that multiply these six basic frequencies according to

$$\omega_i = a_1\omega_\tau + a_2\omega_s + a_3\omega_h + a_4\omega_p + a_5\omega_N + a_6\omega_{pl} \,.$$

Doodson [2.17] developed a shorthand notation for the weights called the Doodson number to define the different partial tides. (For example, the interaction of lunar declination changes and earth rotation leads to a pair of frequencies $\omega_\tau + 4\omega_s$ and $\omega_\tau - 4\omega_s$; which have Doodson number specifications (1 4 0 0 0 0) and (1 −4 0 0 0 0), respectively). The short list of the most important partial tides in Table 2.8 show equilibrium tidal forcing amplitude ratios relative to that of the M_2 tidal constituent – at many sites the most important constituent.

For practical purposes, the astronomical tide-producing forcing is fixed for all time. So the tidal response of the world's oceans, while it is complex spatially, is also fixed for all time. Thus to uniquely determine the tide at a particular location – say sea level $SL(t)$ – the task is to determine the amplitude (H_i) and

Table 2.8 List of the identities, periods, and relative equilibrium tidal amplitudes of the 19 of the most important partial tides (out of the 400)

Name of partial tides	Symbol	Speed (degrees per mean solar hour)	Period (solar hours)	Coefficient ratio $M_2 = 100$
Semidiurnal components				
Principal lunar	M_2	28.98410	12.42	100.0
Principal solar	S_2	30.00000	12.00	46.6
Larger lunar elliptic	N_2	28.43973	12.66	19.2
Luni-solar semidiurnal	K_2	30.08214	11.97	12.7
Larger solar elliptic	T_2	29.95893	12.01	2.7
Smaller lunar elliptic	L_2	29.52848	12.19	2.8
Lunar elliptic second order	$2N_2$	27.89535	12.91	2.5
Larger lunar evectional	η_2	28.51258	12.63	3.6
Smaller lunar evectional	λ_2	29.45563	12.22	0.7
Variational	μ_2	27.96821	12.87	3.1
Diurnal components				
Luni-solar diurnal	K_1	15.04107	23.93	58.4
Principal lunar diurnal	O_1	13.94304	25.82	41.5
Principal solar diurnal	P_1	14.95893	24.07	19.4
Larger lunar elliptic	Q_1	13.39866	26.87	7.9
Smaller lunar elliptic	M_1	14.49205	24.86	3.3
Small lunar elliptic	J_1	15.58544	23.10	3.3
Long-period components				
Lunar fortnightly	M_1	1.09803	327.86	17.2
Lunar monthly	M_m	0.54437	661.30	9.1
Solar semiannual	S_{sa}	0.08214	2191.43	8.0

phase (κ_i) for each of the partial tides (or *tidal constituents* η_i with its own ω_i) given by

$$\eta_i = H_i \cos(\omega_i t + \kappa_i) \; .$$

The harmonic constant H_i/κ_i pairs for each of the partial tides can be obtained by a harmonic analysis of a sufficiently long measured record – in this case sea level $SL(t)$ – from the site of interest. Modern harmonic analyses of measured records are based on a spectral analysis that has been modified to consider the variability at only the special partial tidal frequencies. Once determined, the partial tides can be constructed and summed to produce the total astronomically forced tidal sea level record according to

$$h(t) = \sum_i \eta_i = \sum_i H_i \cos(\omega_i t + \kappa_i) \; .$$

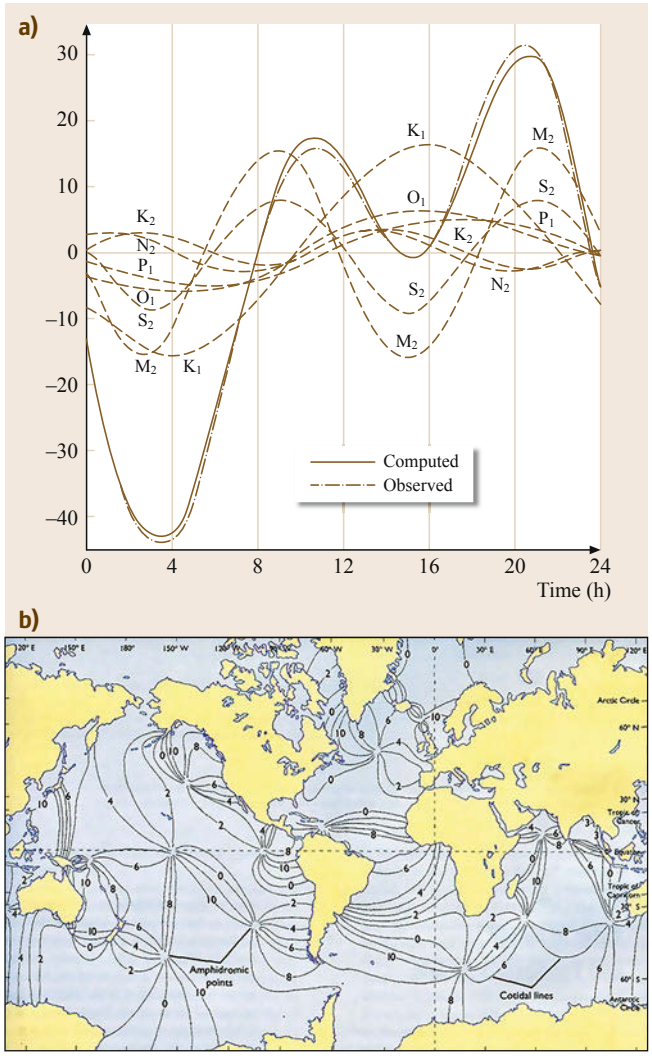


Fig. 2.42 (a) Harmonic decomposition of an observed sea level record (*dash-dot*) into several partial tidal constituents, which when added yield the computed tidal sea level (*solid*) part of the observed sea level record (after [2.10]). **(b)** The M_2 semidiurnal harmonic constants for an world's ocean array of sea level stations are used to construct a cotidal chart for the Atlantic Ocean tide. The solid cotidal lines mark the Greenwich times of high tide (after [2.7])

The harmonic constant H_i/κ_i pairs can be used to predict the future (present and past) tidal variability at the particular site. For most purposes, less than 10 tidal constituents are required to describe the tide adequately at a particular station. Figure 2.42a example shows the results of a tidal harmonic analysis of a sea level record in terms of seven tidal constituents. Note that this method is useful in predicting only that part of the measured sea level fluctuation which (2.1) occurs at astronomical frequencies and (2.2) is *phase locked* to the astronomical forcing.

Harmonic analysis results of a sea level record are widely useful because the principal contribution to sea level variability is usually at tidal frequencies. This helps to explain the relatively small difference between the observed sea level and the computed tidal sea level – a difference called the nontidal *residual sea level*. The residual or tidal signal *noise* can be of interest because it is due to other oceanic processes including weather forced phenomena, hydrodynamic nonlinearities, and wave phenomena at nontidal frequencies.

The maps of the harmonic constants can be used to define the patterns of the individual partial tides. The pattern of the very important M_2 semidiurnal tide is mapped for the Atlantic ocean in Fig. 2.42 – right in terms of its lines of constant tidal range (*cotidal lines*) and constant tidal phase (*copahse lines*). This *cotidal chart* for the M_2 semidiurnal tide reveals the classic signature of an *amphidromic system* in the North Atlantic; in which high tide, ranging from zero tidal range at its central *amphidromic point* to its maximum at its coastal extremes, marches around the basin in 12.42 h. The world's oceans consist of about six major interacting amphidromic systems.

The results of tidal harmonic analyses of ocean currents are more difficult to interpret than those from sea level because significant portions of ocean current variability (2.1) is due to wind – and not tidal forcing, and (2.2) even the tidal frequency variability is often are not phase locked to astronomical forcing. Internal gravity wave-generated currents can represent another major source of noise for the tidal harmonic analysis of currents. It is not unusual for internal wave currents with complex vertical structures to be superimposed on surface (or external) tidal currents with their more depth-

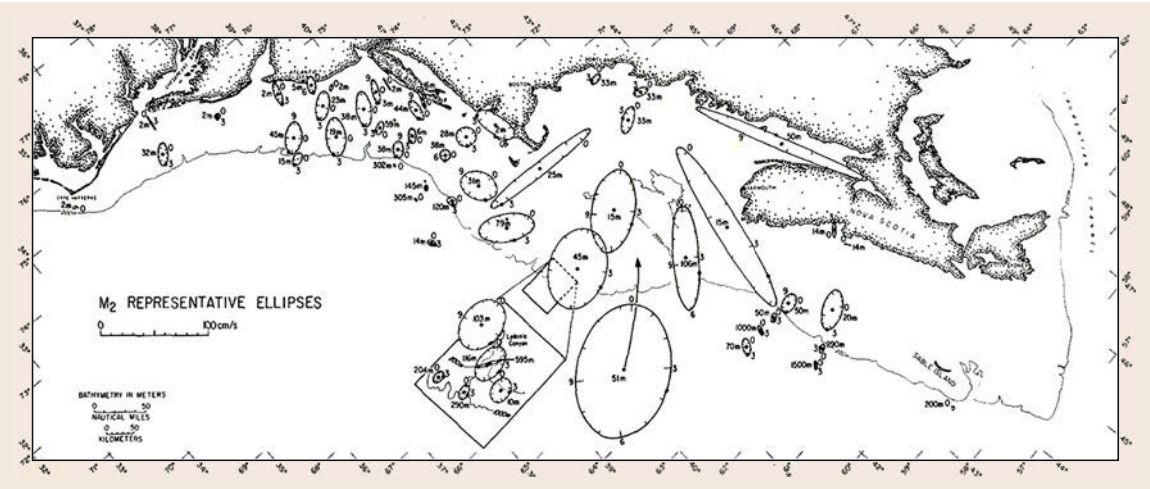


Fig. 2.43 Representative M_2 tidal current ellipses; representative of the mid-water column (water depths given). Ticks on ellipse edge indicated Greenwich hour, with the time of Boston high water indicated by the dot between hours 3 and 4. *Asterisks* identify currents 1 m above bottom, with bathymetry in meters (after [2.18])

independent structures. Therefore, relative to sea level records, current records need to be longer so that uncertainties in the tidal harmonic constants are comparable.

This guideline is less true for places like Georges Bank (at the entrance to the Gulf of Maine) where surface (or external) tidal currents forced by North Atlantic tidal variability dominate. This is because a large volume of water rushes into/out of the Gulf of Maine over the relatively shallow Bank during each semidiurnal tidal cycle. Because of earth rotation effects, semidiurnal tidal currents of more than 50 cm/s (≈ 1 nautical mile/hour or knot) rotate through a clockwise elliptical current vector pattern (for the M_2 semidiurnal ellipse see Fig. 2.43). Georges Bank tidal currents at are strongly sheared near the bottom, where their frictional effects transport sediments.

The amplitudes and ellipse characteristics of tidal currents vary greatly in the coastal ocean because of the effects of bathymetry and continental shelf configuration. The Gulf of Maine M_2 tidal current ellipses are larger than the rest of the northeast shelf region depicted in Fig. 2.43 because the Gulf of Maine length (Georges Bank – head of the Bay of Fundy) makes it is near-resonance with the semidiurnal tidal forcing.

2.9 Oceanic Internal Waves and Tides

With surface gravity waves, the air–sea density difference between air and water (ratio $\approx 1/800$) leads to the gravitational restoring force for fluid that has been displaced vertically (usually by wind). In the stably stratified water column, the similar, though weaker restoring force, makes *internal gravity waves* are possible. The surface expression of internal waves is very small, hence the name.

Consider the case depicted in Fig. 2.44a, where a thin layer (h') of less dense water ρ' overlies a thicker layer (h'') of slightly more dense water ρ''. Theory [2.19] indicates that an internal wave with wavelength L and phase speed of

$$c = \left[g h' \left(\frac{\rho'' - \rho'}{\rho''} \right) \right]^{1/2} ,$$

where $L/20 > h'$ and $L/2 < h''$ can arise. Thus the wave phase speed of a *shallow water* internal wave is considerably less than that of its surface gravity water counterpart because the reduced vertical density difference across the interface decreases the effect of gravity.

In contrast to the 2-layer example above, oceanic density and density gradients vary with depth or -z (Chap. 5). Internal gravity waves propagate in complicated ways throughout most of the ocean. Theory indicates that internal waves can occur at all depths in a stable water column between the frequencies of the inertial frequency $f(\phi) = 2\Omega \sin(\phi)$ (where ϕ is latitude) and the local buoyancy frequency that is given by

$$N(z) = \sqrt{-\frac{g}{\rho} \frac{\partial \rho_{pot}}{\partial z}} \quad \text{in radians/unit time} ,$$

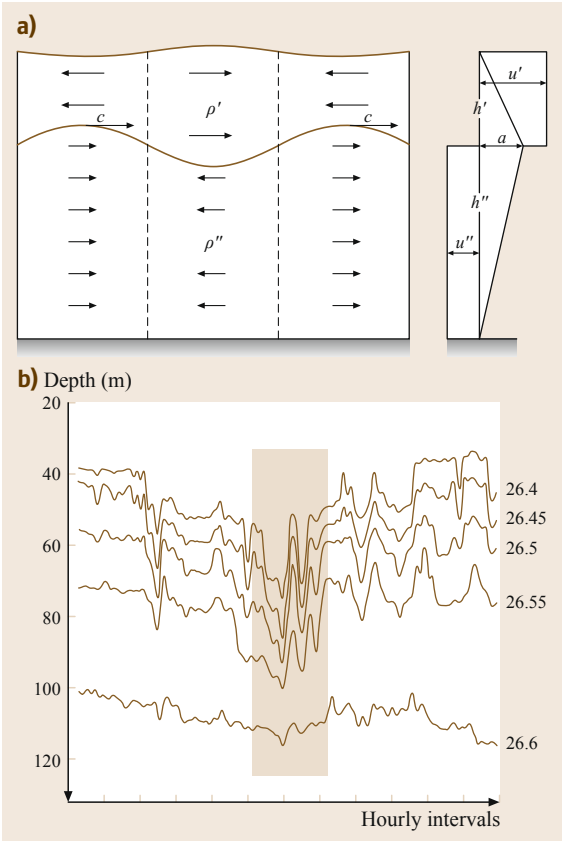


Fig. 2.44 **(a)** The configuration of an internal gravity wave in a 2-layer ocean with one layer of thickness h' and density ρ' overlying the deeper layer of thickness h'' and ρ'' supports an interfacial internal wave with a phase speed c that is independent of wave amplitude a (see text). Note the minor surface expression and that the opposed currents (u' and u'') in the respective layers do depend on the wave amplitude (after [2.8]). **(b)** Observed isopycnal depth time series define two rank-ordered oscillations on the trailing edge of an internal solitary wave of depression – estimated from moored observations on the northern California shelf in 133 m of water in April 21, 1981 (after [2.20])

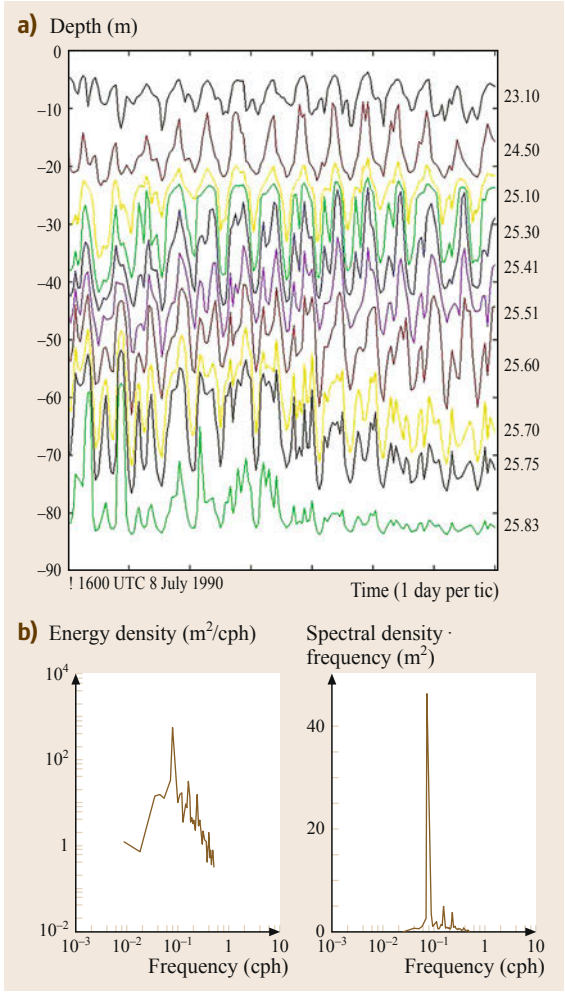


Fig. 2.45 **(a)** Massachusetts Bay isopycnal depth time series – inferred from density measurements – indicate strong internal tidal signatures throughout the water column. **(b)** Isopycnal displacement energy density spectrum. The prominent peak in the semidiurnal (≈ 12 h) frequency band is consistent with the proximity of the measurements to the internal tidal generation zone on Stellwagen Bank with 95% confidence limits indicated (b) Variance-preserving spectrum of the isopycnal displacement energy indicates the dominance of the semidiurnal internal tidal energy in the 12 h frequency band

where $-\partial\rho_{pot}/\partial z$ is the stable local gradient of potential density – a density that has been corrected for pressure effects (Chap. 5).

Intermittent interval wave events were observed by *Howell* and *Brown* [2.20] at a site in 133 m of water on the northern California continental shelf in April 1981. The water column at the site approximated a 2-layer ocean with an upper layer of 60 m. One of the observed events was a internal wave of depression with 29 m downward displacement of the mid-depth isopycnals (Fig. 2.44b), upper layer onshore velocity of 0.14 ± 0.02 m/s, lower layer offshore velocity of 0.14 ± 0.02 m/s, and time scale of 28 ± 4 min; all of which agreed well with internal soliton theory applied to the conditions of the observation site. They concluded that internal solitary waves, generated in the region of the continental slope, evolved into packets of solitons as they propagate shoreward before dissipating their energy of about 8.23×10^5 J per meter of wave crest.

The distribution of internal wave energy is determined by the nature of the different forcing mechanisms including winds and interactions between the surface tidal currents and bathymetry.

Temperature measurements made in the deep ocean (e.g., off Southern California and in the South China Sea) show significant concentrations of energy at semidiurnal frequencies. These temperature fluctuations were due to internal waves of tidal frequency (i. e., internal tides) with vertical displacement amplitudes of nearly 100 m and wavelengths of about 100 km. Evidence indicates that these energetic internal tides were generated at nearby bathymetric slopes.

Widespread evidence indicates that these internal tides represent an important component in the internal wave field many places in the world's oceans. In particular, subsequent interactions between (a) propagating internal tides from multiple generation sites, (b) the normally variable and heterogeneous oceanic density field, and (c) complex bottom bathymetry can lead to evolution and scattering that contribute to a general oceanic internal waves field with a range in periods that range between the local buoyancy (N) and inertial (f) frequencies. Other contributors to the oceanic internal wave field, include natural and well as ocean vehicle interactions with upper ocean stratification, with amplitudes that depend on the specific energetics of their generation processes.

The internal tide is an importantly energetic component of the internal wave field – particularly in the coastal ocean. One important and well-studied example of the internal wave being generated by surface (or external) tidal currents impinging on the bathymetric slopes is found on Stellwagen Bank in Massachusetts Bay. An array of moored measurements near Stellwagen Bank have been used to compute density time series at six levels; from which were produced the suite of isopycnal displacement time series shown in Fig. 2.45.

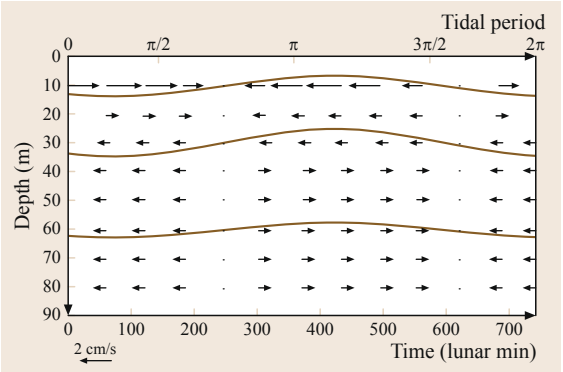


Fig. 2.46 Internal tidal kinematics: The spatial distribution of isopycnal displacement (−) and horizontal velocities associated with one cycle of an M_2 internal tidal wave in an ocean with exponentially decreasing buoyancy frequency

The energy spectra of these isopycnal time series clearly indicate the importance of the internal tide in the region (Fig. 2.45b). An analysis of these data has enabled us to construct the kinematical picture of the isopycnal displacement and current structure in Fig. 2.46. Note the zones of horizontal current convergence in the upper water column (strongest at the surface) and divergence at depth leading the isopycnal wave crest as it propagates from right to left. The current shear associated with such wave motion can become unstable leading to internal wave breaking and mixing.

References

2.1 R.A. Anthes: *Meteorology*, 6th edn. (Macmillian Publ., New York 1992)

2.2 G.L. Pickard, W.J. Emery: *Descriptive Physical Oceanography* (Pergamon, Oxford 1982)

2.3 A.C. Duxbury, A.B. Duxbury: *Introduction to the World's Oceans* (Addison-Wesley, Boston 1984)

2.4 S. Neshyba: *Oceanography Perspectives on a Fluid Earth* (Wiley, New York 1987)

2.5 B. Kinsman: *Wind Waves, Their Generation and Propagation on the Ocean Surface* (Prentice Hall, Englewood Cliffs 1965)

2.6 W.S. von Arx: *An Introduction to Physical Oceanography* (Addison-Wesley, Reading 1974)

2.7 P.R. Pinet: *Invitation to Oceanography*, 2nd edn. (Jones and Bartlett Publ., Sudbury 2000)

2.8 J.A. Knauss: *Introduction to Physical Oceanography* (Prentice Hall, Englewood Cliffs 1978)

2.9 V. Cornish: *Ocean Waves and Kindred Phenomena* (Cambridge Univ. Press, Cambridge, 1934)

2.10 B.W. Pipkin, D.S. Gorsline, R.E. Casey, D.A. Dunn, S.A. Schellenberg: *Laboratory Exercises in Oceanography*, 3rd edn. (Freeman, Macmillan Learning 2000), Online at http://www.macmillianhighered.com

2.11 E. Aguado, J.E. Burt: *Understanding Weather and Climate* (Prentice Hall, Upper Saddle River, 1999)

2.12 BAM, Courtesy of R. Sterner and S. Babin, Johns Hopkins University Applied Physics Laboratory

2.13 E. Bryant: *Natural Hazards*, 2nd edn. (Cambridge Univ. Press, Cambridge 2005)

2.14 MARACOOS: http://maracoos.org/blogs/main/?p=108

2.15 G. Neumann, W.J. Pierson Jr.: *Principles of Physical Oceanography* (Prentice Hall, Englewood Cliffs 1966)

2.16 H. Stommel, A.B. Arons: An idealized model of the circulation pattern and amplitude in oceanic basins, Deep Sea Res. **6**, 217–233 (1960)
2.17 A.T. Doodson: The harmonic development of the tide-generating potential, Proc. Roy. Soc. A **100**, 305–329 (1921)
2.18 J.A. Moody, B. Butman, R.C. Beardsley, W.S. Brown, P. Daifuku, J.D. Irish, D.A. Mayer, H.O. Mofjeld, B. Petrie, S. Ramp, P. Smith, W.R. Wright: Atlas of tidal elevation and current observations on the Northeast American Continental Shelf and Slope, US Geol. Surv. Bull. **1611**, 122 (1984)
2.19 H. Lamb: *Hydrodynamics*, 6th edn. (Cambridge Univ. Press, Cambridge 1932)
2.20 T.L. Howell, W.S. Brown: Nonlinear internal waves on the California continental shelf, J. Geophys. Res. **90**(c4), 7256–7264 (1985)

3. Metocean Extreme and Operating Conditions

George Z. Forristall, Cortis K. Cooper

Metocean stands for *meteorology* and *oceanography*, an acronym that is commonly used in the offshore oil industry to encompass almost all topics involving the quantitative description of the ocean and atmosphere needed to design and operate man-made structures, facilities, and vessels in the ocean or on the coast. The metocean environment controls many aspects of facility design and operation, so errors in quantifying metocean conditions can cascade through the design and operational decisions. Errors can result in damage and lost lives. Conversely, if the variables are overestimated, costs will be overestimated perhaps to the point that the project becomes uneconomic and is never built.

Metocean criteria are typically broken into two categories: operating and extreme. The former involves quantification of metocean conditions in which the facility or vessel should be capable of achieving the routine functions of its primary purpose. Extreme conditions are typically associated with *storms*.

The approach taken by this chapter is to outline the methods commonly used in industry to quantify the most important metocean variables that impact offshore facilities. These methods are drawn largely from the offshore oil and gas industry but they are also readily applicable to other engineering applications involving the design and operation of vessels, coastal structures, offshore wind farms, navigational aids, coastal geomorphology, and pollution studies.

Of course there are a multitude of metocean variables which could be covered but this chapter focuses on winds, waves, and currents (WWC), since these are the variables that most often control extreme loads or operating conditions on man-made facilities. Because of space constraints it is necessary to only briefly cover some of the topics and to provide the reader with references for further reading.

After the introductory section, the next section reviews key physical processes: WWC spectra, wind and current profiles, wave growth and wave breaking. It is followed by two sections that briefly address some of the more important issues that can arise regarding measurements and models. Section 3.5 examines ways to calculate the marginal probability of WWC processes. Section 3.6 describes some of the analysis products that are typically used to quantify operating conditions. Finally, the last section covers the topic of extreme criteria.

3.1 **Quantifying the Metocean Environment** ... 48
3.2 **Overview of WWC Processes** ... 49
3.2.1 Winds ... 49
3.2.2 Waves ... 51
3.2.3 Currents ... 53
3.3 **Measurements** ... 55
3.3.1 Historical Storm Databases ... 55
3.3.2 Satellite Databases ... 56
3.3.3 In Situ Measurements ... 57
3.3.4 Mobile Measurements ... 58
3.4 **Modeling** ... 58
3.4.1 Winds ... 59
3.4.2 Waves ... 59
3.4.3 Currents, Surge, and Tides ... 60
3.5 **Joint Events** ... 61
3.5.1 Response-Based Analysis ... 61
3.5.2 Load Cases ... 62
3.5.3 Environmental Contours ... 63
3.5.4 Inverse FORM ... 63
3.6 **Operational Criteria** ... 64
3.6.1 Probability Distributions ... 64
3.6.2 Persistence ... 65
3.6.3 Currents ... 65
3.7 **Extreme Criteria** ... 66
3.7.1 Risk and Reliability ... 66
3.7.2 The Historical Method ... 67
3.7.3 Synthetic Storm Modeling ... 68
3.7.4 Modeling Versus Measurements ... 68
3.7.5 Accounting for Physical Limits ... 69
3.7.6 Seasonality ... 69
3.7.7 Directionality ... 69
3.7.8 Combining Long and Short-Term Distributions ... 69
3.7.9 Rogue Waves ... 70

3.7.10 Extremely Rare Events 71
3.7.11 Quantifying Uncertainty 71
3.7.12 Stationarity 71
3.8 **Conclusions** 72
References 72

3.1 Quantifying the Metocean Environment

Metocean is an acronym from *met*eorology and *ocean*ography and is commonly used in the offshore oil industry to encompass almost all topics involving the quantitative description of the ocean and atmosphere needed to design and operate man-made structures, facilities, and vessels in the ocean or on the coast. When engineers design a major facility or vessel to operate and survive in the sea, they must consider the loads and other constraints that may affect the structure. If those loads and constraints are underestimated, then damage can result and lives may be lost. Conversely, if loads and constraints are overestimated, then the costs will be overestimated perhaps to the point that the project becomes uneconomic and is never built.

The metocean environment controls so many aspects of facility design and operation that errors in quantifying metocean conditions can cascade though the design and operational decisions. For instance, overestimating a design wave height for a deepwater floating production platform could result in adding too many mooring lines. Since these additional lines would add tons of static load, a larger facility would be needed to provide the necessary buoyancy, thus generating additional capital cost well beyond the cost of the excess mooring lines. In short, the accurate quantification of metocean criteria can have far-reaching effects on the safety and profitability of offshore facilities. For this reason, metocean criteria are usually specified and described in a separate chapter or stand-alone document in a project's design documents. In 2005, the American Petroleum Institute (API) recognized the influence of metocean criteria and began publishing a stand-alone set of recommended practices for the offshore industry [3.1].

Metocean criteria are typically broken into two categories: operating and extreme. The former involves quantification of metocean conditions in which the facility or vessel should be capable of achieving the routine functions of its primary purpose. Examples of routine functions include pumping oil, drilling, receiving or pumping out natural gas, and generating wind energy. Typical products used to quantify operational conditions include a cumulative probability distribution of wave height and a table of wind speed persistence. These products are used in estimating the fatigue lives for components. In contrast, extreme conditions occur rarely and are often generated by episodic events (e.g., storms). During extreme conditions, normal operations are usually suspended – the vessel is slowed, oil or gas production is stopped, wind turbines are feathered, etc. A common example of an often used extreme condition parameter is the 100-year maximum wave height – the largest wave expected over a three-hour period once in 100 years.

With this background in mind, the goal of this chapter can now be stated: it is to outline the methods commonly used in industry to quantify the most important metocean variables that impact offshore facilities. These methods are drawn largely from the offshore oil and gas industry but they are also readily applicable to other engineering applications involving the design and operation of vessels, coastal structures, offshore wind farms, navigational aids, coastal geomorphology, and to some extent, pollution studies. While we attempt to provide some physical insights into the underlying metocean processes, this chapter focuses on the methodology for deriving the key variables, and the nuances of their correct application.

Of course there are a multitude of metocean variables that could be covered in this chapter. Potential topics include water temperature, tides, and salinity. While these variables can be important for some engineering applications such as acoustics, this chapter will focus on winds, waves, and currents (WWC), since these are the variables that most often control extreme loads or operating conditions on man-made facilities. However, even this narrowing leaves countless aspects of WWC that could be covered with far too little space to do them justice. Thus we again have chosen to narrow the frame further by specifically focusing on aspects of WWC that tend to drive capital or operating decisions in large offshore facilities. For those interested in coastal features where shallow-water effects are important, the Coastal Engineering Manual [3.2] serves as an excellent reference.

Much as we the authors have had to narrow the topics, a metocean design basis for a major project must narrow the variables that are covered. This is because the sea and atmosphere are filled with complicated processes, many of which are site specific and poorly understood. If aggressive filtering is not undertaken, then too much time can be spent quantifying variables

that make little difference to the design or operation of the facility. The first and best way to eliminate variables from investigation is to understand the basic responses of the *particular* facility. In other words to answer the question: which metocean variables impact this facility most and which have little or no impact? For example, squalls and their dynamic effects can be especially important in designing the moorings for floating production vessels near the equator, such as off Indonesia. Carefully quantifying squall intensity and its change over time scales of a few minutes and length scales of the order of 50 m is of highest importance. In contrast, quantifying water, storm surge, and air temperatures is not critical.

Section 3.2 is an overview of key processes, including a discussion of WWC spectra, wind and current profiles, and important though arguably tangential discussions of wave growth and wave breaking.

Sections 3.3 and 3.4 briefly address some of the more important issues that can arise regarding measurements and models. Since all metocean criteria are founded on one or both of these inputs, it is important to understand the various sources and databases, and their advantages and limitations. Otherwise one may well suffer the consequences of *garbage in, garbage out*.

Section 3.5 examines ways to calculate the marginal probability of WWC processes. One of the more interesting cases is when two variables are statistically independent (or nearly so) in time or space and yet there is often a non-negligible probability that the two can occur simultaneously and generate loads that exceed the load from an individual process at the same probability level.

Section 3.6 describes some of the analysis products that are typically used to quantify operating conditions. The discussion begins with the simplest approaches such as univariate probability density functions and then moves on to address more sophisticated products to characterize storm and calm persistence, directional dependence, and vertical space variations.

Finally, the last section covers the topic of extreme criteria. It is important because the economic and safety consequences of getting it right are so high. It is also important because there is no general and all encompassing methodology to estimate extremes so the topic is rich in subtleties, complexity, and potential traps.

3.2 Overview of WWC Processes

3.2.1 Winds

Most winds that are important for offshore design and operations come from extra-tropical or tropical storms. Extra-tropical storms are large-scale systems that are well represented on standard meteorological charts. The measurements used to produce these charts are discussed in Sect. 3.3. Tropical storms are relatively small features on common weather charts. Observations in them are scarce. Detailed wind fields in tropical storms are produced using dynamic or kinematic numerical models, which are discussed in Sect. 3.4.

Wind specification requires especially careful attention to definitions. *Richardson* [3.3] made an eloquent statement of the problem many years ago:

> *Does the wind possess a velocity? This question, at first sight foolish, improves on acquaintance . . . let us not think of velocity, but only of various hyphenated velocities.*

Richardson was concerned that $\Delta x/\Delta t$ might not have a limit in a turbulent fluid. Examples of *hyphenated velocities* which do have a clearly defined meaning are the one-hour or three-second wind. The one-hour wind is the wind velocity averaged over an hour. The three-second wind is the maximum three second average velocity in an hour interval unless another interval is stated. The three-second wind gust is about 30% higher than the one-hour average.

Wind speeds also vary with altitude. Friction at the water surface reduces the wind speed near the boundary. Wind speeds increase with height through the atmospheric boundary layer. The speed at 30 m height is about 15% higher than that at the common anemometer height of 10 m. Unless the averaging time and height of a wind measurement are given, that measurement is not very useful.

The standard offshore engineering method for converting wind speeds from one averaging period and height to another is given by Standards Norway (NORSOK) [3.4] and serves as the basis for the ANSI (American National Standards Institute)/API [3.1] recommended practices. These guidelines give the wind speed $u(z,t)_o$ at height z above mean sea level for averaging period t_o as

$$u\left(z,t_o\right)=U\left(z\right)\left[1-0.41I_u\left(z\right)\ln\left(\frac{t}{t_o}\right)\right], \qquad (3.1)$$

where the one-hour mean wind $U(z)$ is given by a modified logarithmic profile that depends on the one-hour

mean wind speed at 10 m elevation, $U(10)$, as given in (3.2) and (3.3)

$$U(z) = U(10)\left[1 - C\ln\left(\frac{z}{10}\right)\right], \tag{3.2}$$

$$C = 0.0573\sqrt{1 + 0.15U(10)}, \tag{3.3}$$

and the wind speed at other averaging periods given in (3.1) depends on the turbulence intensity $I_u(z)$ defined as the standard deviation of the wind speed at height z, $\sigma(z)$, divided by the one-hour mean wind speed at height z, $U(z)$. According to the API standard

$$I_u(z) = 0.06[1 + 0.043U(10)]\left(\frac{z}{10}\right)^{-0.22}. \tag{3.4}$$

Note that the equations use units of meters for height and m/s for velocity. Equations (3.1)–(3.4) are based on an extensive set of wind measurements made from a tower on a small islet off the coast of Norway. While all these measurements were made in extra-tropical storms, the equations are commonly used for tropical storms as well, e.g., [3.1]. However, recent work by *Vickery* et al. [3.5] shows that the equations from *ESDU* (Engineering Sciences Data Unit) [3.6, 7] fit the observations from tropical cyclones noticeably better than the *NORSOK Standards* [3.4] equations.

The original ESDU equations are more complicated than the *NORSOK Standards* [3.4] equations but *Vickery* et al. [3.5] found a number of simplifications which apply in cases of engineering interest and yield the following equations

$$U(z) = \frac{u_*}{k}\ln\left(\frac{z}{z_o}\right), \tag{3.5}$$

where U is the one-hour averaged velocity at a height of z above MSL (mean sea level), k is von Karman's constant (0.4), u_* is the friction velocity, and z_o is the roughness height. The latter two are defined as

$$u_* = U(10)\sqrt{C_d}, \tag{3.6}$$

$$z_o = 10e^{-k/\sqrt{C_d}}, \tag{3.7}$$

where C_d is the drag coefficient at 10 m above sea level. There are many expressions cited in the literature for the drag coefficient but *Vickery* et al. [3.5] chose *Large* and *Pond* [3.8]

$$C_d = 1.2 \quad 4 \le U(10) < 11\,\mathrm{m\,s^{-1}}, \tag{3.8}$$

$$C_d = [0.49 + 0.065U(10)]\,10^{-3} \quad 11 \le U(10) < 25\,\mathrm{m\,s^{-1}}, \tag{3.9}$$

where U must be in units of $\mathrm{m\,s^{-1}}$. For hurricanes, *Vickery* et al. [3.5] suggests restricting the maximum value of C_d based on *Vickery* et al. [3.9] to

$$C_{dMax} = (0.0881r + 17.66)10^{-4}, \tag{3.10}$$

where r is the horizontal distance from the storm center to the site. The value in (3.9) exceeds the value in (3.10) at about $22\,\mathrm{m\,s^{-1}}$ for $r = 20$ km. More will be said shortly about the cap on C_d.

The peak wind gusts for averaging time t_o can be calculated with

$$u(z, t_o) = U(z)\left[1 + g(\vartheta, t_o, z)\,I_u(z)\right], \tag{3.11}$$

where $I_u(z)$ is defined in [3.4]. Using the simplifications described by *Vickery* et al. [3.5]

$$\sigma(z) = \frac{7.5u_*\left[0.538 + 0.09\ln\left(\frac{z}{z_o}\right)\right]}{1 + 0.156\ln\left(\frac{u_*}{fz_o}\right)}, \tag{3.12}$$

where f is the Coriolis parameter.

The peak factor $g(\vartheta, t_o, z)$ is a function of the length of the record (typically 1 h), T_o, and the zero crossing period, ϑ, or

$$g(\vartheta, t_o, z) = \sqrt{2\ln(T_0\vartheta)} + \frac{0.577}{\sqrt{2\ln(T_0\vartheta)}}\frac{\sigma(z, t_o)}{\sigma(z)}, \tag{3.13}$$

where the variables are defined as

$$\sigma(z, t_o) = \sigma(z)\left[1 - 0.193\left(\frac{T_u}{t_o} + 0.1\right)^{-0.68}\right], \tag{3.14}$$

$$\vartheta = \frac{0.007 + 0.213\left(\frac{T_u}{t_o}\right)^{0.654}}{T_u}, \tag{3.15}$$

$$T_u = 3.12z^{0.2}. \tag{3.16}$$

Neither NORSOK or ESDU equations used to calculate wind at various time averages apply to short-lived squalls because the wind speed is not statistically stationary in them. Nor is it clear how well the wind profiles apply.

Squalls are important for engineering design and operations in low latitudes or where the wave fetch is limited by land. Squall lines often originate onshore where convection is strongest and then propagate with the mean winds. When a squall line passes a site, the wind speed rapidly increases and then decays over a few hours, perhaps with some oscillations. Squalls are generally modeled in design analyses as time series scaled up from actual measured squall records.

Compliant structures in deep water can have natural periods much longer than the vibration periods of fixed structures. Resonant oscillations of these structures can

be excited by long period variations in wind speeds. Knowledge of the wind spectrum is required in order to calculate the response. Again, the standard engineering wind spectrum is that given by NORSOK [3.4]. It is

$$S(f,z) = \frac{(320\,\mathrm{m^2\,s^{-1}})(U(10)/U_{\mathrm{ref}})^2(z/z_{\mathrm{ref}})^{0.45}}{\left(1+\tilde{f}^{0.468}\right)^{3.561}}\,, \tag{3.17}$$

where

$$\tilde{f} = (172\,\mathrm{s})f\left(\frac{z}{z_{\mathrm{ref}}}\right)^{2/3}\left(\frac{U(10)}{U_{\mathrm{ref}}}\right)^{-0.75}\,. \tag{3.18}$$

The reference elevation above the mean sea surface z_{ref} is 10 m, and the reference wind speed U_{ref} is $10\,\mathrm{m\,s^{-1}}$.

The drag of wind on the sea surface produces waves and currents, so accurate knowledge of the drag as a function of wind speed is important for modeling waves and currents. The drag coefficient depends on atmospheric stability, but in the high winds that interest us, the equations for neutral conditions usually apply. The wind stress τ is equal to $u_*^2\rho_{\mathrm{o}}$ where u_* is the so-called friction velocity given by (3.6) and ρ_{o} is the air density. The friction velocity is dependent on the drag coefficient, C_{d}. There are many formulations for C_{d} but one of the more popular is from *Large* and *Pond* [3.8] as shown in (3.8) and (3.9). For years, many metocean experts used (3.9) well above the maximum $25\,\mathrm{m\,s^{-1}}$ suggested by *Large* and *Pond* [3.8], but more recent hurricane measurements by *Powell* et al. [3.10] showed that the drag coefficient starts to level off around $30\,\mathrm{m\,s^{-1}}$. They conjecture that high wind speeds create a layer of sea foam and bubbles at the sea surface thus dropping the effective roughness of the sea. This reasoning was supported by the laboratory experiments by *Donelan* et al. [3.11]. *Powell* [3.12] provided additional support from field measurements. *Frolov* [3.13] showed that capping the drag coefficient at 0.0022 was essential to model currents measured in Hurricane Katrina in the Gulf of Mexico.

3.2.2 Waves

Waves grow because of the input of momentum from the wind, but knowledge of the exact mechanism by which this momentum is transferred has remained elusive. The fundamental mechanism, first proposed by *Miles* [3.14], seems to be a resonance interaction between wave-induced pressure fluctuations and the waves. As the waves propagate, they are modified by nonlinear interactions between different frequencies, frictional dissipation and wave breaking. A fuller discussion of wave generation and modeling is given in Sect. 3.4.

Ocean waves are a complex and irregular function of space and time. This complexity is best understood by considering the sea surface to be the superposition of many cosine waves, as shown in Fig. 3.1. Each of the cosine waves is characterized by a period T and an amplitude a. The height of a cosine wave $H = 2a$. Later we will see that this relation is not true for real waves. The wave frequency $f = 1/T$ is the inverse of the wave period. The wave length L between two crests is given by $L = gT^2/2\pi$ in deep water. The phase speed or celerity is given by $c = L/T$. A more detailed discussion of wave kinematics and dynamics is given in Chap. 2.

A wave record measured at a point can be analyzed into its component cosine waves using the Fourier transform. This transform gives the amplitude and phase of each component. It includes all of the information and irregularity of the original record. This is too much detail for most purposes because an individual wave record is a single realization of a random process. We would usually prefer to know the distribution of wave energy with frequency in the underlying process. If $F(f)$ is the Fourier transform of the wave record, its power

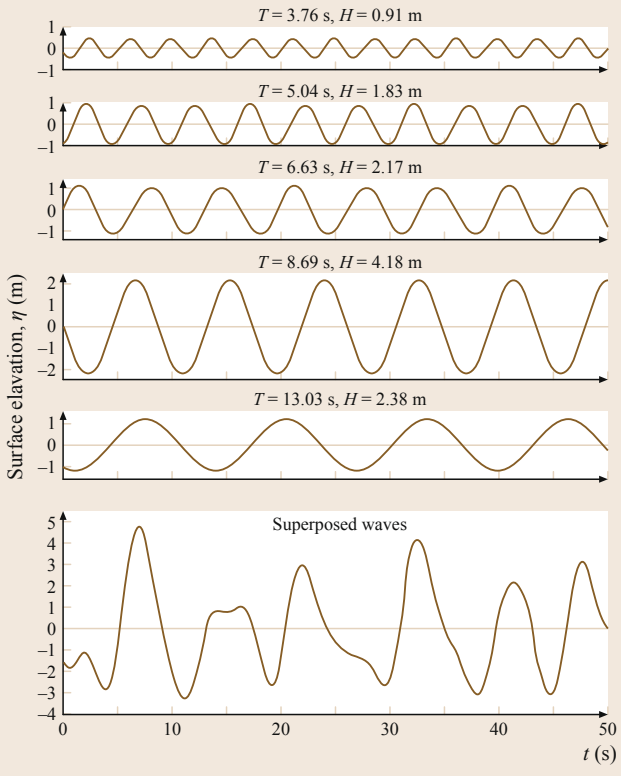


Fig. 3.1 Superposition of cosine waves to make regular waves

spectral density is given by

$$S(f) = 2\,|F(f)|\,/n\,, \tag{3.19}$$

where n is the number of points in the time series. Taking the square of the amplitude of the Fourier transform removes the phase information from the record, but the result is still a very irregular function of frequency. A smooth version of the spectrum is found by filtering $S(f)$ over frequency or averaging spectra from several ensembles. *Glover* et al. [3.15] give a good, practical guide to the details of calculating power spectra.

Once the spectrum is known, the significant wave height H_S is defined as

$$H_S = 4\int S(f)\mathrm{d}f = 4\sigma\,, \tag{3.20}$$

where σ is the variance of the wave record. The peak wave frequency is the frequency at the highest point in the power spectrum. The mean frequency f_m and zero-crossing frequency f_z are given by

$$\begin{aligned} f_m &= m_1/m_0\,,\\ f_z &= m_2/m_0\,. \end{aligned} \tag{3.21}$$

Where the spectral moments are calculated as

$$m_n = \int S(f) f^n \mathrm{d}f\,. \tag{3.22}$$

Wave spectra for design are generally specified in an analytic form. The most popular of these is the Joint North Sea Wave Observation Project (JONSWAP) spectral form. It is given by

$$S(f) = \beta f^{-5}\exp\left[-\frac{5}{4}\left(\frac{f}{f_p}\right)^{-4}\right]\gamma^{\exp\left[-\frac{(f-f_p)^2}{2\sigma^2 f_p^2}\right]}\,, \tag{3.23}$$

where

$$\sigma = \begin{cases}\sigma_a = 0.07 \text{ if } f \le f_p\,,\\ \sigma_b = 0.09 \text{ if } f > f_p\,.\end{cases} \tag{3.24}$$

The JONSWAP spectrum was originally proposed to describe fetch-limited waves, but by adjusting its parameters, it can give a reasonable fit to most single-peaked spectra. Given the significant wave height, peak period, and peak enhancement factor γ, *Goda* [3.16] showed that the scale factor is approximated by

$$\beta = \frac{5}{16}H_s^2 f_p^2\left(1.15 + 0.168\gamma - \frac{0.925}{1.909+\gamma}\right)^{-1}\,. \tag{3.25}$$

The JONSWAP spectrum can be used to describe most spectra with single peaks. However, combinations of sea and swell in storms can result in spectra with two or more peaks. The *Ochi–Hubble* [3.17] spectrum is often used to describe double-peaked spectra in areas subject to tropical storms. It is the sum of two Gamma distributions

$$S(f) = \sum_{j=1}^{2}\frac{H_{Sj}T_{Pj}(\lambda_j+0.25)^{\lambda_j}}{4\Gamma(\lambda_j)(T_{Pj}f)^{(4\lambda_j+1)}}\exp\left(-\frac{(\lambda_j+0.25)}{(T_{Pj}f)^4}\right)\,. \tag{3.26}$$

This spectrum has three parameters for each of the two wave systems, a significant wave height, a peak period, and a shape factor λ.

The *Torsethaugen* and *Haver* [3.18] double-peaked spectrum is also the sum of two Gamma functions. Their paper gives parameters which were fit to measurements made in the North and Norwegian seas.

The spectral representation of waves makes it natural to think of them as a Gaussian random process. The envelope of a Gaussian process has a Rayleigh distribution, and to first order, so do wave and crest heights. However, the trough preceding a large crest is likely to be on a lower part of the envelope. Trough to crest wave height differences are, therefore, slightly smaller than given by the Rayleigh distribution

$$P(h) = \exp\left[-2\left(\frac{h}{H_S}\right)^2\right]\,, \tag{3.27}$$

where H_S is four times the standard deviation of the wave trace.

The empirical distribution suggested by *Forristall* [3.19] accounts for the observed reduction in wave height and has been shown to agree with many observations, including measurements in water depths less than 30 m. It is given by

$$P(h) = \exp\left[-2.263\left(\frac{h}{H_S}\right)^{2.126}\right]\,. \tag{3.28}$$

Crest heights in steep waves are higher than those predicted by Gaussian theory because the waves are nonlinear. The distribution produced from simulations of second-order waves by *Forristall* [3.20] accounts for the most important nonlinearity. It is a Weibull distribution of the form

$$P(\eta_2) = \exp\left[-\left(\frac{\eta}{\alpha H_s}\right)^{\beta}\right]\,, \tag{3.29}$$

where

$$\alpha = \sqrt{\frac{1}{8}} + 0.2568 S_1 + 0.0800 Ur\,,$$
$$\beta = 2 - 1.7912 S_1 - 0.5302 Ur + 0.2824 Ur^2\,. \quad (3.30)$$

The mean steepness and Ursell number are given by

$$S_1 = \frac{2\pi}{g}\frac{H_s}{T_1^2}\,,$$
$$Ur = \frac{H_s}{k_1^2 d^3}\,. \quad (3.31)$$

The wave and crest height distributions in (3.27)–(3.29) do not take into account higher-order nonlinearities that may lead to rogue waves. The evidence for rogue waves and possible theoretical reasons for their existence are discussed in Sect. 3.7.9.

Representing waves as a Fourier series makes the tacit assumption that the waves do not break. A Fourier series, and most wave theories, cannot handle double-valued time series. Yet during a storm, the sea is covered with breaking waves [3.21]. Fortunately, almost all of these breaking events are spilling events that only affect a small portion of the wave crest. Because of this design calculations in deep water typically ignore breaking. Measured forces and the survival of structures in severe storms indicate that neglecting deep water breaking waves does not change wave forces significantly [3.22].

The situation is completely different near the shore. The transformation of wave spectra near the shore is modeled by specialized hindcasting tools such as SWAN (Simulating WAves Nearshore) [3.23]. Shoaling waves can steepen rapidly and form a plunging breaker. *Longuet-Higgins* and *Cokelet* [3.24] succeeded in integrating the equations of motion in a free surface flow past overturning many years ago. Such computations show that particle velocities in the crest of plunging breakers can exceed the phase velocity of the wave and are much higher than particle velocities in non-breaking waves. *Christou* et al. [3.25] used a boundary element method to calculate the particle kinematics in a shoaling wave shown in Fig 3.2. The velocities in the crest are about twice the velocities calculated before the wave breaks.

3.2.3 Currents

Knowledge of ocean currents is important when designing, building, or operating an offshore structure. Wind-driven currents are the most important consideration for structural design because their velocities add to wave particle velocities. Wind stress imparts momentum to the sea surface. Turbulent processes mix the momentum downward. The Coriolis force rotates the resulting currents (to the right in the northern hemisphere and to the left in the southern hemisphere). A fuller discussion of wind-driven current generation and modeling is given in Sect 3.4.

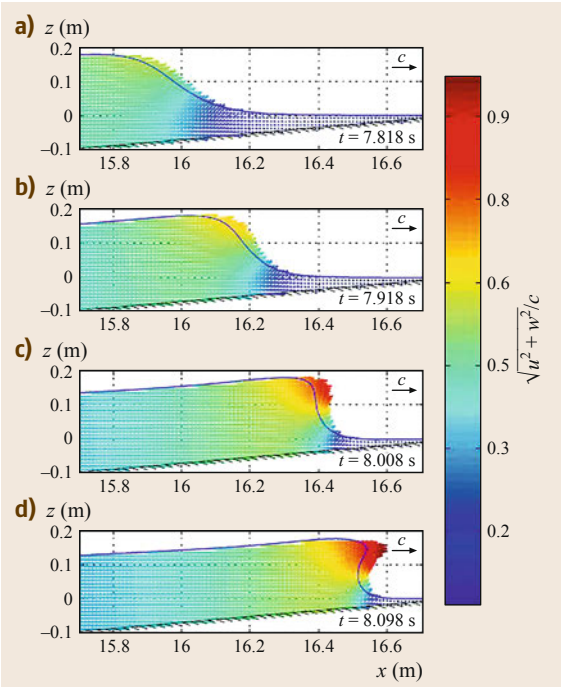


Fig. 3.2a–d Particle velocities in a shoaling breaking wave calculated using a boundary element method (after [3.25])

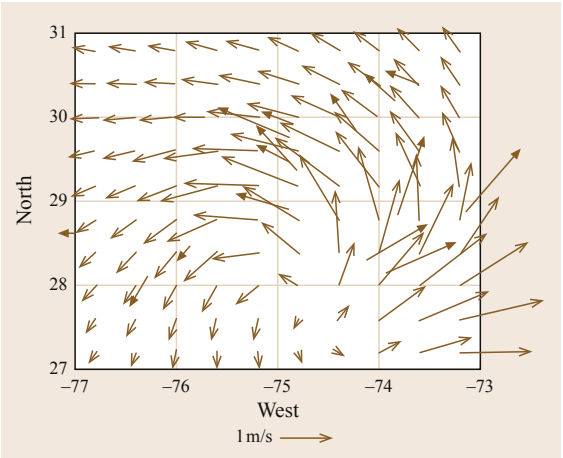


Fig. 3.3 Currents measured near the surface in Hurricane Gloria (1985). The *solid arrows* are measurements from air-dropped expendable current profilers and the *open arrows* are from a one-dimensional current model (after [3.26])

In the northern hemisphere, wind-driven currents are particularly strong on the right-hand side of a hurricane track. There, the current rotation due to Coriolis is often close to resonance with the turning of the wind stress as the hurricane passes, so near-surface currents can exceed 2 m s^{-1}. Figure 3.3 shows an example from Hurricane Gloria in 1985 [3.26]. At the time of the measurements, the hurricane center was at 28.75°N, 74.98°W and moving toward the north-northwest. The closed arrow heads show measurements made with air-dropped expendable current profilers and the open arrow heads show results from a one-dimensional (1-D) numerical model that used the turbulence closure model of *Kantha* and *Clayson* [3.27].

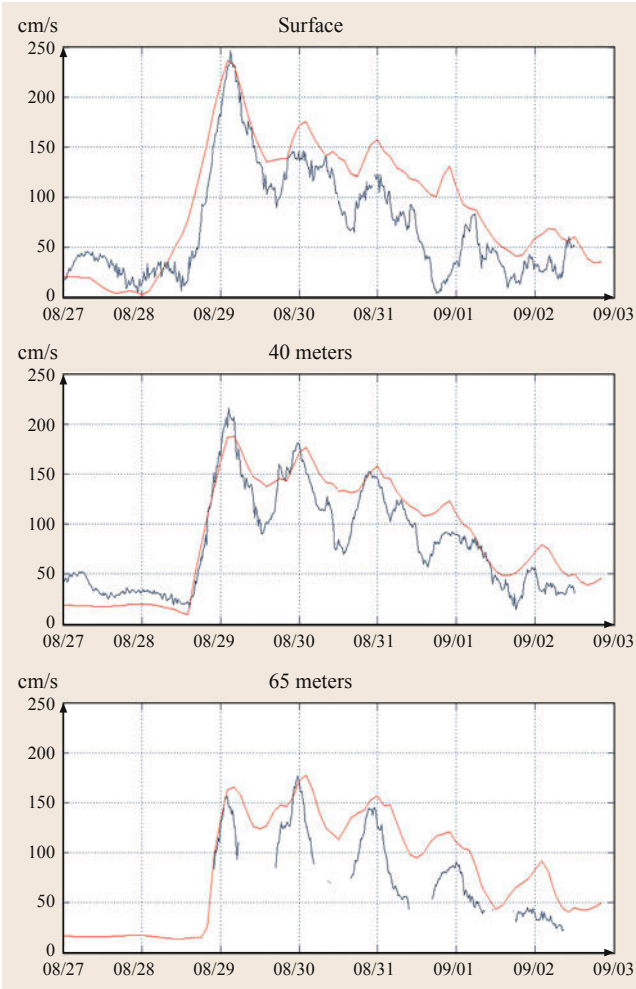


Fig. 3.4 Currents generated by Hurricane Katrina at the Telemark platform in the Gulf of Mexico. The *blue lines* are from the measurements and the *red lines* are from a three-dimensional (3-D) numerical model. The three panels show the current speeds at three depths

The downward propagation of wind-driven momentum is constrained to the upper water column by vertical stratification if it exists. Strong stratification is usually found at most sites in water depths greater than about 30–60 m during the summer months. *Kantha* and *Clayson* [3.28] give a detailed discussion of mixing in vertically stratified flows. Both measured and modeled currents are much stronger on the right-hand side of the storm because the Coriolis rotation is in the same direction as the wind stress rotation. The agreement between measured and modeled currents is good except for a direction difference to the east of the storm center.

Friction in deep water is extremely low, so wind-driven currents can persist as inertial currents for several days after the wind dies out. The rotation of the earth, through the Coriolis force, causes these inertial currents to rotate clockwise (in the northern hemisphere). The rotation period is $\pi/\Omega \sin \Phi$, where Φ is the latitude and Ω is the earth's rotation rate (2π/day). Figure 3.4 shows currents measured during and after Hurricane Katrina in the Gulf of Mexico [3.13]. The peak wind speed was at about the same time as the peak current early on August 29 2005, but the inertial currents persisted for 7 days after that when the wind was essentially calm.

Deep water structures must often contend with the permanent strong current systems that exist near the margins of the ocean. Examples include the Gulf Stream, the Loop Current, the Brazil Current, Kuroshio, and the Somali Current. Figure 3.5 shows these and many others. *Tomczak* and *Godfrey* [3.29] give a good descriptive introduction to these current systems.

Most of these currents are permanent features of the oceanic circulation. However, their position and strength can vary greatly. When they depart from the shelf break, they can often have large meanders and shed eddies than can persist for months [3.30]. Current speeds in these systems can exceed 2 m s^{-1} with speeds over 1 m s^{-1} down to 200 m.

The best design information for these current systems comes from combining remote sensing of the current positions with in situ measurements of current profiles. These techniques have been applied extensively in the Gulf of Mexico to study the Loop Current. The studies have led to the development of a kinematic hindcast model for Loop Current eddies that uses historical eddy positions and shapes as input [3.31].

The external astronomical tide generates weak currents in the deep ocean. Tidal currents are typically less than 10 cm s^{-1} in deep water. In shallower water tidal currents can exceed 2 m s^{-1} and must be considered in the design of facilities such as floating LNG (liquefied natural gas) terminals. Tides and tidal currents are predictable compared to other oceanographic phenomena,

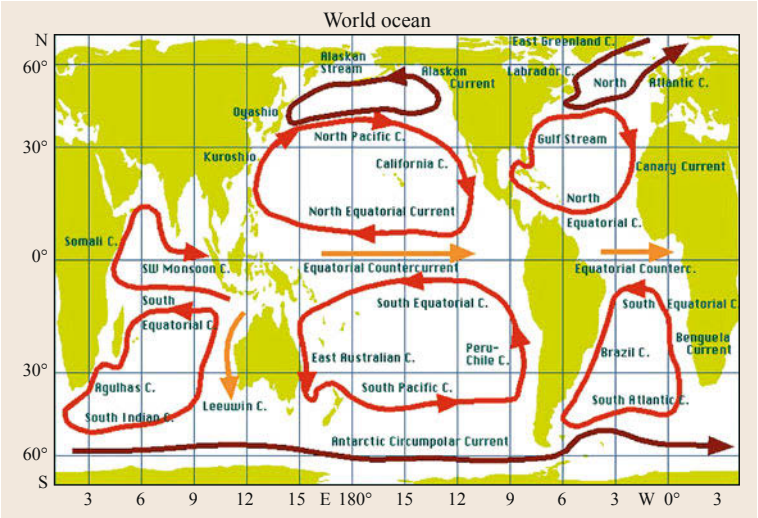


Fig. 3.5 Major current systems of the world

so it only takes a few weeks of measurements followed by harmonic analysis to enable multiyear prediction. The numerical models discussed in Sect. 3.4 also do well predicting tidal currents.

In waters that are strongly stratified in the vertical, the external tide in conjunction with sharp bathymetric features can generate a strong internal tide that is characterized by internal waves and possibly solitons with amplitudes of up to 40 m, phase speeds of about 50 cm s^{-1}, and wavelengths of several kilometers [3.32]. As these waves approach shallower water, they can break and cause high velocities and scouring of the seabed [3.33].

A power spectrum of current measurements typically shows a broad peak at periods of a few days corresponding to the motion of weather systems. There are then sharp peaks at the inertial period and any tidal periods that are important at that site. Measurements in laboratory flumes often show high frequency turbulence with amplitudes of as much as 5% of the mean flow speed. That turbulence strongly affects vortex-induced vibrations of cylinders, so it is important to understand whether such turbulence exists in open ocean currents. *Mitchell* et al. [3.34] made turbulence measurements in a Loop Current eddy using a towed body with a specially configured acoustic doppler current profiler (ADCP). The most energetic events had speed scales of only 1 cm s^{-1}. The typical and average values are more than ten times smaller. *Dhanak* and *Holappa* [3.35] made similar measurements using an autonomous underwater vehicle (AUV). These measurements of low turbulent intensity were made in deep water far from land. Turbulence is expected to be higher in shallow water and near the surface or bottom.

3.3 Measurements

Metocean criteria ultimately trace their roots to measurements or models. While models have become the predominant source data, measurements are still needed to provide boundary conditions and/or initial conditions and to validate or calibrate the model. In the case of small-scale ocean currents or atmospheric storm systems (e.g., squalls), modeling accuracy is problematic in large part because of a lack of understanding of the fundamental physics of geophysical fluids at small length and time scales. For these processes, measurements remain the dominant source of input data for metocean criteria.

Measurements can come from many different sources, including air or satellite-based sensors, vessel-mounted instruments, dedicated moorings, bottom-mounted instruments, Lagrangian drifters, and most recently, automated mobile platforms such as gliders or AUVs. Some of the more common and useful sources are described in further detail below.

3.3.1 Historical Storm Databases

Observations from ships formed the basis for the first database of winds and waves providing nearly global coverage. Wind speeds taken from ships are often measured with an anemometer, but waves are usually based on a sailor's visual observations. As one might expect, the primary challenge in using human observations is

to remove bias and reduce scatter. These wind measurements also have significant error because a ship's superstructure distorts wind flow patterns. *Thomas* et al. [3.36] discuss this issue in detail. Of particular concern with all ship observations is the so-called fair weather bias – the tendency for ships to avoid storms and thus underestimate the true probability of larger waves and winds. A number of attempts were made to remove bias and scatter [3.37] and these efforts eventually culminated in the work of *Hogben* et al. [3.38] who used observations from ships passing close to instrumented buoys to develop corrections that largely removed bias, at least in the North Atlantic. However, as *Hogben* et al. [3.38] point out, they were not nearly as successful in the southern hemisphere where there are far fewer offshore instruments. Nor could they do much about the scatter inherent in subjective human observations.

Another important historical dataset is the so-called HURDAT (National Hurricane Centers HURricane DATabases) best-track data maintained and distributed by NOAA (National Oceanic and Atmospheric Administration) [3.39] and documented by *Jarvinen* et al. [3.40]. It contains the time histories of tracks, peak winds, central pressure, and radius for historical North Atlantic tropical storms from 1851 to the present. NOAA provides similar information for the eastern North Pacific but only from 1949 to the present. The NOAA GTECCA (global tropical and extratropical cyclone climatic atlas) database contains historical track data for global storms up to 1995. For cyclones, the coverage starts as early as 1870 in the North Atlantic but not until 1945 for the other major basins of the world [3.41]. Coverage of Northern Hemisphere extratropical cyclones (winter storms) starts in 1965. Note that these datasets do not include detailed wind fields, just the basic intensity information that can be used to reconstruct the detailed wind fields using various methods such as parametric models [3.42]. Others have used the historical tracks to assimilate into much more sophisticated numerical models providing high resolution gridded wind velocity and pressure [3.43].

The accuracy of early storms in HURDAT have been questioned especially as climate scientists have tried to detect trends in historical storm severity. *Karl* et al. [3.44] provide a fairly recent summary of these findings. In part because of these questions, NOAA recently undertook a re-analysis of the data underlying HURDAT. These efforts have been documented in a series of publications, which are referenced in *Hagen* and *Landsea* [3.45]. Even after their reexamination, the researchers in this effort readily admit that large uncertainties remain in the storms prior to routine airborne observations which started in the early 1950s. *Emanuel* [3.46] and others have noted that there is even more uncertainty in basins outside the North Atlantic, and this uncertainty extends into the post-1950 era because of the lack of reconnaissance flights in most basins.

3.3.2 Satellite Databases

The study of the oceans and winds from space started in the 1970s with the launch of Skylab and Geos-3, which were equipped with a radar-altimeter, windscatterometer, radiometer, and infrared scanner. *Le Traon* [3.47] gives an overview of the state of operational satellites used in oceanography and to a lesser extent, meteorology.

One of the most useful sensors for ocean engineering has proven to be the altimeter. The first of many operational altimeters began with TOPEX (Ocean Topography Experiment)/Poseiden and ERS1 (Earth Resources Satellite) in 1991. Since 1998, there have been as many as four altimeters flying simultaneously because several altimeters are needed to properly resolve the length and time scales of energetic oceanographic phenomena such as mesoscale eddies, storm-driven waves, etc. The most used channels from the altimeter are wind velocity (speed and direction), wave height and period, and sea surface height.

Sea surface height from the operational altimeters is available in near real time and in historical archives from various sites [3.48–50]. Though accuracy varies by satellite, typical RMS (root mean square) errors are less than 3 cm [3.51]. These heights are useful in tracking geostrophic ocean currents and developing comprehensive maps of astronomical tides in deeper water, the latter being prohibitively expensive before the advent of satellite altimeters. *Shum* et al. [3.52] assessed 20 of these tidal models and found many to be accurate to better than 2 cm.

Wind speed and wave height and period measurements from the operational satellite altimeters are available over the web [3.53–55] but these are typically organized by individual tracks for each satellite or statistics from several satellites averaged over large areal blocks. The track data must typically be filtered to eliminate periods with heavy rainfall or close passage to land. Several companies offer commercial products with fully analyzed databases that can be accessed with their proprietary software [3.56, 57].

Numerous researchers have investigated the accuracy of altimeter-derived winds and waves, e.g., [3.58, 59]. These efforts show that there is a systemic bias unique to each satellite but it is easily corrected leaving an RMS difference with buoy current meter measure-

ment of about 30 cm for significant wave height and 1.5 m s^{-1} for wind speed. Some of this difference is due to buoy measurement error.

Synthetic aperture radar (SAR) has been sporadically deployed starting with Seasat. By far the most successful application was the QuikSCAT (quick scatterometer) satellite which launched in 1999 and continued to operate until late 2009. Unlike altimeters which only measure along their track, SARs measure over a wide swath. In the case of QuikSCAT the swath was 1800 km, resulting in the coverage of 90% of the Earth's surface in a single day. Results were widely used to improve forecast models, so archived model results are one of the best ways to use QuikSCAT since native measurements have many hours between samples. Archived QuikSCAT measurements are downloadable from the web [3.60].

SAR from various satellites has also been used to measure ice coverage, oil slicks, waves and currents as described in [3.47]. Unlike the altimeter, SAR measures wave direction in addition to wave height and period. Furthermore, it makes those measurements over a wide swath. Several commercial satellite wave databases include SAR measurements. SAR also has significant theoretical advantages over other sensors when it comes to identifying near-surface currents. Unlike the altimeter, SAR can identify non-geostrophic current fronts (i. e., currents that cause no detectable change in sea surface height). However, analysis of SAR images is complex in part because artifacts can be caused by natural surfactants, and also because a minimum wind threshold is needed. These disadvantages have limited the use of SAR for the measurement of waves and currents.

The wind velocity, wave height/period, sea surface temperature, and sea surface height measurements from satellites are routinely assimilated into ocean models to provide nowcast and forecast products [3.61–65]. It is probably in this form that the satellite results are the most valuable, since the models are able to interpolate between the large gaps in time and space that invariably appear in all sources of satellite measurements. Some of the more popular modeling products that are publicly available are described in Sect. 3.4.

3.3.3 In Situ Measurements

Instruments are commonly deployed at a fixed offshore location using quasi-permanent facilities like oil production jackets, moorings with subsurface or surface buoyancy, or quasi-permanent coastal facilities.

No matter how instruments are deployed, the ocean currents at a particular site are commonly measured by an acoustic doppler current profiler (ADCP). The accuracy and range of the instrument mostly depends on the transmission frequency, which ranges from 38 to 1200 kHz for commercially available instruments. ADCPs offer many advantages over earlier technologies. They are solid state instruments which are not easily fouled by marine growth. Perhaps most attractive of all is their ability to accurately measure at up to 1000 m from the instrument. That said, ADCPs can yield problematic results which may not be obvious to the untrained eye, especially in cases where the primary sources of scatter are mobile like plankton or fish scatterers or where the scatterer is fixed (i. e., risers on an offshore platform). *Hogg* and *Frye* [3.66] and *Magnell* and *Ivanov* [3.67] give some good examples of artifacts that can contaminate ADCP measurements.

High-frequency radar is a somewhat newer and more expensive technology than ADCPs but its use has grown rapidly in the past 5 years, largely because HF radar (High Frequency radar) can map surface currents over areas of the order of 1000 km^2 using only two coastal installations. Dozens of HF radars have been installed along most of the eastern and western coastlines of the US [3.68] and are available in real time from NODC [3.69]. *Paduan* and *Graber* [3.70] discuss the basic technology along with some of its limitations and provide numerous references.

Surface gravity waves can also be accurately measured with ADCPs [3.71] and HF radar [3.70]. However, most historical measurements have been taken with surface-following buoys equipped with accelerometers and perhaps augmented by roll sensors to measure directionality. Many developed countries with coastlines have deployed such instruments for several decades and the US results are available from the National Data Buoy Center (NDBC). *Pandian* et al. [3.72] summarize the limitations of accelerometer-based systems. The most noteworthy is the tendency of the smaller buoys to be pulled under water or be tossed about in larger waves.

Wind velocity has typically been measured with mechanical anemometers using some type of impeller. Most offshore buoys are still equipped with this type of sensor. While these measurements are useful most of the time, questions have been raised about their accuracy during large wave events, especially for the smaller buoys. Better sensors are needed to get detailed profiles. The least expensive of these better sensors are based on LIDAR (light detection and ranging; optical) or SODAR (sonic detection and ranging; sound). Both systems use a Doppler principle to determine velocity and are capable of measuring multiple bins over ranges of roughly 200 m above the sensor. Freeman et al. [3.73] compare a LIDAR profiler to more traditional anemometers and show excellent accuracy. In contrast, *de Noord* et al. [3.74] raise serious con-

cerns about the accuracy and consistency of SODAR measurements, especially in storm conditions or near obstructions.

Regardless of the variable being measured, interference from nearby man-made or natural features must always be considered. *Cooper* et al. [3.75] discuss some of the challenges of taking metocean measurements off of oil-industry facilities like jackets. Similar issues crop up with wind measurements from stations on the coast, and these have been well studied and codified into numerous recommended practices, e.g., ASCE (American Society of Civil Engineers) 7-05 [3.76] and EUROCODE [3.77].

Another issue that frequently arises is the *averaging interval*. As explained in Sect. 3.2, it is especially important for wind velocity because there is considerable wind energy at higher frequencies. Another important factor affecting winds is the elevation of the measurement above the sea or land, which is also explained further in Sect. 3.2. In short, specification of wind velocity should always include a minimum of four variables: speed, direction, averaging interval, and elevation. Similar issues arise with ocean currents, although they tend to be less pronounced because of the inherent difference in the turbulence spectra of winds and currents. In the case of waves, the issue of elevation is irrelevant. However, the temporal scale of the sampling is important when calculating statistical values like significant wave height. This issue will be discussed in more detail in Sect. 3.7.8. A good general rule is that the minimum sampling period for wave spectra should never be less than 20 min.

As indicated above, a great deal of in situ measurements, including wind velocity, are reported in real time and archived at the NDBC. Another good source for measurements from limited duration deployment is the National Oceanographic Data Center (NODC).

3.3.4 Mobile Measurements

Instrumented wind measurements have been taken from vessels for decades. In the 1980s, oceanographers began mounting ADCPs on ships [3.78]. In both cases, correcting for ship motion presents challenges but GPS (global positioning system) has largely resolved these. Vessel-based wind measurements remain susceptible to flow interference from the ship superstructure.

Starting in the 1980s there was a rapid increase in the use of semi-automated or fully automated mobile platforms, starting with Lagrangian drifting buoys whose paths are tracked by satellite, e.g., [3.79]. With the advent of GPS, the drifter position could be precisely tracked and accurate velocities estimated. *Coholan* et al. [3.80] describe the use of drifters to measure the strong currents associated with the Loop Current in the Gulf of Mexico.

Autonomous underwater vehicles (AUV) have not been used much for current measurements because of their cost and limited range. However, as mentioned in Sect. 3.2.3, *Dhanak* and *Holappa* [3.35] made good use of an AUV to measure turbulence.

In the late 1990s gliders became increasingly common thanks to their light weight (50–100 kg), small size (2 m), relatively low cost ($100 k), and lengthy deployment capability (several months). *Rudnick* et al. [3.81] describe the technology in some detail. Though gliders can only progress horizontally at about 1 knot, their long endurance and the ability to remotely pilot them make gliders highly cost effective and adaptable. The present crop of sophisticated gliders can reach 1000 m depth, though this will likely be extended in the near future. Gliders have limited payload and power capacities. They are typically equipped with CTD (conductivity-temperature-depth) sensors, although other sensors have been deployed, including fluorometers, dissolved oxygen, and pH. A time-mean, depth-averaged water velocity can be derived from the surfacing coordinates of a glider. Efforts are underway to incorporate ADCPs into a glider, though power consumption and obtaining an absolute velocity measurement in deeper water remain challenges.

The Global Drifter Program [3.82] began deploying large numbers of drifting buoys in 1999 as a fundamental component of the global ocean observing system (GOOS) and as of August 2011, nearly 11 000 buoys had been deployed worldwide and are available from the GDP website. Unfortunately, there is not yet an equivalent to NDBC for obtaining real-time or archived measurements from other mobile instruments.

3.4 Modeling

Since the advent of relatively cheap computing power in the past 30 years, numerical modeling has started to replace measurements as the primary feedstock for metocean criteria. There are many reasons for this change.

Models are typically much less expensive than measurements and can provide results at a specific site and for durations of many years. In contrast, one rarely has the luxury of having more than a year or two of

measurements at their site of interest. Calculating extreme criteria, say the 100-y event, from short duration measurements will give values with extremely large uncertainty and a high likelihood of major bias. Even 1–2 years of measurements are often inadequate to capture interannual variability.

3.4.1 Winds

Extratropical wind field calculations generally use pressure contours on archived meteorological analysis charts as input information. A balance between the pressure gradient and the Coriolis force gives the wind speed. That calculation must be modified using a boundary layer model to find the desired wind speed and direction at 10 m elevation.

An important modeling hindcast dataset is the NCEP (National Centers for Environmental Prediction) reanalysis product [3.83]. The first phase is documented in *Kalnay* et al. [3.84] and consists of a numerical model hindcast of wind and pressure fields from 1948 to the present. Observations from ships, satellites, and fixed sites have been assimilated into the model. A follow on effort, NCEP/DOE (Department of Energy) Reanalysis II, covered 1979–2010 [3.85] and included far more satellite observations, as well as bias correction and a more refined model. *Saha* et al. [3.86] describe the most recent model and processing.

The NCEP data is on a rather coarse grid, so for storm hindcasts it probably needs to be augmented with an analysis by an experienced meteorologist using all available data. Wind speeds derived from satellite scatterometers can be very helpful in this process.

Hurricanes offer a special challenge since they are small features relative to the scale of regular weather charts. To compensate for this, kinematic or dynamic hurricane models are often used to hindcast hurricane winds. The models typically begin with specification of the atmospheric pressure field. Winds due to that pressure field are found from the gradient wind balance equations. Then the wind is adjusted to 10 m elevation using a boundary layer model. *Holland* [3.42] introduced the radial pressure model

$$p(r) = p_c + \Delta p \exp\left[-\left(\frac{R_{max}}{r}\right)^B\right], \qquad (3.32)$$

where r is the distance from the center of the storm, R_{max} is the radius to maximum winds, Δp is the central pressure deficit, and the Holland B parameter modifies the exponential shape of the pressure curve. If enough data is available, different pressure curves may be used in different storm quadrants. A second exponential function is now often added to account for secondary wind speed maxima. *Cardone* et al. [3.87] give a good description of how the pressure gradient is transformed to boundary layer winds. The Hurricane Research Division HWIND model [3.88] uses these methods to produce wind fields for Atlantic Basin hurricanes.

3.4.2 Waves

Komen et al. [3.89] describe how wave hindcasts solve the transport equation directional wave spectrum $S(f,\theta)$

$$\frac{\partial S(f,\theta)}{\partial t} + v\nabla S(f,\theta) = S_{in} + S_{nl} + S_{ds}\,, \qquad (3.33)$$

where v is the group velocity of the waves, so the left-hand side of the equation represents the advection of wave energy. The right-hand side of the equation schematically lists the source terms for the spectrum: S_{in} represents the input of energy from the wind, S_{nl} represents the nonlinear interactions between wave frequencies, and S_{ds} represents dissipation terms such as bottom friction and wave breaking. Only the nonlinear term is known theoretically, but because its computation is formidable it is greatly simplified in operational models. The other two terms must be parameterized based on experimental data and tuned to observed wave growth. The directional spectrum calculated from the model is summarized as significant wave height, peak and average wave periods, mean wave direction, and wave directional spreading.

The standard wave model, WAM (Wave Modeling Project), was created by an international consortium of wave modelers called the WAMDI group. The development of WAM is thoroughly described by *Komen* et al. [3.89]. WAM has been continually modified, and versions have been installed at many national forecast offices. The NOAA version, WAVEWATCH III, is available for download at ftp://polar.ncep.noaa.gov/pub/wwatch3/v2.22. That site also maintains an archive of forecast and hindcast wave data for US waters.

The accuracy of wave modeling crucially depends on accurate specification of the wind fields. For severe storms, this often requires hand analysis by an experienced meteorologist. Given good wind fields, RMS wave height accuracies of less than 10% can be achieved for extratropical storms [3.90] and for hurricanes [3.91].

The various NCEP reanalysis products have been used to force wave models and generate 50+ year hindcast databases, e.g., [3.92]. The primary limitation of these products (other than NARR (North American Regional Reanalysis)) is the relatively coarse spatial grid (2.5°) and temporal resolution (6 h). *Cardone* et al. [3.93] discuss some of the implications of

this coarse resolution on modeling extreme waves and winds. In short, the reanalysis products suffer from poor resolution of areas of high winds in extratropical storms and in virtually all parts of tropical storms.

3.4.3 Currents, Surge, and Tides

With the advent of relatively fast and inexpensive computers in the 1970s, numerical models of ocean currents in *shallow* water began to proliferate. While the numerical discretization differs, all these models solve a similar set of differential equations conserving mass and momentum, which are often referred to as the shallow water equations. Model output includes the time series of depth-averaged velocity and surface elevation at discrete grid points in the horizontal domain. This class of model is now routinely used to accurately simulate astronomical tides and wind-induced currents and surge in coastal waters where stratification in the water column is not important, typically in 30 m of water or less and beyond the influence of substantial river inflow. *Johnsen* and *Lynch* [3.94] provide numerous examples of these so-called two-dimensional (2-D) models. Several models such as MIKE 21 HD [3.95, 96] and ADCIRC [3.97] have good user interfaces and can be successfully applied by users with modest familiarity with numerical ocean modeling. The accuracy of these models of course depends on the specifics but assuming that the bathymetry and the initial and boundary conditions are specified accurately, modeled currents and surface elevations can achieve a 15% RMS error when compared to measurements.

A similar story can be told for stratified deeper waters but only for some types of forcing such as winds and external tides. Some examples are given in [3.94]. An example for hurricane-generated currents is given by *Frolov* [3.13], who used a 3-D model to accurately simulate the ocean response from Katrina and Georges, both during the initial phase of direct wind forcing and the subsequent phase of inertial oscillations. It is worth noting that during the Katrina simulations, a storm marked by Category 4 winds, Frolov had to cap the surface drag coefficient, as discussed in Sect. 3.2.

Three-dimensional current models can also accurately simulate the longer time and length scales of quasi-geostrophic currents like the Gulf Stream, provided that they have accurate boundary and initial conditions, which are typically provided by satellite altimeters. While these models can achieve less than 20% RMS error on large length scale processes, they have difficulty simulating processes of approximately 100 km or less [3.98, 99]. That is because these processes often exhibit baroclinic instabilities with small length and timescales that are undersampled by the present altimeter array. These limitations can be circumvented to some degree by assimilating fine-scale measurements from ships, drifters, gliders, etc. [3.99]. A number of models are publicly or commercially available, but the skill needed to effectively apply these models is much higher than for the 2-D models, in large part because the underlying physics of a stratified ocean (3-D) are far more complex than an unstratified one (2-D). Examples of generally usable 3-D models include HYCOM [3.100], ROMS [3.101], and MIKE 3 HD. As in the case of 2-D models, the numerical methods employed by 3-D models differ greatly, but the better ones are capable of modeling similar real-world situations with equivalent accuracy.

That is the good news. The bad news is that there are a host of other processes, many of them energetic, where numerical models yield RMS errors of more than 100%. Published examples are hard to find because poor matches tend to go unpublished. However, the authors' experience suggests models have great difficulty simulating internal (baroclinic) tides and solitons, turbidity currents, and river outflows. In these cases, data assimilation is usually impractical because of the short length and time scales of the characteristic processes. In addition, since these cases are dominated by small length scales where turbulence and mixing play an important role, the physics are not well understood. CFD (Computational Fluid Dynamics) may someday be a viable tool but not until computer capacity increases substantially.

A number of extensive data sets of ocean currents have been generated in the last decade and are readily available over the web. These models assimilate data from satellite measurements and sometimes buoy and drifter measurements. Some noteworthy and useful data sets include:

NOAA's RTOFS global model [3.102] provides forecasts up to 7 days. The historical forecasts are *not* downloadable at this time, though that may change in the future (personal communication, Hendrik Tolman, NCEP, Environmental Modeling Center, 23 Aug 2012). The model is based on HYCOM and is composed of curvilinear grid points with variable horizontal sizes spanning 5−17 km. It uses 26 hybrid layers/levels in the vertical.

The HYCOM global model provides forecasts up to 7 days and archives back to 2003, though until 2013 the archive only saved the modeled fields at midnight. The model uses a $1/12^\circ$ grid.

NCOM (Navy Coastal Ocean Model) regional models [3.103] provide forecasts up to 4 days. The historical forecasts are *not* downloadable at this time, though that may change in the future. The models use a $1/36^\circ$ version of the global NCOM model, a version of the

Princeton Ocean Model (POM). Nowcasts have been archived back to 2010. Other regional models are also available: RTOFS for the North Atlantic, MERCATOR for Mediterranean [3.104], and BLUElink for Australia [3.105]. These models assimilate satellite observations within their domain and take their boundary conditions from larger-scale global models.

When utilizing the archive data sets from 3-D models, one must keep in mind the weaknesses described earlier in this section. More specifically, these archived products will not adequately resolve the peak current during tropical cyclones. Nor will they be able to reliably replicate historical mesoscale features, though they may be able to reproduce the statistics of those features (e.g., reproduce the histogram of speed). In summary, if there are energetic ocean current processes with length scales of less than 100 km affecting the site of interest, model archives should only be used with caution. At the very least, several months (preferably much more) of local measurements should be obtained and used to validate and calibrate the model before relying on model results.

3.5 Joint Events

Most ships and offshore facilities are designed to withstand a load with a specific return interval of n years, e.g., the 100-year event. For many decades, offshore designers assumed that the n-y event was created by the simultaneous occurrence of the n-y wind, n-y wave, and n-y current (i. e., the so-called *n-y independent events*), all aligned in the same direction. However, about 30 years ago, metocean researchers started collecting detailed measurements during major storm events and realized that the peaks of winds, waves, and currents, in fact, did not occur simultaneously in direction or time and they began developing various techniques to account for this fact. The more popular ones are described next.

3.5.1 Response-Based Analysis

The simplest and perhaps most accurate way of estimating the n-y response is to feed a time series of wind, wave, current into a response model of the facility and then do an extreme analysis of a key response variable. For example, if a structural engineer designing the legs in an offshore jacket for the 100-y overturning moment (OTM). In the response-based approach, the structural engineer would first develop a fairly simple response function whose input variables include wind, wave, and current and whose output is the OTM. Second, the metocean time series is fed into the response function resulting in a time series of OTM. Third, a peak over threshold (POT) analysis is done, as described in Sect. 3.7.2 and the n-y OTM calculated. Finally, a set of winds, waves and currents that produce the OTM is found and used for detailed analysis.

Ewans [3.106] describes the application of the response-based approach to pipeline stability. *Heideman* et al. [3.107] provide one of the earliest examples of the approach and show that for a jacket-type structure in the North Sea, one can combine the 100-y wind and wave with an *equivalent* current that is 0.25 times the 100-y current to reach the 100-y OTM. Their case is perhaps on the extreme end of potential savings as it is situated in the North Sea where storm winds and waves are weakly correlated to the extreme current. Nevertheless, even in regions dominated by hurricanes where the metocean variables are highly correlated, ANSI/API [3.1] recommends that the 100-y wave can be combined with 0.95 of the 100-y wind speed and 0.75 of the 100-y surface current speed. A further 3% reduction of the current and wind is allowed if directionality is considered.

The response-based approach is versatile and can apply to the calculation of extreme loads like base shear or OTM in a jacket, extreme responses like the n-y heave in a ship, or operating conditions like the marginal probability distribution of pitch and roll. ANSI/API [3.1] recommends the response-based analysis as the preferred alternative. Part of the reason for the rise in popularity of response-based analysis is the increase in computer power, which has made the repetitive solution of fairly complex response functions feasible. Another enabling technology has been the advent of long-duration hindcast datasets of simultaneous wind, wave, and current time series derived from numerical models.

That said, the downside of response-based analysis is the need for a response model with sufficient complexity to accurately reflect the critical response of the facility yet with sufficient computational efficiency to run many thousands of times. Developing such response models can be daunting for complex floating systems like TLPs (tension-leg platform) or spars. Of course there are shortcuts in the analysis that can reduce the computational requirement yet still preserve accurate results. For instance, in the case of calculating extreme loads, the metocean time series can be truncated into a much smaller set of events that only considers the stronger storms. Obviously this approach does not work as well for developing operational criteria.

3.5.2 Load Cases

As discussed in the previous section, a response-based analysis is generally the preferred approach in developing extreme criteria, but in practice the response function is often not easily simplified, so metocean specialists are requested to develop so-called *load cases*. These consist of likely combinations of wind, wave, and current that could cause the *n*-y load or response. For instance, one common load case would be the *n*-y wave and the *associated* wind (the most likely wind velocity to occur simultaneously with the *n*-y wave condition) and associated current. Another analogous case would be the *n*-y current and associated wind and wave. ANSI/API [3.1] makes routine use of load cases in their recommended practice.

There are two main questions to be answered when using load cases: *What combinations of wind, wave, and current can cause the n-y load/response? How is the* associated *value found?* Answering the first question is straightforward for well-studied facilities like offshore jackets. That is because previous work has shown that the *n*-y load for key global responses like base shear and overturning moment occurs during the *n*-y wave and associated wind/current. Other combinations such as the *n*-y wind and associated wave and current, come close but do not exceed the *n*-y wave case. However, for other facility types this may not be true and so there is a risk of missing load combinations with *n*-y recurrence that exceed traditional cases like the *n*-y wave and associated wind/current. One way to mitigate this risk is to provide a broad range of possible load cases, though a firm justification for those cases may be difficult to establish unless a response-based analysis is performed.

Several methods have been developed to answer the second question and these are discussed in the following sections.

Regression Analysis

Using a regression analysis to find the associated values can be straightforward, especially when the primary and secondary variables are well correlated. The analyst starts by estimating the *n*-y value of the primary variable using a peak-over-threshold (POT) method, as described in Sect. 3.7.2. Next, a scatter plot is made of the coincident (in time) primary and secondary variables. If there is some correlation evident in the plot, the data is fit with a curve to derive an equation expressing the secondary variable in terms of the primary one. The associated value can then be found by substituting the *n*-y primary variable into the equation.

Figure 3.6 illustrates this approach for the case where the primary variable is H_s, the significant wave height, and the secondary variable is W, the wind speed.

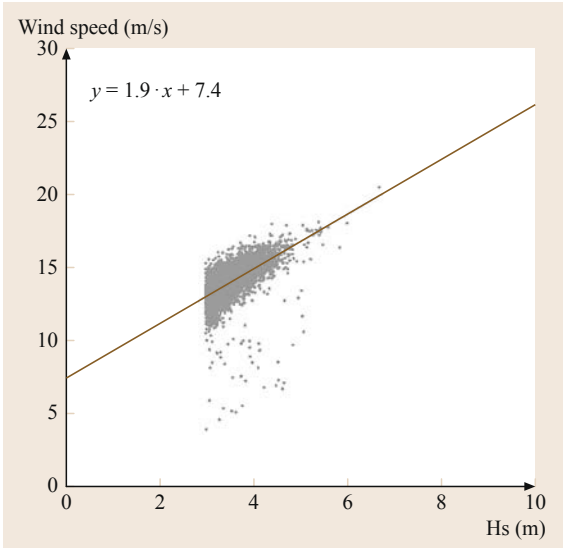


Fig. 3.6 Scatter plot of H_s vs W for all hurricane-generated waves with $H_s > 3$ m. The *red line* shows the least squares fit

The figure suggests that H_s is well correlated to W (correlation coefficient of 0.91) in a linear way. The red line shows the least squares fit with the resulting algebraic expression shown in the upper left-hand corner of the figure. A threshold of $H_s > 3$ m has been applied to remove the weaker winds and waves and make the best-fit curve linear. For this particular dataset, the 100-y H_s is about 9 m, so the red line suggests an associated wind speed of 24.5 m s^{-1}, well less than the 32 m s^{-1} suggested by an independent POT analysis of the 100-y W in this dataset. The 24.5 value represents a mean estimate with a 50% probability of being exceeded. Therefore, one might want to increase that value to reflect the scatter in the data and uncertainty in the fit.

The data shown in the figure was taken from a hurricane dataset, so the highly correlated relationship between the stronger waves and wind is not surprising. However, there are other situations in which the correlation may be weak or nonexistent, such as with currents and wind in deep water. In such cases, it is sometimes reasonable to set the associated value to the mean of the secondary variable. That said, there are subtleties that crop up in certain parts of the world. Consider the derivation of the 100-y wind speed and associated wave off Nigeria where the extreme winds are controlled by squalls that pass quickly and only generate small waves. It would be unconservative to use those squall-generated waves with the squall-generated winds, since much stronger waves are frequently found in the region originating from persistent southeasterlies and/or swell from the Roaring 40s. In this case, a reasonable estimate

of the associated wave for the *n*-y squall case would be the mean for the entire population of wave-producing events during the squall season.

Simulations

Numerical simulations are often one of the best ways of determining associated values. Take, for example, the challenge of estimating the astronomical tide and storm surge to associate with the peak wave crest height. Such a combined event is needed for setting the deck height on jackets. *Fox* [3.108] describes a Monte Carlo approach of numerical simulations to estimate the expected value of the three processes.

Simulations are often the only way to determine the associated value when there are strong nonlinear interactions between the two variables. *Cooper* and *Stear* [3.109] describe an example where the Loop Current and a hurricane simultaneously affected a site in the Gulf of Mexico. While both processes are statistically independent, a tropical cyclone crosses over the Loop or one of its eddies every 3 years, on average, in the deep water Gulf. Most crossings are glancing and of no consequence, but every few decades a hurricane will cross the western half of an eddy or the Loop, resulting in a strong nonlinear interaction that can magnify the subsurface ocean currents by four times the linear superposition of the hurricane-only and Loop-only currents [3.13]. *Cooper* and *Stear* [3.109] estimate the frequency of occurrence of the Loop and hurricane current by shuffling the years from a hindcast historical dataset with a hindcast Loop/eddy database. They then use a lookup table of hindcasted joint hurricane/Loop events to estimate the *n*-y combined current.

3.5.3 Environmental Contours

The largest structural responses may not come from the combination of the largest primary variable and the associated secondary variables. For example, the largest roll response of a floating structure may come from a lower wave height and a wave period that matches the roll period. Those cases can be systematically investigated using environmental contours.

Haver and *Winterstein* [3.110] give a good description of the method and its use. In their example, they fit an extreme value distribution to the significant wave height. Then they fit marginal distributions for peak wave period to ranges of wave height. Finally they fit the parameters of the marginal distributions so they can be extrapolated to low wave height probability levels. This process produces a functional form for the environmental contours of wave height and period.

It is also possible to produce non-parametric environmental contours using a kernel density estimator. In this method, each point in a scatter diagram is replaced by a probability density function. All of those density functions are added together to give a smooth probability density function for the entire data set. According to *Scott* [3.111], if we use a bivariate normal kernel, the optimum standard deviation of the kernel is given by

$$h_i = \sigma_i n^{-1/6} , \tag{3.34}$$

where σ_i is the standard deviation of the data in dimension i, and n is the number of data points.

The resulting probability density can be contoured using standard library functions like MATLAB's contourc.m. The probability levels of the contours are chosen so that the maximum wave heights on the contours equal the independent return period wave height. At low probability levels it may be necessary to limit the steepness of the waves to eliminate waves that are steeper than physically realistic. Figure 3.7 shows significant wave height and peak period contours based on NDBC buoy measurements during hurricanes from 1978–2010 in the Gulf of Mexico.

Similar contours can be calculated for other pairs of parameters such as wave height and wind speed.

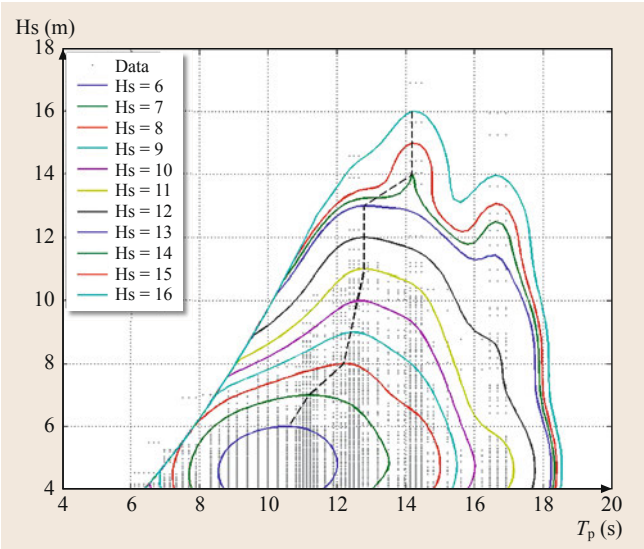


Fig. 3.7 Contours of significant wave height and peak period based on NDBC buoy measurements made in the Gulf of Mexico

3.5.4 Inverse FORM

Once equal probability contours of environmental parameters are calculated, inverse first-order reliability methods (IFORM) provide a general procedure for finding design conditions. The contours are searched for the point which maximizes some response function such

as the roll of a floating system. The environmental parameters at that point then become the design point. *Winterstein* et al. [3.112] give a good explanation of the method.

Inverse FORM maps the environmental contours to standard normal distributions. If the probabilities are expressed as annual extremes and the return period of interest is 100 years, then the probability of exceeding the 100 year value is $p = 1/100$, and the *reliability index* is

$$\beta = \Phi^{-1}(1-p) \,, \tag{3.35}$$

where Φ is the standard normal distribution. The design contour expressed in standard normal variables is then defined by

$$\beta^2 = \sum x_i^2 \,. \tag{3.36}$$

In two dimensions, (3.36) defines a circle. In three dimensions it is a sphere. Each point on the surface has the same probability. The theory extends to hyperspheres in higher dimensions, although the search for the maximum response then becomes much more time consuming.

Suppose the actual environmental parameters are wave height H and period T. If these distributions are independent, then each point on the contour given by (3.36) has the physical parameters

$$\begin{aligned} H &= F_{\mathrm{H}}^{-1}\left[\Phi(x_1)\right] \,, \\ T &= F_{\mathrm{T|H}}^{-1}\left[\Phi(x_2)\right] \,. \end{aligned} \tag{3.37}$$

where F_{H}^{-1} is the inverse wave height distribution and $F_{\mathrm{T|H}}^{-1}$ is the inverse distribution of T given H.

3.6 Operational Criteria

The *operating conditions* are the metocean conditions in which a facility or vessel should be capable of achieving its routine functions. Typical products used to quantify operational conditions include a cumulative probability distribution of wave height or a table of wind speed persistence. These products are used in estimating the fatigue lives for components for which this is a concern. In contrast, extreme conditions rarely occur and are often generated by *storms* of some kind. During extreme conditions, normal operations are usually suspended – the vessel is slowed, oil production may be stopped, windmills feathered, etc. The first two sections below describe several common methods for describing operational criteria of variables that have at least one, highly correlated associated variable, e.g., wind speed and direction. However, the methods are often used even when there are more than one correlated associated variable such as waves, e.g., wave height, wave period, and wave direction.

3.6.1 Probability Distributions

The simplest method for quantifying a variable with a single dimension like wave height is to provide a table or plot of the probability distribution (histogram) as shown in Fig. 3.8a. However, since virtually all metocean variables are vectors, such tables or graphs are typically expressed as joint probability distribution tables, as shown in Fig. 3.8b for the case of wind speed and direction. In this table, each cell shows the probability of the occurrence of wind speed for a given wind direction. A *wind rose* is another way of graphically displaying a vector like wind velocity (Fig. 3.8c). In this case, each *bar* shows the percent occurrence of the speed in discrete bins along the indicated heading. All three images are based on the same dataset, so Fig. 3.8a basically shows a plot of the first column of the table on the x-axis versus the tenth column on the y-axis, while Fig. 3.8c shows the percent occurrence of the speed (binned in $5\,\mathrm{m\,s^{-1}}$ increments) by direction.

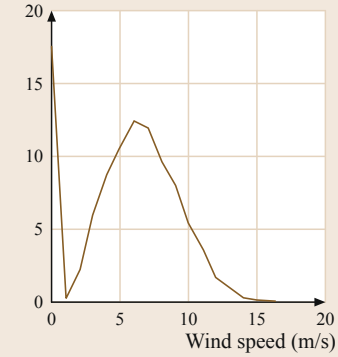
One of the challenges in clearly quantifying the operational environment is dealing with variables that have multiple associated variables that are highly correlated. The section on currents below describes several ways of dealing with this issue for currents.

Waves are typically described by pairing of the associated variables. For instance, one can generate joint probability distribution tables of wave height versus wave period by direction sector. Alternatively, one could generate tables of wave height versus heading by period bin, e.g., a table like that shown in Table 3.1 for all wave periods between 10–12 s.

Many facilities are sensitive to wave fatigue, so designers need the probability distribution of the wave spectra. For regions dominated by single-mode spectra this is straightforward – one can use the tables described in the previous paragraph in conjunction with parametric spectra like JONSWAP. In other words, knowing the probability of a discrete bin of wave height, period, and direction, one can calculate the corresponding spectra at that probability level. Further refinements may be needed if the spectral width and/or directional spreading vary in the region.

Many regions of the world such as Brazil experience sea states characterized by spectra that have multiple peaks, several of which contain substantial en-

a) *P* (Spd = s)



b)

Spd (m/s)	Hdg(from) 0	45	90	135	180	225	270	315	Sub.	Total
2.5	18.61	3.33	7.92	4.89	2.51	1.15	0.94	0.73	40.09	40.09
7.5	0.09	9.59	26.17	7.05	3.27	2.20	0.95	0.56	50.53	90.61
12.5	0.07	1.90	6.07	0.21	0.30	0.21	0.26	0.12	9.13	99.74
17.5	0.01	0.01	0.02	0.02	0.02	0.06	0.05	0.02	0.19	99.93
22.5	0.00	0.01	0.00	0.00	0.02	0.00	0.01	0.01	0.04	99.98
27.5	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.02	99.99
32.5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	100.00
37.5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00
Sub.	19.38	14.83	40.18	12.17	6.12	3.61	2.25	1.45	100	100
Total	19.38	34.22	74.40	86.57	92.69	96.30	98.55	100	100	100

c)

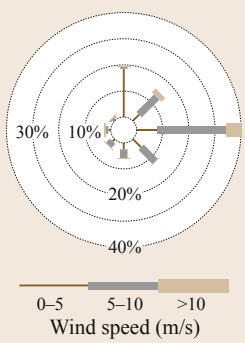


Fig. 3.8a–c Samples of typical methods of displaying the probability distribution. Panel **(a)** shows the probability distribution of wind speed, **(b)** shows tabular marginal distribution of wind speed and direction, and **(c)** shows wind rose of wind velocity. **(a–c)** use the same dataset

ergy. For floating facilities, the weaker secondary or tertiary peaks may be close to resonance of the facility and thus cause far larger forces than the primary peak. In these situations, it can be very unconservative to utilize single-peak spectra like JONSWAP. Perhaps the most widely used dual-peaked spectrum is that of *Ochi–Hubble* [3.17].

3.6.2 Persistence

Certain types of offshore operations require that the metocean environment not exceed a threshold for a specific period of time. If it does, the operation is suspended and there is *downtime*. While estimates of downtime can be made using the probability distributions described in the previous section, such an approach is an oversimplification that can distort the perceived risk. A more accurate method is to scan a time series of the variable of interest and characterize the periods when the variable lies above or below a specified threshold. For example, consider the case where a wind sensitive operation can be completed in 12 h, provided that the wind never exceeds 7.5 m s^{-1}. A simple frequency analysis shows that winds at this site exceed 7.5 m s^{-1} nearly 60% of the time, which at first glance might be discouraging. However, a persistence analysis of the

Table 3.1 Calm persistence for 1-y time series of wind gusts

Threshold [m s^{-1}]	2.50	5.00	7.50	10.00
# of Occur.	0.00	3.00	68.00	70.00
Avg. days	0.00	1.23	2.21	4.56
Max days	0.00	1.46	9.52	30.25
Min days	0.00	1.02	0.54	0.06
Std. dev.	0.00	0.22	1.42	5.87
CDF (% <)	0.00	1.03	41.75	88.72

Table 3.2 Storm persistence for a 1-y time series of wind gusts

Threshold [m s^{-1}]	2.50	5.00	7.50	10.00
# of Occur.	2.00	5.00	68.00	69.00
Avg. days	179.79	71.18	3.08	0.59
Max days	323.30	233.61	43.38	2.96
Min days	36.27	10.19	0.02	0.02
Std. dev.	202.96	92.48	6.01	0.71

events *below* 7.5 m s^{-1} (Table 3.1) indicates that there were 68 calm events in which the wind was less than 7.5 m s^{-1} and all of them lasted more than 12 h (minimum 0.54 days). A closer look at the events *exceeding* 7.5 m s^{-1} (Table 3.2) shows that when the winds did exceed 7.5 m s^{-1}, the events lasted an average of 3.08 days and roughly 85% (mean + standard deviation; 3.08 + 6.01) of these events lasted less than 9.09 days.

Thus by looking at persistence one could conclude that there is an expected downtime of about 3 days for the operation, which is a lot less onerous than might be concluded from looking at the 60% occurrence rate based on the frequency analysis.

While persistence analysis can provide valuable insights, it cannot easily incorporate multiple variables. This is especially limiting for floating systems, which are often dependent on wave height, period, direction, etc. For these cases, numerical simulations using a vessel response function are often preferred, e.g., *Beamsley* et al. [3.113].

3.6.3 Currents

Fatigue damage caused by currents is an important design consideration for oil drilling and production risers in deep water. Deep water current profiles have complicated shapes, and thousands of profiles are often now

available from models or measurements. All of that information must be condensed into a manageable number of cases for analysis. *Prevosto* et al. [3.114] discuss three methods for doing this and compare the results of using them with a full analysis of all profiles.

Empirical orthogonal functions (EOF) provide a method for capturing the important characteristics of current profiles in a few variables. *Forristall* and *Cooper* [3.115] outline the method and give examples. Singular value decomposition permits any matrix **A** to be decomposed as

$$\mathbf{A}_{ij} = \sum_{k=1}^{N} w_k \mathbf{U}_{ik} \mathbf{V}_{jk} \,. \tag{3.38}$$

Each current profile is written as a row in matrix **A** and each column represents the time series at one depth. The columns of **V** are called the EOFs. Each EOF is a vector with a value at each depth in the original data. There are the same number of functions as there are depths. They play the same role as cosine waves in a Fourier analysis. The diagonal elements of **W** are the magnitudes of the EOF modes. They give the relative importance of the modes. The matrix **U** gives the amplitudes of the modes in each current profile. There is one row in **U** for each profile. It gives the amplitudes of each mode at one time.

As it stands, (3.38) is not a more efficient representation of the data. The gain in efficiency comes from the fact that the magnitudes of the first few modes are often much larger than the rest. A good representation of the data can then come by summing over many fewer than N modes. The amplitudes of those modes can then fill a manageable scatter diagram.

There are, however, locations where a few EOF modes fail to describe all the dominant characteristics of the current profiles. The characteristic current profile (CPC) was developed by *Jeans* et al. [3.116] to work with those cases. For each current velocity time series, a number of possible states are defined at each selected depth level, and possible characteristic profiles are constructed from every permutation of these states. The number of measured profiles corresponding to each of these possible characteristic profiles is then counted and percentage occurrence values derived. The reduction in the number of profile shapes is accomplished by selecting a relatively small number of depth levels.

Self-organizing maps (SOM) are useful to better categorize current profiles. The SOM process begins with a two-component EOF analysis. Then, a nonlinear cluster analysis groups the thousands of current profiles into a smaller number of clusters [3.117]. The EOF amplitudes are varied to produce a two-dimensional array of current profiles. Each original profile is assigned to the EOF profile that it best matches. The EOF profiles are modified by taking weighted averages of the neighboring profiles in the grid. Then, the original profiles are re-assigned to the modified profiles that they best match. This process is iterated until the sum of differences between the SOM profiles and the original profiles is minimized. If the array of profiles is small, there can be a lot of variability around some of the weaker SOM profiles. The variability around the SOM profiles decreases when more profiles are used.

Prevosto et al. [3.114] found that using a few hundred profiles calculated by one of these methods gave good accuracy in fatigue damage calculations.

3.7 Extreme Criteria

3.7.1 Risk and Reliability

Metocean design specifications should be set considering the risk and cost of failure. The risk tolerance is different for structures that are not normally manned and structures that are evacuated before severe storm conditions than it is for structures that are manned and not evacuated before severe storms. Gulf of Mexico structures are evacuated upon the approach of a hurricane. North Sea structures remain manned during frequent severe winter storms. For structures that are unmanned or evacuated, the risk calculation is complicated but straightforward. The cost of strengthening the structure is balanced against the monetary cost of structural damage or failure. The cost includes not only repairing or replacing the structure, but sometimes also lost production, pollution-related costs, and damage to corporate image. These costs can be an order of magnitude greater than the cost of replacement.

The failure rate is found by calculating the ultimate strength of the structure and comparing it to the metocean loading at different probability levels. The cost of strengthening the structure is then added to the cost of failure after strengthening. If the total cost is lower, designing to a lower probability of failure is economically justified. For standard steel jacket structures, an annual failure rate near 10^{-3} is generally appropriate. This is consistent with the normal practice of designing for a 100-y storm because steel jackets have considerable reserve strength beyond the first yielding of a member.

Establishing an appropriate failure rate for a manned structure is conceptually more difficult because no

one wants to put a price tag on human life. Rational risk levels can still be set by considering risk levels in other industries and risk on offshore structures due to causes other than structural failure. Those risks include travel, explosions, collisions, and falls. These considerations indicate that the individual risk per annum (IRPA) should be reduced to a level below 10^{-3}. An IRPA lower than 10^{-6} is considered negligible. However, the risk should be reduced below 10^{-3} to a level that is as low as reasonably practical (ALARP). Measures to reduce IRPA should be examined and implemented until the cost of the upgrade becomes grossly disproportionate to the benefit obtained. *Efthymiou* et al. [3.118] discuss the ALARP principle in detail. The cost of stuctural strengthening should be weighed against the cost of lowering other risks. Manned structures in the North Sea are now usually designed for a 10^{-4} annual risk of failure. Providing criteria with such a low probability is a special challenge to the metocean specialist, who must extrapolate conditions far beyond experience levels.

3.7.2 The Historical Method

The traditional way of estimating metocean extreme values is extrapolation from historical data. The data generally come from hindcasts rather than measurements so that more years of data are available. However, even hindcast records are short compared to 1000 or 10 000 years, so extrapolation using extreme value distributions is required.

Extreme value theory assumes a time series of independent events, so the first step is to choose those events. Generally, this is done by finding the peak values over a threshold (POT). The peaks are sorted in ascending order, and their probability is plotted against their magnitude. There are then many choices for choosing an extreme value functional form and fitting it to the data [3.119]. In the limit as the number of points tends to infinity, it can be shown that the extremes follow the generalized Pareto distribution

$$F(y) = 1 - (1 + \xi y/\sigma)^{-1/\xi} \,. \tag{3.39}$$

The parameter ξ controls the shape of the distribution, giving a heavy tail if $\xi > 0$ and a finite upper limit if $\xi < 0$. In practice, we do not know whether the data extends far enough into the tail of the distribution for the limit to hold, and small changes in the data can influence whether an upper limit is predicted. For these reasons, engineers often choose to fit peaks to the Weibull distribution

$$F(y) = 1 - \exp\left[-\left(\frac{y-\gamma}{\alpha}\right)^{\beta}\right] \,. \tag{3.40}$$

The commonly used Weibull plotting position is

$$P_i = \frac{i}{N+1} \,, \tag{3.41}$$

but *Goda* [3.120] showed that the unbiased plotting position is actually

$$P_i = \frac{i + \frac{0.5}{\sqrt{\beta}} - 0.6}{N + 0.2 + \frac{0.23}{\sqrt{\beta}}} \,. \tag{3.42}$$

Gibson et al. [3.121] tested various methods of fitting (3.40) to data simulated from a known Weibull distribution. They found that both least squares and maximum likelihood fits gave good results when

1. The unbiased plotting position in (3.42) was used.
2. The position parameter γ was set to the threshold value.
3. All of the data were used instead of binning the data into ranges of y.
4. The fit was made of y to $\alpha\,[\ln(1-P)]^{\beta} + \gamma$.

There are numerous weaknesses with the historical approach. First and foremost, historical datasets are often short relative to the probability level needed for design criteria. This is especially a problem for tropical storms, which are spatially small and infrequent. *Toro* et al. [3.122] show the 100-y criteria for a particular site in regions like the Gulf of Mexico, is most heavily influenced by how close a few strong storms passed to the west of the site. As a consequence, large differences in the 100-y design condition are often observed at sites separated by only 50 km in deep water far from the coast where there is no physical basis to believe the n-y condition would be any worse at one site than the other. These unrealistic spatial gradients are more apparent at even shorter return periods (e.g., 10 and 25-y) in basins like north Australia and the Gulf of Thailand, where the reliable historical database is shorter or the storm frequency is lower than in the Gulf of Mexico. The spatial gradients in n-y criteria are largely a result of under sampling. There are simply not enough larger storms in the database. Furthermore, it is reasonable to expect that if the database could be extended over a longer period of time, a strong storm would eventually cross near all the sites.

To counter this under sampling, metocean experts often *pool* nearby sites, as described in *Cooper* et al. [3.123]. Pooling basically combines all storm peaks from several nearby sites into a single probability distribution. In essence, pooling adds *synthesized* storms by shifting the tracks of the historical storms. When pooling, the probability distributions from each site are assumed to be statistically independent, even though they are not. However, *Toro* et al. [3.122] show

that this assumed independence does not lead to substantial errors for return intervals of up to roughly 100 years for the Gulf of Mexico hurricane population. That is because the 100-y condition is strongly dependent on the track crossing distance and much less dependent on the other particulars of the storm. Unfortunately, for rarer return periods beyond a few hundred years, pooling starts to yield increasingly biased results because the longer return intervals become more sensitive to the intensity of the stronger storms, and pooling does nothing to increase that population of storms. When pooling it is important not to extend the averaging grid too widely else one can suppress real spatial gradients like those suspected to exist in the Gulf of Mexico hurricane patterns, e.g., *Cooper* [3.124].

Another limitation with the historical method is that it cannot easily be used to develop criteria involving the rare combination of two relatively independent events. A case in point is the superposition of the Loop Current (or one of its detached eddies) and a hurricane. Recent simulations by *Cooper* and *Stear* [3.125] suggest these events happen roughly every 4 years. When they do, a number of potential nonlinear interactions can occur such as wave focusing [3.126], amplification of mid-water currents [3.13], and intensification of the hurricane [3.124]. The first two phenomena depend strongly on the distance between the hurricane and Loop, and there are virtually no comprehensive measurements of the wave and current field in joint events. Hence, historical events are missing.

3.7.3 Synthetic Storm Modeling

The previous section pointed out two major weaknesses of the historical approach. To address these weaknesses, researchers have looked at various means of generating so-called *synthetic* storms; that is, storms that did not actually occur but could have occurred. *Georgiou* et al. [3.127] describe one of the first efforts. They first fit standard storm parameters like intensity and radius to standard distribution functions (e.g., lognormal). They then drew randomly from these distributions to construct synthetic storms whose probability was calculated from the underlying distributions of the storm parameters. Once the combination of storm parameters was selected, these were input into a standard parametric wind model that could calculate the detailed wind field along the historical tracks. Because the probability distributions of each hurricane parameter is at most weakly dependent on the other parameters (e.g., radius to maximum wind is only weakly correlated to intensity), the overall probability of a given synthetic storm scales roughly as the product of the probability of the individual parameters. Hence the method can generate rare (low probability) synthetic storms using a combination of storm parameters that are well away from the tail of their respective probability distributions and hence have relatively low uncertainty.

While the early models went a long way in reducing statistical uncertainty of the longer return period estimates, they continued to utilize historical tracks to estimate the frequency of storm passage and they assumed that the change in storm parameters was independent of that track. This latter assumption is clearly problematic in places like the Gulf of Mexico, where the warm waters of the Loop Current likely affect storm intensity as do nearby land masses. To partially address these limitations, *Vickery* et al. [3.128] used statistical properties of track heading, track speed, and intensity, combined with a regression model to generate synthetic storms. This approach allows for the generation of thousands of years of storms with low statistical uncertainty. *Emanuel* et al. [3.129] investigated stochastic techniques to generate many synthetic storm tracks and a deterministic model to calculate storm intensity along each of those tracks. They investigated two track models. Their first model was conceptually similar to that of *Vickery* et al. [3.128], while their second track generation method accounted for large-scale weather, including vertical shear and steering flow. Once *Emanuel* et al. [3.129] had constructed the tracks, they used a deterministic model to calculate the parameters, including intensity and radius. *Vickery* et al. [3.130] used a track model that accounts for large-scale weather but in a more deterministic fashion than *Emanuel* et al. [3.129], by using NCEP reanalysis. To calculate the storm parameters they used a statistical intensity model that incorporated atmospheric inputs, much as Emanuel et al. [3.129] did, but *Vickery* et al. [3.130] also included ocean temperature feedback.

Perhaps the biggest challenge in using these models is determining whether some of the more extreme synthetic storms are realistic. The next section addresses this point.

3.7.4 Modeling Versus Measurements

In an ideal world, the ocean would be covered with measurement sites that have operated for centuries. In the real world, the metocean specialist is often faced with developing criteria where there are no measurements at the site, or if there are, they may only be a year or less in duration. Extrapolating such a short record to return intervals of a few decades or more will usually result in large statistical uncertainty at best, and at worst, large biases. On the other hand, numerical model hindcasts spanning many decades now cover most of the world, as discussed in Sect. 3.4. Depending on the hor-

izontal resolution of the model grid, the nearest model element to the site of interest is often only a few kilometers away. Thus the best strategy for developing extreme design criteria at site is often to use available measurements to calibrate a hindcast model that has been run for several decades, rather than to do an extreme analysis on the measurements themselves.

3.7.5 Accounting for Physical Limits

Whether one uses the historical method or synthetic modeling to estimate extremes, extrapolation of some form is almost always being used to estimate criteria well beyond any observed storms. This raises the concern that the method can be generating values that cannot be physically attained in the real world. Perhaps the clearest example of this danger is the case where a metocean specialist tries to fit a Weibull distribution (historical method) to waves measured during a 2-y long measurement program in a water depth where wave breaking can occur for the stronger storms. Fitting this kind of data with a classic historical method can yield 100-y estimates that are unrealistically high because those waves would have broken in the real world. The problem, of course, is that the extreme distributions are purely statistical functions with no physical basis.

One solution to this potential problem is to use numerical models that include the necessary physics to account for the limits. This can be a practical and straight forward solution for the example of breaking waves cited above. In that case, hindcasting storms over many decades using a wave model that accounts for breaking is usually a quick and effective solution.

Regrettably, incorporating physical limits into numerical models is not straightforward when the physics are not well understood. A case in point is calculating extremely rare hurricane conditions, say the 10 000 year significant wave height. *Cardone* and *Cox* [3.131] applied a third-generation wave model to strong storms and found that the wave heights trended toward an asymptotic limit. However, it is debatable whether the asymptotic limit is generated by real physical limits or artificial ones imposed by the model equations. There is no way to be sure, as wave and wind measurements during the events considered by Cardone and Cox have not been recorded. Another approach used by *Vickery* et al. [3.130] applied the concept of the maximum probable storm intensity. *Emanuel* [3.132] and others provide evidence that such limits appear to exist.

3.7.6 Seasonality

Fixed offshore facilities are designed for year-round conditions but there are some instances where metocean conditions are needed for seasonal construction or drilling. If the operation is, say, planned for only the three months of summer, then only the metocean conditions for those months need to be considered. More specifically, if a 1%/y risk of failure is desired (expected failure of once every 100-y), then the extreme values of metocean conditions in 100 years of summers should be calculated.

However, caution must be exercised when considering seasonality for a drilling rig or operations which will continue year-around. To illustrate this point, consider the question of how one might combine seasonal criteria to calculate the annual survival rate. Assume that the target reliability is an average 99% survival rate (1% failure rate) each year. One might be tempted to use the 99% probability value for each season, but more careful consideration reveals this will badly overestimate the survivability. That is because the annual survival rate is given by the probability that the rig will survive the summer *and* the fall *and* the winter *and* the spring. It follows that if the extremes in each season are statistically independent, then the annual survival probability is given by the product of the seasonal probabilities, or $0.99^4 = 0.96$; 4% less than the annual target survival rate of 99%. An obvious solution to this shortfall is to use the 0.9975 probability for each of the four seasons, which yields an annual survival rate $99\% = 99.75^4$.

3.7.7 Directionality

Directional metocean specifications are sometimes desired when a structure is considerably stronger or less prone to motion in some directions than others. The considerations in this case are similar to those for seasonal specifications, especially the concept that the total survival probability from all individual directions should not be significantly different from the omnidirectional survival probability. Using the same arguments given in the previous discussion on seasonality, it is clear that using the n-y metocean criteria in each direction bin will give a much lower survival probability than using the n-y omni-directional criteria [3.133].

The simplest way to insure a reasonable result is to make the probabilities in all of the direction bins equal. So, for example, if the target annual survivability is 99% and four direction bins are used, then the target survival probability in each of these four bins should be 99.75%.

3.7.8 Combining Long and Short-Term Distributions

Estimating extreme values of individual wave and crest heights requires combining a long-term extreme value

distribution with the short-term distributions discussed in Sect. 3.2.2. *Borgman* [3.134] showed how the maximum wave and crest height during a storm can be estimated by integrating the short-term wave and crest height distributions over the storm's sea state history. *Tucker* and *Pitt* [3.135] give a thorough description of how the Borgman integral has been applied to the calculation of extreme wave heights. *Forristall* [3.136] validated these methods using long-time series of individual waves.

If the probability that the wave or crest height exceeds η is given by $P(\eta)$, then the probability that the height will not exceed η in N waves is given by

$$P(\eta_{\max} < \eta) = [1 - P(\eta)]^N \ . \tag{3.43}$$

For a sequence of records $i = 1 \ldots k$ during a storm, the probability of non-exceedance becomes

$$P(\eta_{\max} < \eta) = \prod_{i=1}^{k} [1 - P_i(\eta)]^{N_i} \ . \tag{3.44}$$

The calculations can be performed more accurately by taking the logarithm of (3.44) to give

$$\log [P(\eta_{\max} < \eta)] = \sum_{i=1}^{k} N_i \log [1 - P(\eta)] \ . \tag{3.45}$$

Equation (3.45) is applied to calculate the expected maximum wave and crest height in each storm. The sets of maximum heights are then fit to extreme value distributions to determine the return period individual wave and crest heights.

Fitting an extreme value distribution to the most probable maxima does not account for the fact that values higher than the most probable value can occur in any storm. *Tromans* and *Vanderschuren* [3.137] proposed a method for taking account of this short-term variability. If the probability distribution of the most probable maxima in a storm is $P(H_{\text{mp}})$ and the distribution of the maximum given H_{mp} is $P(H|H_{\text{mp}})$, then the distribution of the maximum in a single random storm is

$$P(H|srs) = \int P(H|H_{\text{mp}}) p(H_{\text{mp}}) \mathrm{d}H_{\text{mp}} \ . \tag{3.46}$$

Tromans and Vanderschuren found that $P(H|H_{\text{mp}})$ was very similar from one storm to another, and that its mean could be described by the function

$$P(H|H_{\text{mp}}) = \exp\left(-\exp\left\{-\log N\left[\left(\frac{H}{H_{\text{mp}}}\right)^2 - 1\right]\right\}\right) \ , \tag{3.47}$$

where N is an *equivalent number of waves* in a storm. The value of $\ln N$ can be estimated from the short-term distributions in the historical storms. Values between $\ln N = 8$ and $\ln N = 10$ are typical, and the results are not very sensitive to the exact value. The result is to increase the estimates of extreme maximum wave and crest heights to about 5% more than the most probable values.

3.7.9 Rogue Waves

It is generally agreed that a *rogue wave* is one with a height greater than 2.2 times the significant wave height or a crest greater than 1.25 times the significant wave height. There have been many reports of such waves in the literature in the last few years. The best known is the *Draupner wave*, recorded in the North Sea on January 1, 1995 [3.138]. The crest height of this wave was 1.55 times the significant wave height. Unfortunately, very little is known about the instrumentation used for this measurement. The *Andrea wave* [3.139] is a much better documented case. It was also recorded in the North Sea on November 9, 2007. Essentially, the same wave was recorded by four laser altimeters. Analysis of the intensity of the return signals indicated that there was no sea spray at the wave crest. The height of the Andrea wave was 2.49 times the significant wave height, and the crest was 1.63 times the significant wave height.

The central question in the study of rogue waves is whether they can be explained as a statistical anomaly or whether they require a physical explanation different than second-order theory [3.140]. The statistical explanation for something like the Andrea wave is certainly a stretch. According to second-order statistics, its crest had a probability of 6×10^{-8}. However, *Christou* and *Ewans* [3.141] did a careful study of over 10^8 measured waves and found that the sample crest distribution was only slightly higher than predicted by second-order statistics.

Some processes that produce very high waves are understood reasonably well. Waves traveling into an opposing current can steepen and become much higher. Many ships have been damaged when they encountered such waves in the Agulhas Current south of Africa. Bottom features can refract waves, making them much larger in localized areas. Surfers are well aware of this phenomenon. The more difficult cases to explain are unusually high waves in deep water far from shore.

Theoretical attempts to explain rogue waves involve the integration of nonlinear equations that approximate the development of steep random waves [3.142]. All of these show a modulation of the wave envelope similar to the Benjamin–Feir instability observed in regular

waves. The modulated waves alternate between series of higher and lower waves than predicted by Gaussian statistics. The kurtosis of the wave trace becomes greater than 3.0, and the extreme waves are higher than the Rayleigh distribution. Steep random unidirectional waves in laboratory basins often show this behavior. However, several numerical and laboratory studies, such as that by *Toffoli* et al. [3.143], have shown that modulational instabilities are much less effective in producing large wave groups when the waves are spread.

Rogue waves remain an active area of research and it is too early to draw definite conclusions. Fortunately, rogue waves may not have a big influence on extreme wave heights for design. Almost by definition, they have a low probability. The probability that a rogue wave occurs during one of the sea states far out in the distribution of significant wave height is even lower. *Haver* [3.138] estimated the effect of adding rare rogue waves to the short term distribution and found that it had little effect on the risk of failure.

3.7.10 Extremely Rare Events

Designers now frequently request metocean criteria with return periods of 1000–10 000 years. Deriving these values from a few years of measurements is difficult, if not impossible, to justify. Even deriving such rare events from historical hindcasts is problematic, since reliable historical databases rarely extend beyond 50 years. Synthetic modeling certainly holds the most promise for deriving rare events, but even with this technology it is difficult to overcome our ignorance of the physical limits which probably occur for many metocean phenomena. This issue is discussed in more detail in Sect. 3.7.5.

3.7.11 Quantifying Uncertainty

Uncertainty affecting the calculation of metocean extremes comes primarily from the noise and/or bias in the numerical models or measurements used to generate the peaks, and from the inability of the chosen extreme distribution to fit the peaks – what is often referred to as statistical uncertainty. The impacts of these two types of errors on extreme value uncertainty are discussed in more detail in Sect. 3.7.4.

Statisticians have extensively studied statistical uncertainty and developed numerous ways of quantifying it, as discussed in *Tucker* and *Pitt* [3.135]. If the input peaks come from a short time series and the extrapolation is lengthy (e.g., 2-y of measurements extrapolated to a 100-y return period), then the statistical error can be large.

One often sees fits to extreme distributions that show confidence limits that are based on the statistical uncertainty only. If the peaks are based on site-specific measurements, the statistical uncertainty is fairly representative, but if the peaks come from measurements some distance from the site or from models, then the statistical uncertainty is probably much smaller than the uncertainty from the input data source. Sections 3.4.1 to 3.4.3 can help quantify that error.

3.7.12 Stationarity

Nonstationarities can be thought of as low-frequency processes that have been sampled at far less than their Nyquist frequency. For example, if one has only a few months of data to analyze, then nonstationarities will arise from seasonal, annual, decadal, etc., time scales. Innumerable papers and books have been written on the topic, including a relatively recent one by *Rao* et al. [3.144].

Issues regarding nonstationarities have always plagued metocean analysis. The challenge is perhaps greatest when dealing with the calculation of extreme values (e.g., 100-y wind speed) where stationarity of the underlying time series is assumed in almost any analysis method and nonstationarities in the underlying dataset will tend to be amplified. For storm extremes, important sources of nonstationarities can come from natural oscillations in the atmosphere like the North Atlantic Oscillation or El Nino, which can cause substantial variations in storm severity over periods of several decades [3.145]. In theory, the obvious solution is to include many decades of historical storms in the extreme value analysis, but such long time series are available in only a few regions of the world and even there, data quality from the older decades my be problematic and introduce other forms of bias [3.146].

Global warming is introducing strong nonstationarities in many variables, the most obvious being atmospheric temperature. Projections from the IPCC (Intergovernmental Panel on Climate Change) [3.147] show that these nonstationarities or trends will increase rapidly over the coming decades and for long-lived facilities, the changes will need to be considered. A starting point for estimating nonstationarities is to use projections from numerical models such as those provided by the IPCC [3.147]. However, these projections do not consider all variables of interest to engineers (e.g., waves) and use models with fairly large grid sizes which can fail to capture important regional variability. Fortunately, computer power is continuing to increase, so the limits on grid size are starting to recede, allowing for the development of regional nested models with smaller grid sizes [3.148].

3.8 Conclusions

The metocean environment controls many aspects of facility design and operation so errors in quantifying metocean conditions can cascade though the design and operational decisions. Errors can result in damage and lost lives. Conversely, if the variables are overestimated, costs will be overestimated perhaps to the point that the project becomes uneconomic and is never built.

Metocean criteria are typically broken into two categories: operating and extreme. The former involves quantification of metocean conditions in which the facility or vessel should be capable of achieving the routine functions of its primary purpose. In contrast, extreme conditions occur rarely and are often generated by episodic events (e.g., storms). Both categories may start with the same databases but the analysis techniques and final design specifications will differ substantially.

There are a host of sophisticated methods and tools that can be used to quantify the most important metocean variables that impact offshore facilities. We have suggested methods drawn largely from the offshore oil and gas industry but they are also generally applicable to other engineering applications involving the design and operation of vessels, coastal structures, offshore wind farms, navigational aids, coastal geomorphology, and pollution studies.

When developing a metocean design basis for a major project, the metocean engineer should first identify the variables of primary importance. This is because the sea and atmosphere are filled with complicated processes, many of which are site specific and poorly understood. If aggressive filtering is not undertaken then too much time can be spent quantifying variables that make little difference to the design or operation of the facility. The first and best way to eliminate variables from investigation is to understand the basic responses of the *particular* facility. In other words to answer the question: which metocean variables impact this facility most and which have little or no impact?

Finally, it should be noted that the methods, tools, and databases cited in this chapter reflect a snapshot in time; they are continually being updated and improved. The reader should always consider these citations as a starting point and check the web and journals for updates before proceeding with the analysis.

References

3.1 ANSI/API: *Recommended Practice 2MET. Derivation of Metocean Design and Operating Conditions* (API, Washington 2014)

3.2 U.S. Army Corps of Engineers: Coastal Engineering Manual, Report Num. EM1110-2-1100 http://chl.erdc.usace.army.mil/cem

3.3 L.F. Richardson: Atmospheric diffusion shown on a distance-neighbor graph, Proc. R. Soc. London Ser. A **110**, 709–737 (1926)

3.4 NORSOK Standards: *N-003: Actions and Action Effects. Rev. 2* (Standards Norway, Lysaker 2007)

3.5 P.J. Vickery: Analysis of hurricane winds, Proc. Offshore Tech. Conf. (2014)

3.6 Engineering Sciences Data Unit: *Strong winds in the atmospheric boundary layer, Part 1: Mean hourly wind speed, No. 82026* (ESDU, London 1982)

3.7 Engineering Sciences Data Unit: *Strong winds in the atmospheric boundary layer, Part 2: Discrete gust speeds*, No. 83045 (ESDU, London 1983)

3.8 W.G. Large, S. Pond: Open ocean momentum flux measurements in moderate to strong winds, J. Phys. Oceangr. **11**, 324–336 (1981)

3.9 P.J. Vickery, D. Wadhera, M.D. Powell, Y. Chen: A hurricane boundary layer and wind field model for use in engineering applications, J. Appl. Meteor. Climate **48**, 381–405 (2009)

3.10 M.D. Powell, P.J. Vickery, T.A. Reinhold: Reduced drag coefficient for high wind speeds in tropical cyclones, Nature **422**, 279–283 (2003)

3.11 M.A. Donelan, B.K. Haus, N. Reul, W.J. Plant, M. Stiassnie, H.C. Graber, O.B. Brown, E.S. Saltzman: On the limiting aerodynamic roughness of the ocean in very strong winds, Geophys. Res. Lett. **31**, L18306 (2004)

3.12 M. Powell: New findings on hurricane intensity, wind field extent and surface drag coefficient behavior, 10th Int. Workshop Wave Hindcast. Forecast. Coast. Hazard Symp. (2007) pp. 11–16

3.13 S. Frolov: Ocean response to hurricanes in presence of the Loop Current, Offshore Technol. Conf. (2010)

3.14 J.W. Miles: On the generation of surface waves by shear flows, J. Fluid Mech. **6**, 568–582 (1957)

3.15 D.M. Glover, W.J. Jenkins, S.C. Doney: *Modeling Methods for Marine Science* (Cambridge Univ. Press, Cambridge 2011)

3.16 Y. Goda: *Random Seas and the Design of Marine Structures* (Univ. of Tokyo Press, Tokyo 1985)

3.17 M.K. Ochi, E.N. Hubble: Six parameter wave spectra, Proc. 15th Coast. Eng. Conf. (1976) pp. 301–328

3.18 K. Torsethaugen, S. Haver: Simplified double peaked spectral model for ocean waves, 14th Int. Offshore Polar Eng. Conf. (2004)

3.19 G.Z. Forristall: On the statistical distribution of wave heights in a storm, J. Geophys. Res. **83**, 2353–2358 (1978)

3.20 G.Z. Forristall: Wave crest distributions: Observations and second-order theory, J. Phys. Oceanogr. **30**, 1931–1943 (2000)

3.21 A.V. Babanin: *Breaking and Dissipation of Ocean Surface Waves* (Cambridge Univ. Press, Cambridge 2011)

3.22 J.W. Van de Graaf, P.S. Tromans, M. Efthymiou: The reliability of offshore structures and its dependence on design code and environment, Offshore Technol. Conf. (1994)

3.23 N. Booij, R.C. Ris, L.H. Holthuijzen: A third-generation wave model for coastal regions 1. Model description and validation, J. Geophy. Res. **104**(C4), 7649–7666 (1999)

3.24 M.S. Longuet-Higgins, E.D. Cokelet: The deformation of steep surface waves on water, I. A numerical method of computation, Proc. R. Soc. A **350**, 1–26 (1976)

3.25 M. Christou, C. Swan, O.T. Gudmestad: The description of breaking waves and the underlying water particle kinematics, Proc. Offshore Mech. Arctic Eng. (2007)

3.26 J.F. Price, T.B. Sanford, G.Z. Forristall: Forced stage response to a moving hurricane, J. Phys. Oceangr. **24**, 233–260 (1994)

3.27 L.H. Kantha, C.A. Clayson: An improved mixed layer model for geophysical applications, J. Geophys. Res. **99**, 25235–25266 (1994)

3.28 L.H., Kantha, C.A. Clayson: *Small Scale Processes in Geophysical Fluid Flows* (Academic Press, San Diego 2000)

3.29 M. Tomczak, J.S. Godfrey: *Regional Oceanography: An Introduction* (Pergamon, Oxford 1994)

3.30 M.M. Hall, N.P. Fofonoff: Downstream development of the Gulf Stream from 68° to 55° W, J. Phys. Oceanogr. **23**, 225–249 (1993)

3.31 G.Z. Forristall, R.R. Leben, C.A. Hall: A statistical hindcast and forecast model for the Loop Current, Offshore Technology Conference (2010)

3.32 P. Van Gastel, G.N. Ivey, M. Meuleners, J. Antenucci, O. Fringer: Seasonal variability of the large-amplitude internal wave field on the Australian North West Shelf, J. Cont. Shelf Res. **29**, 1373–1383 (2009)

3.33 J.M. Klymak, S. Legg, M.H. Alford, M. Buijsman, R. Pinkel, J.D. Nash: The direct breaking of internal waves at steep topography, Oceanography **25**(2), 150–159 (2012)

3.34 T.M. Mitchell, R.G. Lueck, M.J. Vogel, R.E. Raye, G.Z. Forristall: Turbulence measurements in a Gulf of Mexico warm-core ring, Proc. OMAE 26th Int. Conf. Offshore Mech. Arct. Eng. (2007)

3.35 M.R. Dhanak, K. Holappa: An autonomous ocean turbulence measurement platform, J. Atmos. Ocean. Technol. **16**, 1506–1518 (1999)

3.36 B.R. Thomas, E.C. Kent, V.R. Swail: Methods to homogenize wind speeds from ships and buoys, Int. J. Climatol. **25**, 979–995 (2005)

3.37 N. Hogben, F.E. Lumb: *Ocean Wave Statistics* (Ministry of Technology, London 1967)

3.38 N. Hogben, N. Dacunha, G. Oliver: *Global Wave Statistics* (Unwin Brothers, London 1986)

3.39 NOAA National Weather Service National Hurricane Center: http://www.nhc.noaa.gov/pastall.shtml#hurdat

3.40 B.R. Jarvinen, C.J. Neumann, M.A.S. Davis: *A Tropical Cyclone Data Tape for the North Atlantic Basin, 1886–1983: Contents, Limitations, and Uses* (National Hurricane Center, Miami 1988)

3.41 US Navy Joint Typhoon Warning Center (JTWC): http://www.usno.navy.mil/NOOC/nmfc-ph/RSS/jtwc/best_tracks/

3.42 G.J. Holland: An analytic model of the wind and pressure profiles in hurricanes, Mon. Weather Rev. **108**(8), 1212–1218 (1980)

3.43 E.F. Thompson, V.J. Cardone: Practical modeling of hurricane surface wind fields, ASCE J. Waterw. Port Coast Ocean Eng. **122**(4), 195–205 (1996)

3.44 T.R. Karl, G.A. Meehl, C.D. Miller, S.J. Hassol, A.M. Waple, W.L. Murray (Eds.): *Weather and Climate Extremes in a Changing Climate. Regions of Focus: North America, Hawaii, Caribbean, and U.S. Pacific Islands. A Report by the U.S. Climate Change Science Program and the Subcommittee on Global Change Research* (Department of Commerce NOAA's National Climatic Data Center, Washington D.C. 2008)

3.45 A.B. Hagen, C.W. Landsea: On the classification of extreme Atlantic hurricanes utilizing mid-Twentieth-Century monitoring capabilities, J. Clim. **25**, 4461–4475 (2012)

3.46 K. Emanuel: Increasing destructiveness of tropical cyclones over the past 30 years, Nature **436**, 686–688 (2005)

3.47 P.Y. Le Traon: Satellites and operational oceanography. In: *Operational Oceanography in the 21st Century*, ed. by A. Schiller, G.B. Brassington (Springer, Berlin 2011)

3.48 NOAA/NESDIS/Star: Laboratory for Satellite Altimetry, http://www.star.nesdis.noaa.gov/sod/lsa/index.php

3.49 Jet Propulsion Lab: EarthData, http://podaac-tools.jpl.nasa.gov/

3.50 CNES AVISO: http://www.aviso.oceanobs.com

3.51 F. Lee-Lueng, A. Cazenave: *Satellite Altimetry and Earth Sciences* (Academic Press, San Diego 2001)

3.52 C.K. Shum, P. Woodworth, O. Andersen, G. Egbert, O. Francis, C. King, S. Klosko, C. Le Provost, X. Li, J. Molines, M. Parke, R. Ray, M. Schlax, D. Stammer, C. Tierney, P. Vincent, C. Wunsch: Accuracy assessment of recent ocean tide models, J. Geophys. Res. **102**(C11), 25173–25194 (1997)

3.53 Delft University of Technology: http://rads.tudelft.nl/rads/data/authentication.cgi

3.54 University of Southampton National Ocean Center: http://www.noc.soton.ac.uk/ooc/SURFACE/SATELLITE_WAVES/waves.php

3.55 Infremer: GLOBWAV, http://globwave.ifremer.fr/

3.56 Fugro OCEANOR Integrated Environmental Monitoring Systems: Ocean Wave Climate, http://www.oceanor.com/Services/wwa_info/

3.57 BMT Argoss Services and Solutions: http://www.bmtargoss.com

3.58 P.D. Cotton, D.J.T. Carter: Cross calibration of TOPEX, ERS-1, and Geosat wave heights, J. Geophys. Res. **99**(C12), 25025–25033 (1994)

3.59 J.F.R. Gower: Inner calibration of wave and wind data from TOPEX/POSEIDON and Mort buoys off the west coast of Canada, J. Geophys. Res. **101**, 3817–3829 (1996)

3.60 Jet Propulsion Lab: PODAAC http://podaac.jpl.nasa.gov/datasetlist

3.61 NOAA National Weather Service Ocean Prediction Center: http://www.opc.ncep.noaa.gov/newNCOM/NCOM_currents.shtml

3.62 HYCOM Consortium for Data Assimilative Modeling: https://hycom.org/ocean-prediction

3.63 Mercator Ocean, Ocean Forecasters: http://bulletin.mercator-ocean.fr

3.64 NOAA National Weather Service Environmental Modeling Center: http://polar.ncep.noaa.gov/waves/index2.shtml

3.65 European Center for Medium Weather Forecasts (ECMWF): Datasets, http://www.ecmwf.int/search/datasets/

3.66 N. Hogg, D. Frye: Performance of a new generation of acoustic current meters, J. Phys. Oceanogr. **37**(2), 148–161 (2007)

3.67 B.A. Magnell, L.I. Ivanov: Performance of the 75 kHz Long Ranger ADCP in a low scattering environment, Curr. Meas. Tech. Conf. (2008)

3.68 Coastal Observing Research and Development Center, Scripps Institution of Oceanography: http://cordc.ucsd.edu/projects/mapping/maps/

3.69 NOAA National Data Buoy Center: http://www.ndbc.noaa.gov/

3.70 J.A. Paduan, H.C. Graber: Introduction to high-frequency radar: Reality and myth, Oceanography **10**(2), 36–39 (1997)

3.71 E.A. Terray, B.H. Brumley, B. Strong: Measuring waves and currents with an upward looking ADCP, Proc. IEEE 6th Work. Conf. Curr. Meas. (1999)

3.72 P.K. Pandian, O. Emmanuel, J.P. Ruscoe, J.C. Side, R.E. Harris, S.A. Kerr, C.R. Bullen: An overview of recent technologies on wave and current measurement in coastal and marine applications, J. Oceanogr. Marine Sci. **11**(1), 1–10 (2010)

3.73 F.H. Freeman, R.M. Huffaker, R.M. Hardesty, M.E. Jackson, T.R. Lawrence, M.J. Post, R.A. Richter, B.F. Weber: Wind measurement accuracy of the NOAA pulsed infrared Doppler lidar, Appl. Opt. **23**(15), 2503–2506 (1984)

3.74 M. de Noord, A. Curvers, P. Eecen, I. Antoniou, H.E. Jørgensen, T.F. Pedersen, S. Bradley, S. von Hünerbein, D. Kindler, H. Mellinghoff, S. Emeis: *WISE: Wind Energy Sodar Evaluation, Final Report, ECN-C-05-044* (ECN, Petten 2005)

3.75 C. Cooper, G.Z. Forristall, R.C. Hamilton, C.C. Ebbesmeyer: Utilization of offshore oil platforms for meteorological and oceanographic measurements, Mar. Tech. Soc. J. **27**, 10–23 (1993)

3.76 ASCE Standards: *Minimum Design Loads for Building and Other Structures* (ASCE, Reston 2005)

3.77 European Committee for Standardization: Eurocode 1: Action on structures. General Actions Part 1–4: Wind Action, 1991-1-4 (ESC, Brussels 2004)

3.78 T.M. Joyce: On in 62 *calibration* of shipboard ADCPs, J. Atmos. Ocean. Technol. **5**, 169–172 (1989)

3.79 G.J. McNally, W.C. Patzert, A.D. Kirwan, Jr., A.C. Vastano: The near-surface circulation of the north Pacific using satellite tracked drifting buoys, J. Geophys. Res. **80**(8), 7507–7518 (1983)

3.80 P.D. Coholan, J.W. Feeney, S.P. Anderson: Life and times of Eddy Zoro: A review of the 2007 Gulf of Mexico Loop Current activity, Proc. Offshore Technol. Conf. (2008)

3.81 D. Rudnick, R.E. Davis, C.C. Eriken, D.M. Frantantoni, M.J. Perry: Underwater gliders for ocean research, Mar. Tech. Soc. J. **38**(1), 48–59 (2004)

3.82 NOAA Global Ocean Observations: The Global Drifter Program, http://www.aoml.noaa.gov/phod/dac/index.php

3.83 NOAA Earth System Research Laboratory Physical Sciences Division: Reanalysis Datasets at PSD, http://www.esrl.noaa.gov/psd/data/gridded/reanalysis/

3.84 E. Kalnay, M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K.C. Mo, C. Ropelewski, J. Wang, A. Leetmaa, R. Reynolds, R. Jenne, D. Joseph: The NCEP/NCAR 40 Year Reanalysis Project, Bull. Am. Meteorol. Soc. **77**(3), 437–471 (1996)

3.85 M. Kanamitsu, W. Ebisuzaki, J. Woollen, S. Yang, J.J. Hinlo, M. Fiorino, G.L. Potter: NCEP-DOE AMIP-II reanalysis (R-2), Bull. Am. Meteorol. Soc. **83**, 631–643 (2002)

3.86 S. Saha, S. Moorthi, H.-L. Pan, X. Wu, J. Wang, S. Nadiga, P. Tripp, R. Kistler, J. Woollen, D. Behringer, H. Liu, D. Stokes, R. Grumbine, G. Gayno, J. Wang, Y.-T. Hou, H.-Y. Chuang, H.-M.H. Juang, J. Sela, M. Iredell, R. Treadon, D. Kleist, P. Van Delst, D. Keyser, J. Derber, M. Ek, J. Meng, H. Wei, R. Yang, S. Lord, H. van den Dool, A. Kumar, W. Wang, C. Long, M. Chelliah, Y. Xue, B. Huang, J.-K. Schemm, W. Ebisuzaki, R. Lin, P. Xie, M. Chen, S. Zhou, W. Higgins, C.-Z. Zou, Q. Liu, Y. Chen, Y. Han, L. Cucurull, R.W. Reynolds, G. Rutledge, M. Goldberg: The NCEP climate forecast system reanalysis, Bull. Amer. Meteor. Soc. **91**(8), 1015–1057 (2010)

3.87 V.J. Cardone, C.V. Greenwood, J.A. Greenwood: *A Unified Program for the Specification of Hurricane Boundary Layer Winds Over Surfaces of Specified Roughness* (US Army Corps of Engineers, Vicksburg 1979)

3.88 NOAA Hurricane Research Division Atlantic Oceanographic & Meteorological Lab: http://www.aoml.noaa.gov/hrd/data_sub/wind.html

3.89 G.J. Komen, L. Cavaleri, M. Donelan, K. Hasselmann, S. Hasselmann, P.A.E.M. Janssen: *Dynamics and Modeling of Ocean Waves* (Cambridge Univ. Press, Cambridge 1994)

3.90 V.J. Cardone, R.E. Jensen, D.T. Resio, V.R. Swail, A.T. Cox: Evaluation of contemporary ocean wave models in rare extreme events: Halloween storm of October, 1991; Storm of the century of March, 1993, J. Atmos. Oceanic Tech. **13**, 198–230 (1996)

3.91 R.E. Jensen, V.J. Cardone, A.T. Cox: Performance of third generation wave models in extreme hurricanes, Proc. 9th Intern. Workshop Wave Hindcast. Forecast. (2006)

3.92 V.R. Swail, A.T. Cox: On the use of NCEP/NCAR Reanalysis Surface Marine Wind Fields for a Long term North Atlantic Wave Hindcast, J. Atmo. Tech. **17**(4), 532–545 (2000)

3.93 V.J. Cardone, A.T. Cox, V.R. Swail: Evaluation of NCEP reanalysis surface marine wind fields for ocean wave hindcasts, Proc. WMO Workshop Adv. Marine Climatol. (CLIMAR99) (1999) pp. 210–222

3.94 M.T. Johnsen, D.R. Lynch: Quantitative Skill Assessment for Coastal Ocean Models. In: *Coastal and Estuarine Studies*, Vol. 47, ed. by D.R. Lynch, A.M. Davies (Am. Geophys. Union, Washington 1995)

3.95 Old Dominion University Center for Coastal Physical Oceanography: The Princeton Ocean Model, http://www.ccpo.odu.edu/POMWEB/

3.96 Danish Hydraulic Institute, MIKE: http://www.dhisoftware.com/

3.97 Univeristy of North Carolina: ADCIRC, http://adcirc.org/

3.98 L.Y. Oey, T. Ezer, G. Forristall, C. Cooper, S. DiMarco, S. Fan: An exercise in forecasting loop current and any frontal positions in the Gulf of Mexico, Geophys. Res. Lett. **32**, L1261 (2005)

3.99 L.H. Kantha, J.-K. Choi, K.J. Schaudt, C.K. Cooper: A regional data-assimilative model for operational use in the Gulf of Mexico. In: *Circulation in the Gulf of Mexico*, ed. by T. Sturges, A. Lugo-Fernandez (AGU, Washington 2005)

3.100 HYCOM Consortium for Data Assimilative Modeling: http://hycom.org/

3.101 Rutgers University Institute of Marine and Coastal Sciences: Regional Ocean Modeling System (ROMS), http://www.myroms.org/

3.102 NOAA National Weather Service Environmental Modeling Center Marine Modeling & Analysis Branch: Global Real-Time Ocean Forecast System, http://polar.ncep.noaa.gov/global?

3.103 NOAA National Weather Service Ocean Prediction Center: Ocean Model Current Areas, http://www.opc.ncep.noaa.gov/newNCOM/NCOM_currents.shtml

3.104 Mercator Ocean, Ocean Forecasters: Ocean monitoring and forecasting, http://www.mercator-ocean.fr/eng/produits-services/produits-numeriques

3.105 Australian Government Bureau of Meteorology: Marine & Ocean, Sea Temperature and Currents: http://www.bom.gov.au/oceanography/forecasts/

3.106 K.C. Ewans: A response-based method for developing joint metocean criteria for on-bottom pipeline stability, J. Offshore Mech. Arct. Eng. **125**(2), 119–125 (2003)

3.107 J.C. Heideman, O. Hagen, C. Cooper, F.-E. Dahl: Joint probability of extreme waves and currents on the Norwegian shelf, J. Waterw. Port Coast. Ocean Eng. **115**(4), 534–546 (1989)

3.108 J.P. Fox: *A Monte Carlo Approach to Joint Probability of Wave, Tide, and Surge in Extreme Water Level Calculations, Res. Rep. R 740* (HSE, Norwich 2009)

3.109 C.K. Cooper, J.D. Stear: Estimating Design currents during joint eddy-TRWs and joint eddy-hurricanes, Proc. Offshore Tech. Conf. (2009)

3.110 S. Haver, S.R. Winterstein: Environmental contour lines: A method for estimating long term extremes by a short term analysis, Trans. Soc. Nav. Archit. Marine Eng. **116**, 116–127 (2009)

3.111 D.W. Scott: *Multivariate Density Estimation* (Wiley, New York 1992)

3.112 S.R. Winterstein, T.C. Ude, C.A. Cornell, P. Bjerager, S. Haver: Environmental parameters for extreme response: Inverse FORM with omission factors, Proc. 6th Int. Conf. Struct. Saf. Reliab. (1993)

3.113 B.J. Beamsley, P. McComb, D. Johnson, J. Silk: Estimating whether downtime for ocean engineering using sequential downtime analysis, Proc. Coasts Ports (2007)

3.114 M. Prevosto, G.Z. Forristall, G. Jeans, C. Herry, G. Harte, L. Harrington-Missin, P. Dooley: Worldwide approximation of current profiles for steel riser design – The WACUP project, Proc. OMAE 31st Int. Conf. Offshore Mech. Arct. Eng. (2012)

3.115 G.Z. Forristall, C.K. Cooper: Design current profiles using empirical orthogonal functions and inverse FORM methods, Proc. Offshore Technol. Conf. (1997)

3.116 G. Jeans, C. Grant, G. Feld: Improved current profile criteria for deepwater riser design, Proc. OMAE 21st Int. Conf. Offshore Mech. Arct. Eng. (2002)

3.117 T. Kohonen: *Self-Organizing Maps*, Springer Series in Information Sciences, Vol. 30 (Springer, New York 2001)

3.118 M. Efthymiou, J.W. van de Graaf, P.S. Tromans, I.M. Hines: Reliability based design for fixed steel offshore platforms, Proc. Offshore Mech. Arctic Eng. (1996)

3.119 A.C. Davidson: *Statistical Models* (Cambridge University Press, Cambridge 2003)

3.120 Y. Goda: *Random Seas and Design of Marine Structures* (World Scientific, Singapore 2000)

3.121 R. Gibson, C. Grant, G.Z. Forristall, R. Smyth, P. Owrid, O. Hagen, I. Leggett: Bias and uncertainty in the estimation of extreme wave heights and crests, Proc. Offshore Mech. Arct. Eng. (2009)

3.122 G.R. Toro, C.A. Cornell, V.J. Cardone, D.B. Driver: Comparison of Historical and Deductive Methods for the Calculation of Low-Probability Seastates in the Gulf of Mexico, Proc. OMAE 23rd Int. Conf. Offshore Mech. Arct. Eng. (2004)

3.123 C. Cooper, J. Stear, J. Heideman, M. Santala, G. Forristall, D. Driver, P. Fourchy: Implications of Hurricane Ivan on deepwater Gulf of Mexico metocean design criteria, Offshore Tech. Conf. (2005)

3.124 C. Cooper: A Preliminary Case for the Existence of Hurricane Alleys in the Gulf of Mexico, Offshore Technol. Conf. (1992)

3.125 C. Cooper, J. Stear: Estimating design currents during Joint Eddy-TRWs and Joint Eddy-Hurricanes, Offshore Tech. Conf. (2009)

3.126 D.W. Wang, A.K. Liu, C.Y. Peng, E.A. Meindl: Wave-current interaction near the Gulf Stream during the Surface Wave Dynamics Experiment, J. Geophys. Res. **99**(C3), 5065 (1994)

3.127 P.N. Georgiou, A.G. Davenport, P.J. Vickery: Design wind speeds in regions dominated by tropical cyclones, J. Wind Eng. Ind. Aerodyn. **13**, 139–152 (1983)

3.128 P.J. Vickery, P.F. Skerjl, L.A. Twisdale: Simulation of hurricane risk in the US using empirical track model, J. Struct. Eng. **126**, 1222–1237 (2000)

3.129 K. Emanuel, S. Ravela, E. Vivant, C. Risi: A statistical deterministic approach to hurricane risk assessment, Bull. Amer. Meteorol. Soc. **87**, 299–314 (2006)

3.130 P.J. Vickery, D. Wadhera, J.D. Stear: A synthetic model for Gulf of Mexico hurricanes, Proc. Offshore Technol. Conf. (2010)

3.131 V.J. Cardone, A.T. Cox: Modeling very extreme sea states (VESS) in real and synthetic design level storms, ASME 30th Int. Conf. Ocean Offshore Arct. Eng. (2011) pp. 531–535

3.132 K. Emanuel: The dependence of hurricane intensity on climate, Nature **326**(6112), 483–485 (1987)

3.133 G.Z. Forristall: Use of directional wave criteria, J. Waterw. Port Coast. Ocean Eng. **130**, 272–275 (2004)

3.134 L.E. Borgman: Probabilities for highest wave in hurricane, J. Waterw. Harbors Coast. Eng. Div. **99**(2), 185–207 (1973)

3.135 M.J. Tucker, E.G. Pitt: *Waves In Ocean Engineering* (Elsevier, Amsterdam 2001)

3.136 G.Z. Forristall: How should we combine long and short term wave height distributions?, Proc. OMAE 27th Int. Conf. Offshore Mech. Arct. Eng. (2008)

3.137 P.S. Tromans, L. Vanderschuren: Response based design conditions in the North Sea: Application of a new method, Proc. Offshore Technol. Conf. (1995)

3.138 S. Haver: Design of offshore structures: Impact of the possible existence of freak waves, Rogue Waves, Proc. Aha Huliko'a Hawaii. Winter Workshop (2005)

3.139 A.K. Magnusson, M.A. Donelan: The Andrea wave: Characteristics of a measured North Sea rogue wave, Proc. OMAE Int. Conf. Offshore Mech. Arct. Eng. (2012)

3.140 G.Z. Forristall: Understanding rogue waves: Are new physics really necessary?, 14th Aha Huliko'a Hawaii. Winter Workshop (2005)

3.141 M. Christou, K. Ewans: Examining a comprehensive dataset with thousands of freak wave events Part 2 – analysis and findings, Proc. 30th Conf. Offshore Mech. Arct. Eng. (2011)

3.142 A.R. Osborne: *Nonlinear Ocean Waves and the Inverse Scattering Transform* (Academic Press, Oxford 2011)

3.143 A. Toffoli, O. Gramstad, K. Trulsen, J. Monbaliu, E. Bitner-Gregersen, M. Onorato: Evolution of weakly nonlinear random directional waves: Laboratory experiments and numerical simulations, J. Fluid Mech. **664**, 313–336 (2010)

3.144 A.R. Rao, K.H. Hamed, H. Chen: *Nonstationarities in Hydrologic and Environmental Time Series* (Kluwer, Boston 2003)

3.145 C. Landsea: El Nino-Southern oscillation and the seasonal predictability of tropical cyclones. In: *El Nino: Impacts of Multiscale Variability on Natural Ecosystems and Society*, ed. by H.F. Diaz, V. Markgraf (Cambridge Univ. Press, Cambridge 2000)

3.146 C. Cooper, J. Stear: Hurricane climate in the Gulf of Mexico, Offshore Technol. Conf. (2006)

3.147 Intergovernmental Panel on Climate Change: Fifth Assessment Report 2014, http://www.ipcc.ch/publications_and_data/publications_and_data_reports.shtml

3.148 J. Done, G.J. Holland, C. Bruyere, A. Suzuki-Parker: Effects of climate variability and change on Gulf of Mexico tropical cyclone activity, Proc. Offshore Technol. Conf. (2011)

4. Mechanics of Ocean Waves

James M. Kaihatu, Palaniswamy Ananthakrishnan

This chapter reviews mechanics of water waves and wave–body interactions pertaining to ocean and coastal engineering based on linear and weakly nonlinear wave theories. Numerical methods based on Green's theorem and mixed Eulerian–Lagrangian formulation for fully nonlinear wave and wave–body interaction problems are also discussed. The discussion also covers methods to determine the wave forces on fixed and floating structures, including the viscous drag force.

4.1 **Ocean Surface Waves** 77
4.2 **Wave Theories** 78
4.2.1 Potential Flow Formulation 78
4.3 **Properties of Small Amplitude Gravity Waves** 80
4.3.1 Linear Dispersion Relation 80
4.3.2 Phase Speed 80
4.3.3 Group Speed 80
4.3.4 Amplitude Modulation of Water Waves 81
4.3.5 Average Wave Energy Density 81
4.3.6 Propagation of Wave Energy 81
4.3.7 Water Particle Trajectory 82
4.3.8 Spatio-Temporal Evolution of Waves 82
4.3.9 Shoaling and Refraction of Waves 83
4.3.10 Closing Remarks to the Section 83
4.4 **Weakly Nonlinear Deep Water Wave Theories** 83
4.4.1 Properties of Weakly Nonlinear Deep Water Waves 84
4.4.2 Evolution of Weakly Nonlinear Deep Water Waves 85
4.5 **Shallow Water Wave Theories** 87
4.5.1 Properties of Weakly Nonlinear Shallow Water Waves 88
4.5.2 Evolution of Weakly Nonlinear Shallow Water Waves 89
4.6 **Transformation of Waves Approaching Land** 90
4.7 **Computational Method for Fully Nonlinear Waves** 93
4.8 **Wave Forces on Fixed and Floating Structures** 94
4.8.1 Incident Wave Force: Froude–Krylov Force 94
4.8.2 Morison Force on a Stationary Body 95
4.8.3 Wave Diffraction over a Body 96
4.8.4 Wave Radiation Force on an Oscillating Body 96
4.9 **Concluding Remarks** 97
References 98

4.1 Ocean Surface Waves

The interface between the atmosphere and water, when disturbed, results in the generation of surface waves. In an open domain, the disturbance has to continue for the waves to persist as the propagating waves radiate energy. There is a limit to the energy contained by waves; once exceeded, wave breaking occurs, at which point energy is dissipated by turbulence during the breaking process. The limit depends on parameters such as wave height to wave length (or steepness) and wave height to water depth ratios. Wave instabilities or damping are also governed by air flow and separation, surface tension and surfactant effects, and instability of the free surface boundary layers. The physics of water waves is unique in many respects compared to other wave motions in fluids because of the dispersive nature of the water waves. Water wave problems are among the earliest topics attempted in applied mathematics, as illustrated, for example, by the classical Cauchy–Poisson problem on transient waves, tackled as early as in 1815 [4.1] and *Michell*'s theory of ship wave resistance developed in 1898 [4.2, 3]. A historical account of the development of classical water wave theory is given in [4.1]. Today, from an engineering viewpoint, understanding the properties of waves and wave–body

interactions is essential for efficient design and performance prediction of surface ships, offshore platforms, coastal structures, beach erosion mitigation measures, and beach fill configurations.

This chapter of the Handbook deals with the mechanics of surface waves, in particular:

1. Kinematic properties of surface waves
2. Weakly nonlinear deep water waves
3. Transformation of waves as they approach land from the deep ocean
4. Shallow water waves
5. Evolution of deep and shallow water waves
6. Breaking waves
7. Nearshore long wave generation, and
8. Wave forces on floating and fixed ocean structure.

Linear, weakly nonlinear, and fully nonlinear theories, and results on above topics are reviewed. The formulation of nonlinear wave theories in deep and shallow water are outlined, including the spectral evolution of these waves. Methods to solve wave–body interaction problems and to determine wave forces are discussed. The subject is presented in a manner so that it can serve as a reference for practicing engineers and researchers in ocean engineering. In each section of the chapter, an overview of the mathematical theory and formulations is given along with references for details.

4.2 Wave Theories

With exact equations governing water waves, wave transformations and wave–body interactions being nonlinear and involving arbitrary boundaries which make analysis difficult, numerous approximate theories have been developed over the years. These include linear Airy wave theory, Stokes weakly nonlinear theory, Boussinesq weakly nonlinear long wave theory and Korteweg–de Vries (KdV) theory for shallow water waves. Recent developments include computational methods to solve the fully nonlinear wave problem with the notable one by *Longuet-Higgins* and *Cokelet* based on mixed Eulerian–Lagrangian formulation [4.4].

Table 4.1 Key wave and flow variables

A	wave amplitude
C_g	wave group speed
C_p	wave phase speed
g	acceleration of gravity
h	water depth
H	wave height
k	wave number
L	wave length
p	pressure field
s	surface tension
t	time variable
T	wave period
$\boldsymbol{u}$	velocity field
(x, y, z)	Inertial earth-fixed coordinates against g with $z = 0$ on the calm surface
γ	specific weight of water
η	wave elevation
$\boldsymbol{\phi}$	velocity potential
ρ	water density
ω	wave radian frequency
$\boldsymbol{\Omega}$	vorticity

The above theories are based on the potential flow formulation, which assumes water to be inviscid and flow irrotational. The wave boundary conditions are, however, still nonlinear; they may be linearized for the case of small amplitude waves as in Airy's water wave theory. Despite the idealization and assumptions involved, linear theory captures many properties of wave phenomena and measures wave effects quite reasonably in most cases. Of course, for practical engineering applications, knowledge of large amplitude waves, including transformation over rapidly changing bathymetry, is essential; here linear theory has limited application and one has to consider weakly and fully nonlinear wave models. This chapter reviews these theories and the corresponding wave properties.

4.2.1 Potential Flow Formulation

We begin with a brief overview of the mathematical formulation of the water wave problem. The notations used in this chapter for key wave and flow variables are shown in Table 4.1. Additional notations used locally in the text are explained in the context.

A typical wave–body interaction and wave transformation problem encountered in ocean engineering and considered in this chapter is illustrated in Fig. 4.1. The body S_B may either be stationary or moving with velocity $\boldsymbol{U}$ (translation) and $\boldsymbol{\Omega}$ (rotational), resulting in a normal velocity $V_n = \boldsymbol{U} \cdot \hat{n} + \boldsymbol{\Omega} \cdot (\boldsymbol{r} \times \hat{n})$ with $\boldsymbol{r}$ denoting the position vector from the axis of rotation.

Neglecting effect of viscosity and consequently assuming the flow to be irrotational (i. e., $\boldsymbol{\omega} \equiv (\nabla \times \boldsymbol{u} = 0)$), one can define the flow in terms of velocity potential so that

$$\boldsymbol{u} = \nabla \boldsymbol{\phi} \, . \tag{4.1}$$

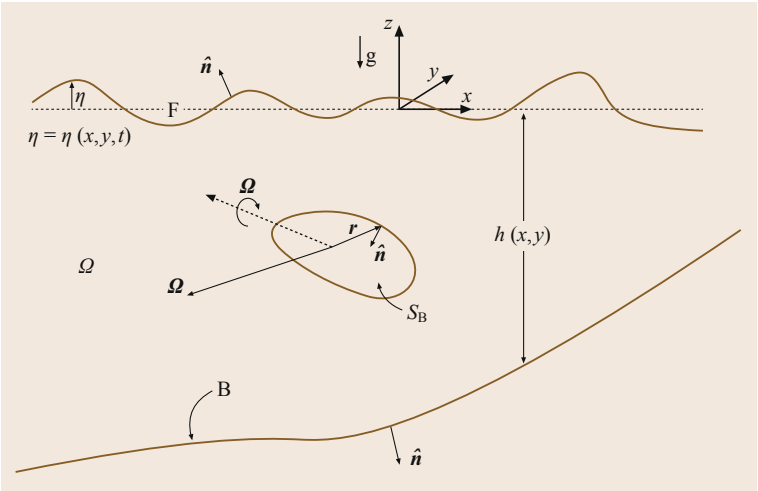


Fig. 4.1 Illustration of wave–body interaction and wave transformation over changing bathymetry

For incompressible fluid, per equation of continuity,

$$\nabla \cdot \boldsymbol{u} = 0\,, \tag{4.2}$$

which in terms of $\boldsymbol{\phi}$ is the Laplace equation

$$\nabla^2 \boldsymbol{\phi} = 0\,. \tag{4.3}$$

With $\boldsymbol{u} = \nabla\boldsymbol{\phi}$ in the Euler equation motion and through spatial integration one obtains the following Euler's integral for the pressure

$$p = -\rho g z - \rho \frac{\partial \boldsymbol{\phi}}{\partial t} - \frac{\rho}{2}|\nabla \boldsymbol{\phi}|^2, \tag{4.4}$$

with the first term on the right-hand side representing the static pressure and the last two terms denoting the dynamic pressure.

The bottom and body boundary conditions are the no-flux conditions given by

$$\frac{\partial \boldsymbol{\phi}}{\partial n} = 0, \quad \text{on the bottom B}\,, \tag{4.5}$$

and

$$\frac{\partial \boldsymbol{\phi}}{\partial n} = V_n, \quad \text{on the body surface } S_B\,. \tag{4.6}$$

On the free surface, we have the following two conditions, one based on the kinematics of free surface motion and the other on the continuity of pressure across the free surface

$$\frac{\partial \eta}{\partial t} + \frac{\partial \eta}{\partial x}\frac{\partial \boldsymbol{\phi}}{\partial x} + \frac{\partial \eta}{\partial y}\frac{\partial \boldsymbol{\phi}}{\partial y} = \frac{\partial \boldsymbol{\phi}}{\partial z} \quad \text{on } z = \eta\,, \tag{4.7}$$

$$\frac{\partial \boldsymbol{\phi}}{\partial t} + \frac{1}{2}|\nabla \boldsymbol{\phi}|^2 + g\eta = 0, \quad \text{on } z = \eta\,. \tag{4.8}$$

The former is called the free surface kinematic condition and the latter the free surface dynamic condition. Both are specified on $z = \eta$, which in itself is also an unknown and to be solved as a part of the problem. These conditions make the problem nonlinear and, therefore, difficult to solve.

For small amplitude waves (i. e., one in which the wave amplitude is much smaller than the wave length), the free surface boundary conditions can be linearized as

$$\frac{\partial \eta}{\partial t} + \frac{\partial \boldsymbol{\phi}}{\partial z} \quad \text{on } z = 0\,, \tag{4.9}$$

$$\frac{\partial \boldsymbol{\phi}}{\partial t} + g\eta = 0 \quad \text{on } z = 0\,. \tag{4.10}$$

Eliminating η from the above two conditions, one obtains the following linearized combined free surface condition

$$\frac{\partial^2 \boldsymbol{\phi}}{\partial t^2} + g\frac{\partial \boldsymbol{\phi}}{\partial z} = 0 \quad \text{on } z = 0\,. \tag{4.11}$$

In the case of spatially periodic waves with wave length L in, say, x direction, $\boldsymbol{\phi}(x,y,z,t) = \boldsymbol{\phi}(x \pm nL, y, z, t)$ and $\eta(x,y,t) = \eta(x \pm nL, y, t)$ and for time harmonic case with period T, $\boldsymbol{\phi}(x,y,z,t) = \boldsymbol{\phi}(x,y,z,t \pm nT)$ and $\eta(x,y,t) = \eta(x,y,t \pm nT)$ where n denotes positive integer. The wave length L and wave period T are related by the dispersion relation as discussed later in the chapter.

4.3 Properties of Small Amplitude Gravity Waves

In the absence of a body in the fluid, a periodic two-dimensional (long crested) progressive small amplitude (linear) wave solution to the linearized problem (as formulated above) can be written as

$$\phi = \frac{H}{2}\frac{\mathrm{g}}{\omega}\frac{\cosh k(z+h)}{\cosh kh}\sin(kx-\omega t)\,,$$
$$\eta = \frac{H}{2}\cos(kx-\omega t)\,, \tag{4.12}$$

where $k \equiv 2\pi/L$ denotes the wave number, H the wave height (which the vertical distance between the crest and trough), and h the mean water depth (taken here to be constant). In the case of a deep water wave, i. e., in the limit $kh \to \infty$ with $z \in [0, -h]$ the above solution becomes

$$\phi = \frac{H}{2}\frac{\mathrm{g}}{\omega}\mathrm{e}^{kz}\sin(kx-\omega t)\,,$$
$$\eta = \frac{H}{2}\cos(kx-\omega t)\,. \tag{4.13}$$

For the above velocity potential, the velocity field ($\boldsymbol{u} = \nabla\phi$) is given by

$$u = \frac{H}{2}\mathrm{g}k\omega\frac{\cosh k(z+h)}{\cosh kh}\cos(kx-\omega t)\,,$$
$$v = 0\,,$$
$$w = \frac{H}{2}\mathrm{g}k\omega\frac{\sinh k(z+h)}{\cosh kh}\sin(kx-\omega t)\,.$$

From above expressions one finds that the horizontal velocity field is in phase with the wave elevation. Its slip on the bottom is to be interpreted as the velocity at the outer edge of a bottom boundary layer. Using the Euler integral one can obtain the following expression for the pressure

$$p = -\rho\mathrm{g}z + \rho\mathrm{g}\frac{H}{2}\frac{\cosh k(z+h)}{\cosh kh}\cos(kx-\omega t)$$
$$= -\rho\mathrm{g}z + \rho\mathrm{g}\eta\frac{\cosh k(z+h)}{\cosh kh}\,.$$

In the case of shallow water waves – i. e., for $kh \to 0$ with $z \in [0, -h]$ (or approximately, $kh \le \pi/10$) –, the expression for pressure reduces to

$$p = -\rho\mathrm{g}(z-\eta).$$

The above expression looks very similar to the familiar hydrostatic pressure equation for homogeneous water. The static pressure corresponds to the depth measured from the calm surface while total pressure to the depth measured from the actual free surface with the difference contributes to the dynamic pressure.

4.3.1 Linear Dispersion Relation

As per the linear solution obtained above, the wave number $k \equiv 2\pi/L$ and wave frequency $\omega \equiv 2\pi/T$ are related by the dispersion relation

$$\omega^2 = \mathrm{g}k\,\tanh kh\,, \tag{4.14}$$

which approximates to

$$\omega^2 = \mathrm{g}k,\ \text{for}\ kh \ge \pi\,, \tag{4.15}$$

and to

$$\omega^2 = \mathrm{g}k^2h,\ \text{for}\ kh \le \pi/10\,. \tag{4.16}$$

The dispersion relation thus provides a basis for defining deep water, shallow water, and intermediate waves. Waves satisfying $kh \ge \pi$, which is the same as $h \ge L/2$, are referred to as deep water (or short) waves; that satisfying $kh \le \pi/10$, which is same as $L \ge 20h$, are referred to as shallow water (or long) waves and that in between (i. e., $\pi/10 < kh < \pi$) as the intermediate waves.

4.3.2 Phase Speed

The phase speed (also known as the wave speed or wave celerity) for linear waves becomes

$$C_\mathrm{p} \equiv \frac{L}{T} = \frac{\omega}{k} = \sqrt{\frac{\mathrm{g}}{k}\tanh kh}\,, \tag{4.17}$$

which reduces to

$$C_\mathrm{p} = \sqrt{\frac{\mathrm{g}}{k}\tanh kh}, \text{for}\ \pi/10 < kh < \pi$$
(intermediate waves)
$$= \sqrt{\frac{\mathrm{g}}{k}}, \text{for}\ kh \ge \pi$$
(deep water or short waves)
$$= \sqrt{\mathrm{g}h}, \text{for}\ kh \le \pi/10$$
(shallow water or long waves)\,. \tag{4.18}

As can be observed in the above expressions, longer deep water waves travel faster. In the case of shallow water waves the wave speed depends only on the local depth; the phase speed decreases with decreasing depth.

4.3.3 Group Speed

Often it is the group speed that govern many important wave properties, including the speed of energy propa-

gation, among others. The group speed is defined as

$$C_g \equiv \frac{d\omega}{dk}\,, \tag{4.19}$$

which in view of the dispersion relation for linear gravity waves become

$$\begin{aligned} C_g &= \frac{\omega}{2k}\left(1+\frac{2kh}{\sinh 2kh}\right), \text{ for } \pi/10 < kh < \pi \\ &= \frac{\omega}{2k} = \frac{C_p}{2} = \frac{1}{2}\sqrt{\frac{g}{k}}, \text{ for } kh \geq \pi \\ &\quad \text{(deep water waves)} \\ &= \frac{\omega}{k} = C_p = \sqrt{gh}, \text{ for } kh \leq \pi/10 \\ &\quad \text{(shallow water waves)}\,. \end{aligned} \tag{4.20}$$

So, for linear gravity waves, except in the limit of shallow water (long) waves, the group speed is different from the phase speed, and this fact has a bearing on some unique properties of shallow water waves to be discussed in later sections.

4.3.4 Amplitude Modulation of Water Waves

A wave consisting of two periodic progressive waves of equal amplitude and nearly equal wave number, and because of the continuous dispersion relation nearly equal frequency, will undergo an amplitude modulation with the modulation itself propagating as a wave but at the group speed rather than the phase speed. In other words,

$$\begin{aligned} \eta &= \eta_1 + \eta_2 \\ &= \frac{H}{2}\cos(k_1 x - \omega_1 t) + \frac{H}{2}\cos(k_2 x - \omega_2 t)\,, \\ &\quad \text{where } k_1 = k + \frac{\delta k}{2},\ k_2 = k - \frac{\delta k}{2}, \\ &= H\ \cos(\delta k\, x - \delta\omega\, t)\ \cos(kx - \omega t), \\ &\quad \omega_1 = \omega + \frac{\delta\omega}{2},\ \omega_2 = \omega - \frac{\delta\omega}{2}, \\ &= A\ \cos(kx - \omega t), \\ &\quad \text{where the amplitude } A \equiv H\ \cos(\delta k\, x - \delta\omega\, t)\,. \end{aligned} \tag{4.21}$$

The amplitude envelope propagates at speed $\delta\omega/\delta k$ and in the limit of $\delta\omega, \delta k \to 0$, at the group speed C_g! To an observer moving in the direction of wave propagation, the amplitude will appear stationary (constant); however, the observer will not be in phase with the wave, unless if it is a shallow water wave in which case $C_g = C_p$. Thus we observe that in the case of amplitude-modulated waves, the wave amplitude is conserved at the group speed.

4.3.5 Average Wave Energy Density

Through straightforward integration and averaging over a wave period, it can be shown that the average wave potential energy in a water column of unit cross area is given by

$$\overline{PE} = \frac{1}{16}\rho g H^2\,.$$

The word *density* in the present context corresponds to the unit area of the cross-section of the water column. Similarly, using the expressions for velocity and the definition of kinetic energy one can show that the average wave kinetic energy density is given by

$$\overline{KE} = \frac{1}{16}\rho g H^2\,,$$

and, therefore, the average wave mechanical energy density

$$\bar{E} \equiv \overline{PE} + \overline{KE} = \frac{1}{8}\rho g H^2\,.$$

To particularly note in the above expression is the fact that the wave energy is proportional to the square of the wave height.

4.3.6 Propagation of Wave Energy

Using the work-energy theorem, it can be shown that the average amount of energy propagating across a surface of unit crest length is given by

$$\bar{\mathcal{F}} = \bar{E} C_g\,,$$

meaning that the wave energy propagates at the group speed! For example, in the case of a wave front advancing in deep water the waves will appear to disappear at the wave front because $C_p = 2C_g$; in other words, propagation of energy cannot keep up with phase to sustain the wave. The above expression is quite useful to estimate the power required to generate waves and the wave resistance of vehicles moving over a free surface and wave energy conversion. In the case of confluence or divergence of waves, as due to bottom bathymetry or power take off, one has to multiply the above expression by the respective crest widths. The reader may refer to the topic of the *antenna effect* in the chapter on wave energy conversion of this Handbook for more on this aspect.

4.3.7 Water Particle Trajectory

For a propagating wave with velocity known (from above) one can determine the trajectory of water particles through time integration. For small amplitude waves, the trajectory is elliptical and closed

$$\frac{(X-X_o)^2}{A^2}+\frac{(Z-Z_o)^2}{B^2}=1\,,$$

where (X_o, Z_o) denote the coordinates of the center of the ellipse, A the major semi-axis

$$A\equiv\frac{H}{2}\frac{gk}{\omega^2}\frac{\cosh k(Z_o+h)}{\cosh kh}\,,$$

and B the minor semi-axis

$$B\equiv\frac{H}{2}\frac{gk}{\omega^2}\frac{\sinh k(Z_o+h)}{\cosh kh}\,.$$

As to be expected, near the bottom the particle will oscillate back and forth parallel to the bottom. In the case of deep water waves ($kh\to\infty$, or for practical calculations $kh\geq\pi$), above expressions for semi-axes reduce to

$$A=B=\frac{H}{2}\frac{gk}{\omega^2}e^{kZ_o}=\frac{H}{2}e^{kZ_o}\,.$$

In other words, the trajectory is a circle for small amplitude deep water waves. As the particle paths are closed, the average mass transport is zero. However, in the case of finite amplitude waves, this is not the case. In the case of finite amplitude waves, the trajectory is not quite closed, meaning that a particle will not return to the original position after one wave period but to a point slightly ahead of the original point. This is referred to as the Stokes drift. More on finite amplitude wave properties are given in later sections of this chapter.

4.3.8 Spatio-Temporal Evolution of Waves

Consider the generation of waves by (for example) a mudslide into water or by a seismic activity and the resulting waves propagating over water of varying depth. Here both wave numbers and frequency will vary with space and time. Let the direction of the wave propagation be along the x direction. Here the wave number and frequency may vary both in x and in time. In other words, the phase function will be

$$\Theta=(k(x,t)x-\omega(x,t)t)\,.$$

With the so-called slowly varying wave assumption, the wave number and wave frequency can be written as

$$k\approx\frac{\partial\Theta}{\partial x};\ \omega\approx-\frac{\partial\Theta}{\partial t}\,.$$

Cross-differentiating the above expressions we obtain

$$\frac{\partial k}{\partial t}+\frac{\partial\omega}{\partial x}=\frac{\partial^2\Theta}{\partial x\partial t}-\frac{\partial^2\Theta}{\partial t\partial x}=0\,. \tag{4.22}$$

In the special case of the steady-state limit, i. e., $k=k(x)$ and $\omega=\omega(x)$, from above we find ω to be constant. Thus, in the case of steady-state evolution of waves, the wave number (hence wave length) may change spatially in x, but the frequency will be the same everywhere.

The above equation for k and ω can be re-written as

$$\frac{\partial k}{\partial t}+C_g\frac{\partial k}{\partial x}=0,\ \text{where } C_g=\text{Group Speed}=\frac{d\omega}{dk}\,. \tag{4.23}$$

The above equation is called the wave number conservation equation; its solution implies that k will appear constant to an observer moving at the group speed. One could draw several other important conclusions using this equation. For example, assume waves are propagating in deep water (i. e., so that $kh\geq\pi$). At a fixed field point away from the wave source, as per the wave number conservation equation, the wave frequency and wave number will increase with respect to time, meaning that as time progresses shorter and shorter waves will reach the field point. To deduce this fact from the wave number conservation equation, consider the time that will be taken by a wave of wave number k generated at the source to reach the field point

$$t=\frac{d}{C_g}\ \to\ C_g=\frac{d}{t}\,,$$

where d denotes the distance between the source and the field point. Since $C_g=0.5\,g/\omega$ for deep water gravity waves, the above becomes

$$\frac{1}{2}\frac{g}{\omega}=\frac{d}{t}\to\omega=\frac{1}{2}\frac{gt}{d}\,.$$

In other words, the frequency of waves reaching the field point increases as the time increases. With $\omega^2=gk$ for deep water waves, the above equation can be written as

$$k=\frac{1}{4}\frac{gt^2}{d^2}\,,$$

meaning that the number of waves reaching the field point will also increase with time.

4.3.9 Shoaling and Refraction of Waves

One can generalize the derivations and findings of the previous section to allowing the waves to propagate along any arbitrary direction. In the case of slowly varying, small amplitude steady-state waves approaching an ideal coast (straight and long shore line with all bottom bathymetry contours parallel to the shore line) the direction of wave propagation is governed by Snell's law

$$\frac{\sin\theta}{C_\mathrm{p}} = \text{constant}\,, \tag{4.24}$$

where θ denotes the local angle of wave incidence measured from the shore normal, with $\theta = 0$ corresponding to crests parallel to the shore and $\theta = 90°$ to crests normal to the shore. The change of direction of wave propagation is called wave refraction. As one may recall from classical physics, the path of waves as predicted by the above Snell's law corresponds to the shortest time of travel between two points in a variable velocity field. If the velocity is constant, then θ will also be constant, meaning that the path of wave propagation will be a straight line.

As the waves approach the shore with depth $h \to 0$, $C_\mathrm{p} \to \sqrt{gh} \to 0$, then θ should also $\to 0$ for the ratio $\frac{\sin\theta}{C_\mathrm{p}}$ to remain constant; in other words, the wave rays will be normal to the shoreline irrespective of the angle of incidence of the wave while offshore.

Note that in the case of slowly varying steady-state waves, the wave frequency will remain constant and the wave number evolve satisfying the dispersion relation

$$gk\tanh kh = \text{constant}\,.$$

Assuming no wave breaking (which would enhance dissipation through turbulence) and with conservation of mechanical energy, one can show that in the case of a slowly varying, steady-state, nonbreaking wave approaching an ideal shore, the wave height will vary, satisfying

$$H\sqrt{\cos\theta}\,\sqrt{C_\mathrm{g}} = \text{constant} \;\to\; H = \frac{\text{constant}}{\sqrt{C_\mathrm{g}\,\cos\theta}}\,. \tag{4.25}$$

4.3.10 Closing Remarks to the Section

The linear (Airy) wave theory enables us to understand the basic and unique properties of surface waves. There are several books that deal with the linear theory of water waves, which the reader may refer to for details. These include [4.5–9]. The linear potential flow theory will hold good only for small amplitude waves. For finite and large amplitude waves, one has to use weakly and fully nonlinear wave theories, as reviewed in the following sections.

4.4 Weakly Nonlinear Deep Water Wave Theories

While linear theory has proved to be quite useful, it is often insufficient to investigate many of the details of wave propagation. Bulk properties of waves (some measure of wave height) are generally well predicted by linear theory in many instances. However, the small amplitude assumption inherent in linear wave theory is often invalidated (when waves approach their breaking condition), so nonlinear wave theory becomes essential.

The nonlinearity in water wave theory arises from the treatment of the surface boundary conditions. The kinematic and dynamic free surface boundary conditions are both nonlinear (they involve the products of the dependent variables $\boldsymbol{\phi}$ and η) and apply at a surface whose position is unknown a priori, as mentioned above. To move beyond simple linearization of the free surface condition (which involves neglecting the nonlinear terms and applying the boundary conditions on the surface $z = 0$), successive approximations to the nonlinear boundary conditions are required. This requires the establishment of physical scales, which become the basis of these approximations.

We first discuss deep water nonlinear wave theory, or Stokes wave theory (Stokes [4.10]). In deep water, the relative depth kh (where k is the wave number and h the water depth) is large. This can be used to adjust the velocity potential toward its deep water asymptote

$$\boldsymbol{\phi} = \frac{H}{2}\frac{\mathrm{g}}{\omega}\frac{\cosh k(z+h)}{\cosh kh}\sin(kx-\omega t) \approx \frac{\mathrm{g}H}{2\omega}\times O(1)\,, \tag{4.26}$$

which reflects the observation that the hyperbolic cosine term approaches 1 as kh becomes large. The particle velocities u and w similarly become

$$u = \frac{\partial\boldsymbol{\phi}}{\partial x} = \frac{\mathrm{g}kH}{2\omega}\frac{\cosh k(z+h)}{\cosh kh}\cos(kx-\omega t) \approx \frac{\mathrm{g}ka}{\omega}\times O(1) = \frac{\mathrm{g}\epsilon}{\omega}\times O(1)\,, \tag{4.27}$$

$$w = \frac{\partial \phi}{\partial z} = \frac{gkH}{2\omega}\frac{\sinh k(z+h)}{\cosh kh}\sin(kx-\omega t) \approx \frac{gka}{\omega}\times O(1) = \frac{g\epsilon}{\omega}\times O(1)\,, \tag{4.28}$$

where $\epsilon = ka$ and where the amplitude $a = H/2$. It is evident that both u and w are of the same scale of variation, and as such there is no separation between the horizontal and vertical length scales. This allows both the horizontal and vertical scales to be the same. The time variable is scaled using the deep water approximation to the linear dispersion relation. These scales are then applied to the water wave boundary value problem. While the governing equation and bottom boundary conditions remain unchanged, the free surface boundary conditions are transformed to reflect this new scaling. The dynamic free surface boundary condition is now

$$g\eta + \frac{\partial \phi}{\partial t} + \frac{\epsilon}{2}\left[\left(\frac{\partial \phi}{\partial x}\right)^2 + \left(\frac{\partial \phi}{\partial z}\right)^2\right] = 0 \quad \text{on } z = \epsilon\eta\,, \tag{4.29}$$

and the kinematic free surface boundary condition is

$$\frac{\partial \eta}{\partial t} - \frac{\partial \phi}{\partial z} + \epsilon\frac{\partial \eta}{\partial x}\frac{\partial \phi}{\partial x} = 0 \quad \text{on } z = \epsilon\eta\,, \tag{4.30}$$

(It is noted that these boundary conditions are expressed in dimensional form, with the small parameter ϵ modifying the free surface elevation variable η). We further use Taylor series expansion about $z = 0$, which generates additional powers of ϵ. The dynamic free surface boundary condition would then become

$$g\eta + \frac{\partial \phi}{\partial t} + \frac{\epsilon}{2}\left[\left(\frac{\partial \phi}{\partial x}\right)^2 + \left(\frac{\partial \phi}{\partial z}\right)^2\right] + \epsilon\eta\frac{\partial}{\partial z}\left\{g\eta + \frac{\partial \phi}{\partial t} + \frac{\epsilon}{2}\left[\left(\frac{\partial \phi}{\partial x}\right)^2 + \left(\frac{\partial \phi}{\partial z}\right)^2\right]\right\} + \ldots \text{ on } z = 0\,, \tag{4.31}$$

and the kinematic free surface boundary condition

$$\frac{\partial \eta}{\partial t} - \frac{\partial \phi}{\partial z} + \epsilon\frac{\partial \phi}{\partial x}\frac{\partial \eta}{\partial x} + \epsilon\eta\frac{\partial}{\partial z}\left(\frac{\partial \eta}{\partial t} - \frac{\partial \phi}{\partial z} + \epsilon\frac{\partial \phi}{\partial x}\frac{\partial \eta}{\partial x}\right) + \ldots \text{ on } z = 0\,. \tag{4.32}$$

Finally the dependent variables ϕ and η are expanded in a power series in terms of the parameter ϵ

$$\phi = \sum_{n=1}^{\infty}\epsilon^{n-1}\phi_n = \phi_1 + \epsilon\phi_2 + \ldots\,, \tag{4.33}$$

$$\eta = \sum_{n=1}^{\infty}\epsilon^{n-1}\eta_n = \eta_1 + \epsilon\eta_2 + \ldots\,, \tag{4.34}$$

and substituted into the boundary value problem. The problem can then be separated into orders and solved sequentially. The order of the theory is denoted by the parameter ϵ; second-order Stokes theory retains terms up to $O(\epsilon)$, third-order Stokes theory up to $O(\epsilon^2)$, etc. The solutions at each higher order are dependent on lower-order solutions.

4.4.1 Properties of Weakly Nonlinear Deep Water Waves

The solution for ϕ at second order in ϵ can be found to be

$$\phi = Kx + \frac{gH}{2\omega}\frac{\cosh k(h+z)}{\cosh kh}\sin(kx-\omega t) + \frac{3\omega H^2}{32\sinh^4 kh}\cosh 2k(h+z)\sin 2(kx-\omega t)\,, \tag{4.35}$$

where $Kx = U$ is a mean current that accounts for any nonperiodic components of the solution (this is quantified below), and

$$\eta = \frac{H}{2}\cos(kx-\omega t) + \frac{kH^2\cosh kh(2\cosh^2 kh+1)}{16\sinh^3 kh}\cos 2(kx-\omega t) - \frac{kH^2}{8\sinh 2kh}\,. \tag{4.36}$$

It can be seen that the free surface elevation is comprised of a fundamental (or first harmonic) oscillating at a frequency ω, a second harmonic component oscillating at a frequency 2ω, and a mean set down. The two oscillating components are in phase at the crests and 180° out of phase at the troughs; this has the effect of sharpening the crests and flattening the troughs of the combined Stokes wave. The mean set down arises due to the choice of the Bernoulli constant to equal zero, and in this instance is relevant to the case of waves propagating up a slope in an otherwise infinite ocean.

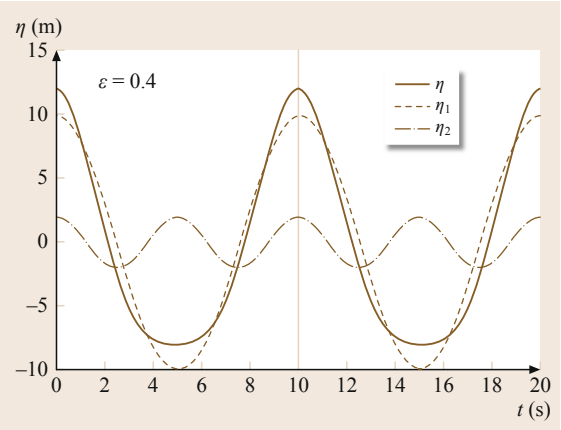


Fig. 4.2 Free surface of a second-order Stokes wave with $T = 10\ s$, along with the two component waves

However, for situations in which the motion is restricted to a finite domain (e.g., standing waves in a basin) the mean sea surface is necessarily zero, which would lead to a nonzero Bernoulli constant. Figure 4.2 shows the individual components and the total free surface profile for a second-order Stokes wave.

The particle velocities under a second-order Stokes wave can be derived from the velocity potential

$$u = \frac{\partial \phi}{\partial x} = K + \frac{gHk}{2\omega}\frac{\cosh k(h+z)}{\cosh kh}\cos(kx-\omega t) + \frac{3\omega H^2 k}{16\sinh^4 kh}\cosh 2k(h+z)\cos 2(kx-\omega t)\ , \tag{4.37}$$

$$w = \frac{\partial \phi}{\partial z} = \frac{gHk}{2\omega}\frac{\sinh k(h+z)}{\sinh kh}\sin(kx-\omega t) + \frac{3\omega H^2 k}{16\sinh^4 kh}\sinh 2k(h+z)\sin 2(kx-\omega t)\ . \tag{4.38}$$

The dispersion relation remains that of linear theory; the effects of nonlinearity on the wavelength appear at third order. Integrating the horizontal velocity u over depth and averaging over a wave period will allow an evaluation of the net mass flux (or Stokes drift). The value of the constant K is dictated by whether zero mass flux is indicative of the scenario at hand. For a large domain, the constant $K = 0$ and the mass flux is the value of the Stokes drift, while in confined situations the mass flux is necessarily zero and the mass flux would be carried by K. This also has implications for the definition of the phase speed C:

- The phase speed C can be defined as the speed of the wave relative to a stationary water column. In this case, $K = 0$, there is no net motion below the trough, and the phase speed

$$C = \sqrt{\frac{g}{k}\tanh kh}\ .$$

  The Stokes drift defines the net mass flux in the direction of wave propagation.
- The phase speed C can alternatively be defined as the speed of the wave relative to a stationary observer seeing waves with no mass flux. The phase speed would then be reduced by an amount proportional to the Stokes drift.

The pressure in the water column can be derived from the Bernoulli equation, and is

$$p = -\rho g z + \frac{\rho g H}{2}\frac{\cosh k(h+z)}{\cosh kh}\cos(kx-\omega t) - \frac{\rho g k H^2}{16\sinh kh\cosh kh}\cosh 2k(h+z) + \left[\frac{3\rho g k H^2}{16\cosh kh\sinh^3 kh}\cosh 2k(h+z) - \frac{\rho g h H^2}{16\sinh kh\cosh kh}\right]\cos 2(kx-\omega t)\ . \tag{4.39}$$

It is evident that the pressure contains a term that does not oscillate and another that is constant with depth. Evaluation of the pressure at the bottom ($z = -h$) reveals a term that reduces the static pressure due to the mean set down.

The basis of Stokes wave theory is that each harmonic of the fundamental frequency represents an additional order in the theory, and as such each harmonic amplitude becomes correspondingly smaller. It is difficult to establish a shallow water limit on the validity of Stokes theory based on the mathematical development, though a limit can be established on physical grounds. For example, one criterion could be that the second-order amplitude remain sufficiently small so that a bump does not appear in the trough of the wave; this implies that

$$kh > \sqrt{\frac{3a}{h}}\ , \tag{4.40}$$

for a bump to not appear in the trough.

4.4.2 Evolution of Weakly Nonlinear Deep Water Waves

Weakly nonlinear, deep water waves do not generally travel as permanent form waves, but evolve as a result of interactions with other waves of different frequencies and directions. These interactions can consist of

so-called *bound* waves (forced waves that do not satisfy the linear dispersion relation and, thus, can grow only a limited amount) and *resonant* waves (forced waves that do satisfy the linear dispersion relation and are thus able to grow without bound as the forcing is secular).

The expansion of the surface boundary conditions in powers of ϵ leads to products of unknowns that increase as the order of the expansion increases. Thus, given

$$\eta \approx e^{i\psi}, e^{-i\psi} , \tag{4.41}$$

(where $\psi = \mathbf{k}\cdot\mathbf{x} - \omega t$), higher-order terms lead to

$$\eta^2 \approx e^{2i\psi}, 1, e^{-2i\psi} , \tag{4.42}$$

$$\eta^3 \approx e^{3i\psi}, e^{i\psi}, e^{-i\psi}, e^{-3i\psi} , \tag{4.43}$$

and so on, until the desired truncation. Terms that result from this operation will remain bound to the primary harmonic and not experience unbounded growth. However, the η^2 will also generate terms proportional to $e^{i\psi}$ and $e^{-i\psi}$, and η^3 will generate terms proportional to $e^{2i\psi}$, 1, and $e^{-2i\psi}$, or (in both cases) terms proportional to the next lower-order solution. Solution of the resulting boundary value problem will generate terms that will grow linearly in either space or time without bounds (secular behavior). To ameliorate this, the method of multiple scales is often used to treat these problems. The application of this method allows the potentially secular behavior to be split into motions with different degrees of variation. In many cases, the fast degree of variation is associated with the waveform itself (the periodic motion) and the slow degree of variation with another property of the wave (for example, the amplitude of the wave form), which can then be linked to variations in some external field (e.g., the bathymetry).

The formalism in the previous section is most relevant to the case where a single wave component gives rise to harmonics, all of which can evolve during propagation to comprise a changing wave form. A coordinate system moving with the group velocity of the wave is defined; the resulting coordinates are defined as τ (time coordinate defining the slow temporal variation) and ξ (space coordinate defining the slow spatial variation). Carrying this analysis to third order will lead to the cubic nonlinear Schrödinger equations [4.11]

$$-i\frac{\partial A}{\partial \tau} - \frac{1}{2}\frac{\partial^2 \omega}{\partial k^2}\frac{\partial^2 A}{\partial \xi^2} + \beta|A|^2 A + \gamma A = 0 , \tag{4.44}$$

where

$$\beta = \frac{\omega k^2}{2} D + \frac{g^2 k^2}{2\omega(C_g^2 - gh)}\left(1 + \frac{kC_g}{2\omega\cosh^2 kh}\right)^2 , \tag{4.45}$$

and

$$\gamma = kS(\tau)\left(1 + \frac{kC_g}{2\omega\cosh^2 kh}\right) , \tag{4.46}$$

where $S(\tau)$ is an integration constant, and

$$D = \frac{\cosh 4kh + 8 - 2\tanh^2 kh}{8\sinh^4 kh} . \tag{4.47}$$

However, in most cases in the ocean, a description of wave propagation via the harmonics of a single wave is inapplicable, and we must turn toward a description comprised of a summation of waves of different frequencies and directions with random phases

$$\boldsymbol{\phi} = \sum_{n=1}^{\infty} -\frac{i}{2} B_n \frac{\cosh k_n(h+z)}{\cosh k_n h} e^{i(\mathbf{k_n}\cdot\mathbf{x} - \omega_n t + \epsilon_n)} + c.c. , \tag{4.48}$$

where B_n is the (complex) amplitude of the n-th component of $\boldsymbol{\phi}$ and ϵ_n is a random phase associated with this component. The corresponding expressions for the free surface elevation is

$$\eta = \sum_{n=1}^{\infty} \frac{A_n}{2} e^{i(\mathbf{k_n}\cdot\mathbf{x} - \omega_n t + \epsilon_n)} + c.c. , \tag{4.49}$$

where $c.c.$ refers to complex conjugate.

Substituting these expressions into the second-order boundary value problem will lead to expressions for the second-order contributions to the random sea, as shown by *Sharma* and *Dean* [4.12]

$$\begin{aligned}\boldsymbol{\phi} = &\sum_{n=1}^{\infty} -\frac{ig}{2\omega_n} A_n \frac{\cosh k_n(h+z)}{\cosh k_n h} e^{i(\mathbf{k_n}\cdot\mathbf{x} - \omega_n t + \epsilon_n)} \\ &+ \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} -\frac{ig^2 A_i A_j D_{ij}^{+}}{8\omega_i\omega_j(\omega_i + \omega_j)} \frac{|\mathbf{k}_i + \mathbf{k}_j|(h+z)}{\cosh|\mathbf{k}_i + \mathbf{k}_j|h} \\ &\times e^{i(\psi_i + \psi_j)} \\ &+ \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} -\frac{ig^2 A_i A_j D_{ij}^{-}}{8\omega_i\omega_j(\omega_i - \omega_j)} \frac{|\mathbf{k}_i - \mathbf{k}_j|(h+z)}{\cosh|\mathbf{k}_i - \mathbf{k}_j|h} \\ &\times e^{i(\psi_i - \psi_j)} + c.c. ,\end{aligned} \tag{4.50}$$

where i and j are two arbitrary frequency components in the spectrum. The coefficients D_{ij}^{+} and D_{ij}^{-} are

$$D_{ij}^{+} = \frac{\left\{\begin{array}{c}(\sqrt{R_i} + \sqrt{R_j})\left[\sqrt{R_j}(k_i^2 - R_i^2) + \sqrt{R_i}(k_j^2 - R_j^2)\right] \\ +2(\sqrt{R_i} + \sqrt{R_j})^2(\mathbf{k}_i\cdot\mathbf{k}_j - R_i R_j)\end{array}\right\}}{(\sqrt{R_i} + \sqrt{R_j})^2 - k_{ij}^{+}\tanh k_{ij}^{+} h} , \tag{4.51}$$

$$D_{ij}^{-} = \frac{\left\{(\sqrt{R_i}-\sqrt{R_j})\left[\sqrt{R_j}(k_i^2-R_i^2)-\sqrt{R_i}(k_j^2-R_j^2)\right]\right\} \\ \left\{+2(\sqrt{R_i}-\sqrt{R_j})^2(\mathbf{k}_i\cdot\mathbf{k}_j+R_iR_j)\right\}}{(\sqrt{R_i}-\sqrt{R_j})^2-k_{ij}^{-}\tanh k_{ij}^{-}h} , \tag{4.52}$$

where

$$k_{ij}^{\pm} = |\mathbf{k}_i \pm \mathbf{k}_j| \tag{4.53}$$

$$R_n = \frac{\omega_n^2}{g} . \tag{4.54}$$

The associated expression for the free surface elevation is

$$\begin{aligned}\eta = &\sum_{n=1}^{\infty}\frac{A_n}{2}e^{i(\mathbf{k_n}\cdot\mathbf{x}-\omega_n t+\epsilon_n)} \\ &+\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\frac{1}{8}A_iA_j\left[\frac{D_{ij}^{+}-\mathbf{k}_i\cdot\mathbf{k}_j+R_iR_j}{\sqrt{R_iR_j}}+R_i+R_j\right] \\ &\times e^{i(\psi_i+\psi_j)} \\ &+\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\frac{1}{8}A_iA_j\left[\frac{D_{ij}^{-}-\mathbf{k}_i\cdot\mathbf{k}_j-R_iR_j}{\sqrt{R_iR_j}}+R_i+R_j\right] \\ &\times e^{i(\psi_i-\psi_j)} .\end{aligned} \tag{4.55}$$

The terms $\psi_i+\psi_j$ and $\psi_i-\psi_j$ denote the sum and difference interactions, respectively. Of particular importance is the difference interaction $\psi_i-\psi_j$, a much longer wave than those wave components responsible for its generation and which is considered to have a role in harbor oscillation and seiching.

The waves generated by this second-order mechanism, as mentioned above, will not grow without limit, as they do not satisfy the linear dispersion relation and, thus, are not free waves. Resonant interactions, which generate waves that do satisfy the linear dispersion relation, are a particular subset of interactions. It can be shown [4.13] that no resonant interactions appear at second order (save for one trivial set), and one must go to third order to determine these interactions. The application of multiple scale techniques to address these resonances was first performed by *Benney* [4.14], and the resulting evolution equations are quite complicated; however, the resonance conditions can be written generically as follows

$$\psi_l + \psi_m - \psi_p = \psi_n , \tag{4.56}$$

where l, m, and p are indices that represent wave components which interact with ψ_n. These quartet interaction terms have been approximated for wind wave generation models [4.15] and are a primary engine for energy transfer within the spectrum in wind wave generation.

4.5 Shallow Water Wave Theories

In contrast to deep water, there is a distinct difference in the scale of variability between horizontal and vertical motions in shallow water (small kh). The horizontal motions are scaled similarly to deep water, but the vertical motions for small kh are

$$\begin{aligned}w = \frac{\partial\boldsymbol{\phi}}{\partial z} &= \frac{gkH}{2\omega}\frac{\sinh k(z+h)}{\cosh kh}\sin(kx-\omega t) \\ &\approx \frac{gHk^2(z+h)}{\omega}\times O(1) ,\end{aligned} \tag{4.57}$$

since $\sinh k(z+h) \approx k(z+h)$ for small kh. Using the shallow water asymptote of the linear dispersion relation and rearranging the shallow water asymptote of the vertical velocity w

$$\begin{aligned}w &\approx \frac{gHk^2(z+h)}{\omega} = \frac{gak^2h}{\omega}\left(1+\frac{z}{h}\right) \\ &= \frac{a}{h}\frac{\omega}{k}kh\left(1+\frac{z}{h}\right) = \delta\mu\left(\frac{\omega}{k}\right)\left(1+\frac{z}{h}\right) .\end{aligned} \tag{4.58}$$

As is the case with ϵ in deep water, two small parameters ($\delta = a/h$ and $\mu = kh$) are evident and are used for scaling. However, in this case, the vertical motions are much smaller than the horizontal motions. The shallow water asymptote of the linear dispersion relation, in addition to the scale difference between horizontal and vertical motions, is now applied to the boundary value problem. Unlike the deep water case, the governing equation here is altered

$$\frac{\partial^2\boldsymbol{\phi}}{\partial z^2}+\mu^2\frac{\partial^2\boldsymbol{\phi}}{\partial x^2} = 0 \quad -h \le z \le \eta , \tag{4.59}$$

and the disparity between horizontal and vertical motions is clear. The bottom boundary condition remains unchanged. The dynamic free surface boundary condition becomes

$$\mu^2\left(g\eta+\frac{\partial\boldsymbol{\phi}}{\partial t}\right)+\frac{\delta}{2}\left(\mu^2\left(\frac{\partial\boldsymbol{\phi}}{\partial x}\right)^2+\left(\frac{\partial\boldsymbol{\phi}}{\partial z}\right)^2\right) = 0$$

$$\text{on } z = \eta , \tag{4.60}$$

and the kinematic free surface boundary condition becomes

$$\mu^2\left(\frac{\partial\eta}{\partial t}+\delta\frac{\partial\eta}{\partial x}\frac{\partial\boldsymbol{\phi}}{\partial x}\right)=\frac{\partial\boldsymbol{\phi}}{\partial z}\quad\text{on } z=\eta\,. \tag{4.61}$$

Rather than using Taylor series about $z=0$ to address the unknown position of the free surface (as was done for Stokes theory), a power series for the depth dependence is used. This takes advantage of the fact that vertical variation of dynamical variables is weak in shallow water

$$\boldsymbol{\phi}(x,z,t)=\sum_{n=0}^{\infty}\left(\frac{z}{h}+1\right)^n\boldsymbol{\phi}_n(x,t)\,. \tag{4.62}$$

This expression for the velocity potential $\boldsymbol{\phi}$ can further be altered to satisfy the governing equation and the bottom boundary condition. The power series can then be expressed in terms of $\boldsymbol{\phi}_0$, the velocity potential at $z=-h$

$$\boldsymbol{\phi}=\boldsymbol{\phi}_0-\frac{\mu^2}{2}\left(\frac{z}{h}+1\right)^2\frac{\partial^2\boldsymbol{\phi}_0}{\partial x^2}+\frac{\mu^4}{24}\left(\frac{z}{h}+1\right)^4\frac{\partial^4\boldsymbol{\phi}_0}{\partial x^4}+O(\mu^6)\,. \tag{4.63}$$

This is then substituted into the kinematic and dynamic free surface boundary conditions and then integrated over depth. The result is a set of mass and momentum conservation equations, which are expressed in terms of powers of μ and δ, consistent with a weakly dispersive, weakly nonlinear assumption. Retaining terms up to $O(\delta,\mu^2)$ leads to the Boussinesq equations [4.16]

$$\frac{\partial\eta}{\partial t}+\frac{\partial}{\partial x}\left([h+\eta]\,\overline{u}\right)=0 \tag{4.64}$$

$$\frac{\partial\overline{u}}{\partial t}+\overline{u}\frac{\partial\overline{u}}{\partial x}+g\frac{\partial\eta}{\partial x}-\frac{h^2}{3}\frac{\partial^3\overline{u}}{\partial^2 x\partial t}=0\,, \tag{4.65}$$

where $\overline{u}$ is a depth-averaged velocity.

4.5.1 Properties of Weakly Nonlinear Shallow Water Waves

The Boussinesq equations can be modeled numerically for general wave propagation conditions. However, they can be further transformed into a single wave equation for η. This equation is known as the Korteweg–deVries (KdV) equation

$$\frac{\partial\eta}{\partial t}+c\frac{\partial\eta}{\partial x}+\frac{3c}{2h}\eta\frac{\partial\eta}{\partial x}+\frac{ch^2}{6}\frac{\partial^3\eta}{\partial x^3}=0\,, \tag{4.66}$$

where $c=\sqrt{gh}$. Due to an ambiguity in the derivation of the KdV equation, there are actually eight formally identical forms, but the form above is generally the version used.

The KdV equation is, in fact, exactly integrable. Two general analytic solution forms are possible with the KdV equation, each representing a wave of permanent form. The first is known as a solitary wave and is a wave consisting of an isolated hump of water with no trough. This wave is often used as a proxy for a tsunami propagating away from its origin. The free surface of a solitary wave is

$$\eta(x,t)=\eta_{\max}\operatorname{sech}^2\left[\frac{\sqrt{3}}{2}\left(\frac{\eta_{\max}}{h^3}\right)^{1/2}(x-ct-x_0)\right], \tag{4.67}$$

where $\eta_{\max}$ is the (specified) maximum free surface elevation and x_0 is the location of $\eta_{\max}$. One property of the solitary wave is that the wave form becomes narrower and more peaked as $\eta_{\max}$ increases. The phase speed c of the solitary wave is

$$c=\sqrt{gh}\left(1+\frac{\eta_{\max}}{h}\right)\,. \tag{4.68}$$

Figure 4.3 shows the free surface profile of a solitary wave for different values of η_{max}. The second form is a periodic wave known as a *cnoidal* wave, so named due to its dependence on the Jacobian *cn* function. Some of the calculable properties of cnoidal waves are listed below

$$c^2=gh\left(1+\frac{H}{h}\left[-1+\frac{2}{m}-\frac{3E(m)}{mK(m)}\right]\right)\,, \tag{4.69}$$

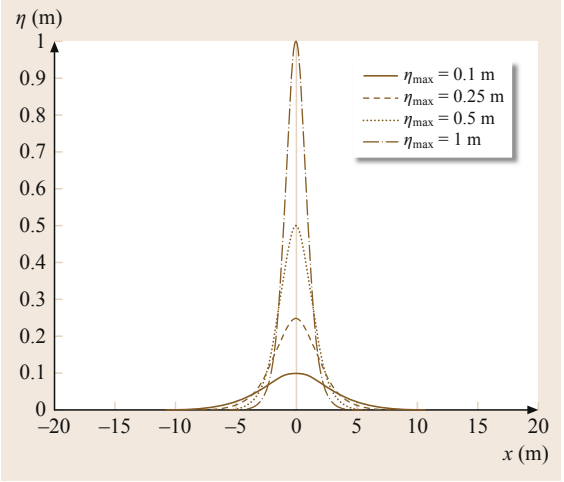


Fig. 4.3 Free surface profiles of a solitary wave with different values of $\eta_{\max}$

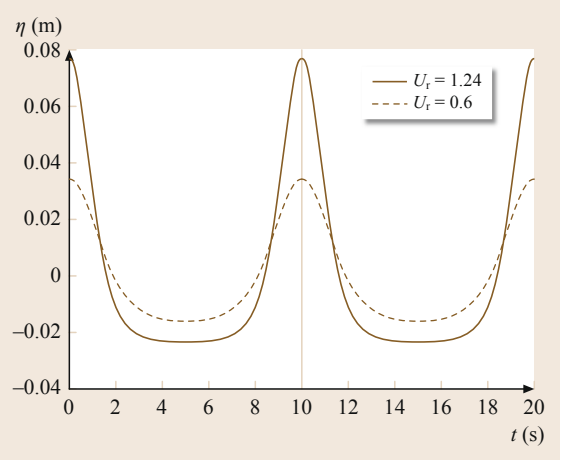


Fig. 4.4 Free surface profile of a cnoidal wave for two different values of the Ursell number U_r

$$\lambda = 4kh\left(\frac{m}{3\frac{H}{h}}\right)^2 , \tag{4.70}$$

$$T = \left(\frac{h}{g}\right)^2 \frac{4K(m)\left[\frac{h}{3H}\right]^2}{1+\frac{H}{h}\left(\frac{1}{2m}\right)\left(-m+2-3\frac{E(m)}{K(m)}\right)} , \tag{4.71}$$

$$\eta = \eta_2 + Hcn^2\left(\frac{2K(m)}{\lambda}(x-ct)\right) , \tag{4.72}$$

In these equations, $K(m)$ and $E(m)$ are the complete elliptic integrals of the first and second kind, respectively; m is a parameter, λ is the wave length, T is the wave period, and η_2 is the distance from the wave trough to the bottom at $z = -h$. The parameter m can be related to wave characteristics and determines the shape of the cnoidal wave; $m = 0$ recovers a sinusoidal wave, while $m = 1$ leads to a solitary wave. *Wiegel* [4.17] developed a straightforward procedure for calculating cnoidal wave profiles based on specified wave characteristics. Figure 4.4 shows cnoidal wave profiles for a wave with $t = 10$ *s* for two different values of the Ursell number $U_r = \delta/\mu^2$.

4.5.2 Evolution of Weakly Nonlinear Shallow Water Waves

In the context of irregular waves, this description of nearshore wave evolution can be described in terms of *wave–wave interaction*, a mechanism whereby waves of different frequencies in the spectrum can trade energy to alter the shape of the spectrum. This was mentioned earlier as a means by which wind sea transforms to long-period swell in the deep ocean. In the nearshore, the net effect of wave–wave interaction is to transfer energy from low frequencies to high ones; this accounts for changes in the spectral shape as waves evolve over sloping bathymetry; this was first described in detail by *Freilich* and *Guza* [4.18].

As with deep water waves, wave–wave interaction in shallow water is controlled by resonance between components of the wave spectrum. Unlike deep water, shallow water nonlinearity is dictated by *near*-resonant interaction among triads of waves. Given any arbitrary triad of wave components in a spectrum, the interaction triad is defined as

$$f_3 = f_2 + f_1 , \tag{4.73}$$

$$k_3 = k_2 + k_1 + O(\delta) , \tag{4.74}$$

where δ is a (small) parameter that describes the detuning away from perfect resonance. At the shallow water limit ($kh \to 0$) the parameter $\delta \to 0$, but even with finite δ significant energy transfer can take place. While the quartet interactions drive nonlinear energy transfer that occurs over spatial scales on the order of tens of kilometers, triad interactions force energy exchange, which can occur over tens of meters, drastically changing the shape and characteristics of the waves. These nonlinear interactions have been embedded into wave shoaling models, starting with *Freilich* and *Guza* [4.18], who derived a nonlinear shoaling model from the Korteweg–deVries equation, resulting in

$$\frac{\partial A_n}{\partial x} + \frac{1}{4h}\frac{\partial h}{\partial x}A_n - \frac{in^3k^3h^2}{6}A_n = \left[\sum_{l=1}^{N-1} A_l A_{n-l} + 2\sum_{l=1}^{N-n} A_l^* A_{n+l}\right] , \tag{4.75}$$

where A_n is the complex amplitude of the n-th Fourier component

$$\eta(x,y,t) = \sum_{n=1}^{N} \frac{A_n}{2} e^{i\int k_n \mathrm{d}x - \omega_n t} + c.c. , \tag{4.76}$$

and

$$nk = \frac{n\omega}{\sqrt{gh}} . \tag{4.77}$$

One limitation of this shoaling equation is the weakly dispersive assumption, which can be problematic for high frequency waves even in shallow water. This has been addressed via finite depth wave theory [4.19–22], among others.

4.6 Transformation of Waves Approaching Land

As waves approach the shoreline they undergo significant transformation in response to the arbitrarily varying bathymetry. Refraction and shoaling (as discussed above) will occur, but cannot be described by the Snell's law formulation. In addition, wave diffraction (the flow of energy along a wave crest due to high local gradients of wave height, and most often associated with waves encountering breakwaters) is also possible over bathymetric shoals.

The mild slope equation, developed by *Berkhoff* [4.23], has been used to simulate wave propagation over arbitrarily varying bathymetry

$$\nabla \cdot \left(CC_g \nabla \boldsymbol{\phi}\right) + k^2 CC_g \boldsymbol{\phi} = 0 \,, \tag{4.78}$$

where ∇ refers to differential operations in the horizontal (x, y) directions. It can be shown that the equation reduces to the Helmholtz equation for a flat bottom, and that it can also be transformed into a set of coupled equations for wave refraction by neglecting curvature terms. The equation is elliptic, which requires prespecification of the boundaries in advance. This is straightforward for enclosed harbors but challenging for open coasts, in which the location of wave breaking is unknown in advance. This has been addressed in the model's elliptic form [4.24]. However, the model has been transformed into a parabolic form [4.25], which affords a more straightforward numerical solution. *Kirby* [4.26] outlines several limitations of the parabolic form and offers some measures for ameliorating these limitations. The mild slope equation has been augmented with nonlinear interaction terms [4.20, 22, 27].

In the nearshore area, wave breaking is primarily controlled by the proximity of the bottom. The wave transformation process will force the wave to reach a limiting wave height, beyond which it cannot sustain its form. At this point the wave breaks. The breaking process and resultant generation of white water in the nearshore (the *surf zone*) is a major engine of nearshore processes (current generation, rip currents, infragravity waves, sediment transport, etc.)

The location of the initial breaking (or *incipient breaking*) of the wave demarcates the outer edge of the surf zone. It is important to identify conditions leading to incipient breaking for several reasons. For example, it is important in numerical modeling of wave propagation in the nearshore, since the wave breaking and dissipation terms are only activated within the surf zones, so it is essential to know when they need to be active.

One of the first incipient breaking criteria was developed by *McCowan* [4.28], who adapted the deep water solution of *Stokes* [4.10] to shallow water. Both studies assumed that a wave broke when the particle acceleration at the crest exceeds gravitational acceleration g, and showed that this condition resulted in an included crest angle of 120°. *McCowan* [4.28], however, also showed that, at this breaking condition

$$\frac{H_b}{h_b} = 0.78 \,, \tag{4.79}$$

or that the wave height reaches approximately 80% of the water depth at breaking. This criterion was developed using the properties of a solitary wave over a flat bottom, which would not generally fit the model of oscillatory motions in the surf zone. In an alternative treatment, *Miche* [4.29] developed a breaking criterion connected with wave steepness

$$\left(\frac{H_b}{L_b}\right) = \frac{1}{7} \tanh kh_b \,. \tag{4.80}$$

Note that this criterion approaches the maximum steepness criterion of *Stokes* [4.10] in deep water (large kh) and $H_b/h_b = 0.9$ in shallow water. A more complete development of the breaking criterion was developed by *Weggel* [4.30], who examined results from a number of laboratory experiments and determined that the bottom slope played an important role in the initiation of breaking. The resulting criterion is

$$\frac{H_b}{h_b} = b(m) - a(m)\frac{H_b}{gT^2} \,, \tag{4.81}$$

where

$$a(m) = 43.8\left(1 - e{-}19m\right) \,, \tag{4.82}$$

$$b(m) = \frac{1.56}{\left(1 + e^{-19.5m}\right)} \,. \tag{4.83}$$

This criterion reduces to (4.79) when the slope $m = 0$. Equation (4.81) is implicit for the breaking wave height H_b, so iteration is required.

Once the wave has broken, a description of the decay of the wave energy in the surf zone is needed. In this section, we discuss wave height decay mechanisms for both monochromatic and random waves. In general, the assumption of a spilling breaker is used; random wave formulations are typically determined by marrying this assumption to a probability distribution of some kind.

The equivalence between the energy decay in a spilling breaking wave and that of a hydraulic jump has been well established [4.31]. *Horikawa* and *Kuo* [4.32] conducted laboratory tests on breaking waves and determined the existence of a *stable* energy flux, which defines a state at which waves no longer break. This concept was further developed into a general description of wave evolution in the surf zone [4.33], with the hypothesis that the change in energy flux in the surf

zone would follow

$$\frac{\partial EC_g}{\partial x} = -\frac{K}{h}\left[EC_g - EC_{gs}\right] , \tag{4.84}$$

where K is a dimensionless decay coefficient, h is the still water depth, EC_g is the energy flux and the subscript s refers to the stable energy flux sought by the wave. Using shallow water wave theory and assuming that the stable wave height H_{stable} is a linear function of the water depth

$$H_{stable} = \Gamma h , \tag{4.85}$$

where Γ is a dimensionless coefficient, (4.84) can be transformed,

$$\frac{\partial\left[H^2\sqrt{h}\right]}{\partial x} = -\frac{K}{h}\left[H^2\sqrt{h} - \Gamma^2 h^{5/2}\right] . \tag{4.86}$$

Dally et al. [4.33] determined that $K = 0.15$ and $\Gamma = 0.4$ for minimum error to the data of *Horikawa* and *Kuo* [4.32]. *Dally* et al. [4.33] went on to develop analytical solutions for various bathymetric profiles, but (4.86) was the primary result.

Breaking of irregular waves in the surf zone requires a different approach than that for monochromatic waves, accommodating the random nature of waves by marrying the physics of wave breaking with probability theory. Early work by *Battjes* [4.34] and *Goda* [4.35] addressed the irregular nature of waves but did not allow for their evolution in the surf zone, constraining the statistical wave height measure to be some ratio between the wave height and the water depth.

Battjes and *Janssen* [4.36] developed a random wave breaking model based on energy flux conservation principles

$$\frac{\partial\overline{EC_g}}{\partial x} = -D , \tag{4.87}$$

where the overbar refers to averaged quantities and D is a dissipation rate. *Battjes* and *Janssen* [4.36] developed this dissipation rate from that of a dissipating bore, and introduced random wave heights by using the assumption that the heights of unbroken nearshore waves follow a Rayleigh distribution [4.37] for a range of heights that vary between zero and the maximum realizable wave height H_{max} at a particular water depth. The resulting dissipation rate D is

$$D = \frac{\alpha}{4} Q_b \overline{f} \rho g H_{max}^2 , \tag{4.88}$$

where H_{max} is the maximum wave height, Q_b is the percentage of breaking waves in a population of waves, $\overline{f}$ is an average frequency, and the coefficient α is of order 1. The maximum wave height is based on the *Miche* [4.29] criterion for maximum wave height, with some modification for the random nature of waves. This was then combined with the Rayleigh distribution for nearshore wave heights and further manipulated to lead to an expression for the fraction of breaking or broken waves Q_b

$$\frac{1 - Q_b}{\ln Q_b} = -\left(\frac{H_{rms}}{H_{max}}\right)^2 , \tag{4.89}$$

which is transcendental in Q_b. The model thus uses the offshore estimate of H_{rms} (from a measurement or a model) to calculate H_{max}, then Q_b from (4.89), and finally D from (4.88). With this estimate of the dissipation rate, (4.87) is then used to calculate the energy (and H_{rms}) at the new position.

Thornton and *Guza* [4.38] commented that the formulation of *Battjes* and *Janssen* [4.36] was, in effect, the implementation of a sharp cutoff value of the probability distribution of wave heights at H_{max}. They argue that waves in a group can momentarily exceed H_{max}, so a more gradual cutoff of the probability distribution near the maximum wave height is required. They then developed two weightings for this region of the probability distribution, either of which would allow wave heights above the theoretical maximum to occur. Using the same dissipative bore paradigm as *Battjes* and *Janssen* [4.36] but with different parameters, using the weightings, and integrating over the Rayleigh probability distribution, they determined two different dissipation rates for random waves. The first,

$$D = \frac{3\sqrt{\pi}}{16}\rho g \frac{B^3\overline{f}}{\gamma^4 h^5} H_{rms}^7 , \tag{4.90}$$

was less accurate when compared to data but leads to an analytical solution, while the second,

$$D = \frac{3\sqrt{\pi}}{16}\rho g B^3 \overline{f} \frac{H_{rms}^5}{\gamma^2 h^3}\left[1 - \frac{1}{\left(1 + \left(\frac{H_{rms}}{\gamma h}\right)^2\right)^{5/2}}\right] , \tag{4.91}$$

compared relatively well to data. The parameter B is nominally defined as the proportion of the front face of a breaking wave covered in foam, while γ is the ratio of the wave height to water depth in the surf zone; both are generally calibrated to data. *Thornton* and *Guza* [4.38] have shown that $\gamma = 0.42$ and $1.3 \le B \le 1.7$ work well for field data.

Both mechanisms have been used as a basis for the incorporation of random wave breaking into nonlinear phase-resolving models [4.20, 39]. The model of *Battjes* and *Janssen* [4.36] has been further extended for steep slopes [4.40], following work by *Baldock* et al. [4.41].

There is ample evidence for the existence of very long period of motion in the nearshore. These motions are generally thought to be largely the result of time variation of wave heights and breaking locations [4.42], though contributions from nonlinear wave–wave interactions [4.43] (see also *Bowen* and *Guza* [4.44]) may also be responsible. These long waves (often termed *infragravity* waves) have periods on the order of minutes and play a significant role in the evolution of the beach face.

The basic equations governing these long wave motions can be derived from the uniform depth equations of motion; the details can be found in [4.45]. The result can be expressed as an inhomogeneous wave equation

$$\frac{\partial^2\overline{\eta}}{\partial t^2} - \frac{\partial}{\partial x}\left(gh_O\frac{\partial\overline{\eta}}{\partial x}\right) - \frac{\partial}{\partial y}\left(gh_O\frac{\partial\overline{\eta}}{\partial y}\right) = F\,, \quad (4.92)$$

where h_O is a representative water depth and $\overline{\eta}$ is a mean (wave-averaged) free surface elevation. The term F represents forcing of the long wave motion.

In many cases [4.46], the long wave motion in the nearshore results from the transformation of long wave energy from offshore; the long wave propagates as a free wave. In this case, $F = 0$ in (4.92). The resulting motion in the nearshore is called an *edge wave*. The mean free surface elevation is assumed to have a longshore periodicity

$$\overline{\eta} = A(x)e^{i(k_y y \pm \omega t)}\,, \quad (4.93)$$

where $k_y = k\sin\theta$, or the longshore component of the wave number; the wave approach direction θ is assumed to be from shore normal. Substituting (4.93) into (4.92), and further assuming a plane beach, leads to a solution for $A(x)$ [4.47]

$$A(x) = a_n e^{-k_y x}L_n(2k_y x)\,, \quad (4.94)$$

where the subscript n denotes a mode of the solution (matching the number of zero crossings of the cross-shore structure of the edge wave) with amplitude a_n. The cross-shore structure is given by $L_n(2k_y x)$, which is the Laguerre polynomial of order n. The associated dispersion relation for this motion is

$$\omega^2 = gk_y\sin\left[(2n+1)\beta\right]\,, \quad (4.95)$$

where β is the angle of the beach from horizontal. Since there is no dissipation, the edge wave will fully reflect from the shoreline. For $n = 0$, all edge waves will propagate in the longshore direction, with crests perpendicular to the shoreline. For $n > 0$, however, the edge wave is affected by refraction; the wave reflects seaward from the shoreline, refracts to the point where the crests are perpendicular to the shoreline, then continues refracting back toward the shoreline. Such an edge wave is said to be *trapped*. On the other hand, a long wave which is *not* refracted back toward shore is said to be *leaky*. The phenomenon of *surf beat* [4.48, 49] is considered a leaky mode.

The discussion of edge waves in the previous section is useful as an introduction to nearshore long waves. However, because the motion is free, these long wave frequencies must be present at reasonable energy levels in the incident wave spectrum. The more significant source of long wave energy in the nearshore and surf zones comes from forced motion [4.46]. In this case, the forcing term F in (4.92) would be nonzero. If we dictate that the forcing is (as for nearshore circulation) dependent on the gradients of radiation stress [4.50], then the term F becomes

$$F = \frac{1}{\rho}\frac{\partial^2 S_{xx}}{\partial x^2}\,, \quad (4.96)$$

where the analysis is limited to processes in the cross-shore direction only (longshore derivatives are zero in (4.92)). In a wave group, the changes in radiation stress are assumed to be due to the spatial variations of the wave heights within the group. Using this assumption and limiting the domain to a constant water depth h_O gives the following solution for the mean sea surface

$$\overline{\eta} = -\frac{S_{xx}(x,t)}{\rho(gh_O - c_g)}\,, \quad (4.97)$$

where it can be seen that the mean sea surface is in antiphase with the gradient of radiation stress. Besides the limitation to constant depth, this solution is problematic in that $\overline{\eta}$ becomes quite large as the group velocity of the short waves in the group begins to approach the shallow water asymptote.

Later approaches sought to remedy these issues and move the area of interest closer to the surf zone. *Symonds* et al. [4.42] represented the wave group as a prescribed wave height variation about a mean amplitude and propagating over a sloping bottom, then used a constant breaking height-to-depth ratio *breaking index* inside the surf zone. The varying wave height led to a moving breakpoint, which served as a varying boundary condition for the generation of the long wave. The assumption of a constant breaking index was used to constrain the wave height in the surf zone, but destroyed any remaining group structure therein, which was later shown to be at least partially incorrect [4.51]. *Foda* and *Mei* [4.52] and *Schäffer* and *Svendsen* [4.53] perused an alternative formulation, in which the breakpoint was fixed but allowed for the group structure to remain in the surf zone.

The growth of the infragravity waves after this initial generation must now be considered. *Elgar*

et al. [4.54] showed that the infragravity wave energy tends to be a function of $h^{1.1}$, an almost linear dependence. However, coupling (4.97) with Green's Law shoaling yielded a growth rate closer to h^{-5}. The discrepancy is due to how the wave environment in the surf zone is treated. *Van Dongeren* and *Svendsen* [4.55], using a quasi-3-D (three-dimensional) nearshore circulation model, showed that the growth rate of infragravity waves can be dictated by manipulating the phase difference between the bound wave (locked with the wave group) and a free wave in the surf zone. Later, *Janssen* et al. [4.56] developed an analytical solution for the phase shift, leading to growth rates more consistent with measurements [4.57].

Oltman-Shay et al. [4.58], examining wave-like structures in nearshore current data taken at Duck, NC (USA), found that these structures did not correspond to any known wave theory. *Bowen* and *Holman* [4.59] applied stability theory to the equations governing nearshore circulation (with several simplifying assumptions) and determined that the most unstable modes had frequencies that were in the range of those observed in [4.58]. These instabilities were termed *shear waves* since they seemed to be caused by the shear instabilities of the longshore current. These motions only appeared when the longshore current was present and sufficiently energetic. The speed of these waves was independent of their frequency (a nondispersive phenomenon) and their signature in frequency longshore wave number space is both distinct and distinctly different from gravity wave phenomena.

Since the original papers there have been numerous studies on these shear waves. *Reniers* et al. [4.60] recreated these waves in the laboratory. Analytical means of studying the growth of these phenomena were advanced by *Dodd* and *Falques* [4.61], *Shrira* et al. [4.62], and *Feddersen* [4.63], among others. Additionally, the modeling of shear waves became a motivating factor for the development and application of numerical models of nearshore circulation ([4.64, 65], among others).

4.7 Computational Method for Fully Nonlinear Waves

With the development of powerful computers, numerical methods have been developed to tackle nonlinear wave problems and problems involving arbitrary body boundaries. Boundary integral methods are found to be more efficient for solving wave problems formulated using the potential flow theory, in particular to handle steep and overturning waves and complex body boundary shapes. For linear problems, boundary integral methods based on the free surface Green's function have been developed for analysis in both the frequency and time domains. Frequency domain analysis using a simple Rankine source has also been used to studying linear wave–body interaction problems in the frequency domain. An excellent review on numerical methods for free surface flows is given in [4.66].

Longuet-Higgins and *Cokelet* [4.4] developed a boundary integral method to solve the fully nonlinear inviscid wave motion problem. The method involves solution of Green's theorem, which is based on the Eulerian description of flow and the nonlinear free surface boundary conditions in the Lagrangian form; the method is, therefore, considered to be based on the mixed Eulerian–Lagrangian (MEL) formulation. To illustrate the method, let us consider a wave–body interaction problem such as that depicted in Fig. 4.1. Let the lateral extent of the domain be truncated by an open boundary Σ. Let us say that the flow has been started from rest with the initial condition being $\phi = 0$ and $\eta = 0$ at time $t = 0$. Since Σ is not a physical boundary, it has to be modeled so that waves incident on it may pass through without any reflection. There are several ways to achieve that approximately, as through use of nonlinear wave equations, free surface damping etc. Here let us consider a simple model by which it is assumed that $\phi = 0$ on Σ during the duration of flow simulation; in other words, simulation will be carried out only until the radiating waves reach the vicinity of the open boundary Σ.

Per Green's theorem,

$$2\pi\phi(P) + \int_{\partial\Omega} \phi \frac{\partial}{\partial n}\frac{1}{r_{PQ}} \mathrm{d}\partial\Omega - \int_{\partial\Omega} \frac{1}{r_{PQ}}\frac{\partial\phi}{\partial n}\mathrm{d}\partial\Omega = 0\,, \tag{4.98}$$

where $\partial\Omega$ is the union of all boundaries; i. e., $\partial\Omega = B + S_\mathrm{B} + \mathcal{F} + \Sigma$. Here Green's function $1/r_{PQ}$ corresponds to the potential at P due to the point source at Q. On B and S_B the normal velocity $\partial\phi/\partial n$ is known based on the no-flux condition, while ϕ is not known. On the open boundary, ϕ is known, here set to be zero, while $\partial\phi/\partial n$ is not known. On the free surface, one can time integrate the fully nonlinear free surface conditions at each time step to determine the free surface deformation and the velocity potential on the free surface; in other words, time integrate the dynamic condition

$$\frac{D\phi}{Dt} = \frac{1}{2}|\nabla\phi|^2 - \mathrm{g}Y\,,$$

to advance ϕ from discrete time n to $n+1$ and time integrate the free surface (material surface) kinematic

condition

$$\frac{D\boldsymbol{X}}{Dt} = \nabla\boldsymbol{\phi}\ ,$$

to advance position of free surface nodes $\boldsymbol{X} = (X, Y, Z)$ from discrete time n to $n+1$. Algorithms such as the fourth-order Runge–Kutta and Adams–Bashforth schemes [4.67, 68] may be used for the time integration. Thus, knowing $\boldsymbol{\phi}$ on the free surface, one may rewrite the above Green's theorem with known terms on the right-hand side and unknown terms on the left-hand side

$$\begin{aligned}
&2\pi\boldsymbol{\phi}(P \in B + S_B) \\
&+ \int\limits_{B+S_B} \boldsymbol{\phi}\frac{\partial}{\partial n}\frac{1}{r_{PQ}}\mathrm{d}\partial\Omega - \int\limits_{\Sigma+\mathcal{F}} \frac{1}{r_{PQ}}\frac{\partial\boldsymbol{\phi}}{\partial n}\mathrm{d}\partial\Omega \\
&= -2\pi\boldsymbol{\phi}(P \in \mathcal{F} + \Sigma) \\
&\quad - \int\limits_{\mathcal{F}+\Sigma} \boldsymbol{\phi}\frac{\partial}{\partial n}\frac{1}{r_{PQ}}\mathrm{d}\partial\Omega + \int\limits_{B+S_B} \frac{1}{r_{PQ}}\frac{\partial\boldsymbol{\phi}}{\partial n}\mathrm{d}\partial\Omega\ . \qquad (4.99)
\end{aligned}$$

The above integral equation is discretized and the resulting algebraic (matrix) equation solved either directly or iteratively for $\boldsymbol{\phi}$ on solid boundaries B and S_B, and for $\partial\boldsymbol{\phi}/\partial n$ on the free surface $\mathcal{F}$ and open boundary Σ. Upon determining $\boldsymbol{\phi}$ on the body, one can use the Euler integral (unsteady Bernoulli's equation) to determine pressure and through integration of pressure the hydrodynamic force on the body. The solution is thus advanced in time.

The mixed Eulerian–Lagrangian formulation has become a standard approach for solving fully nonlinear inviscid wave and wave–body interaction problems and it has been adopted in field discretization methods such as the finite difference method. Works on nonlinear wave and wave–body interaction problems based on the MEL formulation include those by *Vinje* and *Brevig* [4.69], *Grosenbaugh* and *Yeung* [4.67], *Dommermuth* et al. [4.68], *Saout* and *Ananthakrishnan* [4.70], *Ananthakrishnan* [4.71] and Xue et al. [4.72].

4.8 Wave Forces on Fixed and Floating Structures

In this section, the methods to determine the wave exciting force (which consists of incident and body-diffracted wave forces) on a body, and in the case of a freely floating body the additional wave radiation force due to the body motion generated waves are presented. Both theoretical and numerical methods to determine the wave forces will be discussed. We shall take the mean forward of the bodies to be zero here. One can find the nonzero forward speed cases in the literature on ship hydrodynamics and naval architecture. Empirical and exact methods to determine the viscous drag force is also discussed. Particular emphasis is given to parameters that govern ratios of various wave component and viscous drag forces.

4.8.1 Incident Wave Force: Froude–Krylov Force

Let us consider a body (submerged or floating) in a wave field as illustrated in Fig. 4.1. Let the incident wave be of small amplitude and be propagating in the positive x direction with elevation and potential given by

$$\eta^{\mathrm{i}} = \frac{H^{\mathrm{i}}}{2}\cos(kx - \omega t)\ ,$$

$$\boldsymbol{\phi}^{\mathrm{i}} = \frac{H^{\mathrm{i}}}{2}\frac{\cosh k(z+h)}{\cosh kh}\sin(kx - \omega t)\ ,$$

where the superscript i denotes incident wave. Using the Euler integral, one can find the dynamic pressure of the incident wave as

$$p^{\mathrm{i}} = -\rho\frac{\partial\boldsymbol{\phi}^{\mathrm{i}}}{\partial t} = \rho g\frac{H^{\mathrm{i}}}{2}\frac{\cosh k(z+h)}{\cosh kh}\cos(kx - \omega t)\ .$$

By integrating the incident wave pressure about the body surface (mean surface if the body is undergoing oscillation) one can determine the incident wave force, which is also known as the Froude–Krylov force, $\boldsymbol{F}^{\mathrm{i}}$

$$\begin{aligned}
\boldsymbol{F}^{\mathrm{i}} &= \int\limits_{S_o} p^{\mathrm{i}}\mathrm{d}S_o \\
&= \int\limits_{S_o} \rho g\frac{H^{\mathrm{i}}}{2}\frac{\cosh k(z+h)}{\cosh kh}\hat{n}\cos(kx - \omega t)\,\mathrm{d}S_o\ ,
\end{aligned}$$

where S_o denotes the body surface and $\hat{n}$ the unit normal vector into the body. Using the Gauss theorem one may write the above as a volume integral

$$\boldsymbol{F}^{\mathrm{i}} = \int\limits_{S_o} p^{\mathrm{i}}\hat{n}\mathrm{d}S_o = -\int\limits_{\forall_o} \nabla p^{\mathrm{i}}\mathrm{d}\forall_o\ ,$$

where $\forall_o$ denotes the volume occupied by the body. In the component form, the incident wave forces are then

$$\begin{aligned}
F_x^{\mathrm{i}} &= \rho g k\frac{H^{\mathrm{i}}}{2}\frac{1}{\cosh kh} \\
&\quad \times \int\limits_{\forall_o} \cosh k(z+h)\sin(kx - \omega t)\mathrm{d}\forall_o\ ,
\end{aligned}$$

and

$$F_z^{\mathrm{i}} = -\rho g k \frac{H^{\mathrm{i}}}{2} \frac{1}{\cosh kh} \times \int_{\forall_{\mathrm{o}}} \sinh k(z+h) \cos(kx - \omega t) \mathrm{d}\forall_{\mathrm{o}} .$$

If the body is really small, or more precisely spans small distances along x and z compared to the incident wave length, then the above integral may be further approximated by replacing x and z in the integrals by the x and z coordinates of the centroid. The Froude–Krylov force will then simply be

$$F_x^{\mathrm{i}} = \rho g k \frac{H^{\mathrm{i}}}{2} \frac{\cosh k(\bar{z}+h)}{\cosh kh} \sin(k\bar{x} - \omega t) \ \forall_{\mathrm{o}} ,$$

and

$$F_z^{\mathrm{i}} = -\rho g k \frac{H^{\mathrm{i}}}{2} \frac{\sinh k(\bar{z}+h)}{\cosh kh} \cos(k\bar{x} - \omega t) \ \forall_{\mathrm{o}} ,$$

where $(\bar{x}, \bar{z})$ denote the coordinates of the centroid of the body (submerged part of the body if the body were floating). It is thus a straightforward calculation to determine the small amplitude incident wave force, in particular if the body size is small compared to the incident wave length.

From the above integrals, one can also estimate the order of magnitude of the incident wave force. For a near surface body, the incident wave force is of the order of magnitude

$$|\boldsymbol{F}^{\mathrm{i}}| \equiv F^{\mathrm{i}} = O(\rho g \forall_{\mathrm{o}} k \frac{H^{\mathrm{i}}}{2}) = \Delta O(k \frac{H^{\mathrm{i}}}{2}) ,$$

where Δ denotes the weight (displacement) of the body.

Note that $kH^{\mathrm{i}}/2$ denotes the slope of the incident wave. In other words, the Froude–Krylov force of the order of body weight times the wave slope.

4.8.2 Morison Force on a Stationary Body

The incident wave force given by the Froude–Krylov force does not account for the viscous drag force, which could be significant even if the body size is small compared to the incident wave length. Computing the viscous drag force exactly would require solving the incompressible Navier–Stokes equation with free surface conditions, which is a formidable task. *Morison* et al. [4.73] proposed an empirical method to determine the wave force on a body including the drag force. Decomposing the hydrodynamic force into inertia and drag components, which is exact for force on a submerged body without a free surface, Morison proposed to determine the wave force as

$$F = F_{\mathrm{inertia}} + F_{\mathrm{drag}} .$$

In terms of inertia and drag coefficients,

$$F = C_{\mathrm{I}} \rho \dot{u} \forall + C_{\mathrm{d}} \frac{\rho}{2} u|u| A_{\mathrm{p}} ,$$

where F denotes the x component of the force, u the x component of fluid velocity, $\dot{u}$ the x component of fluid acceleration, $\forall$ the displaced volume of the body, and A_{p} the projected area of the body normal to the x-axis. The inertia and drag coefficients C_{i} and C_{d} are empirically obtained; scaled with respect to volume and projected area, they are both $O(1)$. The above is referred to as the Morison equation for the wave force on a body [4.8]. In the Morison equation method the fluid velocity and acceleration are determined using wave theories. As per linear Airy wave theory, as seen in an earlier section, the amplitudes of u and $\bar{u}$ are given by

$$|u| = \frac{H^{\mathrm{i}}}{2} \frac{gk}{\omega} \frac{\cosh k(z+h}{\cosh kh} ,$$
$$|\dot{u}| = \frac{H^{\mathrm{i}}}{2} gk \frac{\cosh k(z+h}{\cosh kh} .$$

The above Morison equation method thus provides a practical method to determine the viscous incident wave force on a body. The decomposition also allows one to determine the relative significance of the inertia and drag components of the incident wave force. For a body in a wave under wave influence (i. e., $kz \approx 0$)

$$\frac{\text{Drag}}{\text{Inertia}} = \frac{C_{\mathrm{d}} \rho u|u| A_{\mathrm{p}}/2}{C_{\mathrm{I}} \rho \dot{u} \forall} = O\left(\frac{H^{i2} g^2 k^2 D^2}{\omega^2 H^{\mathrm{i}} g k D^3}\right)$$
(here D denotes body length)
$$= O\left(\frac{H^{\mathrm{i}} g k}{\omega^2 D}\right) = O\left(\frac{H^{\mathrm{i}}}{D} \frac{1}{\tanh kh}\right)$$
(using the dispersion relation) .

From the above, it is clear that the drag force is more significant for a large wave height to body length ratio and/or in shallow water (i. e., small kh) [4.8]. Using wave kinematics, one can easily establish that in the case of deep water waves, the above ratio is related to the Keulegan–Carpenter number KC of oscillating flows [4.74]

$$KC \equiv \frac{UT}{D} = \frac{\left(\frac{H^{\mathrm{i}}}{2\,\omega}\right)\left(\frac{2\pi}{\omega}\right)}{D} = \frac{H^{\mathrm{i}}}{D} .$$

From the above discussion, the following points are worth recapitulating:

- The ratio of drag force to inertia force depends on (i) the wave height to body diameter ratio and (ii) wave length to water depth ratio.
- In the case of small wave height to body diameter ratio H^{i}/D in deep water waves (i. e., tanh kh = 1), the Morison force will be equal to the Froude–Krylov force for appropriate value of C_i.
- When using the Morison equation method for wave force on a moving body, the velocity and accelerations are taken to be relative to the body motion.

If the body dimension is large compared to wave height and not small compared to the incident wave length, then the drag force will not be important, but the Morison equation cannot be used to determine the inertial force because the scattering (diffraction) of waves by the body will become significant. One can use the potential flow theory to solve the diffraction problem (as done in [4.75] for a vertical cylinder) and also determine the wave diffraction force. The wave exciting force will then be the sum of the Froude–Krylov and the diffraction wave forces. The diffraction wave problem pertaining to linear wave–body interaction is discussed in the next section.

4.8.3 Wave Diffraction over a Body

Let the body shown in Fig. 4.1 be stationary. The mere presence of the body will cause the incident waves to scatter. For small amplitude waves, governed by linearized free surface conditions, one may solve the diffraction problem separately and determine the total potential as

$$\boldsymbol{\phi} = \boldsymbol{\phi}^{\mathrm{i}} + \boldsymbol{\phi}^{\mathrm{d}} ,$$

where the superscripts i and d denote incidence and diffraction, respectively. The solution of the incident wave potential is simply that of a free periodic wave as presented in Sect. 3.2. The diffraction potential is governed by the following set of equations

$$\begin{aligned}
\nabla^2\boldsymbol{\phi}^{\mathrm{d}} &= 0 , \\
\frac{\partial\boldsymbol{\phi}^{\mathrm{d}}}{\partial z} &= 0, \text{ on the sea bottom } z = -h , \\
\frac{\partial\boldsymbol{\phi}^{\mathrm{d}}}{\partial n} &= -\frac{\partial\boldsymbol{\phi}^{\mathrm{i}}}{\partial n} \text{ on the body surface } S_B , \\
\frac{\partial^2\boldsymbol{\phi}^{\mathrm{d}}}{\partial t^2} + \mathrm{g}\frac{\partial\boldsymbol{\phi}^{\mathrm{d}}}{\partial z} &= 0, \text{ on the mean free surface } z = 0 .
\end{aligned}$$

Moreover, in the far field, the diffraction potential must satisfy the Sommerfeld radiation condition [4.9]

$$\sqrt{R}\left(-i\omega\boldsymbol{\phi}^{\mathrm{d}} + C\frac{\partial\boldsymbol{\phi}^{\mathrm{d}}}{\partial R}\right) = 0, \text{ as } R \to \infty ,$$

where $R = \sqrt{x^2 + y^2}$, which denotes the radial distance from the body and C the wave phase speed. The above diffraction problem was solved by *McCamy* and *Fuchs* for the case of a vertical circular cylinder [4.75]. The x component wave exciting force with diffraction is given by

$$F_{\mathrm{x}} = \frac{\rho \mathrm{g}\frac{H}{2}D^2\tanh(kh)}{k^2R^2H_1^{(2)'}(kR)} ,$$

where $H_1^{(2)}$ denotes the Hankel function of the second kind and order 1 and $R = D/2$ the radius of the cylinder.

In the case of small H/D and not so small L/D, as in the case of waves incident on a supertanker or a large gravity platform, the diffraction force will be more significant than the drag force. On the other hand, drag force will be the predominant part of the wave exciting force for a mooring cable (small diameter) in a similar sea.

4.8.4 Wave Radiation Force on an Oscillating Body

In the case of a compliant or freely floating body, the body will undergo oscillatory motion when subject to the wave exciting force. The force due to waves caused by body motion is referred to as the wave radiation force [4.9]. Now let the body considered in Fig. 4.1 undergo rigid body motion such that the normal velocity of the body may be written as

$$\begin{aligned}
V_{\mathrm{n}} &= \boldsymbol{U}\cdot\hat{n} + (\boldsymbol{\Omega}\times\boldsymbol{r})\cdot\hat{n} \\
&= \boldsymbol{U}\cdot\hat{n} + \boldsymbol{\Omega}\cdot(\boldsymbol{r}\times\hat{n}) \\
&= \sum_{i=1}^{i=6} U_i n_i ,
\end{aligned}$$

where $\hat{n}$ denotes the unit normal vector on the body surface and r the position vector from the axis of rotation through the center of gravity. Moreover, $U_i = U_1, U_2, U_3, \Omega_1, \Omega_2, \Omega_3$ and $n_i = n_1, n_2, n_3, (\boldsymbol{r}\times\hat{n})_1, (\boldsymbol{r}\times\hat{n})_2, (\boldsymbol{r}\times\hat{n})_3$ for $i = 1, 2, 3, 4, 5, 6$, respectively, corresponding to the sixth degree of freedom rigid body motion. For the linear small amplitude body and wave motion problem, the radiation wave potential can be decomposed as per Kirchoff modal decomposition [4.9]

and written as

$$\Phi^{\mathrm{r}} = \sum_{j=1}^{j=6} A_j \boldsymbol{\phi}_j^{\mathrm{r}}\, e^{-i\omega t}\,,$$

where Φ^{r} denotes the total wave radiation potential and $\boldsymbol{\phi}_i^{\mathrm{r}}$ the i-th mode of radiation potential per unit amplitude of body motion. The equations governing the unit radiation potentials are given by [4.9]

$$\begin{aligned}
\nabla^2 \boldsymbol{\phi}_j^{\mathrm{r}} &= 0\,,\\
\frac{\partial \boldsymbol{\phi}_j^{\mathrm{r}}}{\partial z} &= 0, \text{ on the sea bottom } z = -h\,,\\
\frac{\partial \boldsymbol{\phi}_j^{\mathrm{r}}}{\partial n} &= -i\omega n_j \\
&\quad \text{on the equilibrium body surface } S_{\mathrm{Bo}}\,,\\
-\omega^2 \boldsymbol{\phi}_j^{\mathrm{r}} + \mathrm{g}\frac{\partial \boldsymbol{\phi}_j^{\mathrm{r}}}{\partial z} &= 0, \,, \\
&\quad \text{on the mean free surface } z = 0\,.
\end{aligned}$$

At infinity, the radiation potential must satisfy the Sommerfeld radiation condition

$$\sqrt{R}\left(-i\omega \boldsymbol{\phi}_j^{\mathrm{r}} + C\frac{\partial \boldsymbol{\phi}_j^{\mathrm{r}}}{\partial R}\right) = 0, \text{ as } R \to \infty\,,$$

where $R = \sqrt{x^2 = y^2}$, which denotes the radial distance from the body and C the phase speed of the radiating waves (which, for example, for the deep water wave case is g/ω). Upon solving the above radiation problem, for example by using the simple source distribution method originated by *Yeung* [4.66], one can determine the complex hydrodynamic coefficient

$$f_{ij} = \rho \int_{S_{\mathrm{Bo}}} \boldsymbol{\phi}_i^{\mathrm{r}} n_j \mathrm{d}S_{\mathrm{Bo}}\,,$$

which can be decomposed into components that are proportional to body acceleration (known as the added mass force) and to velocity (known as the wave damping force)

$$f_{ij} = -\omega^2 \mu_{ij} - i\omega\lambda_{ij}\,,$$

where μ_{ij} denotes the added mass force coefficient and λ_{ij} the wave damping force coefficient, both for force/moment along direction i for the j-th mode of motion. Using Green's identity one can establish that the coefficients are symmetric: $f_{ij} = f_{ji}$. For bodies with symmetry one can show that wave damping and wave excitation forces can be related using the Haskind relation [4.9]. The linear wave–body interaction theory is thus quite useful from a practical viewpoint to determine the wave forces. The theory is also rich in classical mathematics. Thus both theoreticians and practical engineers are attracted to the subject. The reader may refer to classical texts such as those by *Wehausen* and *Laitone* [4.7], *Newman* [4.9], and *Mei* [4.76] for detailed accounts of formulation and analysis of wave–body interactions.

In the case of a linear wave body interaction problem involving freely floating bodies, one then has to solve rigid body dynamics problem to determine the body response to wave forces consisting of incident, diffraction, and radiation wave force. In the case of the fully nonlinear wave–body interaction problem, both the hydrodynamic and body dynamic problems have to solved simultaneously, as each affect the other through the boundary conditions of hydrodynamic problems and through hydrodynamic force and moment of the body dynamics problem. The hydrodynamic problem may be solved using the mixed Eulerian–Lagrangian method of *Longuet-Higgins* and *Cokelet* [4.4] discussed earlier. Between the linear and fully nonlinear wave–body interaction theories, there are also weakly nonlinear theories developed for wave forces, as for example in [4.77], which are not discussed here. As the fully nonlinear wave theories are computationally intensive, linear and weakly nonlinear theories remain useful for engineering solutions to problems involving wave and body motions.

4.9 Concluding Remarks

The fundamentals of the mechanics of ocean wave theory and wave–body interactions were presented in this chapter. It began with an overview of linear wave theory, including the assumptions and limitations inherent in its use. Weakly nonlinear deep and shallow water wave theories were then outlined, including both permanent form waves (classical *Stokes* and cnoidal wave theories), followed by a discussion of wave spectral evolution and nonlinear wave–wave interactions. The transformation of waves over arbitrarily varying bathymetry was then detailed, touching on the mild slope equation for water wave propagation, nearshore wave breaking, infragravity waves, and waves caused by instabilities of nearshore circulation. Computational

methods for calculating fully nonlinear waves and the resultant forces on bodies were discussed. The interaction of waves with a submerged or floating body was also considered, including discussions of the Froude–Krylov force of the incident wave and the Morison equation method for determining the inertia and drag-components of the wave exciting force.

Related topics such as wind wave generation, sea spectra, and wave energy conversion are discussed in other chapters of the Handbook.

References

4.1 A.D.D. Craik: The origins of water wave theory, Annu. Rev. Fluid Mech. **36**, 1–28 (2004)
4.2 J.H. Michell: The wave-resistance of a ship, Philos. Mag. **45**(5), 106–123 (1898)
4.3 E.O. Tuck: *Wave resistance of thin ships and catamarans*, Tech. Rep., Vol. T8701 (University of Adelaide, Adelaide 1987)
4.4 M.S. Longuet-Higgins, E.D. Cokelet: The deformation of steep surface waves on water: I. A numerical method of computation, Proc. R. Soc. L. A **350**, 1–26 (1976)
4.5 H. Lamb: *Hydrodynamics* (Dover, Mineola 1945)
4.6 J.J. Stoker: *Water Waves – The Mathematical Theory with Applications* (Wiley, New York 1958)
4.7 J.V. Wehausen, E.V. Laitone: Surface waves. In: *Fluid Dynamics, Encyclopedia fo Physics*, Vol. 3, ed. by C. Trusdell (Springer, Berlin, Heidlberg 1960)
4.8 R.G. Dean, R.A. Dalrymple: *Water Wave Mechanics for Engineers and Scientists*, Advanced Series on Ocean Engeneering, Vol. 2 (World Scientific, Singapure 1991)
4.9 J.N. Newman: *Marine Hydrodynamics* (MIT Press, Cambridge 1977)
4.10 G.G. Stokes: On the theory of oscillatory waves, Trans. Camb. Philos. Soc. **8**, 441–455 (1847)
4.11 V.E. Zakharov: Stability of periodic waves of finite amplitude on the surface of a deep fluid, J. Appl. Mech. Tech. Phys. **9**, 190–194 (1968)
4.12 J.N. Sharma, R.G. Dean: Second order directional seas and associated wave forces, Soc. Petroleum Eng. J. **21**, 129–140 (1981)
4.13 O.M. Phillips: *The Dynamics of the Upper Ocean* (Cambridge Univ. Press, Cambridge 1980)
4.14 D.J. Benney: Nonlinear gravity wave interactions, J. Fluid Mech. **14**, 577–584 (1962)
4.15 S. Hasselmann, K. Hasselmann: Computations and parameterizations of the nonlinear energy transfer in a gravity-wave spectrum. Part 1: A new method for efficient computations of the exact nonlinear transfer integral, J. Phys. Oceanogr. **15**, 1369–1377 (1985)
4.16 D.H. Peregrine: Long waves on a beach, J. Fluid Mech. **27**, 815–827 (1967)
4.17 R.L. Wiegel: A presentation of cnoidal wave theory for practical application, J. Fluid Mech. **7**, 273–286 (1960)
4.18 M.H. Freilich, R.T. Guza: Nonlinear effects on shoaling surface gravity waves, Philos. Trans. R. Soc. Lond. A **311**, 1–41 (1984)
4.19 Y. Agnon, A. Sheremet, J. Gonsalves, M. Stiassnie: Nonlinear evolution of a unidirectional shoaling wave field, Coast. Eng. **20**, 29–58 (1993)
4.20 J.M. Kaihatu, J.T. Kirby: Nonlinear transformation of waves in finite water depth, Phys. Fluids **7**, 1903–1914 (1995)
4.21 H. Bredmose, Y. Agnon, P.A. Madsen: Wave transformation models with exact second order transfer, Eur. J. Mech. B/Fluids **24**, 659–682 (2005)
4.22 T.T. Janssen, T.H.C. Herbers, J.A. Battjes: Generalized evolution equations for nonlinear surface gravity waves over two-dimensional topography, J. Fluid Mech. **552**, 393–418 (2006)
4.23 J.C.W. Berkhoff: Computation of combined refraction-diffraction, Proc. 13th Int. Conf. Coast. Eng. (1972) pp. 471–490
4.24 Z. Demirbilek, V. Panchang: *CGWAVE: A Coastal Surface Water Wave Model of the Mild Slope Equation*, Tech. Rep., Vol. CHL-98-26 (US Army Engineer Research and Development Center, Vicksburg 1998)
4.25 A.C. Radder: On the parabolic equation method for water wave propagation, J. Fluid Mech. **95**, 159–176 (1979)
4.26 J.T. Kirby: Higher-order approximations in the parabolic equation method for water waves, J. Geophys. Res. **91**, 933–952 (1986)
4.27 Y. Tang, Y. Ouellet: A new kind of nonlinear mild-slope equation for combined refraction-diffraction of multi-frequency waves, Coast. Eng. **31**, 3–36 (1997)
4.28 J. McCowan: On the highest wave of permanent type, Philos. Mag. **38**, 351–358 (1894)
4.29 R. Miche: Mouvements ondulatoires de la mer en profondeur constante ou decroissante, Ann. Ponts Chaussees **114e**, 42–78 (1944)
4.30 R.J. Weggel: Maximum breaker height, J. Waterw. Port Coast. Ocean Eng. **98**, 529–548 (1972)
4.31 B. Le Mehaute: On non-saturated breakers and the wave runup, Proc. 8th Int. Conf. Coast. Eng. (1962) pp. 77–92
4.32 K. Horikawa, C.-T. Kuo: A study of wave transformation inside surf zone, Proc. 7th Int. Conf. Coast. Eng. (1966) pp. 217–233
4.33 W.R. Dally, R.G. Dean, R.A. Dalrymple: Wave height variation across beaches of arbitrary profile, J. Geophys. Res. **90**, 11917–11927 (1985)
4.34 J.A. Battjes: Set-up due to irregular waves, Proc. 13th Int. Conf. Coast. Eng. (1972) pp. 1993–2004
4.35 Y. Goda: Irregular wave deformation in the surf zone, Coast. Eng. Japan **18**, 13–26 (1975)
4.36 J.A. Battjes, J.P.F.M. Janssen: Energy loss and set-up due to breaking of random waves, Proc. 16th Int. Conf. Coast. Eng. (1978) pp. 569–587
4.37 M.S. Longuet-Higgins: On the statistical distributions of sea waves, J. Mar. Res. **11**, 245–265 (1952)
4.38 E.B. Thornton, R.T. Guza: Transformation of wave height distribution, J. Geophys. Res. **88**, 5925–5938 (1983)

4.39 Y. Eldeberky, J.A. Battjes: Spectral modeling of wave breaking: Application to Boussinesq equations, J. Geophys. Res. **101**, 1253–1264 (1996)

4.40 T.T. Janssen, J.A. Battjes: A note on wave energy dissipation over steep beaches, Coast. Eng. **54**, 711–716 (2007)

4.41 T.E. Baldock, P. Holmes, S. Bunker, P. van Weert: Cross-shore hydrodynamics within an unsaturated surf zone, Coast. Eng. **34**, 173–196 (1998)

4.42 G. Symonds, D.A. Huntley, A.J. Bowen: Two-dimensional surf beat: Long wave generation by a time-varying breakpoint, J. Geophys. Res. **87**, 492–498 (1982)

4.43 B. Gallagher: Generation of surf beat by nonlinear wave interactions, J. Fluid Mech. **49**, 1–20 (1971)

4.44 A.J. Bowen, R.T. Guza: Edge waves and surf beat, J. Geophys. Res. **83**, 1913–1920 (1978)

4.45 I.A. Svendsen: *Introduction to Nearshore Hydrodynamics* (World Scientific, Singapore 2006)

4.46 D.A. Huntley, R.T. Guza, E.B. Thornton: Field observations of surf beat. 1. Progressive edge waves, J. Geophys. Res. **86**, 6451–6466 (1981)

4.47 F. Ursell: Edge waves on a sloping beach, Proc. R. Soc. Lond. Ser. A **214**, 79–97 (1952)

4.48 W.H. Munk: Surf beats, Trans. Am. Geophys. Union **30**, 849–854 (1949)

4.49 M.J. Tucker: Surf beats: Sea waves of 1 to 5 min period, Proc. R. Soc. Lond. Ser. A **202**, 565–573 (1952)

4.50 M.S. Longuet-Higgins, R.S. Stewart: Radiation stresses in water waves – A physical discussion, with applications, Deep Sea Res. **11**, 529–562 (1964)

4.51 J.H. List: A model for the generation of two-dimensional surf beat, J. Geophys. Res. **97**, 5623–5635 (1992)

4.52 M.A. Foda, C.C. Mei: Nonlinear excitation of long-trapped waves by a group of short swells, J. Fluid Mech. **111**, 319–345 (1981)

4.53 H.A. Schäffer, I.A. Svendsen: Surf beat generation on a mild-slope beach, Proc. 21st Int. Conf. Coast. Eng. (1988) pp. 1058–1072

4.54 S. Elgar, T.H.C. Herbers, M. Okihiro, J. Oltman-Shay, R.T. Guza: Observations of infragravity waves, J. Geophys. Res. **97**, 15573–15577 (1992)

4.55 A.R. Van Dongeren, I.A. Svendsen: Nonlinear and 3D effects in leaky infragravity waves, Coast. Eng. **41**, 467–496 (2000)

4.56 T.T. Janssen, J.A. Battjes, A.R. van Dongeren: Long waves induced by short-wave groups over a sloping bottom, J. Geophys. Res. **108**, 3252–3264 (2003)

4.57 J.A. Battjes, H.J. Bakkenes, T.T. Janssen, A.R. van Dongeren: Shoaling of subharmonic gravity waves, J. Geophys. Res. **109**, C02009 (2004)

4.58 J. Oltman-Shay, P.A. Howd, W.A. Birkemeier: Shear instabilities of the mean longshore current: 2. Field observations, J. Geophys. Res. **94**, 18031–18042 (1989)

4.59 A.J. Bowen, R.A. Holman: Shear instabilities of the mean longshore current: 1. Theory, J. Geophys. Res. **94**, 18023–18030 (1989)

4.60 A.J.H.M. Reniers, J.A. Battjes, A. Falques, D.A. Huntley: A laboratory study on the shear instability of longshore currents, J. Geophys. Res. **102**, 8597–8609 (1997)

4.61 N. Dodd, A. Falques: A note on spatial modes in longshore current shear instabilities, J. Geophys. Res. **101**, 22715–22726 (1996)

4.62 V.I. Shrira, V.V. Voronovich, N.G. Kozhelupova: Explosive instability of vorticity waves, J. Phys. Oceanogr. **27**, 542–554 (1997)

4.63 F. Feddersen: Weakly nonlinear shear waves, J. Fluid Mech. **372**, 71–91 (1998)

4.64 D.N. Slinn, J.S. Allen, P.A. Newberger, R.A. Holman: Nonlinear shear instabilities of alongshore currents over barred beaches, J. Geophys. Res. **103**, 18357–18379 (1998)

4.65 H.T. Özkan-Haller, J.T. Kirby: Nonlinear evolution of shear instabilities of the longshore current: A comparison of observations and computations, J. Geophys. Res. **104**, 25953–25984 (1999)

4.66 R.W. Yeung: Numerical methods for free-surface flows, Annu. Rev. Fluid Mech. **14**, 395–442 (1982)

4.67 M.A. Grosenbaugh, R.W. Yeung: Nonlinear free-surface flow at a two-dimensional bow, J. Fluid Mech. **209**, 57–75 (1989)

4.68 D.G. Dommermuth, D.K.P. Yue, W.M. Lin, R.J. Rapp, E.S. Chan, W.K. Melville: Deep-water plunging breakers: A comparison between potential theory and experiments, J. Fluid Mech. **189**, 423–442 (1988)

4.69 T. Vinje, P. Brevig: Nonlinear ship motion, Proc. 3rd Int. Conf. Numer. Ship Hydrodyn., Paris (1981)

4.70 O. Saout, P. Ananthakrishnan: Hydrodynamic and dynamic analysis to determine the stability of an underwater vehicle near a free surface, Appl. Ocean Res. **33**, 158–167 (2011)

4.71 P. Ananthakrishnan: Effects of viscosity and free-surface nonlinearity on the wave motion generated by an oscillating twin-hull, Proc. OMAE 31st Int. Conf. Ocean Arct. Offshore Eng., Rio de Janeiro (2012) pp. 279–288

4.72 M. Xue: Xu H. Yue D. K. P.: Computations of fully nonlinear three dimensional wave-wave and wave-body interaction: Part 1. Dynamics of steep three dimensional wave, Journal of Fluid Mechanics **438**, 11–39 (2001)

4.73 J.R. Morison, M.P. O'Brien, J.W. Johnson, S.A. Schaaf: The force exerted by surface waves on piles, Petroleum Trans. **189**, 149–154 (1950)

4.74 T. Sarpkaya, M. Isaacson: *Mechanics of Wave Forces on Offshore Structures* (Van Nostrand Reinhold Co., New York 1981)

4.75 R. McCamy, R. Fuchs: *Wave forces on piles: A diffraction theory*, Tech. Memo, Vol. 69 (US Army Corps of Engineers, Washington 1954)

4.76 C.C. Mei: *The Applied Dynamics of Ocean Surface Waves* (World Scientific, Singapore 1989)

4.77 P. Ferrant, K. Pelletier: Second order wave diffraction patterns about complex offshore structures, Proc. 10th Int. Conf. Offshore Polar Eng., Vol. 3 (2000) pp. 686–693

5. Physical Properties of Seawater

Wendell S. Brown

This chapter presents the definitions of the principal physical properties of the seawater, including pressure, temperature, salinity, density, density anomaly, specific heat, and thermal expansion. The typical global ocean profiles and distribution maps of surface temperature, salinity, and density anomaly are presented. Other oceanic properties, including the stability of oceanic water columns, use of temperature–salinity diagrams in identifying water masses, sea water freezing, as well as oceanic sound transmission and light are treated.

Sea water can be characterized by its temperature, salinity (the dissolved solids), and pressure. From these three quantities and the equation of the state of sea water, the ocean scientist or engineer can calculate other desired quantities, such as the density, sound velocity, heat capacity, or electrical conductivity. The oceanographic tables of yesteryear have been replaced by the microcomputer for the calculation of these derived quantities.

5.1 Hydrostatic Pressure 101
5.2 Temperature 101
5.3 Salinity 103
5.4 Density 104
5.5 Temperature–Salinity Relationships 105
5.6 Specific Heat 106
5.7 Freezing of Sea Water and Sea Ice 106
5.8 Coefficient of Thermal Expansion 106
5.9 Sound Velocity 107
5.10 Acoustic Ambient Noise 107
5.11 Light Transmission 108
References 109

5.1 Hydrostatic Pressure

Assuming a static ocean (zero velocity), the vertical equation of motion reduces to a static force balance between the pressure gradient and water parcel weight according to

$$\frac{\partial p}{\partial z} = -\rho g \,, \tag{5.1}$$

where z is elevation positive upward, ρ is local water density, and g is the acceleration due to gravity. Upon integration of (5.1) from the surface (where gauge pressure is zero by definition) to depth $-z$, we obtain the hydrostatic pressure p_h relation

$$p_h(-z) = -\rho g z \,, \tag{5.2}$$

in which ρ is the average surface to depth $-z$ density. Hydrostatic pressure dominates the total oceanic pressure field. The hydrostatic relation (5.2) enables us to convert measured pressure to depth, if one assumes the density of an average ocean profile. For example, a pressure of 4062 decibars (5891 psi) converts to a depth of 4000 m (*Sanders* and *Fofonoff* [5.1], for further details and limitations).

5.2 Temperature

The temperature T of a water parcel is expressed in degrees Celsius (°C) and gives an indication of the energy or work that has been done on or associated with that water parcel. Temperature is now being easily and accurately measured by electronic thermometers employing thermistors or platinum resistance probes as the sens-

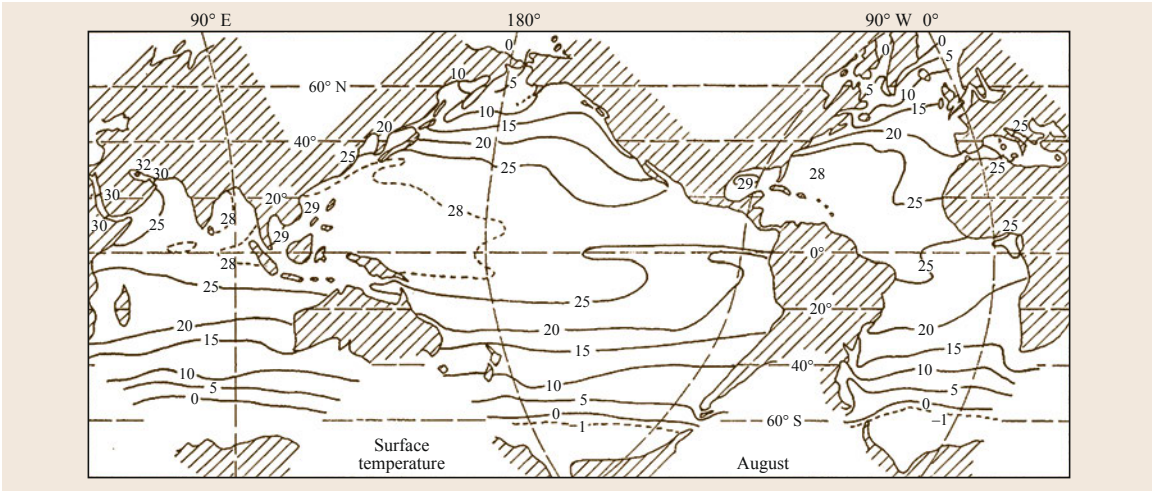


Fig. 5.1 Global ocean surface temperatures in August (after [5.2])

ing elements. These sensors are capable of resolving micro-degree temperature fluctuations, and are stable within millidegrees over a period of months. Oceanographic temperatures are referenced to the international practical temperature scale of 1968 (IPTS-68). (See *Mackenzie* [5.3] for a discussion of the IPTS-68 and the older IPTS-48 standards and the conversion between the two.)

Oceanographic temperature tends to vary systematically with depth and latitude. In the tropics, where there is an excess of the primarily short wave incoming solar radiation over outgoing long-wave radiation, the surface waters are warmer (with maxima of 25–30 °C) than those at adjacent latitudes. By contrast, surface waters at higher latitudes are cooler (minimums of ≈ −2 °C; the freezing point of saltwater), because energy is lost to the atmosphere and outer space by way of the outgoing long-wave radiation heat flux, which exceeds the incoming solar radiation heat flux at those latitudes. The typical global ocean sea surface temperature maps (Fig. 5.1) reflect that tropical warming and polar cooling. However, the distribution of temperature with latitude varies zonally, due to the general oceanographic circulation patterns, which vary with seasons. (See *Talley* et al. [5.4] for more detailed global ocean surface temperature distributions.)

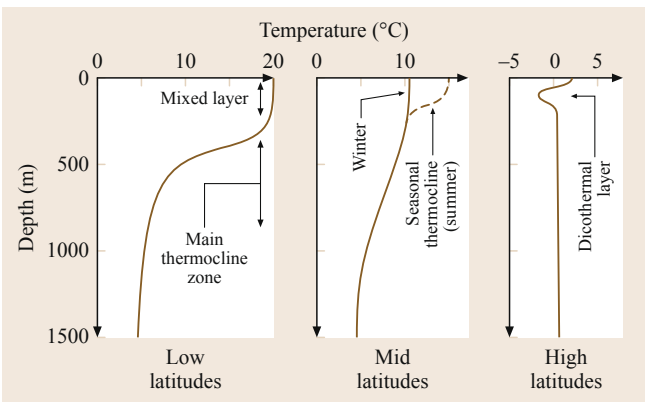


Fig. 5.2 Typical mean temperature profiles in the open ocean (after *Pickard* [5.2])

Typical vertical distributions of oceanic temperature at the low-, mid- and high-latitudes are presented in Fig. 5.2. Low- and mid-latitude oceans often feature a turbulence-induced mixed layer of relatively uniform temperatures that extend downward to a few 10 s of meters in depth. Below the mid-latitude mixed layer, temperature decreases rapidly through the *seasonal thermocline* – typically to a depth of about 100 m. While typical seasonal thermocline gradients are about 0.05 °C/m, they can vary in intensity and vertical extent depending on regional weather and seasonal climate. A permanent feature called the *main* thermocline, with typical vertical gradients of (order 0.02 °C/m), is found in the 100–1000 m depth range. The main thermocline is maintained by global ocean-scale surface heating/mixing and overturning processes. The deep water below the main thermocline is almost isothermal, featuring typically very small temperature gradients of order 0.001 °C/m.

Water is slightly compressible, so that the volume of a parcel of water moved from the surface to the ocean floor (5000 m depth), where the pressure is over 7000 psi (almost 500 times atmospheric pressure), will decrease by about 2%. If a sinking water parcel is compressed adiabatically (i. e., without loss or gain of heat), then the work done by the increasing pressure in forcing molecules closer together will raise its

measured or in-situ temperature T at a rate of about 1.4×10^{-4} °C/m – the *adiabatic temperature gradient*. For example, if a parcel with a temperature of 5.000 °C at the surface were lowered to a 4000 m depth, the temperature would increase to 5.438 °C. Conversely, if a parcel of sea water with an in-situ temperature of 5.000 °C at a 4000 m depth were raised adiabatically to the surface, then its in-situ temperature would decrease to 4.574 °C. (The difference in these two thought experiments is related to the compression process nonlinearity; something that is reflected in the equation of state of the water.) This pressure-induced temperature gradient is imposed on all water parcels uniformly and thus does not contribute to the buoyancy of a particular water parcel relative to the other surrounding water parcels.

So as to focus on the buoyancy-related temperature differences of water parcels, oceanographers define a *potential temperature* θ, which is the in-situ temperature that has been corrected mathematically for the compressibility effects described above. Generally, oceanographers correct in-situ temperatures to the ocean surface pressure which is zero gauge pressure. However, when comparing the temperatures of two water parcels at great depths, it is sometimes more accurate (because of the process nonlinearities) to correct to a common or a standard nonzero pressure or *zeroperfect*.

5.3 Salinity

Historically, ocean salinity was determined by the chemical titration of a sample collected in a bottle lowered on a wire from the ship. This titration-determined the chlorinity Cl of sea water (in units of parts per thousand – ‰) was directly related to salinity by the empirical Knudsen relation

$$S\,[\text{psu}] = 0.03 + 1.805\,\text{Cl}\,[‰]\,.$$

In the 1970s, considerable research was conducted on both the definition/measurement of conductivity and equation of state of sea water from which to calculate salinity. This work led to the adoption of the practical salinity scale of 1978 (PSS-78). (For details [5.5]) While modern computers enable the computation of salinity from these well-known complicated set of relationships, we find for practical purposes that they can be approximated by

$$S\,[\text{psu}] = 3.55 + 1.02\,C\,[\text{S/m}] - 0.73\,T\,[^\circ\text{C}]\,,$$

where *psu* is the practical salinity unit, T is the in-situ temperature measurements T (in units of degree Centigrade) and C is the in-situ conductivity measurement of (in units of Siemens/meter; a Siemen being an inverse ohm or the SI (international system of units) unit of conductivity). We find that seawater electrical conductivity varies around a value of 4. This technological advance has greatly expanded the measurement of ocean salinity.

The distribution of oceanic salinity (Fig. 5.3) is largely controlled by evaporation and precipitation pro-

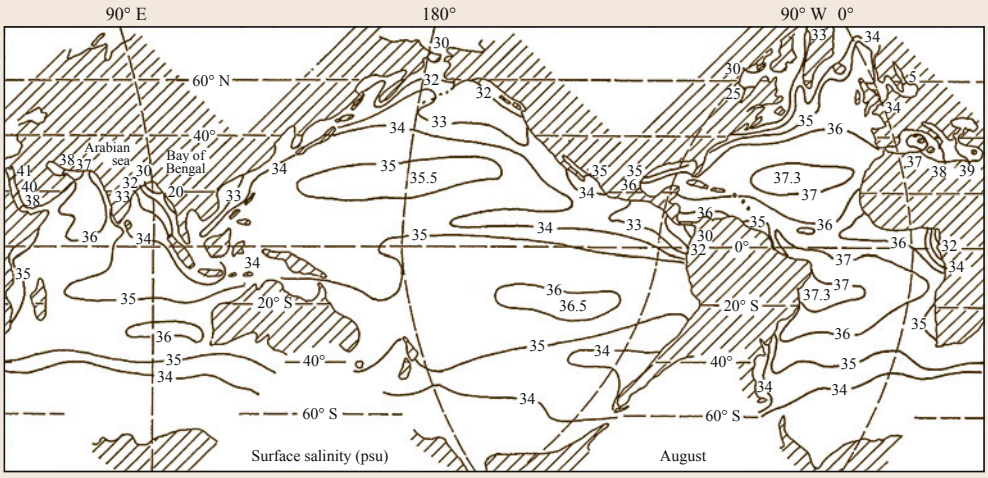


Fig. 5.3 Surface salinities of the oceans in August (after [5.2])

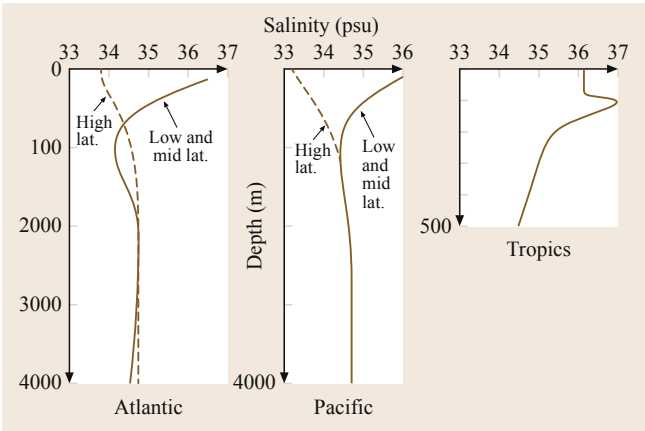


Fig. 5.4 Typical mean salinity profiles in the open ocean (after [5.2])

cesses in open ocean regions, fresh river runoff in coastal zones, and ice formation and melting in polar regions (both sea ice and glaciers). The lowest salinities (0–30 psu) are found in estuaries and polar regions. Coastal salinities are typically higher (30–34 psu), but less than open ocean salinities (33–37 psu). The average oceanic salinity is 34.7 psu. Semiclosed evaporation basins, such as the Mediterranean and the Red Sea, exhibit high salinities of 39 and 41 psu, respectively. The outflow of the very salty Mediterranean waters makes the Atlantic slightly saltier than the Pacific. The highest open ocean salinities are found in the center of the oceanic gyres (Chap. 2), while the lowest open ocean salinities are found along the equator due to excess rainfall. (See *Talley* et al. [5.4] for more detailed global ocean surface salinity distributions).

Typical vertical distributions of salinity are shown in Fig. 5.4. As with temperature, there is a surface mixed layer with relatively low salinity gradients. Below the mixed layer, there is a strong halocline where the salinities decrease; with a destabilizing influence. The depth of the halocline roughly agrees with the depth of the thermocline. There tends to be a salinity minimum around 800–1000 m at mid-latitudes. At low latitudes there tends to be a high surface salinity, while low surface salinities are found at high latitudes.

5.4 Density

The most important dynamic property of sea water is its *density* ρ, which is sometimes expressed as inverse of density or *specific volume* α_{STP}. Density is generally computed from the sea water equation of state, which is expressed as a function of in-situ temperature T, salinity S, and pressure P. (Oceanographers generally express pressure in units of decibars, since 1 decibar is nearly equal to the pressure due to 1 m of sea water. Note that 1 decibar $= 10^{-4}$ pascals, the SI unit of pressure.) Typical values of surface oceanic density vary from 1020 to 1030 $\mathrm{kg/m^3}$. Because oceanic density variations are relatively small, oceanographers have defined a quantity called density anomaly σ

$$\sigma(T,S,p) = \rho(T,S,p) - 1000\,, \tag{5.3}$$

which depends on in-situ temperature T, salinity S, and pressure P. The corresponding quantity for specific volume anomaly is

$$\delta = \alpha_STp - \alpha_{35,0,p}\,, \tag{5.4}$$

where $\alpha_{35,0,p}$ is the specific volume of the *standard ocean* with properties $T = 0°\mathrm{C}C$; $S = 35psu$, and p. The 1980 equation of state of sea water (EOS-80) (as given by *UNESCO* [5.5]) is the currently accepted standard for computing sea water properties.

Differing oceanic pressures induce small but significant volume differences in a water parcel, and hence in-situ density differences throughout the water column. For example, a parcel with a density anomaly of 28.106 $\mathrm{kg/m^3}$ at the surface will have a density anomaly of 46.644 $\mathrm{kg/m^3}$ at 4000 m. The significant pressure effect on the density anomaly in the deep ocean can obscure the effects of temperature and salinity variability.

Therefore, to compare the densities of two water parcels more conveniently, oceanographers have defined two related density anomaly quantities called σ_t and σ_θ. σ_t, which is evaluated according to

$$\sigma_t = \rho(T,S,p=0) - 1000\,, \tag{5.5}$$

is the density anomaly for a water parcel that has been raised to the ocean's surface adiabatically, essentially removing the largest effect of the pressure on the water parcel volume. Potential density anomaly (or sigma-theta σ_θ) is the density anomaly of a parcel of water for which the pressure effects on the volume and in-situ temperature removed. σ_θ is evaluated according to

$$\sigma_\theta = \rho(T=\theta,S,p=0) - 1000\,, \tag{5.6}$$

by using the potential temperature θ and zero pressure. Thus, value of potential density of a water parcel, which

is computed according to

$$\rho_{\text{pot}} = \sigma_\theta + 1000\,, \tag{5.7}$$

is based solely in terms of its surface temperature and salinity.

One of the most important features of the ocean is its vertical density gradient or stratification. In general, oceanic density increases with depth (i. e., $-z$) because there is the tendency for more dense parcels to sink below less dense parcels. This density environment supports a restoring force (due to the density difference and gravity) on a water parcel that has been displaced from its equilibrium position, such that it tends to its equilibrium position. The restoring force, whose strength is related to the vertical potential density gradient, gives rise to a class of oscillations called internal waves (discussed elsewhere). Oceanographers express the strength of this restoring force in terms of the buoyancy frequency, N or natural frequency of oscillation, given by

$$N(z) = \sqrt{-\frac{\mathrm{g}}{\rho}\frac{\partial \rho_{\text{pot}}}{\partial z}} \quad \text{(in units of rad/s)}\,. \tag{5.8}$$

The vertical density gradient – corrected for the effects of compressibility – is the relevant quantity here. If the more dense water is deeper, then the potential density gradient is negative, the water column is stable, N is positive, and internal waves are possible. If the water column density is uniform, then the potential density gradient and N are zero, the water column is neutrally stable, and oscillations are not possible. If more dense water sits atop less dense water, then the potential density gradient is positive, N is imaginary, the water column is unstable, and short-term water column overturning is inevitable.

Open ocean density is largely controlled by temperature, except in polar regions where salinity variations become significant. These relationships can be seen in the latitudinal variations of surface temperature, salinity, and density anomaly (σ_t), as shown in Fig. 5.5. The sampling of representative vertical profiles of density anomaly in Fig. 5.6 features the high-density gradient (high N and stability) zone called a pycnocline in the tropical/equatorial ocean. These upper ocean zones strongly resist vertical displacement, but support internal wave variability.

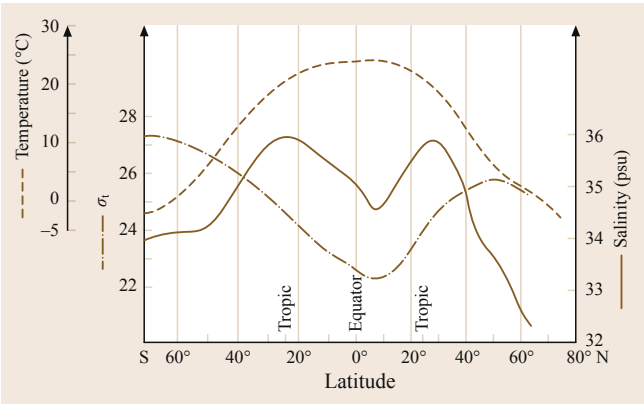


Fig. 5.5 Latitudinal variations of average ocean surface temperature, salinity and density anomaly σ_t (after [5.2])

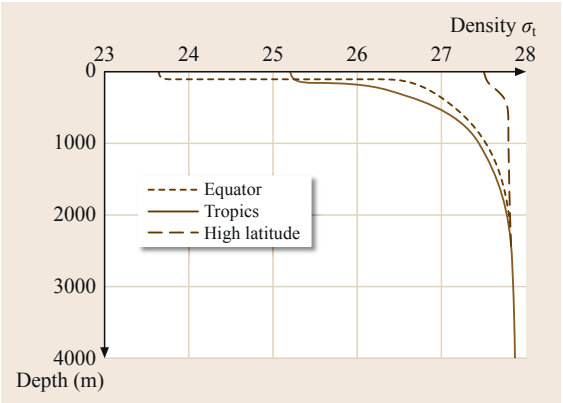


Fig. 5.6 Typical density/depth profiles at low and high latitudes (after [5.2])

5.5 Temperature–Salinity Relationships

Waters from different parts of the ocean can be identified by their distinctively different potential temperature and salinity characteristics as defined by their θ–S *relationships*, as revealed on a θ–S diagram like that in Fig. 5.7. The different curves in Fig. 5.7 are the distinctive θ–S *relationships* of water masses that have different mixing histories. Oceanographers use their knowledge of θ–S relationships to interpret the origin of waters that make up the local water column. Superimposed on the θ–S diagram are lines of equal σ_t. Oceanographers use such relationships to study water motion and mixing, while engineering designers can use water mass distributions to define design criteria.

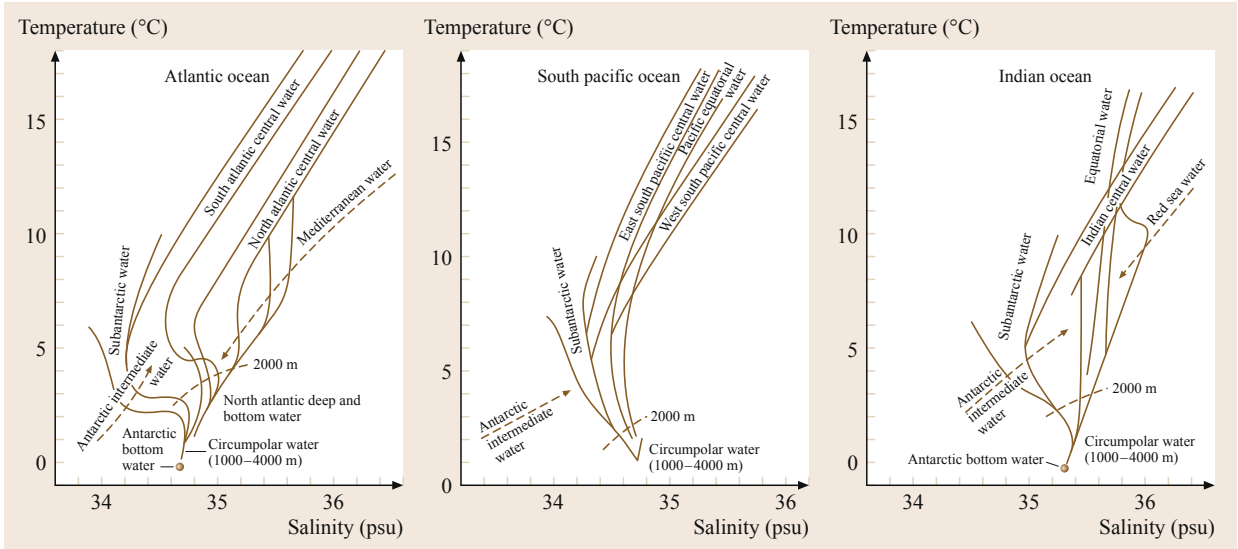


Fig. 5.7 Each of these schematic θ–S relationships that were derived from measurements in different ocean basins are characterized by their distinctly different water mass structures (after [5.6])

5.6 Specific Heat

The specific heat of sea water C_p is defined by the amount of heat in J that is required to raise the temperature of 1 kg of sea water 1 °C at a constant pressure; and thus has units of J/(kg°C). For sea water, the specific heat increases with temperature and decreases with salinity and pressure; with typical values $\approx 4000\,\mathrm{J/(kg°C)}$. The best empirical fit at zero pressure is given by *Millero* et al. [5.7] which extends to low temperatures. The pressure effect has not been directly measured, but has been estimated in UNESCO [5.8, p. 32–35]

5.7 Freezing of Sea Water and Sea Ice

The freezing temperature of sea water depends on the salinity and pressure of water. *Millero* and *Leung* [5.9] (or *UNESCO* [5.8, p. 29]) give an empirical form to laboratory measurements at low pressure. For example, sea water of 35 psu at the surface freezes at a temperature of −2.54 °C. As sea water freezes, the salt settles out of the forming ice and becomes part of a cold, salty water mass below the ice. Conversely, low salinity or freshwater layers are found at the surface in regions of melting ice.

5.8 Coefficient of Thermal Expansion

One of the unique properties of fresh water is the reversal in the sign of the coefficient of thermal expansion at 4 °C. Thus, water starting to freeze at 0 °C is less dense than water at 4 °C. As the salinity increases, this temperature of maximum density decreases. At a salinity of 24.7 psu, the freezing point and temperature of maximum density are equal at −1.33 °C. For salinities greater than 24.7 psu, the water continues to decrease in density with decreasing temperature until the freezing point is reached. A typical value of the coefficient of the thermal expansion is 2×10^{-4} °C. This value increases with temperature and pressure.

5.9 Sound Velocity

Communication in the ocean is done primarily with sound, because the ocean is opaque to all forms of electromagnetic radiation (from long-radio waves through short ultraviolet). Relative to electromagnetic radiation the ocean is relatively transparent to sound transmission. This is because sea water is compressible and so can support waves of compression and expansion, or small fluctuations in pressure about the mean – or *sound*. In the ocean, sound travels with a velocity of about 1500 m/s or about five times faster than in air.

Acoustics are a valuable tool for communication and getting a *view* through the ocean. The velocity of sound is a function of the temperature, salinity, and pressure, increasing with all three factors. (The work by *Chen* and *Millers* [5.10] used standard sea water and so it is both consistent with the Practical Salinity Scale of 1978 and in good agreement with values computed from the 1980 equation of state of sea water.) Temperature effects dominate the sound velocity profile in the upper ocean causing a decrease in sound velocity with depth. Pressure effects dominate in the deep ocean; increasing sound velocity with depth. Therefore, the resulting sound speed profile has a minimum at about 1200 m depth. The variations in sound velocity with depth have little effect on vertical or near-vertical transmissions of sound. However, the sound velocity profile and the minimum or *sound channel* have important implications for horizontal or near-horizontal transmission of sound.

Sound traveling from a point source will spread out spherically, so that the sound energy will decrease as the square of the distance from the source. The limit to the distance for effective sound transmission is reached when the energy density equals that of the ambient noise level (see below). If the sound energy can be confined between two vertical surfaces, then the reduction in energy is dictated by cylindrical spreading to be proportional to the distance. Under those circumstances, sound will travel further than in the point source case.

Sound travels along paths or *rays* whose orientation θ (angle from vertical) is dictated by sound speed C differences in two oceanic layers according to Snell's law

$$C - 1 \sin\theta_2 = C_2 \sin\theta_1 \,, \tag{5.9}$$

where the indices refer the 2 different oceanic layers.

From this relationship, it is obvious that the rays are bent or refracted toward lower velocity. Thus, a ray path from a source at the depth of the sound channel will be refracted toward the sound channel. The energy contained within a solid angle marked by four rays being emitted from the source is constant. Thus when the energy is concentrated in the sound channel, the energy density loses are dictated more by cylindrical spreading than spherical spreading. Hence, it is not surprising that large distances can be obtained by transmitting near the depth of the sound velocity minimum.

Sound transmission in the upper ocean is more difficult. Consider the case of a mixed layer where the sound velocity is controlled by pressure and increases with depth. Any sound transmitted into this region will be bent toward the surface. If this layer is above a layer with decreasing temperature which dominates over pressure, then rays penetrating into this layer will be refracted toward the bottom. This results in a shadow zone where acoustic communication or detection is impossible (Fig. 5.8).

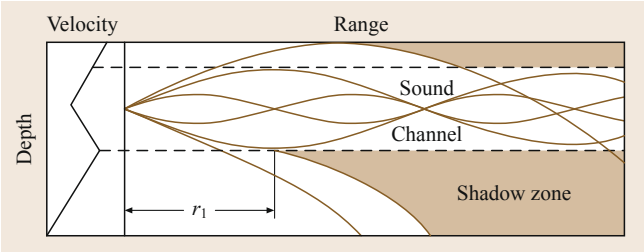


Fig. 5.8 Given the shape of the sound speed profile on the left, this is an acoustic ray trace showing sound channel and shadow zones (which is displaced from the sound source by r_1)

5.10 Acoustic Ambient Noise

The ambient noise level limits acoustic transmission and detection in the ocean. The background noise spectrum (Fig. 5.9) shows that oceanic noise levels decrease with increasing frequency. The main sources of low frequency noise (between 1 and 100 Hz) are seismic activity and explosions. At mid-frequencies (10–1000 Hz), ship noise dominates except when the noise of the rain hitting the surface becomes significant. At high frequencies (100–10 000 Hz), wind-generated noise, which depends on wind speed, dominates.

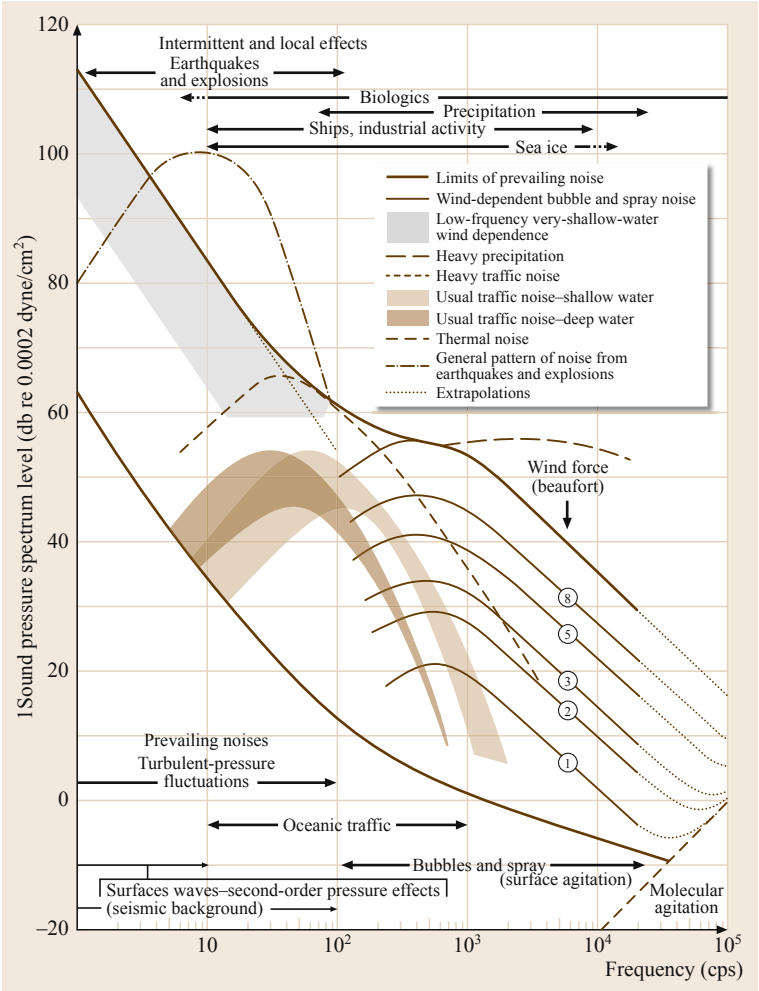


Fig. 5.9 Oceanic acoustic noise spectra

5.11 Light Transmission

Sunlight striking the ocean surface occupies a very narrow part of the spectrum of electromagnetic radiation; namely from 390 to 760 nm (or $390–760 \times 10^{-7}$ cm). The visible light part of this spectrum is defined by the familiar colors: violet, blue, green yellow, orange, and red. As light passes downward through the ocean waters, it is attenuated through absorption and scattering by water molecules, ions, and other living, dead, and inorganic suspended particles. The intensity of light at any depth $-z$ can be estimated from computations using Beer's law in the following form,

$$I_z = I_0 e^{+kz} ,$$

where I_0 is the intensity of light at the ocean's surface and k is the attenuation coefficient, which depends on the clarity of the water. For example in the open ocean, about 50% of the entering light is extinguished by a depth of 10 m and nearly 100% by a depth of 100 m (Fig. 5.10a). In typical coastal waters, nearly 100% of the light is extinguished by a depth of 10 m. However, the rate that light is attenuated differs with its color (i. e., wavelength). For example, in Fig. 5.10b, the longer wavelength red light is clearly attenuated more rapidly than the blue-violets.

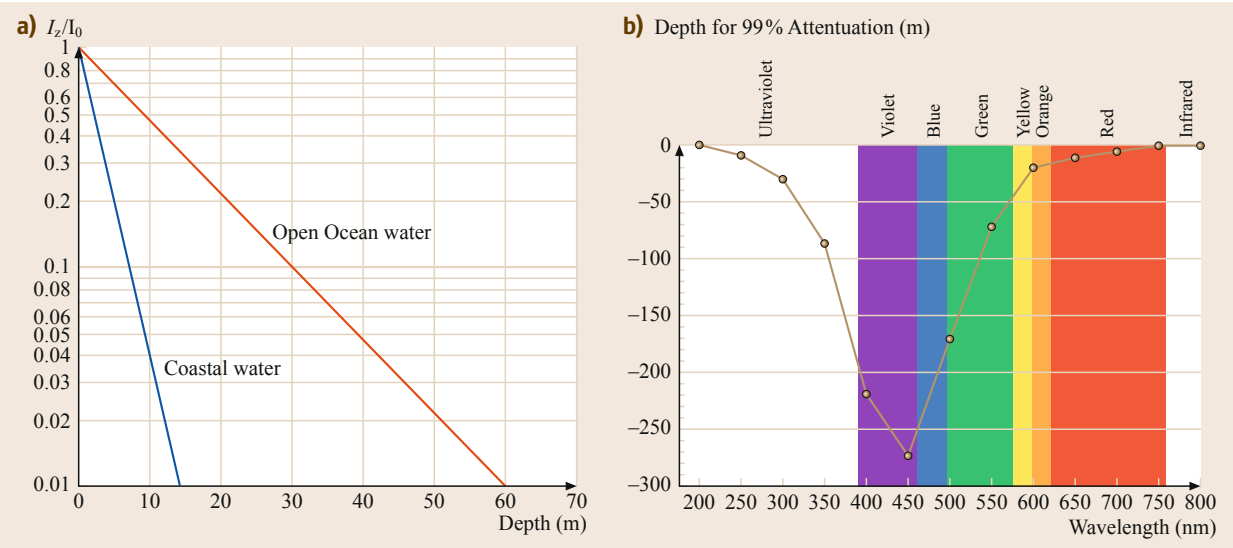


Fig. 5.10 **(a)** Depth versus light attenuation ratio I_z/I_0 for typical open ocean and coastal waters. **(b)** The open ocean 99% attenuation depths of various electromagnetic radiation wavelengths in which visible light color is indicated

References

5.1 P.M. Sanders, N.P. Fofonoff: Conversion of pressure to depth in the ocean, Deep-Sea Res. **23**, 109–111 (1976)

5.2 G.L. Pickard: *Descriptive Physical Oceanography* (Pergamon, Oxford 1975)

5.3 K.W. Mackenzie: A decade of experience with velocimeters, J. Acoust. Soc. Amer. **50**, 1321–1333 (1971)

5.4 L.D. Talley, G.L. Pickard, W.J. Emery, J.H. Swift: *Descriptive Physical Oceanography – An Introduction*, 6th edn. (Academic, London 2011)

5.5 UNESCO: *10th Report of the Joint Panel on Oceanographic Tables and Standards*, Technical Papers in Marine Science, No. 36 (United Nations Educational, Scientific and Cultural Oranisation, Paris, 1981)

5.6 J.A. Knauss: *Introduction to Physical Oceanography*, 2nd edn. (Waveland, Long Grove 2005)

5.7 F.J. Millero, G. Perron, J.F. Desnoyers: Heat capacity of seawater solutions from 5° to 35°C and 0.05 to 22 ‰ chlorinity, J. Geophys. Res. **78**(21), 4499–4506 (1973)

5.8 UNESCO, *Algorithms for computation of fundamental properties of seawater*, Technical Papers in Marine Science, No. 44 (United Nations Educational, Scientific and Cultural Oranisation, Paris, 1983)

5.9 F.J. Millero, W.H. Leung: The thermodynamics of seawater at one atmosphere, Am. J. Sci. **276**, 1035–1077 (1976)

5.10 C.T. Chen, F.J. Millero: Speed of sound in seawater at high pressures, J. Acoust. Soc. Amer. **62**(5), 1129–1135 (1977)

6. Principles of Marine Corrosion

Robert E. Melchers

Seawater is an aggressive environment for materials particularly in warmer and polluted waters. In this environment, steel infrastructure, such as offshore platforms, pipelines, and tanks may corrode quickly, and for engineering and management purposes, the rate at which this occurs is of interest. Here, recently developed understanding of the corrosion processes over extended exposure periods is reviewed and recently developed science-based mathematical models for the prediction of corrosion and pitting of structural steels are described. The effect of various influencing factors is reviewed, including that of seawater quality and the effect of microorganisms on corrosion. Also, short discussions of the corrosion of stainless steels, aluminum, and copper-nickels, all materials that are used under marine exposure conditions, are given. This is followed by a description of recent observations of the corrosion of steel reinforcing bars in reinforced concrete, another widely used construction material in marine environments. It shows that reinforced concrete can be a highly successful construction material, even in long term and under marine conditions, but that severe loss of reinforcement may occur without obvious external signs detectable by visual inspection.

6.1 Chemical and Physical Composition of Seawater 111
6.2 Materials Used in Marine Environments .. 113
6.3 Marine Corrosion of Steel 113
6.4 Modeling Longer Term Corrosion of Steel 116
6.5 Other Influences on Steel Corrosion 118
6.6 Pitting Corrosion of Steel 119
6.7 Some Other Important Materials 121
6.7.1 Stainless Steel 121
6.7.2 Aluminum 121
6.7.3 Copper-Nickels 121
6.7.4 Reinforced Concrete 121
6.8 Conclusion 123
References 123

Seawater is a hostile environment – not only for people. Often it is considered to be the most severe environment to which materials can be exposed. Warmer waters accelerate the rate of corrosion, and warmer climate seawater is particularly aggressive. This was found to be a major problem during the later part of the Pacific campaigns in the Second World War when military equipment was found to corrode unexpectedly quickly in the Tropics [6.1]. However, even in colder waters, corrosion can be severe, as described in the following.

6.1 Chemical and Physical Composition of Seawater

On average, both the chemical and physical compositions of seawater are surprisingly uniform in the oceans around the world (Table 6.1). There are also similar small variations with depth, even for very deep waters [6.3]. The annual variation in water temperature (which typically lags about 3 months behind the seasons) is largely responsible for cyclic changes in many parameters, such as dissolved oxygen (DO) in seawater. It is also responsible in the variation in biological components, including the microbiological components, such as algae, archaea, and bacteria [6.4].

In enclosed seas and estuaries, and generally closer to shore, there are usually deviations in average water quality from those in open sea conditions. This applies particularly to nutrients, such as organic carbon, nitrogenous materials, phosphates, and sulfates. Typically, these variations are caused, in the natural environment, by natural, usually cyclic, annual processes

Table 6.1 Typical major components and physical properties of seawater in selected oceans

Salinity (predominantly Na^+ and Cl^-) [ppt]	34.4 (range mainly 31–38%)
Density [mg/l]	1.025 (average, at surface) 1.05 or more at depth
pH	8.1 typical (range 7.5–8.4)
Ions	**Weight [ppt]**
Chlorine (Cl^-)	18.98
Sulfate (SO_4^-)	2.65
Magnesium (Mg^+)	1.272
Bicarbonate (HCO_3^-)	0.14
Calcium (Ca^{2+})	0.40
Potassium (K^+)	0.38
Bromine	0.067
Carbon	0.028
Dissolved gasses	**Weight [ppm]**
Oxygen (O_2)	
At surface	7.5 (approx.)
−500 m	1.5
−2000 m	3.0
Carbon dioxide (CO_2)	
At surface	1.5
−500 m	4.0
−2000 m	6.0

Many of these are functions of water temperature and/or pressure (after [6.2])

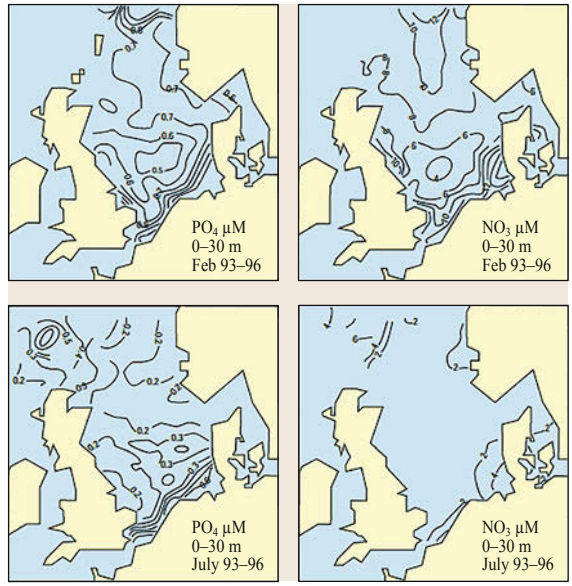


Fig. 6.1 Example of the summer–winter variations in phosphates (PO_4^-) and ammonia (NO_3^-) in the Greater North Sea (after [6.10])

involving marine growth and death [6.3]. However, in some locations, there may be a substantial discharge of nutrients or fertilizers or sewage effluents. Examples include highly populated areas or cities, where adjacent coastal water quality may be lower than that in open sea conditions. In particular, water quality can show elevated concentrations of nutrients. For example, the quality of seawater in the North Atlantic and that of the adjacent North Sea is considerably different from the usual open ocean conditions [6.5]. Figure 6.1 shows an example of the summer–winter variations in dissolved inorganic nitrogen (which in practice is mainly nitrate NO_3^-) and phosphates (PO_4^-). Largely these variations are caused by discharge, from the major rivers in that part of Europe, of sewage effluent and fertilizers (typically rich in N and P) and also some industrial wastes, although all such discharges increasingly have become subject to considerable restrictions [6.6]. Elevated nutrient levels have also been associated with areas usually considered to be almost pristine, such as Florida Keys [6.7] and the Great Barrier Reef [6.8]. Perhaps less surprising is that the elevated nutrient levels have been noted for offshore oil production facilities [6.9].

Elevated nutrient levels are seldom such that they are noticeable in a superficial examination of the water, such as, for example, at a bathing beach, although there may be some possible health effects for humans. Usually, the concentration levels of nutrients are quite low, and they change the overall chemical water properties only very slightly and, thus, have negligible effect on ordinary corrosion kinetics governed by electro chemical reactions. However, elevated nutrient levels can have major effects on living organisms and thus can have a major effect on marine growth and microbiological activity [6.11]. This applies for seawater in oceans, seas, and estuaries, and also in seawater harbors. Elevated nutrient levels relative to ambient can cause unpleasant effects, such as algal blooms in estuaries and unexpected fouling of ships and harbor facilities.

One example of the importance of nutrients in marine microbiology is the experiment in which iron filings were discharged from a vessel in the Pacific Ocean [6.12]. In a very short time, the seawater around the boat turned bright red, the result of the almost immediate proliferation of algae. This illustrates the critical importance even of a minor nutrient, in this case iron. Although iron is a requirement for biological activity, it usually lacks in seawater. The experiment also showed the fast response time of microbiology to changing conditions. For example, bacteria can double in numbers within a few hours [6.13]. From this follows the conclusion that simply counting bacteria at any point in time, as has been a traditional measure of microbiological activity, is largely irrelevant. The critical issues are: (i) the availability of sufficient energy for mi-

crobiological activity and (ii) the availability of essential nutrients to permit microbiological activity to occur. As will be discussed further, nutrients and thus microbiological activity can also have an important effect on the deterioration of metals in seawater and in freshwater.

The component of seawater most often associated with its (deleterious) effect on materials is salinity. Again, apart from the Baltic (where salinity tends to be lower than average) and the Dead Sea (where it is very high), salinity is remarkably similar in open ocean waters. Typically, salinity is reduced in estuaries and harbors, as a result of the inflow of river water. In chemistry, common salt (NaCl) is classed as one of the salts supplying chloride ions (Cl^-). It is one of the halides, a small group of chemicals that usually act as aggressive ions in chemical reactions. An often overlooked property of common salt is that it is hygroscopic, meaning that it attracts water, as is easily verified by attempting to dry out a seawater-soaked garment, such as a T-shirt without washing-out the salt. In other than dehumidified air, it will not dry. This means that once a metal object has been exposed to seawater, or has been sprayed with salt particles, and then not washed, it will attract moisture and thus tend to corrode at a much greater rate than otherwise would be the case, simply because the surface remains *wet* for longer. This is particularly important for the atmospheric corrosion of steel in coastal environments.

The other important component of seawater is its calcium carbonate content. It is supersaturated with it. For the usual pH of seawater (about 8.2), it occurs as calcium bicarbonate. Seashells (largely composed of calcium carbonate) exist because they are able to maintain their pH at just a sufficiently high level to avoid dissolution. The precise mechanisms involved are not of concern here, but it is known that an increase in carbon dioxide in the atmosphere will cause the pH of seawater to drop slightly, thereby upsetting the calcium carbonate balance in the sea, with, as is well known, potentially serious effects on seashells and molluscs. This is one of the well-recognized concerns associated with climate change. As will become clear in the following, it also can have a significant effect on marine corrosion of metals such as steels.

6.2 Materials Used in Marine Environments

Natural constructional materials are restricted largely to rocks, some pure metals (e.g., gold), many oxidized metals (e.g., iron ore), and various plant forms, including wood. However, there is an incredibly large (and growing) range of man-made materials available for commercial and industrial use. They include steels of various types including chromium steels, reinforced concrete, aluminum, glass and carbon fiber composites, nylon and other man-made organic materials. To a greater or lesser extent all of these are used in marine environments. Which material actually is used, or which combinations, depends largely on the economics involved and often this includes the first cost and, increasingly, also the anticipated maintenance and replacement costs. In some cases, it may include the consequential costs should failure occur. In practice, the materials most used in major marine infrastructure applications are steel and reinforced concrete. For mechanical equipment such as harbor facilities and shipping steel is predominant, although aluminum and fiber composites also are used for many applications. Herein attention is focused on steel structures.

Good quality information and prediction tools for likely long-term performance are scarce. Comprehensive compendia of the corrosion performance of many metals and alloys and for many nonmetals are available [6.14, 15] but the approach is largely anecdotal. Until recently this was also the case for steels, as described in more detail in the following. For reinforced concrete, the principles governing the corrosion of steel reinforcement and the performance of the concrete itself in marine environments are established and readily available [6.16, 17], but there are some aspects not yet fully explained, as also described in more detail in the following.

6.3 Marine Corrosion of Steel

Steel is used extensively in industrial, institutional, and commercial projects, and in structural systems, and has a long history of such uses. Applications include offshore structures (floating and fixed), commercial ships, defense force and commercial ships, mooring chain, submerged pipelines, coastal jetties, sheet and other piling, cranes and other loading and offloading facilities, storage tanks, bridges, multistory car-parks, electric-

Table 6.2 Typical steel properties

	Composition [%]									Mechanical Properties			
Steel type	Description	C	Mn	S max	P max	Si	Cu	Ni	Cr	Fsy [ksi]	Fsy [MPa]	Fult [ksi]	Fult [MPa]
Carbon steel A36	General purpose structural	0.26	< 1.65	< 0.05	< 0.04	< 0.6	< 0.6			36	250	58–80	400–550
HSLA A992	Ditto	0.23	0.5–1.5	0.045	0.035	0.4	0.6	0.45	< 0.35	50–65	340–450	65	450
HSLA A588	Structural weathering steel	0.15–0.2	0.75–1.35	0.05	0.04	0.15–0.4	0.2–0.5	< 0.5	0.3–7	42–50	290–350	63–70	435–480

Fsy = yield strength; Fult = ultimate strength; HSLA = high strength low alloy; Conversion: 1 MPa = 0.145 ksi (after [6.18])

ity transmission towers, and industrial and commercial buildings. Mostly these are constructed from mild or low alloy structural steels. Compared to more specialized steels, they are relatively cost-effective and readily available. Table 6.2 shows the typical chemical composition and physical properties of some steel types used in industrial infrastructure.

To ensure steel structures are sufficiently safe against collapse or loss of serviceability, much attention has been given to understanding and defining the capability of steel structures to resist various loading conditions. The results have been formulated into mathematical models and design requirements. Mostly these are set out in design codes and similar documents, increasingly at an international level. In the main the rules in the design codes are based on probabilistic concepts. This recognizes that perfect safety is unattainable in practice and that some element of risk always remains, no matter how well understood the materials and the loading conditions are [6.19].

Components of steel infrastructure may fail in a number of ways, including in tension or in compression, or by buckling, fatigue, corrosion or wear or some combination. Of these, tension, compression, buckling failure modes, and fatigue have received much research attention over a long period of time and comprehensive design rules now exist [6.21]. Fatigue typically displays a significant level of uncertainty and as a result the factors of safety against failure by fatigue tend to be higher than for most other failure modes. In addition, for critical structures, such as high-rise cranes and certain offshore structures, periodic in-situ inspection is warranted. Less well understood are stress corrosion cracking and wear. This is reflected in the wide variations in observed effects and the poor prediction capabilities of existing analytical models.

Corrosion, the subject of this chapter, often is considered well understood, at least from a scientific perspective. But design rules for practical implementation are not readily available, although, as described in the following, some progress has been made recently. In particular, models based on sound scientific principles are now available for the progression of corrosion loss and of corrosion pitting. These are described in Sect. 6.4. First, it is necessary to review some fundamental aspects of the marine corrosion of steel.

Many practical applications in marine environments ensure that steels are protected in some way. By far the most common methods are cathodic protection by impressed current or by sacrificial anodes and the use of protective coatings such as paints. When properly applied and if well maintained, these systems often perform very well, and under these conditions, the progression of corrosion of steel is not of a significant concern. However, in practice, protective coatings are not always applied correctly or with sufficient care or, more typically, are not maintained. Similarly, cathodic protection systems are sometimes neglected. In other cases, these protective measures are not technically feasible, or are deemed uneconomic. As a result, infrastructure located in or near seawater invariably shows some level of corrosion and perhaps structural damage resulting from corrosion.

For the design of new steel structures likely to be subject to corrosive conditions, the usual approach is to add a nominal sacrificial corrosion allowance on top of the minimum thickness required by structural design considerations. This is meant to allow for expected future loss of material. For example, shipping Classification Society rules for commercial vessels typically have a 10% plate thickness allowance for *wastage*. Once this is reached, the element with this degree of corrosion loss must be replaced, usually at considerable cost. In other cases, such as in holds of bulk carrier ships and for chains, surface coatings are ineffective or insufficiently robust and cathodic protection is problematic. Only a sacrificial corrosion allowance is used. Particularly in these cases, there is considerable interest in the progression of corrosion loss and of pitting with time and as a function of various influencing factors.

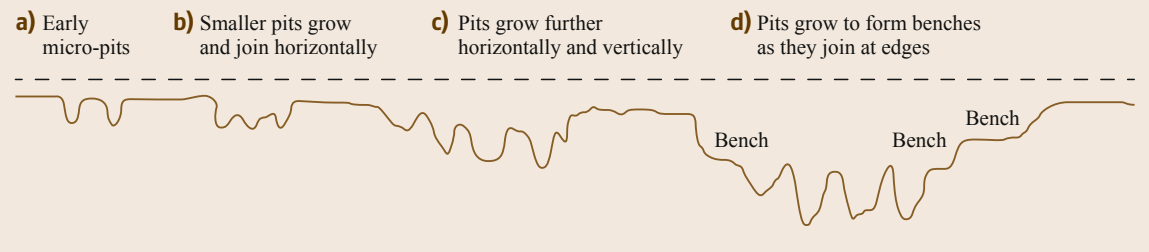


Fig. 6.2a–d Development of pitting as a function of time, showing initial pitting brooding out to form a rough plateau on which new pits then form, giving eventually a series of stepped pit plateaus (after [6.20])

Corrosion of steel invariably produces rusts. Usually, they are much more bulky than the steel lost. They also tend to hide the state of the remaining metal underneath and makes assessment of the state of the remaining steel more difficult. Commonly, ultrasonic measurements at randomly selected spot locations are used to estimate the thickness of the remaining steel. Other techniques based on electrochemistry may also be suitable in some circumstances. However, all techniques are expensive, invasive, and do not provide estimates of the rate of corrosion, or the likely corrosion loss in the future. As shown in below, simple linear extrapolation is usually very conservative.

Information about the corrosion likely to occur under given conditions in the future can be gleaned from corrosion handbooks [6.14, 15] and in some specific application cases estimated using models of the relationship between corrosion loss and time and the various factors that influence corrosion. In the main, such models are empirical and have a high degree of uncertainty, despite considerable efforts to obtain relevant data. Unfortunately, many corrosion texts describe corrosion loss or pit depth observations obtained under accelerated (electro-chemical) conditions in laboratory experiments using artificial seawater. Although this may add an insight into fundamental corrosion processes and conditions, the results do not relate well to field experience [6.22]. How then does steel actually corrode in seawater?

The corrosion of steel of interest under ocean conditions is corrosion in *wet* environments. It can occur only in the presence of water and requires oxygen or some other electron acceptor. It is distinguished from a purely chemical reaction in that it usually involves the physical transport of electrons from one site (anode) to another (cathode) with parts of the chemical reaction occurring at each site. Of course, the sites may be very close to each other (nanometers) or meters apart, provided there is electrical conductivity between the sites. Water provides such an electrolyte but steel itself has much greater conductivity. The initiation of corrosion invariably is by pitting at a micro-scale, set up by very small differences in the surface topography and grain structure of a steel surface. This permits very small (electrical) potential differences to develop and eventually lead to pit initiation [6.23, 24].

In real seawater, invariably there will be a colonization also by biofilms and very soon afterward also by bacteria and various other microorganisms present in seawater. Some microorganisms, such as the sulfate reducing bacteria, can produce hydrogen sulfide, known to be very corrosive for steels and typically also producing localized corrosion or pitting [6.4, 24]. The net result is that within hours of being exposed to seawater a steel surface begins to form very small pits. Many, if not most, of these very small early pits stop growing soon after formation and are essentially overtaken by the others. These continue to grow in depth for some time (days, weeks) but then slow down and grow mainly in width. It follows that microscopic examination of a corroded steel surface invariably reveals a complex mix of larger and smaller pits as well as unaffected regions (cathodes), at least for some time. Figure 6.2 shows a schematic view of the development of pitting with increased exposure time. In particular, it shows that the initial major pits stop growing in depth but amalgamate to form shallow depressions and that later new pitting develops on the depression surfaces. The net result is that a series of depressions are formed and that there is always a range of pit depths and sizes. Importantly, unlike the conventional wisdom about pit growth, it is not a continuous single function process. Figure 6.3 shows some microscopic photographs (at the same scale) of the progression of pitting [6.20].

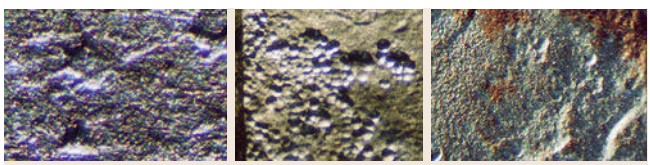


Fig. 6.3 Successive views of pitted surface of steel coupons (after [6.20])

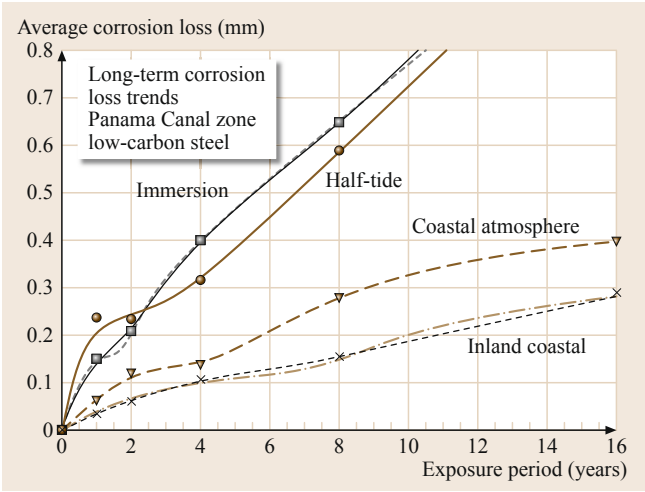


Fig. 6.4 Corrosion as measured by mass loss for different exposure zones in the Panama Canal zone, based on reported data with trends curves added (after [6.25, 26])

It is thus seen that the so-called *uniform* corrosion is an erroneous but still a convenient concept, particularly for engineering design. In practice, it is obtained usually from the differences in mass loss obtained from coupons exposed for different periods of time. Figure 6.4 shows some typical curves for the development of mass loss with time, for four different exposure environments.

Sometimes other types of corrosion are also defined, but mostly they are special cases of pitting. For example, crevice corrosion can be considered a special case of pitting corrosion that occurs in sharp, narrow deviations from the surface, but that involves the same fundamental mechanisms involved in pitting corrosion [6.23]. Another type of corrosion is *galvanic* corrosion, indicating that it is caused by a difference in the electrochemical potential of different materials (such as galvanizing zinc compared to steel). Again, this is a special, macro-level, case of the potential differences observed at the micro-level on metal surfaces and involving slight differences in composition or grain structure or both.

Table 6.3 Selection of factors known to influence corrosion

Factor E_i	Importance
Bacteria	Very high, long term
Biomass	Likely low
Oxygen supply	Very high, shorter term
Carbon dioxide	Low
Salinity	None by itself
pH	High
Carbonate solubility	Low
Pollutants	Varies
Temperature	Very high
Pressure	None
Suspended solids	None
Wave action	High
Water velocity	High

The factors that often are considered to influence the rate at which corrosion occurs in seawaters are summarized in Table 6.3, together with subjective estimates of their impact on corrosion and pitting.

6.4 Modeling Longer Term Corrosion of Steel

The many factors in Table 6.3 make science-based modeling complex, since the relationships between the factors and corrosion are not all known with a high degree of certainty. A more practical approach is to restrict data sources and factors to the main influences. Thus, model development might be restricted initially to a seemingly simple case – steel immersed relatively close to the surface of unpolluted, protected coastal seawater. Under these conditions, full aeration could be expected and all factors other than water temperature and bacteria can be eliminated from initial consideration. Using basic concepts drawn from accepted corrosion science theory, it was then hypothesized that when a steel plate (say) is immersed in seawater it will be subjected to immediate oxidation using oxygen drawn from the water in the immediate surroundings, thereby setting up an oxygen concentration gradient. As this develops, the rate of oxygen supply will reduce slowly but a more noticeable effect will begin once the rust products start to build up. Eventually, these will be such that the rate of diffusion of oxygen through them will become very slow. This gives rise to a gradually reducing rate of corrosion. These two processes may be idealized as sequential and can be represented as phases 1 and 2, respectively, in the model shown in Fig. 6.5. Phase 0 of the model refers mainly to the processes that occur immediately on immersion. This includes pit initiation and colonization of the metal surface by biofilm materials and bacteria.

Figure 6.5 shows that in phase 2, there is a gradually declining rate of corrosion and that after time, shown as t_a, further corrosion increases at a considerable rate (phase 3) which then slowly declines to a near steady state (phase 4). The rationalization for phase 3 lies in

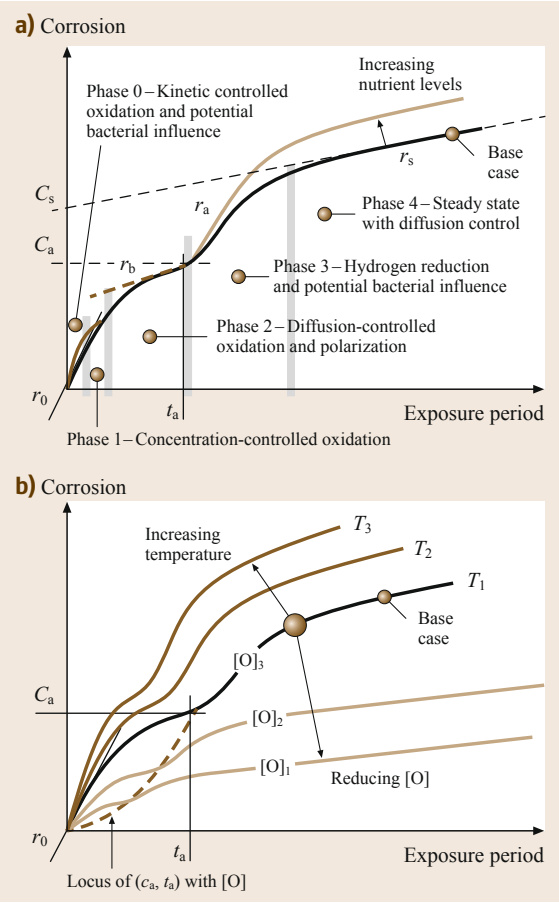


Fig. 6.5a,b Corrosion loss as a function of exposure time, showing also the effect of nutrients on microbiologically influenced corrosion **(a)**, of water temperature and of reduced oxygen concentration in seawater **(b)** (after [6.27])

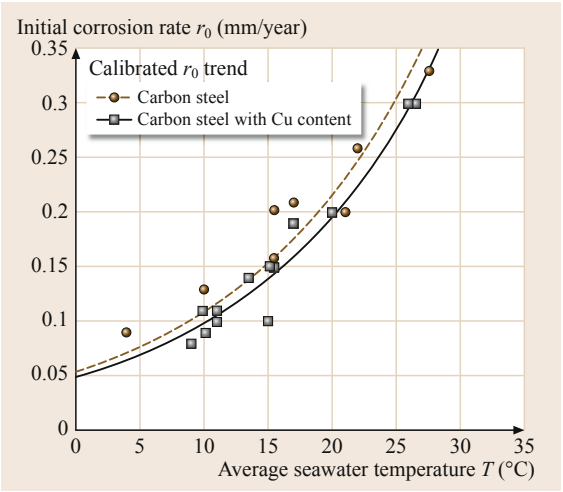


Fig. 6.6 Initial corrosion rate r_0 as a function of mean seawater temperature for carbon steel and for copper-bearing carbon steel

two components – the autocatalytic corrosion processes that form crevices and pits under substantial rust deposits and that occur within anoxic local regions, and the development of microbiological activity and microbiologically influenced corrosion in anoxic regions under anaerobic conditions. Both were hypothesized from the observed behavior of steel corrosion (Fig. 6.4) and verified by the observations of substantial pitting and the presence of bacteria, both under well-developed rusts [6.28]. The fact that pitting occurs also under sterile or clean water corrosion conditions supports the occurrence of two related but independent processes – abiotic and biotic corrosion.

The basic model of Fig. 6.5 must be adjusted to account for the effects summarized in Table 6.2. Figure 6.5 shows, schematically, some of these effects [6.27]. Generally, an increased availability of nutrients will increase the microbiological activity and hence corrosion losses and pitting severity. The effect of seawater temperature, is shown schematically in Fig. 6.5a based on average water temperatures in the range 4–28 °C. It is evident that as the water temperature increases, t_a decreases and that phases 0–2 become shorter and show less corrosion in the interval $0 - t_a$. At higher water temperatures the later phases, 3 and 4, are dominant. This shows the complexity of the effect of water temperature on corrosion. It is not simply a matter of considering the Arrhenius relationship – the latter applies strictly only to the initial corrosion rate r_0. Similarly, lower concentrations of oxygen in the seawater will reduce corrosion losses (Fig. 6.5). The effect of other influences is summarized in Table 6.4.

Mathematical formulations have been given for the most important phases in Fig. 6.5. Further details are available in the technical literature [6.29, 30]. More importantly, the conceptual model of Fig. 6.5 has been calibrated to a multitude of field observations reported by a considerable number of authors from quite separated (and essentially independent) field investigations. Table 6.5 summarizes the functional relationships that have been established for the model parameters. As an example, the calibration curves for the initial corrosion rate r_0 as a function of average seawater temperature are shown in Fig. 6.6 [6.31].

Figure 6.5 shows that it is misleading to use the initial corrosion rate r_0 for long-term corrosion. Unfortunately, short-term corrosion tests, often used for estimating longer term corrosion, typically estimate

Table 6.4 Some factors influencing seawater corrosion of steel

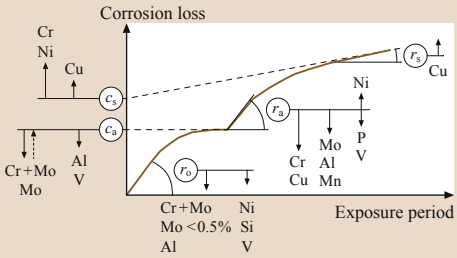
Factor E_i	Importance
Water velocity	Tends to increase corrosion in the first few weeks as rusts build up, pushing standard function upward and thereafter little additional effect
Composition (shows effect on model parameters)	
Salinity	Important but effect also influenced by water hardness. Some soft fresh waters are highly corrosive
Timing	Steel exposed, say in autumn or winter has initially a slower rate of corrosion. This pushes model (to the right) in time by about 6 months
Pressure	No evidence that increased water pressure has any noticeable effect on corrosion. Any depth effects are due primarily to temperature, DO and nutrient levels
Depth	Any depth effects are due primarily to temperature, DO and nutrient levels
Size/area	Precise shape or area involved in corrosion not discernible as important variable, for moderate-sized objects, i. e., with distances insufficient for the water properties to change

Table 6.5 Values for model parameters based on calibration to field data (after [6.30, 31])

Model parameter	General corrosion
r_0 [mm/year]	$r_0 = 0.076\exp(0.054T)$
c_a [mm]	$c_a = 0.32\exp(-0.038T)$
t_a [year]	$t_a = 6.61\exp(-0.088T)$
r_a [mm/year]	$r_a = 0.066\exp(0.061T)$
c_s [mm]	$c_s = 0.141 - 0.00133\,T$
r_s [mm/year]	$r_s = 0.039\exp(0.0254\,T)$

only r_0, and this clearly is not indicative of the longer term behavior. From a practical perspective, the most important parts of the model of Fig. 6.5 are the short-term corrosion rate r_0 and the long-term rate r_s together with its intercept c_s on the vertical (corrosion loss) axis. Table 6.5 gives the relevant values as a function of average seawater temperature. Similar curves and parameters are available for pitting corrosion of steel in seawaters [6.32].

6.5 Other Influences on Steel Corrosion

The size and orientation of the steel being corroded can be important. Most information about the corrosion of steels has been obtained from coupons, that is, relatively small pieces of flat steel exposed in the marine environment. It is known that in immersion and tidal conditions, the size of the sample is not critical to measuring mass loss (general corrosion), but this is not necessarily the case for atmospheric corrosion. Also, long steel strips (or sheet piling or similar) extending upward from the immersion zone, through the tidal zone and into the splash zone or even the atmosphere zone, will show corrosion characteristics quite different to individual, electrically isolated, coupons extending over an identical vertical range. Figure 6.7 shows a typical plot for mass loss for strips and coupons, standardized to an equivalent coupon size. It is clear that while corrosion in the tidal zone is high for coupons, it is quite low for strips. This demonstrates the importance of the orientation of steel equipment relative to the tidal zone. It also shows that electrical connectivity can be very important.

Figure 6.7 also shows the effect of higher nutrient levels in the seawater where steel strips or piling is located. Typically, under higher levels of nutrients in natural seawater, microbiological activity is enhanced such that in the zone immediately below mean low tide there is increased corrosion. This is known as *accelerated low water corrosion*. It is a major problem in seawater harbors located in regions where high levels of seawater pollution occur [6.33], has been associated with microbially induced corrosion (MIC) [6.34, 35] and has been shown to be closely correlated with the level of dissolved inorganic nitrogen in the seawater. Typically, this is caused by fertilizers that reach the seawater through agricultural runoff or from sewage treatment plants [6.36]. Sim-

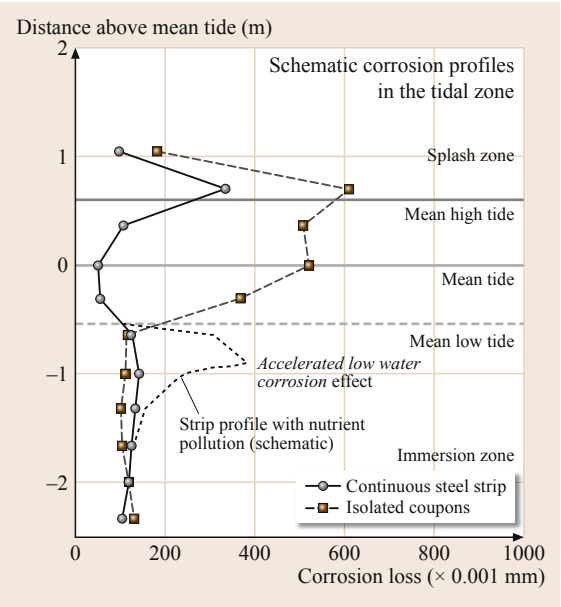


Fig. 6.7 Profile of corrosion intensity with elevation relative to mean tide, for coupons and for strips, showing for strips also the effect of elevated nutrient concentration in the seawater (based on classical data as interpreted in [6.36])

ilar problems have been noted for ships moored for extensive periods of time in polluted marine waterways [6.37].

Bacteria also have been held to be involved in corrosion of steel at considerable depths in the oceans. The most famous example is the corrosion that is occurring, albeit slowly, of the Titanic in the North Atlantic at a depth of nearly 4 km. At that depth, the dissolved oxygen concentration is very low, the water temperature is also low and there is only low water velocity. Under these conditions, conventional corrosion involving oxygen is unlikely. Instead, corrosion under localized anoxic conditions is highly likely and this then permits the possibility of microbiologically influenced corrosion with sulfate reducing bacteria (i. e., corrosion equivalent to that in phases 3 and 4 of Fig. 6.5). One possible restriction on the severity of the rate of corrosion is that MIC requires an adequate rate of nutrient supply and also an adequate rate of supply of energy. At these deep locations, it is not immediately obvious how these requirements can be met. The presence of bacteria at the corrosion sites and of rusts often associated (but not exclusively so) with MIC are not sufficient evidence to guarantee that MIC plays a major role, or even that MIC is involved at all [6.38, 39].

6.6 Pitting Corrosion of Steel

Although pitting usually is considered separately from the so-called general or uniform corrosion, the two are intimately linked. This may be seen from observations of the details of the corrosion of surfaces of mild steel coupons as a function of exposure time (Figs. 6.2 and 6.3). As noted earlier, when exposed to seawater, the steel very quickly develops localized corrosion and pitting and after only 1–2 days pits up to 100 µm deep have been observed [6.40]. Soon the deepest pits then almost stop growing in depth but tend to grow sideways and this continuous for some time. Eventually, however, when sufficient rust products have formed on the surface, localized anoxic conditions develop and the maximum pit depth increases sharply (Fig. 6.8).

Because of the close links between pitting and general corrosion, it is not surprising that Fig. 6.8 shows the same general characteristic functional form as seen for general corrosion loss as a function of time (Fig. 6.5). However, maximum pit depth in phases 3 and 4 are much greater relative to pit depths in phase 2 than the corresponding mass losses. This reflects the fact that the deepest pits occur over relatively small areas of a steel surface (Fig. 6.3). The apparent randomness of the location of the deepest pits in the early stages of corrosion is related to the apparently random na-

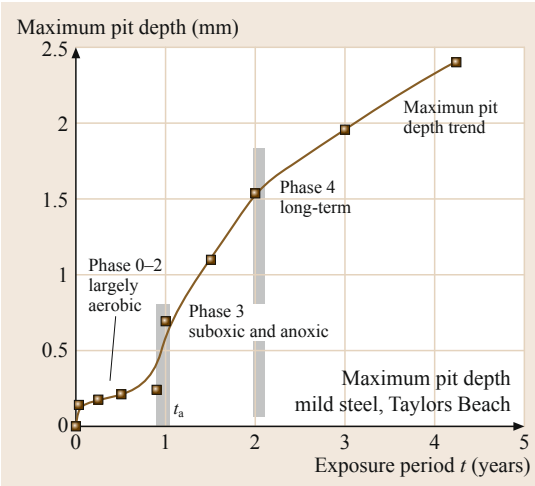


Fig. 6.8 Typical growth of pit depth with increased exposure time as inferred from observations of feel coupons exposed to coastal seawater (after [6.32])

ture of the defects in a metallic crystal structure but somewhat later it is influenced also by the cessation in the growth of some pits (meta stable pits) in favor of others [6.41]. This can be considered to influence subsequent pit growth and the development of pitting, and giving the appearance of randomness also for later pitting. However, studies of the correlation structure of larger corroded surfaces show a high degree of uniformity in pit depth with only a small proportion of greater depth than the remainder [6.42]. Although this finding tends to undermine statistical analyses (such as those based on extreme value theory) that are based on each pit depth being independent from all other pit depths, it is, from a practical perspective, not unexpected. A high degree of correlation can be expected since the corrosive environment is common to the whole steel surface and the surface itself is largely homogeneous except for the microscopic crystal structure and its associated defects. As noted, the latter become almost irrelevant as corrosion proceeds.

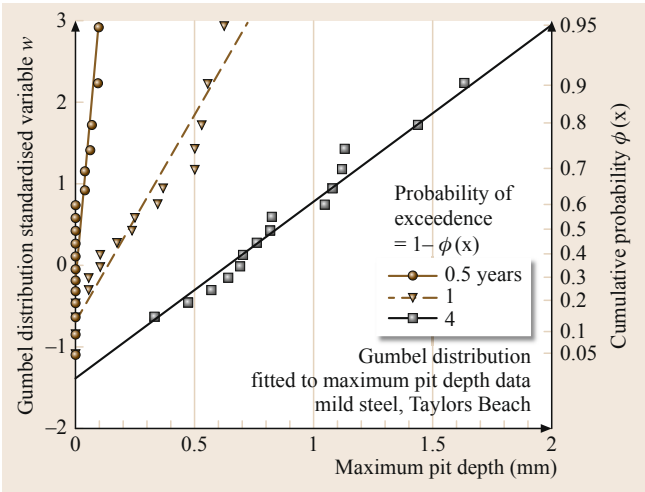


Fig. 6.9 Gumbel extreme value plot showing several sets of data for pit depth

Fig. 6.10 Localized corrosion on a stainless steel plate formed by sucker footprint of a marine mollusk (*centre*) and at one of the identification holes (2 mm diam.) (*top left*) for one quarter of a 100 × 200 mm coupon, after 9 months exposure in coastal seawater. Note that this is a stainless steel plate for which the rest of the surface has no significant corrosion (although fine pitting is widespread over the surface when examined under the microscope) and thus the effect of localized corrosion due to marine growth can be seen more clearly

Fortunately, for statistical analyses, a certain degree of dependence can be tolerated in extreme value analysis, provided it can be assumed that the maximum pit depths are asymptotically independent, that is, independence of the deepest pit depths can be assumed as the total number of pits increases [6.43]. Usually, it is assumed that the deepest pits follow the Gumbel extreme value distribution and also that the shallowest pits are of a different statistical distribution and follow an exponential distribution. The concept is shown, schematically, in Fig. 6.9. This is a Gumbel plot. Its vertical axis, representing (cumulative) probability, is deliberately distorted such that if the data were indeed truly Gumbel distributed, they would plot in a straight line (in the general direction south-west to north-east, with north at the top of the plot). Extrapolation of the upper end (right hand) of the distribution then allows the estimation of the probability of exceedence of a given depth of pit. This is important in estimating the probability of failure through perforation of the wall of a pipe or tank or other steel containment structure.

The approach represented by Fig. 6.9 and as described in many papers has a long history. It is adequate for relatively shallow pits or, similarly, for relatively short-term exposures, that is, within phases 0–2 of the model in Fig. 6.8. However, since the character of pit growth changes with the commencement of phase 3, the probabilistic properties will also change. As a result, it has been shown that for longer exposures, the extreme value probability distribution will deviate from Gumbel [6.44]. This observation is important for estimating the probability of failure due to pitting particularly for infrastructure that can pit severely. Thus, it is directly relevant, for example, for assessing the life of steel pipelines used in offshore oil installations.

As in general corrosion, microbiologically influenced corrosion can play a major role in the rate of pit growth. This can occur under biofilms and under marine growth but also, for example, under the footprint of marine organisms that settle on steel surfaces. Figure 6.10 shows an example of the localized corrosion that has occurred after 9 months exposure in seawater. The example given here is for stainless steel plate as this has negligible general or uniform corrosion in seawater (at least for shorter exposure periods) and thus shows the effect of marine organism footprints more clearly. Similar high rates of localized corrosion can oc-

cur also under corrosion products, under washers and fasteners, and under marine growth. While the general concepts are well understood, quantitative estimation of the depth and extent of pitting under these conditions awaits model development similar to that for general corrosion. This includes developing relationships between seawater nutrient content and the likely depth of the extreme pits.

6.7 Some Other Important Materials

6.7.1 Stainless Steel

Stainless steels often are considered the ideal solution to corrosion problems involving seawater conditions. This is correct for exposures in the atmosphere, particularly for marine grade stainless steel such as grade 316. In the atmosphere, the high chromium content of 316 ensures that the exterior of the steel is always covered with a highly resistant oxide film. Any steel that is exposed to the atmosphere as a result of scratches or damage to the oxide film quickly oxidizes to form more protective oxide film. Under aggressive conditions, such as near the coast, or inside swimming pool halls, a small amount of corrosion may occur, leaving the so-called *tea-staining* – small areas of brown rust stains. In the tidal zone, the oxide film is somewhat less effective and under immersion conditions, it can be quite ineffective, with long-term corrosion approaching that of mild steel. The reason is that marine biofilms and marine biological growth can inhibit the recovery of oxide film, even though the water may be highly oxygenated. Particularly, under fasteners such as bolt heads or nuts, or under marine growth (Fig. 6.10), localized or pitting corrosion can be severe [6.45]. Also, other grades of stainless steel, such as 304, contain lower amounts of chromium and usually are much less resistant to atmospheric marine corrosion than is 316. Conversely, a range of highly alloyed steels is available, ostensibly with very good corrosion resistance. In many cases, longer term experience and quantitative data are lacking, although behavior similar to Fig. 6.5 has been proposed [6.46]. At this time, the use of stainless steel under conditions where rapid oxidation may not be possible requires expert advice. Stainless steels are considerably more expensive than mild and low alloy steels.

6.7.2 Aluminum

A variety of aluminum alloys are available and these can have quite different corrosion characteristics [6.47]. In general, however, they can be summarized by noting that in air aluminum alloys form a protective oxide that, similar to that for stainless steel, is *self-healing* when damaged. As a result, aluminum tends to be very resistant to atmospheric corrosion. In seawater, it forms a noticeable gray oxide and is known to be subject to pitting. The latter is not always superficially obvious since the pits tend to undercut the metal surface [6.48]. The corrosion resistance of aluminum in freshwater environments is complicated by the fact that it is pH dependent, with lower and higher pH waters being very aggressive. Unlike steels, including stainless steel, the available evidence indicates that aluminum is not subject to microbiological attack under seawater conditions [6.49]. Again, the application of aluminum alloys requires expert advice.

6.7.3 Copper–Nickels

There is a long tradition of using copper-nickels and various brasses for fittings on naval vessels and for smaller bore pipework. The common grades are 90 : 10 and 70 : 30 copper-nickels, with the former being both less resistant to corrosion and cheaper than the latter, reflecting the cost of the nickel component. Both general corrosion and pitting tend to be relatively low, but copper-nickels are prone to microbiologically influenced corrosion, despite the fact that copper is used in antifouling treatments as a biocide. A considerable body of literature exist that deal with the practical aspects of corrosion of copper–nickel alloys for infrastructure applications [6.14, 15, 50] and with their resistance to microbiologically influenced corrosion.

6.7.4 Reinforced Concrete

Reinforced concrete structures are used in many marine locations. Local economics of material availability and the easy of construction compared to steel structures as well as maintenance issues usually are decisive in the choice. Reinforced concrete structures are generally considered to have good resistance to marine exposure conditions, largely because the steel reinforcing bars are protected to some degree by the concrete. Typically, the steel bars are located 30–50 mm or more inside the concrete. However, not all modern reinforced concrete structures exposed to marine conditions perform well, and there are many cases where the concrete has cracked along the steel bars, exposing them to sea-

Fig. 6.11 Typical cracking of concrete cover along a reinforcing bar as a result of rust build-up in a marine environment (after [6.52])

water or sea-spray and then causing severe corrosion (Fig. 6.11). Conventionally, this is considered to result from the ingress of chlorides (from seawater, or from brine where used for de-icing roads) reaching the steel bars and initiating corrosion [6.16, 17]. Much research has been carried out on this topic to estimate the rate of ingress of chlorides and to find means to eliminate or reduce the rate of chloride diffusion to the steel reinforcement bars. The conventional wisdom is that denser, less permeable concretes, greater concrete cover thickness to the bars, better quality control for both cover thickness and concrete quality and the use of various additives to reduce concrete permeability are all beneficial. Galvanizing the reinforcement bars, or coating them with a plastic (epoxy) material, or even substituting them with stainless steel also have been tried, with mixed success [6.51].

Despite the conventional wisdom, many examples are available where the rules have not been followed and yet the concrete structures have shown, over many decades, very good performance, with little or no evidence of reinforcement corrosion [6.52–54]. Importantly, this is despite evidence that in many older structures the concrete was made with seawater rather than freshwater as is mandatory under current design requirements. It should be clear that in these cases, the chloride concentration next to the steel reinforcement was at a very high level right from the beginning, as also verified by the high concentration of chlorides in many older reinforced concrete structures for which no corrosion is evident, even when the steel bars are exposed after many years [6.55, 56]. Thus, the rate of ingress of chlorides is irrelevant. If the conventional wisdom is correct, these structures should have shown serious reinforcement corrosion already quite early in their lives.

According to fundamental corrosion thermodynamics, corrosion of steel can initiate only when the pH of the solution environment immediately adjacent to the steel is sufficiently low. The pH of new concrete is around 13 as a result of the alkalis formed during the concrete setting reaction. Such pH is too high for corrosion initiation despite the high concentration of chlorides. In fact, the chloride/alkali ratio has been recognized as the important criterion for initiation [6.57]. It follows that in the cases where corrosion has not occurred, even with high chloride levels, the alkali content has remained high. It follows also that corrosion initiation in this case requires reduction in alkalis. Some leach out relatively easily but the main alkali, calcium hydroxide, formed by the cement reaction, is considered to be relatively insoluble. However, there is evidence that it leaches out much more quickly in the presence of chlorides [6.58, 59]. Taken to its logical conclusion, these observations imply that chlorides facilitate alkali leaching and thus alkali reduction at the reinforcing bars rather than directly causing corrosion initiation. It follows also that reinforcement corrosion can be delayed simply by ensuring a very high reserve of alkalis (i. e., increased *alkalinity* or *buffering* capacity) in the concrete. This correlates well with a study of older reinforced concrete structures that showed that those made with alkali-rich aggregates (including finer aggregates and sand) showed longer times before corrosion commenced and longer service lives [6.53]. Of course, any reduction in the rate of ingress of chlorides, if not already present, will also delay the rate of alkali leaching and, therefore, be beneficial for durability.

Fig. 6.12 Interior of a concrete element broken open after some 65 years exposure to the North Sea, showing the almost complete loss of the reinforcing bar inside the concrete (at *top*). All that remains to show where the bar had been is the thin black line. Note the watery rust stains and also the severe local corrosion (lower center) of the reinforcing bar that is almost uncorroded only a short distance away (at *right*)

Fig. 6.13 Severely corroded end of a 6 mm diameter reinforcing bar shown in Fig. 6.12

The marine environment is, nevertheless, severe for concrete structures. Under suitable conditions, the chlorides in seawater can combine with the ferrous ions released by the steel reinforcement in the corrosion process to produce highly soluble ferrous chlorides ($FeCl_2$), and with the sufficient availability of oxygen, very soluble ferric chloride ions ($FeCl_3$). These can leach out of concretes very readily and leave merely brown watery-looking rust stains rather than the usual build-up of solid rust products responsible for cracking, such as shown in Fig. 6.11. The corrosion associated with $FeCl_2$ production can be very severe. Figure 6.12 shows a case where apparently sound concrete, some 65 years old, without any visible external rust staining or cracking was broken open as part of a routine scientific investigation. It revealed almost complete loss of about 300 mm length of reinforcement bars inside the concrete. Only some magnetite (the black material) was left to show the location of the bar. Figure 6.13 shows the severely corroded end of the bar, that some small distance away, was left intact within the adjacent concrete [6.52]. Several other, similar cases have are now available, some in summary form [6.55, 56] and detailed investigations are proceeding. Finally, the marine environment can also be a problem for the concrete itself [6.16, 17], but modern practices largely have eliminated this as an issue.

These observations show that reinforced concrete can be a very successful construction material in marine environments, provided a high level of alkali material is embodied in the concrete mass, giving long-term buffering against alkali loss through leaching. However, it is possible for severe localized corrosion loss of reinforcement to occur without obvious external warning signs. This latter aspect is currently under urgent investigation as the implications for existing reinforced concrete infrastructure as significant.

6.8 Conclusion

Seawater is a complex mix of many different chemicals and microorganisms, even though it is remarkably uniform for most parts of the world. It is usually considered to be a hostile environment for materials and this is evident in the research efforts devoted to understanding the initiation of corrosion and its propagation once it starts. The present chapter has outlined some of the current research in this area, focusing mainly on engineering rather than purely scientific investigations. The research is driven largely by the needs of industry but anchored in fundamental science and the basic understanding of the processes involved. It is focused on developing tools and techniques that can be applied in the offshore, shipping, and coastal environment engineering industries. The present chapter gives an overview of this work and some of its principal findings and models.

References

6.1 C.R. Southwell, J.D. Bultman, C.W. Hummer: Estimating service life of steel in seawater. In: *Seawater Corrosion Handbook*, ed. by M. Schumacher (Noyes Data Corp., Park Ridge 1979) pp. 374–387

6.2 J.M. Wright, A. Colling, G. Bearman: *Open University: Seawater: Its Composition, Properties and Behaviour* (Butterworths-Heinemann, Oxford 1995)

6.3 H.U. Svedrup, M.W. Johnson, R.H. Fleming: *The Oceans: Their Physics, Chemistry and General Biology* (McGraw-Hill, New York 1942)

6.4 B.J. Little, J.S. Lee: *Microbiologically Induced Corrosion* (Wiley, New York 2007)

6.5 OSPAR: *Regional Quality Status Report II for the Greater North Sea* (OSPAR Comm., London 2000)

6.6 H.A. Cole: Pollution of the sea and its effects, Proc. R. Soc. B **205**(1158), 17–30 (1979)

6.7 A.M. Szmant, A. Forrester: Water column and sediment nitrogen and phosphorous distribution patterns in the Florida Keys, Coral Reefs **15**, 21–41 (1996)

6.8 M.J. Furnas, J. Brodie: Current status of nutrient levels and other water quality parameters in the great barrier reef. In: *Downstream Effects of Land Use*, ed. by H.M. Hunter, A.G. Eyles, B. Rayment (Department of Natural Resources, Queensland 1996) pp. 9–21

6.9 J.M. Odom: Industrial and environmental activities of sulfate-reducing bacteria. In: *The Sulfate-Reducing Bacteria: Contemporary Perspectives*, ed. by J.M. Odom, R. Singleton Jr. (Springer, New York 1993) pp. 189–210

6.10 G. Radach, J. Genkeler: *Gridding of the NOWESP Data Sets*, Ber. Zent. Meere. Klimatforsch. B: Ozeanogr., Vol. 27 (Inst. Meeresk., Hamburg 1997)

6.11 A.F. Carlucci: Nutrient and microbial response to nutrients in seawater. In: *Effect of Ocean Environments on Microbial Activities*, ed. by R.R. Colwell, R.Y. Morita (Univ. Park Press, Baltimore 1974)

6.12 K.H. Coale, K.S. Johnson, S.E. Fitzwater, R.M. Gordon, S. Tanner, F.P. Chavez, L. Ferioli, C. Sakamoto, P. Rogers, F. Millero, P. Steinberg, P. Nightingale, D. Cooper, W.P. Cochlan, M.R. Landry, J. Constantinou, G. Rollwagen, A. Trasvina, R. Kudela: A massive phytoplankton bloom induced by an ecosystem-scale iron fertilization experiment in the equatorial pacific ocean, Nature **383**, 495–501 (1996)

6.13 B. Austin: *Marine Microbiology* (Cambridge Univ. Press, Cambridge 1988), Chap. 5

6.14 G. Kreysa, R. Eckermann: *Dechema Corrosion Handbook*, Vol. 11 (VCH, New York 1992)
6.15 M. Schumacher (Ed.): *Seawater Corrosion Handbook* (Noyes Data Corp., Park Ridge 1979)
6.16 O.E. Gjorv: *Durability Design of Concrete Structures in Severe Environments* (Taylor Francis, London 2009)
6.17 M.G. Richardson: *Fundamentals of Durable Reinforced Concrete* (Spon Press, London 2002)
6.18 J.E. Bringas: *Handbook of Comparative World Steel Standards* (ASTM, Mayfield 2004)
6.19 R.E. Melchers: *Structural Reliability Analysis and Prediction* (Wiley, Chichester 1999)
6.20 R. Jeffrey, R.E. Melchers: The changing topography of corroding mild steel surfaces in seawater, Corros. Sci. **49**, 2270–2288 (2007)
6.21 J. Heyman: *Elements of the Theory of Structures* (Cambridge Univ. Press, Cambridge 1996)
6.22 US Office of Naval Research: *Second World Congress on Corrosion in the Military, Naples* (Trans-Tech, Pfäffikon 2007)
6.23 D. Jones: *Principles and Prevention of Corrosion*, 2nd edn. (Prentice-Hall, Upper Saddle River 1996)
6.24 H.H. Uhlig, R.W. Revie: *Corrosion and Corrosion Control*, 3rd edn. (Wiley, New York 1985)
6.25 B.W. Forgeson, C.R. Southwell, A.L. Alexander: Corrosion of metals in tropical environments. Part 3. Underwater corrosion of ten structural steels, Corrosion **16**(3), 87–96 (1960)
6.26 C.R. Southwell, B.W. Forgeson, A.L. Alexander: Corrosion of metals in tropical environments. Part 2. Atmospheric corrosion of ten structural steels, Corrosion **14**(9), 53–59 (1958)
6.27 R.E. Melchers, R. Jeffrey: The critical involvement of anaerobic bacterial activity in modelling the corrosion behaviour of mild steel in marine environments, Electrochim. Acta **54**, 80–85 (2008)
6.28 R. Jeffrey, R.E. Melchers: Bacteriological influence in the development of iron sulphide species in marine immersion environments, Corros. Sci. **45**(4), 693–714 (2003)
6.29 R.E. Melchers: Mathematical modelling of the diffusion controlled phase in marine immersion corrosion of mild steel, Corros. Sci. **45**(5), 923–940 (2003)
6.30 R.E. Melchers, P.A. Wells: Models for the anaerobic phases of marine immersion corrosion, Corros. Sci. **48**(7), 1791–1811 (2006)
6.31 R.E. Melchers: Modeling of marine immersion corrosion for mild and low alloy steels – Part 1: Phenomenological model, Corrosion **59**(4), 319–334 (2003)
6.32 R.E. Melchers: Pitting corrosion of mild steel in marine immersion environment – 1: Maximum pit depth, Corrosion **60**(9), 824–836 (2004)
6.33 J.E. Breakell, K. Foster, M. Siegwart: *Management of Accelerated Low Water Corrosion in Steel Maritime Structures*, Rep. C634 (CIRIA, London 2005)
6.34 I.B. Beech, S.A. Campbell: Accelerated low water corrosion of carbon steel in the presence of a biofilm harbouring sulphate-reducing and sulphur-oxidising bacteria recovered from a marine sediment, Electrochim. Acta **54**, 14–21 (2008)
6.35 J.M.R. Genin, A.A. Olowe, B. Resiak, N.D. Benbouzid-Rollet, S. L'Haridon, M. Confente, D. Prieur: Microbially induced corrosion of steel sheet piles in a harbour and the presence of green rust two compound. In: *Progress in the Understanding and Prevention of Corrosion*, ed. by J.M. Costa, A.D. Mercer (Inst. Met., London 1993) pp. 1177–1184
6.36 R.E. Melchers, R.J. Jeffrey: Corrosion of long vertical steels strips in the marine tidal zone and implications for ALWC, Corros. Sci. **65**, 26–36 (2012)
6.37 M.H. Peterson, L.J. Waldron: Investigation of mild steel corrosion rate in San Diego harbor, Corrosion **17**(4), 188t–190t (1961)
6.38 I.B. Beech, J. Sunner: Biocorrosion: Towards understanding interactions between biofilms and metals, Curr. Opin. Biotechnol. **15**, 181–186 (2004)
6.39 G. Cragnolino, O.H. Tuovinen: The role of sulphate-reducing and sulphur-oxidizing bacteria on the localized corrosion of iron-base alloys – A review, Int. Biodeterior. **20**(1), 9–26 (1984)
6.40 G. Butler, P. Stretton, J.G. Beynon: Initiation and growth of pits on high-purity iron and its alloys with chromium and copper in neutral chloride solutions, Br. Corros. J. **7**(7), 168–173 (1972)
6.41 Z. Szklarska-Smialowska: *Pitting Corrosion of Metals* (NACE, Houston 1986)
6.42 R.E. Melchers, M. Ahammed, J. Jeffrey, G. Simundic: Statistical characterization of corroded steel surfaces, Mar. Struct. **23**, 274–287 (2010)
6.43 J. Galambos: *The Asymptotic Theory of Extreme Order Statistics*, 2nd edn. (Krieger, Malabar 1987)
6.44 R.E. Melchers: Extreme value statistics and long-term marine pitting corrosion of steel, Probab. Eng. Mech. **23**, 482–488 (2008)
6.45 S.C. Dexter, G.Y. Gao: Effect of seawater biofilms on corrosion potential and oxygen reduction of stainless steel, Corrosion **44**(10), 717–723 (1988)
6.46 R.E. Melchers, R. Jeffrey: Modeling of long-term corrosion loss and pitting for chromium-bearing and stainless steels in seawater, Corrosion **64**(2), 143–154 (2008)
6.47 C. Vargel: *Corrosion of Aluminium* (Elsevier, London 2004)
6.48 Z. Szklarska-Smialowska: Pitting corrosion of aluminum, Corros. Sci. **41**, 1743–1767 (1999)
6.49 J.E. Amonette, C.K. Russell, K.A. Carosino, N.L. Robinson, J.T. Ho: Toxidity of Al to *Desulfovibrio desulfuricans*, Appl. Environ. Microbiol. **69**(7), 4057–4066 (2003)
6.50 R.E. Melchers: Temperature effect on seawater immersion corrosion of 90 : 10 copper-nickel alloy, Corrosion **57**(4), 440–451 (2001)
6.51 J. Tinnea, L. Li, W.H. Hartt, A.A. Sagüés, F. Pianca, B. Chandler, G. Pedersen: Corrosion in bridges and highways. In: *ASM Handbook. Volume 13C, Corrosion: Environments and Industries*, ed. by S.D. Cramer, B.S. Covino Jr. (ASM Int., Materials Park 2006) pp. 559–597
6.52 R.E. Melchers, C.Q. Li: Reinforcement corrosion in concrete exposed to the North Sea for over 60 years, Corrosion **65**(8), 554–566 (2009)

6.53 R.E. Melchers, C.Q. Li: Reinforcement corrosion initiation and activation times in concrete structures exposed to severe marine environments, Cem. Concr. Res. **39**, 1068–1076 (2009)
6.54 R.E. Melchers: Carbonates, carbonation and the durability of reinforced concrete marine structures, Aust. J. Struct. Eng. **10**(3), 215–226 (2010)
6.55 R.E. Melchers, T.M. Papé: Aspects of long-term durability of reinforced concrete structures in marine environments, Eur. J. Environ. Civ. Eng. **15**(7), 969–980 (2011)
6.56 R.E. Melchers, T.M. Pape: The durability of reinforced concrete structures in marine environments, Proc. Aust. Struct. Eng. Conf., Sydney (2012)
6.57 D.A. Hausmann: Steel corrosion in concrete, how does it occur, Mater. Perform. **6**(11), 19–23 (1967)
6.58 A.A. Sagues, E.I. Moreno, C. Andrade: Evolution of pH during in-situ leaching in small concrete cavities, Cem. Concr. Res. **27**(11), 1747–1759 (1997)
6.59 G.J. Verbeek: *Mechanisms of Corrosion of Steel in Concrete, Corrosion of Metals in Concrete, SP-49* (Am. Concr. Inst., Farmington Hills 1975) pp. 21–38

7. Hydromechanics

Karl Dietrich von Ellenrieder, Manhar R. Dhanak

Ocean engineering is a systems engineering field and requires a relatively broad background in several aspects of engineering. The role of hydrodynamics is central to all ocean engineering activities and despite an ocean engineer's specialist designation, sooner or later some aspects of hydrodynamics will encroach upon the problem he or she is trying to solve. The intent of this chapter is not to provide an exhaustive treatment of marine hydrodynamics, but instead to act as a reference source to provide enough working knowledge to solve practical problems, or at least a good idea of where to start looking. A range of topics is covered including dimensional analysis, static and dynamic flows, potential and viscous flows, laminar and turbulent flows, boundary layers, wakes, jets and shear layers, and drag and lift forces.

7.1 **Dimensional Analysis, Basic Estimation, and Model Testing** 127
7.1.1 Drag Force on a Sphere 127
7.1.2 Physical Significance of the Dimensionless Π Groups 129
7.1.3 Similitude 134
7.1.4 Skin-Friction Drag 135
7.1.5 Estimating the Drag on a Ship from Model Testing 136
7.1.6 Hydrofoil Lift and Drag 138
7.1.7 Screw Propellers 143
7.1.8 Air and Wind Resistance 146
7.1.9 Hydrodynamic Characterization of Marine Surface Vessels 146
7.2 **Fluid Statics** 153
7.2.1 Pressure Forces on Surfaces 153
7.2.2 Static Stability 154
7.3 **Hydrodynamics** 155
7.3.1 Flow Kinematics 155
7.3.2 The Navier–Stokes Equations 156
7.3.3 Flow of an Ideal Fluid 159
7.3.4 Flow of a Viscous Fluid 169
References 174

7.1 Dimensional Analysis, Basic Estimation, and Model Testing

At a fundamental level, dimensional analysis and similitude allow the development of physical laws and equations, *from-scratch* or based on observation and measurement.

Dimensional analysis is the starting point for understanding many physical phenomena. It is most often used for:

- Estimating the effects of different parameters
- Interpreting experimental data
- Defining scaling laws that can be used to predict the performance of full-sized prototypes using the data from measurements conducted on engineering models.

In hydromechanics, the use of dimensional analysis and similitude is especially important as there are a small number of *exact* solutions to the equations governing fluid flow and numerical solutions are limited to flows with fairly simple geometries or to restricted ranges of conditions.

The use of dimensional analysis might be best explained through use of a simple example.

7.1.1 Drag Force on a Sphere

Here we wish to find an estimate for the drag force D acting on a spherical body completely immersed in a flowing fluid and far from any boundaries. We might start by asking ourselves what physical properties of the flow and spherical body the drag might depend upon. One could argue that the size of the sphere is likely important and pick the diameter A of the sphere as being a characteristic or representative length scale of the body. The speed of the flow will also likely play a role and so the freestream speed U of the flow can be used as a characteristic value. One could argue that part of the

force felt on the sphere comes from the change in momentum which occurs when the flowing fluid impacts the spherical body. As the momentum is related to both the specific mass and speed of the fluid, we can add the density of the fluid ρ as another important characteristic parameter that should be taken into account (we already have speed). The viscosity of the fluid might also be expected to set up shear stresses in the flow, which creates additional drag on the spherical body. Thus, the dynamic (molecular) viscosity of the flow might also be expected to be an important parameter that should be included in our analysis. Thus, we get some sort of functional relationship of the form

$$D = f(A, U, \rho, \mu) \,. \tag{7.1}$$

In order to determine this functional relationship, one could propose performing an experiment in which the drag is measured for 10 different values of A, 10 different values of U, 10 different values of ρ, and 10 different values of μ. This process would require 10^4 experiments and would be infeasible in terms of cost and time for most! How can this process be made more tractable?

If one thinks of the functional relation in (7.1) as being an equation, D is the dependent variable and A, U, ρ, and μ are the independent variables. If we could somehow reduce the number of independent variables, we can reduce the number of experiments that would need to be performed to characterize the functional dependence in (7.1). Dimensional analysis is the process by which we will do this. However, before we embark on a description of how to use dimensional analysis, we will need to consider two basic concepts: dimensional homogeneity and the Buckingham-Π theorem.

Dimensional Homogeneity

Any equation that describes a basic law of physics must have the same dimensions (units) on every term and on both the right hand side and the left-hand side of the equation. The units of all additive terms in an equation must be the same, and the dimensions will remain homogeneous after any mathematical operation on the equations, such as differentiation or integration. The basic dimensions are mass M, length L, and time T – all physical quantities can be quantified by combinations of these units. In practice, it may be convenient to use nonfundamental units in dimensional analysis. For example in engineering, the temperature Θ is sometimes taken as a basic unit and in physics the luminosity L is often used as a basic unit of measure, although strictly speaking, temperature is a measure of energy and luminosity is a measure of power flux, both of which are composed of the more fundamental M, L, and T units. Here, we will use square brackets to denote *dimensions of*, e.g.,

$$[U] = \frac{L}{T} \,.$$

Buckingham-Π Theorem

The Buckingham-Π theorem can be used to determine the number of dimensionless groups that can be found in a functional relation, such as (7.1). If we let n be the number of variables we believe it is involved in our functional dependence and m be the number of basic physical dimensions involved in those n variables, the Buckingham-Π theorem states that $(n-m)$ dimensionless groups can be found by combining variables. The combinations of variables are referred to as Π groups and are used to re-write the sought after functional dependence as a function of Π groups.

$$\Pi_1 = \phi(\Pi_2, \Pi_3 \ldots, \Pi_{n-m})$$

Getting back to the spherical body, we see that the number of variables involved in (7.1) is $n = 5$ and the number of basic dimensions can be found by looking at the units of the five variables.

$$[D] = \frac{ML}{T^2} \quad [U] = \frac{L}{T} \quad [\mu] = \frac{M}{LT} \,,$$
$$[A] = L \quad [\rho] = \frac{M}{L^3} \,, \tag{7.2}$$

where we see that $m = 3$. Thus, we expect to be able to reduce (7.1) to a functional dependence between $n - m = 2$ Π groups

$$\Pi_1 = \phi(\Pi_2) \,, \tag{7.3}$$

which, as one might expect, will vastly reduce the number of experiments required to characterize the functional relation. There are several different approaches that can be used to identify the Π groups in (7.3). Here two approaches will be discussed.

The Step-by-Step Elimination Method [7.1]

Each variable in a functional dependence such as (7.1) is combined with other variables by multiplying or dividing it into the other variables. One starts by *eliminating* the simplest or purest variables first, meaning those with the fewest number of basic units. In general, one should not eliminate the dependent variable. Each elimination can be thought of as being the same as removing one of the basic dimensions, length L, mass M, or time T. Going back to our example, from (7.2) we can see that the variable with the simplest dimensions is the diameter A, so we begin by using A to eliminate the length dimension from our problem:

i. Eliminate length

$$\left[\frac{D}{A}\right] = \frac{M}{T^2}\,, \quad \left[\frac{U}{A}\right] = \frac{1}{T}\,,$$
$$[\rho A^3] = M\,, \quad [\mu A] = \frac{M}{T}\,.$$

We can see that the dimension of length has been removed from the combined variables thus far and that each step of the process will reduce the total number of variables by 1, until $n-m$ of them remain. We will next repeat the process for the dimensions of M and T. Looking at the dimensions of the combined variables we see that both U/A and ρA^3 have simple dimensions of only one type of unit each. Thus, we have a choice of using either set of combined variables to remove one of the remaining dimensions. Here we proceed next by eliminating time.

ii. Eliminate time

$$\left[\frac{DA}{U^2}\right] = M\,, \quad \left[\frac{\mu A^2}{U}\right] = M\,, \quad [\rho A^3] = M\,.$$

The only dimension remaining in the combined variables is now mass. Remembering that we never want to use the set of variables that contains the dependent variable, we must next choose which set of combined variables we will use to eliminate mass. This example illustrates one of the advantages of using the step-by-step elimination method over other nondimensionalization approaches – it is possible to examine the resulting combined variables at each step of the process and if it becomes clear that any of them could be reduced to one of the well-known dimensionless numbers, such as Reynolds number or Froude number, we can elect to steer the elimination process in that direction. Here it can be seen that if the set ρA^3 is used to eliminate mass, the resulting sets will be the inverse of the well-known Reynolds number and a set similar to what is known as a drag coefficient.

iii. Eliminate mass

$$\Pi_1 = \frac{D}{\rho U^2 A^2}\,, \quad \Pi_2 = \frac{\mu}{\rho U A}\,.$$

Thus, we have identified the dimensionless Π groups (combined variables) which give a simpler functional dependence (7.3) than the original (7.1). This analysis only specifies that there is a functional dependence and which groups are likely to be involved, not the actual function or form of the dependence, which must typically be determined from physical observation or experimental data. Therefore, we can equally say that perhaps $\Pi_1 = \phi(1/\Pi_2)$, or that

$$C_\mathrm{D} = \frac{D}{\frac{1}{2}\rho U^2 A^2} = \phi\left(\frac{\rho U A}{\mu}\right) = \phi(\mathrm{Re})\,, \tag{7.4}$$

where C_D is the common form of the coefficient of drag and Re is defined as the Reynolds number. A plot of the relationship between C_D and Re for a sphere can be seen in Fig. 7.1.

The Exponent Method

If we can combine the variables involved in our expected functional dependence (7.1) as a product of exponentials, the resulting equation must be dimensionally homogeneous

$$[D] = [U^a \rho^b \mu^c A^d]\,. \tag{7.5}$$

Thus, solving for the exponential terms a, b, c, d which make the relation (7.5) dimensionally homogeneous should provide the nondimensional Π groups we seek. Writing out the units in the terms of (7.5) gives

$$\frac{ML}{T^2} = \left(\frac{L}{T}\right)^a \left(\frac{M}{L^3}\right)^b \left(\frac{M}{LT}\right)^c L^d\,,$$

which results in having to solve the following system of equations so that (7.5) is dimensionally homogeneous

$$\begin{aligned} L:&\quad 1 = a - 3b - c + d \\ M:&\quad 1 = b + c \\ T:&\quad -2 = a - c\,. \end{aligned} \tag{7.6}$$

This is an underdetermined system of equations. We can solve for three of the variables in terms of a fourth. It is convenient to choose that variable which appears most often, which is c in this case. Solving the system of equations (7.6), we find that $a = 2-c$, $b = 1-c$, and $d = 2-c$. Re-writing (7.5) using these exponential values gives

$$D = \rho U^2 A^2 \left(\frac{\mu}{\rho U A}\right)^c\,. \tag{7.7}$$

Thus, we can see that (7.7) is of the functional form (7.4).

7.1.2 Physical Significance of the Dimensionless Π Groups

Π Groups can be interpreted as the ratio of typical values of two different, competing, physical effects. In

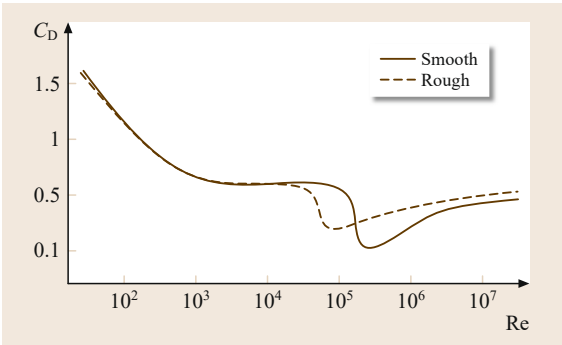


Fig. 7.1 Drag coefficient C_D vs. Reynolds number Re

order to determine which of the two physical effects might be the dominant effect in a particular situation, we use an order of magnitude analysis approach. Here, the following common rule of thumb will be used to determine the order of magnitude $\mathcal{O}$ of a Π group.

$$\begin{aligned} &\vdots \\ 0.3 \le \Pi \le 3 \quad &\mathcal{O}(1) \\ 3 \le \Pi \le 30 \quad &\mathcal{O}(10) \\ 30 \le \Pi \le 300 \quad &\mathcal{O}(100) \\ &\vdots \end{aligned}$$

In order to conduct an order of magnitude analysis one must think abstractly, for example, think of an object as having a single characteristic length scale, say L, which may be moving through a fluid moving with a characteristic velocity scale U. Generally, the velocity will vary throughout a flow, especially near objects in the flow or at the flow's boundaries. Thus, the most appropriate velocity scale is often taken to be the freestream velocity, far upstream or downstream of an object in the flow where the flow is relatively unaffected by the presence of a body.

If we approximate the forces acting on our object in an abstract way we can find the following:

Inertial force – from Newton's second law – this will vary as the product of object mass and acceleration

$$F_i = ma \approx \rho L^3 \frac{U}{T} ,$$

where T is the time it will take a particle of fluid to traverse the object $T \approx L/U$ – using this expression for T in the above gives

$$F_i \approx \rho U^2 L^2 .$$

The shear stress τ acting on an object varies as the product of dynamic viscosity of the fluid μ and the gradient of the velocity $\partial u/\partial y$ at the object's surface

$$\tau = \mu \frac{\partial u}{\partial y} .$$

When the boundary layer region around an object is relatively large, we can approximate the gradient term as

$$\frac{\partial u}{\partial y} \approx \frac{U}{L} .$$

If the surface area of the object varies as L^2, the viscous force F_v acting on the object will vary as

$$F_v \approx \mu \left(\frac{U}{L}\right) L^2 = \mu U L .$$

If an object has a pressure differential Δp acting across it, the force from this pressure F_p will vary as

$$F_p \approx \Delta p L^2 .$$

Lastly, when an object is neutrally buoyant, its weight will be equal to the weight of water that it displaces. We can approximate the force related to gravity F_g as

$$F_g \approx \rho g L^3 .$$

As mentioned above, each of the Π groups can be interpreted as the ratio of competing physical effects:

a) Reynolds number.
The Reynolds number is very important in determining when viscous effects can be neglected in a flow and for predicting (in an approximate way) when a boundary layer flow may be turbulent or laminar. The Reynolds number is the ratio of inertial to viscous forces

$$\mathrm{Re} = \frac{F_i}{F_v} \approx \frac{\rho U^2 L^2}{\mu U L} = \frac{\rho U L}{\mu} = \frac{U L}{\nu} , \qquad (7.8)$$

where $\nu \equiv \mu/\rho$ is the kinematic viscosity of the fluid.

b) Pressure coefficient.
The ratio of the pressure force to the inertial force is known as the pressure coefficient and plays a major role in the characterization of the flow over hydrofoils

$$C_p = \frac{F_p}{F_i} \approx \frac{\Delta p L^2}{\rho U^2 L^2} = \frac{\Delta p}{\rho U^2} . \qquad (7.9)$$

c) Froude number.(FR)
The ratio of the inertial to gravity forces is the square of the Froude number

$$\mathrm{Fr}^2 = \frac{F_\mathrm{i}}{F_\mathrm{g}} \approx \frac{\rho U^2 L^2}{\rho \mathrm{g} L^3} = \frac{U^2}{\mathrm{g}L} \rightarrow \mathrm{Fr} = \frac{U}{\sqrt{\mathrm{g}L}} \,. \tag{7.10}$$

The Froude number is most often used to characterize the flow at a free surface and is related to the celerity (phase speed) of a gravity wave. It can also be used to predict when wave drag is important for ships and other vessels. As the equilibrium position of subsurface buoys represents a balance between buoyancy, weight, and drag, which varies as the square of the freestream flow speed, Froude scaling can also be used to model their dynamic responses, even when they are deeply submerged [7.2].

d) Weber number.
The Weber number is the ratio of inertial forces to surface tension forces

$$\mathrm{We} = \frac{\rho U^2 L^2}{T_\mathrm{s} L} = \frac{\rho U^2 \mathrm{L}}{T_\mathrm{s}} \,, \tag{7.11}$$

where T_s is the surface tension ($T_\mathrm{s} = 0.073\,\mathrm{N\,m^{-1}}$ for water). The Weber number can be used to estimate the size of droplets in a spray, the thickness of a spray sheet (such as at the chine of a planing hulled vessel), and is also important in formation of ventilated cavities behind surface piercing propellers [7.3, 4]. Note that it is not generally possible to match both the Weber number and the Froude number in a flow.

e) Cavitation number.
This is the ratio of the difference between the local pressure in a flow and the vapor pressure and the dynamic pressure of the flow $\frac{1}{2}\rho U^2$

$$\sigma \equiv \frac{p_0 - p_\mathrm{v}}{\frac{1}{2}\rho U^2} \,, \tag{7.12}$$

where the vapor pressure of water at 20 °C is $p_\mathrm{v} = 2.336\,\mathrm{kPa}$. Cavitation may generally occur when $\sigma < 0.1$ and the inception will tend to depend on a number of factors including water purity, water depth, and background pressure – which can necessitate the use of pressured model test facilities in practice. As the local pressure over a submerged surface approaches the local vapor pressure of the water, the water essentially starts to boil and small pockets of water vapor will form over the surface. The process is unsteady and when the air pockets begin to collapse they can damage the surface. Cavitation differs from ventilation, in which there is a connection or small tunnel of air between the submerged surface and the free surface of the water. Both cavitation and ventilation can lead to a loss of lift on submerged hydrofoils or propeller blades.

f) Force coefficients.
The two-dimensional (2-D) drag coefficient is typically defined using the drag per unit span of a wing, d, whereas the three-dimensional (3-D) drag coefficient is defined using the total drag on a wing, D

$$\text{2-D} \qquad C_\mathrm{d} \equiv \frac{d}{\rho U^2 L} \,, \tag{7.13}$$

$$\text{3-D} \qquad C_\mathrm{D} \equiv \frac{D}{\rho U^2 L^2} \,. \tag{7.14}$$

The lift coefficients are defined similarly, where l is the drag per unit span and L is the total lift

$$\text{2-D} \qquad C_\mathrm{l} \equiv \frac{l}{\rho U^2 L} \,, \tag{7.15}$$

$$\text{3-D} \qquad C_\mathrm{L} \equiv \frac{L}{\rho U^2 L^2} \,. \tag{7.16}$$

g) Strouhal number.
The Strouhal number represents the ratio of a characteristic flow time to a characteristic oscillation time. For example, if the frequency of vortex shedding in the wake of a cylinder of diameter L in a flow with a speed U is f, then the Strouhal number will be

$$\mathrm{Sr} \equiv \frac{\text{char. flow time}}{\text{char. oscillation time}} \approx \frac{L/U}{1/f} = \frac{fL}{U} \,, \tag{7.17}$$

the Strouhal number relates the frequency of vortex shedding in a flow and is important for ocean engineering problems, such as the strumming of underwater cables and structures in a current and the swimming of marine animals [7.5–7]. Incidentally, the Strouhal number is proportional to the tip speed ratio (TSR) of underwater ocean energy extraction turbines [7.2, 8] and varies as the inverse of the advance ratio J (often used to characterize propellers) [7.5].
In the motion of a bluff body such as a cylinder of diameter d, accelerating from rest to moving at a constant speed U normal to its axis, the body experiences an adverse pressure gradient on its

downstream side, leading to a backward flow and flow separation from the aft surface of the body. The detached boundary layer rolls up aft of the body into two coherent counter-rotating vortices that are shed from the body once they reach a certain size for a given local convective flow. The vortices are typically shed asymmetrically from the two sides of the body and continue to form and get shed alternately at a frequency f, leading to a staggered array of counter-rotating vortices, known as a von Karman vortex street, in the wake of the body. The vortex shedding induces an alternating lift force normal to the direction of the flow and the body axis. The lift force per unit span of the bluff body may be estimated as

$$l(t) = \rho U \Gamma(t) \,,$$

where ρ is the density of the fluid and $\Gamma(t)$ is the instantaneous circulation around the body. Γ is related to the net vorticity shed from body at any instance. $l(t)$ therefore fluctuates at a frequency f and amplitude $l_{\max}$ as $\Gamma(t)$ varies between $\pm\Gamma_{\max}$, where $\Gamma_{\max}$ is the amplitude of the fluctuating circulation. The lift coefficient and the nondimensional frequency of the fluctuation

$$C_{l\max}(\mathrm{Re}) = \frac{l_{\max}}{0.5\rho U^2 d} = \frac{\Gamma_{\max}}{0.5 U d} \,,$$
$$S(\mathrm{Re}) = \frac{fd}{U} \,,$$

are functions of the Reynolds number (and the shape of the body). For a cylinder S has been shown to be approximately $S \approx 0.22{-}0.27$ for Reynolds numbers between 10^4 and 10^7 [7.9]. The lift coefficient is approximately given by $C_{l\max} \approx 0.5$ [7.10]. If the natural frequency of vibration of the body is close to f, then the body undergoes transverse resonant vibrations [7.10]. Such vibrations are called vortex-induced vibrations (VIV). Strumming oscillations of cables in crosswinds and transverse vibrations of pilings are examples of the phenomena. Use of a splitter plate at the back of the cylinder or of helical ribbon fairings on the cylinder can be made to minimize VIV.

Example 7.1

Dispersion of surface waves on water: assume that the wave speed c depends on the other five parameters listed in Table 7.1. Find the Π groups, note that the wavenumber is defined as $k \equiv 2\Pi/\lambda$, where λ is the wavelength of a wave.

Table 7.1 Potential dispersion parameters of interest

Parameter	Dimension
c Wave speed	$[c] = \frac{L}{T}$
g Gravity	$[\mathrm{g}] = \frac{L}{T^2}$
γ Surface tension	$[\gamma] = \frac{M}{T^2}$
k Wave number	$[k] = \frac{1}{L}$
h Water depth	$[h] = L$
ρ Density	$[\rho] = \frac{M}{L^3}$

Solution 7.1

6 variables, 3 dimensions $\rightarrow 3\Pi$-groups.

- Eliminate length:

$$[ck] = \frac{1}{T} \,, \quad [\mathrm{g}k] = \frac{1}{T^2} \,,$$
$$[\gamma] = \frac{M}{T^2} \,, \quad [hk] = 1 \,, \quad [\rho k^3] = M \,.$$

- Eliminate time:

$$\left[\frac{ck}{\sqrt{\mathrm{g}k}}\right] = \left[c\sqrt{\frac{k}{\mathrm{g}}}\right] = 1 \,,$$
$$\left[\frac{\gamma}{\mathrm{g}k}\right] = M \,, \quad \left[\rho k^3\right] = M \,.$$

- Eliminate mass:

$$\left[\frac{\gamma k^2}{\rho \mathrm{g}}\right] = 1$$
$$\therefore \quad c\sqrt{\frac{k}{\mathrm{g}}} = \varphi\left(kh, \frac{\gamma k^2}{p\mathrm{g}}\right) .$$

This can be rewritten as

$$c^2 = \frac{\mathrm{g}}{k}\varphi^2\left(kh, \frac{\gamma k^2}{\rho \mathrm{g}}\right) .$$

As shown in [7.11], the phase speed of deep water waves (gravity is the restoring force) is

$$c^2 = \frac{\mathrm{g}}{k}\tanh(kh) \,,$$

and the phase speed of capillary waves (surface tension is the restoring force) is

$$c^2 = \frac{\mathrm{g}}{k}\left(1 + \frac{\gamma k^2}{\rho \mathrm{g}}\right)\tanh(kh) \,.$$

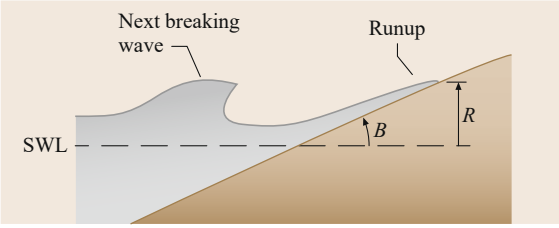


Fig. 7.2 Definition of the main parameters of beach runup

Example 7.2

Wave *runup* R is the height above the still water line (SWL) that an ocean wave will wash up onto a beach (Fig. 7.2). R has been found to depend on waveheight H, deep water wavelength λ, and beach angle β. Use dimensional analysis to find the associated Π groups and explicitly write the resulting functional dependence of R on the other terms.

Solution 7.2

$$R = f(H, \lambda, \beta)\,, \quad [R] = L\,, \quad [H] = L\,,$$
$$[\lambda] = L\,, \quad [\beta] = 1\,.$$

4 variables, 1 dimension $\rightarrow$ 3 Π-groups.

- Eliminate length:

$$\left[\frac{R}{H}\right] = 1\,, \quad \left[\frac{\lambda}{H}\right] = 1\,.$$
$$\therefore \quad \frac{R}{H} = \varphi\left(\frac{\lambda}{H}, \beta\right).$$

Empirical solutions for wave runup have been found by [7.12]

$$\frac{R}{H} = \sqrt{\frac{2\pi}{\beta}}\,,$$

and [7.13]

$$\frac{R}{H} = \sqrt{\frac{\lambda}{H}}\tan\beta\,.$$

Example 7.3

Longshore currents flow parallel to the shoreline and are typically confined to the area between the shore and the zone of breaking waves (Fig. 7.3). The current forms due to the oblique approach of waves to the shore at angle α_b, with the volume rate of flow depending primarily on the height of the incoming waves H_b (which break under the influence of gravity) and the slope of the beach m. Use dimensional analysis to find an expression for the functional dependence of the longshore current speed u.

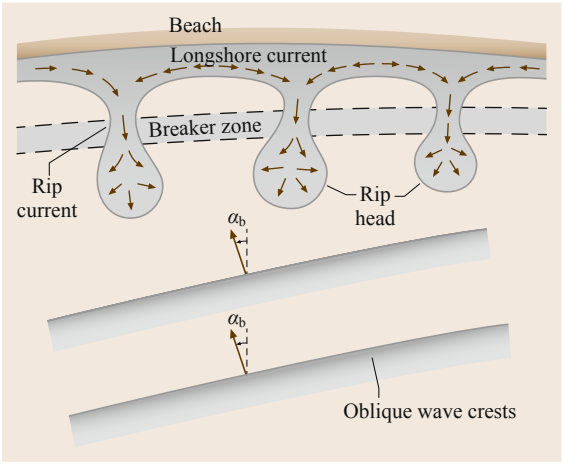


Fig. 7.3 Longshore current and wave approach (after [7.14])

Solution 7.3

$$u = f(\alpha_b, H_b, g, m)\,, \quad [u] = \frac{L}{T}\,, \quad [\alpha_b] = 1\,,$$
$$[H_b] = L\,, \quad [g] = \frac{L}{T^2}\,, \quad [m] = 1$$

5 variables, 2 dimensions $\rightarrow 3\Pi$-groups. There are already two dimensionless terms $\Pi_2 = \alpha_b$, $\Pi_3 = m$.

- Eliminate length:

$$\left[\frac{u}{H_b}\right] = \frac{1}{T}\,, \quad \left[\frac{g}{H_b}\right] = \frac{1}{T^2}\,,$$

- Eliminate time:

$$\Pi_1 = \frac{u}{H_b}\sqrt{\frac{H_b}{g}} = \frac{u}{\sqrt{gH_b}}\,.$$
$$\therefore \quad \boxed{u = \sqrt{gH_b}\,\varphi(\alpha_b, m)\,.}$$

From theory and experiment [7.15]

$$u = 20.7m\sqrt{gH_b}\,\sin(\alpha_b)\,.$$

Example 7.4

The waterplane area is the horizontal cross section formed by the intersection of the free surface of the water and the hull of a vessel. A catamaran (two hulls) has

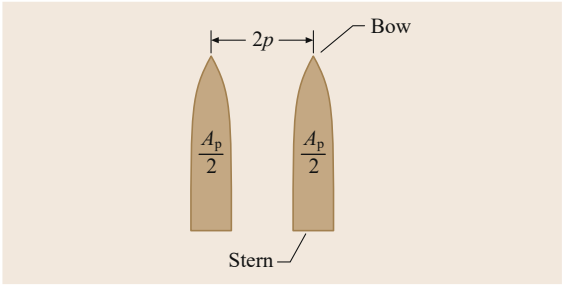


Fig. 7.4 Waterplane area of catamaran hulls

a waterplane area A_P and a displacement (weight) Δ (Fig. 7.4). The two hulls are separated by a transverse distance $2p$. If the density of seawater is ρ, use dimensional analysis to estimate:

a) The natural heave (vertical – z direction) period of the vessel for gravity-induced motion.
b) The natural roll (about the x-axis) period of the vessel for gravity-induced motion.

In each case you must explain the physical rationale behind your selection of the relevant parameters.

Solution 7.4

a) As the motion is gravity induced, gravity must be included in the dimensional analysis. The catamaran will oscillate like a mass–spring system with buoyancy acting as the restoring (spring) force. The dimensions of each of the physical variables are

$$[A_P] = L^2\,, \quad [\Delta] = \frac{\mathrm{ML}}{T^2}\,,$$
$$[\rho] = \frac{M}{L^3}\,, \quad [\mathrm{g}] = \frac{L}{T^2}\,,$$

the natural heave (vertical – z direction) period T_h, where $[T_h] = T$. Assume that the ship's motion is completely vertical, in which case the variables of interest are $A_P, \Delta, \rho, \mathrm{g}, T_h \rightarrow n = 5; m = 3; n-m = 2\Pi$ groups:

- Eliminate mass:

$$[A_P] = L^2, \left[\frac{\Delta}{\rho}\right] = \frac{L^4}{T^2}, [g] = \frac{L}{T^2}, [T_h] = T\,.$$

- Eliminate length:

$$\left[\frac{\Delta}{\rho A_P^2}\right] = \frac{1}{T^2}, \left[\frac{\mathrm{g}}{\sqrt{A_P}}\right] = \frac{1}{T^2}, [T_h] = T\,.$$

- Eliminate time:

$$T_h\sqrt{\frac{\Delta}{\rho A_P^2}} = \Pi_1, \frac{\rho \mathrm{g} A_P^{3/2}}{\Delta} = \Pi_2\,.$$

Thus, it should be possible to express the heave period as

$$\therefore \boxed{T_h = \sqrt{\frac{\rho A_P^2}{\Delta}} f\left(\frac{\rho \mathrm{g} A_P^{3/2}}{\Delta}\right),}$$

where f is a function to be identified through experimental analysis or numerical simulation. In practice, it can be shown that the natural period varies like

$$T_h = \sqrt{\frac{M + A_{33}}{2\Pi\rho \mathrm{g} A_p}}\,,$$

where M is the mass of the vessel and A_{33} is the added mass of the vessel. If we let

$$f\left(\frac{\rho \mathrm{g} A_P^{3/2}}{\Delta}\right) = \frac{\Delta}{\rho \mathrm{g} A_P^{3/2}}\,,$$

then

$$T_h = \sqrt{\frac{\rho A_P^2}{\Delta}} f\left(\frac{\rho \mathrm{g} A_P^{3/2}}{\Delta}\right) = \frac{1}{\mathrm{g}}\sqrt{\frac{\Delta}{\rho A_P}}\,.$$

As you can see the period will vary as the square root of the vessel mass and the inverse of the square root of the waterplane area.

b) The natural roll (about the x-axis) period T_r of the vessel for gravity-induced motion: Here the variables of interest will be the same, but with the addition of the side hull separation distance p. Now $[p] = L$ and $n = 6, m = 3, n-m = 3\Pi$-groups. The dimensional analysis proceeds as above, except with the addition of eliminating the length p. This gives a third group $\Pi_3 = p/\sqrt{A_p}$. So the functional dependence for roll period is

$$\boxed{T_r = \sqrt{\frac{\rho A_P^2}{\Delta}} f\left(\frac{\rho \mathrm{g} A_P^{3/2}}{\Delta}, \frac{p}{\sqrt{A_p}}\right).}$$

7.1.3 Similitude

Similitude is the art and practice of predicting prototype performance from model tests. Basically, dimensionless Π groups are used to scale test results:

a) Geometric similitude.
The model is an exact, but scaled replica of the prototype. Usually, all linear dimensions are scaled by the same factor.

b) Dynamic similitude.
Forces acting on model masses must have the same ratio everywhere in the flow field. To achieve this, *relevant* dimensionless numbers are equated so that each type of force (e.g., viscous, momentum, gravitational, etc.) that is considered to be dominant in determining the flow will act in a proportion determined by the Π-groups.

The *relevant* Π-groups are those which are most likely to play a role in governing a particular flow phenomenon. For example, when examining the flow in a completely enclosed space with no free surface, one would not generally need to consider the possible effects of either gravity or capillary waves. Hence the Froude number and Weber number are unlikely to play a role. The interplay between viscous and momentum forces is likely to dominate the flow. Hence, the Reynolds number would be likely to play a role.

When determining which Π groups might play a role in governing a flow, it is often helpful to examine the magnitude of the Π groups. When a Π group is much larger than unity $\Pi \gg 1$ or much smaller than unity $\Pi \ll 1$, the effects of the Π group can sometimes be neglected.

Solving similitude problems:

i. Think about the dominant Π groups that are likely to apply. When a free surface is present, the Froude number, Weber number, and Reynolds number are likely to play a role. When there is no free surface, the Reynolds number is likely to be an important parameter. When vortex shedding is present, the Strouhal number will play role.
ii. Equate the Π groups that are felt to be important and use geometric similitude to solve for the related length and velocity scales.
iii. Equate force and pressure coefficients to find the corresponding forces on the prototype, when model forces are known.

Example 7.5

An offshore wind turbine is supported with a cylindrical monopile. The system is deployed in currents of $U_\mathrm{p} = 2.5\ [\mathrm{m\,s^{-1}}]$ and waves with a $T_\mathrm{p} = 8.8$ [s] period and $H_\mathrm{p} = 1.9$ [m] height. If a 1 : 40 scale model is tested in a wave flume, what are the model current speed U_m, model wave period T_m and model wave height H_m?

Solution 7.5

The geometry of the model wave can be scaled by $L_\mathrm{m}/L_\mathrm{p} = 1/40$, where L_m and L_p are characteristic length scales for the model and prototype flows, respectively

$$H_\mathrm{m} = \frac{H_\mathrm{p}}{60} = 4.75\,\mathrm{cm}\,.$$

As the flow past the monopile will create surface gravity waves, the Froude number scaling can be used to determine the model current speed and wave period. Using the definition of the Froude number,

$$\mathrm{Fr} \equiv \frac{U}{\sqrt{gL}}\,,$$

we can develop relations for the model current U_m

$$\frac{U_\mathrm{m}}{\sqrt{gL_\mathrm{m}}} = \frac{U_\mathrm{p}}{\sqrt{gL_\mathrm{p}}}$$

$$\rightarrow \boxed{U_\mathrm{m} = U_\mathrm{p}\sqrt{\frac{L_\mathrm{m}}{L_\mathrm{p}}} = 0.40\,\mathrm{m\,s^{-1}}\,.}$$

The celerity (phase speed) of a wave is its wavelength λ divided by its period T. Froude scaling can also be used to relate the celerity of the prototype and model waves

$$\frac{\lambda_\mathrm{m}/T_\mathrm{m}}{\sqrt{gL_\mathrm{m}}} = \frac{\lambda_\mathrm{p}/T_\mathrm{p}}{\sqrt{gL_\mathrm{P}}} \rightarrow \boxed{T_\mathrm{m} = T_\mathrm{p}\sqrt{\frac{L_\mathrm{m}}{L_\mathrm{p}}} = 1.39\,\mathrm{s}\,,}$$

where we have taken $\lambda_\mathrm{m}/\lambda_\mathrm{p} = L_\mathrm{m}/L_\mathrm{p}$.

7.1.4 Skin-Friction Drag

Breakdown of a Laminar Boundary Layer

The transition to turbulence over a flat plate is mainly governed by the Reynolds number, but can be affected by the amount of freestream turbulence present in the flow and the surface roughness of the plate. Above a critical Reynolds number of about $\mathrm{Re}_x \equiv Ux/\nu \approx 5\times10^5$ the flow becomes turbulent. Here the length scale for the Reynolds number is based on the distance from the leading edge of the plate x so that the Reynolds number increases with an increase in the downstream distance from the leading edge. The sequence of events that happen in the transition to turbulence in a flat plate boundary layer is shown in Fig. 7.5.

Near the leading edge of the flat plate, the flow is laminar and the full Navier–Stokes relations hold and $\mathrm{Re}_x \approx 1$. Downstream when $\mathrm{Re}_x \gg 1$, the flow is still laminar, but the approximate boundary layer equations hold [7.16]. A similarity solution for the Blasius boundary layer can be found (Fig. 7.6).

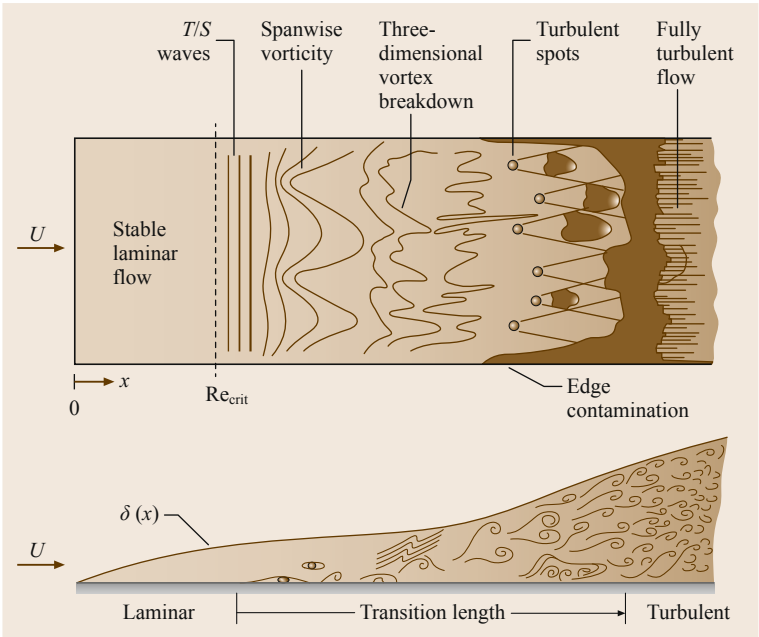


Fig. 7.5 Sequence of events occurring in transition to turbulence (after [7.16])

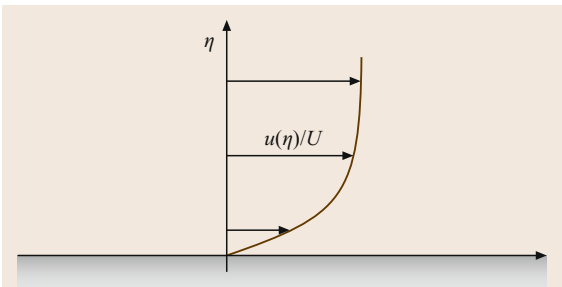


Fig. 7.6 Velocity distribution in a boundary layer over a flat plate. Here, η is the normalized height above the boundary layer, U is the freestream velocity and u is the local velocity. Both u and U are parallel to the plate

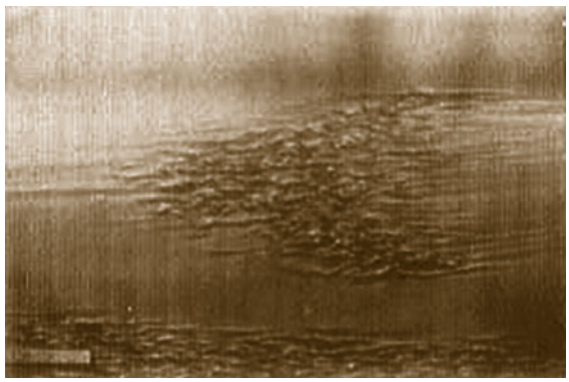


Fig. 7.7 A turbulent spot (after [7.17], courtesy of Cambridge University Press)

When $\mathrm{Re}_x \approx 5 \times 10^5$, instability starts to set in and the flow exhibits the development of two-dimensional Tollmein–Schlichting waves. As the Reynolds number continues to increase, the flow transitions to three-dimensional turbulent spots (Fig. 7.7), which are intermittent in nature. Lastly, the flow will transit to a fully three-dimensional turbulent flow (Fig. 7.8).

7.1.5 Estimating the Drag on a Ship from Model Testing

The results of model testing can be extended to approximate the drag on full-scale surface vessels. Here the prototype system is the full-scale system and the model system is typically geometrically similar, but smaller, and tested under controlled conditions in a towing tank or model basin. To avoid multiple scaling issues, test models do not usually include any appendages, such as fins or rudders; propulsion systems are also generally not included in model testing. The total resistance (drag) of surface vessels has been found to be a function of both the Reynolds number and the Froude number $C_\mathrm{T} = C_\mathrm{T}(\mathrm{Re}, \mathrm{Fr})$. However, it is not possible to simultaneously match both the model and prototype Reynolds and Froude numbers.

To circumvent this issue, Froude's hypothesis is used in practice. It is assumed that the total drag force can be split into two components, a frictional part F_v which is a function of the Reynolds number

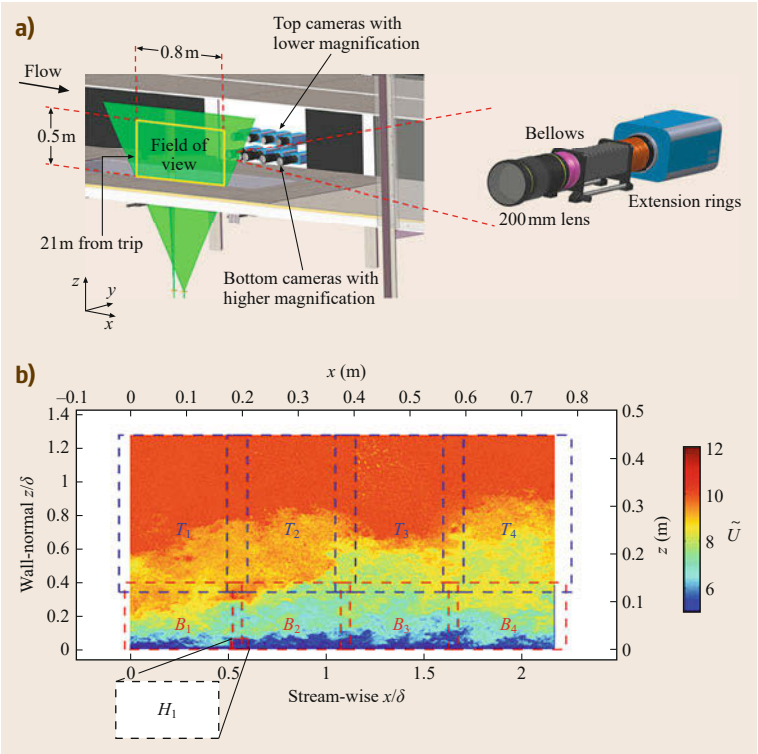


Fig. 7.8a,b Experimental measurement of a flat plat turbulent boundary layer. **(a)** Experimental setup; **(b)** spatial distribution of $\widetilde{U}$, the instantaneous velocity component in the streamwise direction (in meters per second). Here δ is the boundary layer thickness and the dashed rectangles represent the areas imaged by each of the cameras in **(a)**. The rectangle H_1 indicates an area where high resolution was used to image the near-wall region of the boundary layer (after [7.18], courtesy of AIP publishing)

only

$$C_F(\mathrm{Re}) \equiv \frac{F_v}{\frac{1}{2}\rho U^2 S}\,, \tag{7.18}$$

where S is the wetted surface area of the hull, and a residuary drag component F_R, which is a function of the Froude number only

$$C_R(\mathrm{Fr}) \equiv \frac{F_R}{\frac{1}{2}\rho U^2 S}\,. \tag{7.19}$$

The residuary part of the drag is assumed to arise from normal pressure forces acting perpendicularly to the hull (wave drag, viscous pressure resistance, and flow separation) and the frictional component is assumed to arise from shear stresses acting tangentially along the hull (skin friction). The total drag is simply the sum of the frictional and residuary components

$$C_T(\mathrm{Re},\mathrm{Fr}) = C_F(\mathrm{Re}) + C_R(\mathrm{Fr})\,. \tag{7.20}$$

Let a subscript m refer to the model hull and the subscript p refer to the corresponding prototype hull. The procedure for using Froude's hypothesis is as follows:

1. Use Froude scaling (7.10) to determine the model test speed from the anticipated operational speed of the prototype. Using the known model and prototype lengths, match the model and prototype Froude numbers and solve for the required test speed

$$\mathrm{Fr}_m = \frac{U_m}{\sqrt{gL_m}} = \frac{U_p}{\sqrt{gL_p}} = \mathrm{Fr}_p \rightarrow U_m = U_p\sqrt{\frac{L_m}{L_p}}\,.$$

Generally, the length on the waterline of the model and prototype are used as the reference lengths L_m and L_p, respectively.
2. The total drag on the model is experimentally measured and used to determine the total drag coefficient C_{Tm}.
3. The coefficient of residuary resistance of the model C_{Rm} is assumed to be equal to the coefficient of residuary resistance of the prototype C_{Rp} when the model and prototype Froude numbers are equal. The residuary resistance of the model is determined from Froude's hypothesis as $C_{Rm} = C_{Tm} - C_{Fm} = C_{Rp}$.
4. The coefficient of friction of both the model and prototype are determined using Reynolds number dependent empirical relations for the drag on a flat plate. In practice, the 1957 ITTC (International Towing Tank Conference) line [7.19] is often used

$$C_F = \frac{0.075}{[\log_{10}(\mathrm{Re}) - 2.0]^2}\,. \tag{7.21}$$

5. Then the total resistance of the prototype is

$$C_{Tp} = C_{Fp} + C_{Rp} = C_{Fp}(Re_p) + C_{Tm} - C_{Fm}(Re_m) .$$

Note that the ITTC line assumes that the flow is turbulent. For most models, even those of the modest length (about 1 m), towed at slow speeds (2 kn), the flow will be turbulent over at least part of the model. However, the geometrically scaled thickness of the boundary layer about the model will tend to be much thicker than that about the prototype and the difference can lead to some discrepancy in the drag predicted using Froude's hypothesis. To mitigate this, turbulence stimulators (tripwires, sandpaper, thick tape, etc.) are often utilized in model testing to ensure that the boundary layer transitions from laminar to turbulent flow as quickly as possible at the bow of the model. The required thickness of a trip t can be estimated using the following relation for the critical Reynolds number for transition [7.20]

$$Re_{crit} = \frac{Ut}{\nu} \approx 826 . \qquad (7.22)$$

Note that the selection of trip thickness is a balancing act between forcing the flow to transition to turbulence and affecting the pressure distribution about the hull and thereby affecting the global flow distribution.

7.1.6 Hydrofoil Lift and Drag

Strictly speaking, a hydrofoil is the cross-sectional shape of a larger rudder, fin or underwater wing. Any two-dimensional streamlined body that generates lift can be thought of as a hydrofoil. Here, the lift force per unit span l is perpendicular to the direction of the freestream flow (flow undisturbed by the presence of the hydrofoil) and the drag force d is always parallel to the direction of the freestream (Fig. 7.9). The hydrofoil chord c is defined as the straight line segment that connects the leading edge of the hydrofoil to its trailing edge. The angle of attack is the angle between the freestream flow direction and the chord line of the foil.

As mentioned above, the extension of a hydrofoil into three dimensions gives a fin or rudder. The dimension perpendicular to the hydrofoil cross section is called the span s and the projected area perpendicular to the fin or rudders lift force is known as the planform area A (Fig. 7.10).

The aspect ratio of the wing is defined as $\Lambda \equiv s^2/A$. When there is little twist about the span-wise axis of a rudder or fin, the flow can be approximated as two-dimensional when $\Lambda \gg 1$. In practice, when the aspect ratio Λ is larger than about 5–7, the flow around a hydrofoil can be approximated as two-dimensional flow for modeling, design, or experimental purposes. When performing experiments, endplates are often added to the ends of a fin or rudder to help ensure two-dimensionality of the flow at measurement locations near the center span.

The basic geometry of a hydrofoil is shown in Fig. 7.11. The thickness of a hydrofoil is its maximum dimension in a direction perpendicular to the chord line. The mean camberline is defined as the mean of the upper and lower positions of the hydrofoil in a direction perpendicular to the chord line. The *camber* of the hydrofoil is the maximum distance between the camberline and the chord line. Both the thickness and camber of a hydrofoil are typically specified as a fraction of the chord length. For example, a common convention for defining hydrofoil shape is use of the 4-digit NACA (National Advisory Committee for Aeronautics) desig-

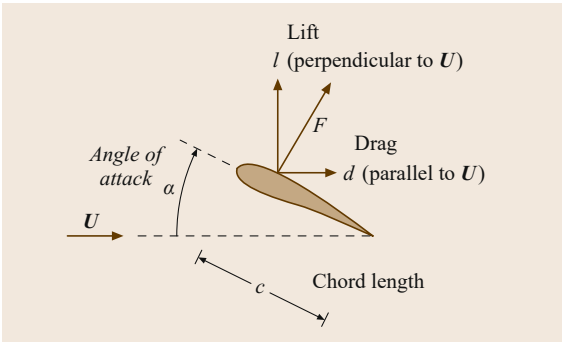


Fig. 7.9 Lift and drag on a foil

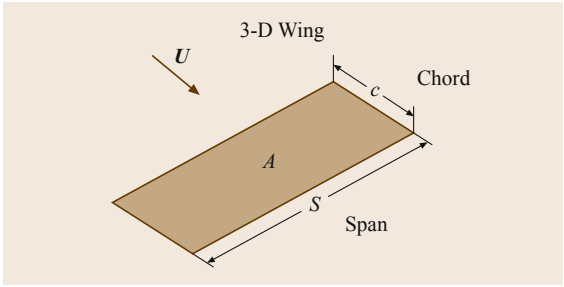


Fig. 7.10 Fin or rudder

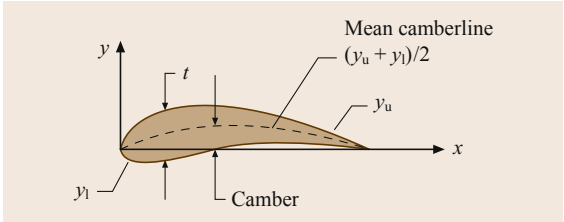


Fig. 7.11 Hydrofoil geometry

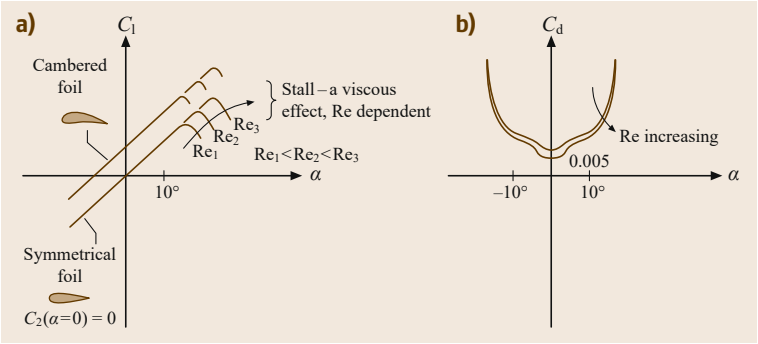


Fig. 7.12a,b Example hydrofoil lift coefficient **(a)** and drag coefficient**(b)** versus angle of attack

nation [7.21, 22]. The geometry of simple hydrofoils can be specified using the NACA convention, as follows [7.21]:

1. The first digit represents the camber of the foil as a percentage of the chord length.
2. The second digit represents the location of maximum camber with respect to the leading edge, as a percentage of the chord length divided by 10.
3. The last 2 digits represent the thickness of the foil as a percentage of the chord length.

For example, a NACA2412 hydrofoil has 2% camber, the location of maximum camber is 40% of the chord length from the leading edge, and the thickness of the foil is 12% of the chord length.

As shown in Fig. 7.12, the lift and drag coefficients of the hydrofoil have been found to be functions of the Reynolds number Re and angle of attack α of the foil. The Reynolds number dependence of the lift curve is a result of flow separation, which is a viscous effect. When the flow separates from the surface of the hydrofoil, the pressure difference across the foil is drastically reduced and there is a rapid loss of lift with increasing angle of attack. As can be seen in Fig. 7.12a, so long as there is no separation, the Reynolds number dependence of the two-dimensional lift coefficient can be ignored

$$C_l \equiv \frac{l}{\rho U^2 c} = C_l(\mathrm{Re}, \alpha) \approx C_l(\alpha) \,, \tag{7.23}$$

$$C_d \equiv \frac{d}{\rho U^2 c} = C_d(\mathrm{Re}, \alpha) = C_F(\mathrm{Re}) + C_P(\alpha) \,. \tag{7.24}$$

For thin hydrofoils, approximately when $t/c < 0.14$, the two-dimensional lift coefficient can be approximated as $C_l = 2\pi\alpha$, where α is in radians, using the *flat plate approximation*. In practice, this approximation works fairly well for estimating the lift produced by a hydrofoil in initial design estimates or simple simulations. Note that, cambered foils will produce lift when $\alpha = 0$.

The corresponding drag coefficient curves are shown in Fig. 7.12b. It is a common practice to separate the effects of Reynolds number and angle of attack into two, additive components. The first component C_F is a function of Reynolds number only and is related to the viscous friction of the hydrofoil. The second component C_P is a function of the angle of attack and is related to the pressure coefficient of the foil. The large increase in drag coefficient that occurs when magnitude of the angle of attack is large is due to stall effects.

A convenient approximation for the two-dimensional drag coefficient on a hydrofoil or submerged streamlined strut at zero angle of attack is given by *Hoerner*'s formula [7.23]

$$C_d = 2C_F[1 + 2\left(\frac{t}{c}\right) + 60\left(\frac{t}{c}\right)^4] \,, \tag{7.25}$$

where C_F is the skin friction coefficient, as given by the 1957 ITTC line (7.21).

Three-Dimensional Effects

Owing to the pressure difference across the top and bottom of a low aspect ratio, lift-generating fin (or rudder or sail), the streamlines at the tip of the fin will roll up to create wingtip vortices, as shown in Fig. 7.13.

These wingtip vortices induce a small velocity behind the fin in a direction normal to its planform. The induced velocity changes the local angle of attack of the flow at the hydrofoil and effectively causes the local lift vector rotate toward the downstream direction, effectively increasing the drag on the fin – this is known as *induced drag* or can sometimes be called *drag due to lift*. For thin fins with an elliptical planform shape, the aspect ratio effects of the wingtip vortices can be accounted for in the lift coefficient C_L and the induced drag coefficient C_{Di} as follows

$$C_L = \frac{2\pi\alpha}{1 + 2/\Lambda} \,, \tag{7.26}$$

$$C_{Di} = \frac{4\pi\alpha^2\Lambda}{(\Lambda+2)^2} = \frac{C_L^2\Lambda}{\pi(\Lambda+2)^2} \,, \tag{7.27}$$

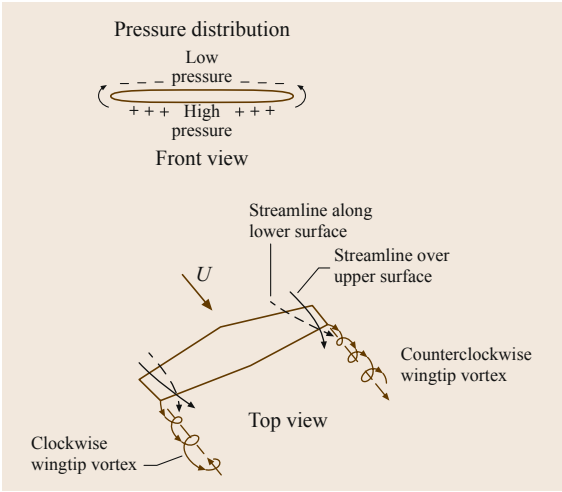


Fig. 7.13 Formation of wingtip vortices at the tips of a fin or rudder

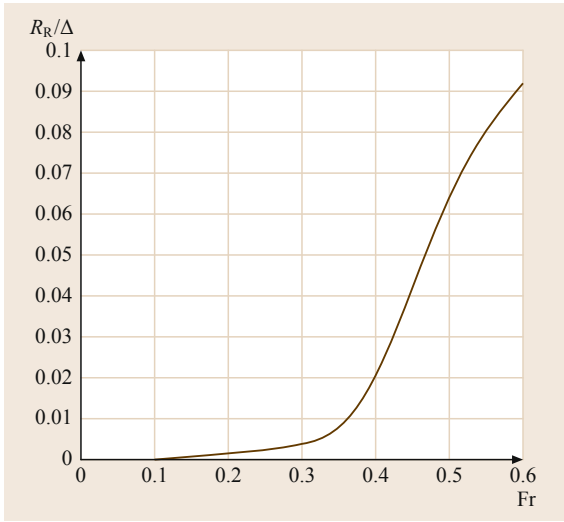


Fig. 7.14 Upright (no heel) residuary resistance R_R as a fraction of displacement Δ versus Fr on the model hull from experimental measurements

where the total drag coefficient is the sum of the two-dimensional drag coefficient C_d and the induced drag coefficient $C_{Dtot} = C_d + C_{Di}$.

Example 7.6

The following questions pertain to a prototype sailboat designed for oceanographic measurement [7.24–28]. Some of the model and prototype dimensions are shown in Table 7.2; properties of air and seawater at design conditions are given in Table 7.3.

The following questions may be posed:

1. The hullform:
   a) The model and prototype sailboats are geometrically similar:
      i. Fill-in the missing prototype values in Table 7.2.
      ii. The prototype will be operated at Fr = 0.375, where the relevant length scale is taken as the length on the waterline, LWL (Table 7.2). What is the design speed V_S of the prototype?
   b) Use Fig. 7.14 to estimate the C_D for the prototype sailboat (heel angle is not included) – calculate the weight of the prototype Δ from the volume of water displaced by the hull $\nabla = 8.0\,m^3$. Approximate the frictional drag coefficient of the vessel using the 1957 ITTC Line based on LWL (Table 7.2).
   c) What is the estimated hydrodynamic drag F_H on the prototype sailboat hull? We are ignoring the drag on the keel and rudder, any heel angle, the leeway angle, etc.
2. The sail:
   As shown in Fig. 7.12, the sailboat must be able to sail in a wind of $V_t = 20.0\,kn$ ($10.29\,m\,s^{-1}$) at an angle of $\psi = 40°$. A wing having a symmetrical airfoil cross section, a span of $s_s = 10\,m$ and a chord of $l_s = 2\,m$ will be used as the sail. Refer to Fig. 7.15 and treat the sail as a 2-D wing by ignoring the three-dimensional effects at the wingtip:
   a) Determine the angle β and speed V_A of the *apparent wind*.
   b) Determine the equation of motion (for constant V_S) along the boat's course.
   c) Assume F_H lies along the sailboat's course. If $\alpha = 15°$, compare the component of the sail drag F_D along the boat's trajectory with F_H – can this component of F_D be ignored?
   d) Estimate the slope of the lift curve ($\partial C_L/\partial\alpha$) from the airfoil data provided (Fig. 7.16).
   e) Using the slope of the lift curve in your equation of motion, determine the angle of attack α required for the sail. Will it stall?

Table 7.2 Model and prototype dimensions

Quantity	Model	Prototype
Length overall (LOA) [m]	2.47	11
Length on waterline (LWL) [m]	2.00	
Wetted surface area [m^2]	0.965	

Table 7.3 Properties of water and air at 15 °C

Medium	Sea water	Air
Density [$kg\,m^{-3}$]	1025.9	1.226
Kinematic viscosity [$m^2\,s^{-1}$]	1.19×10^{-6}	1.45×10^{-5}

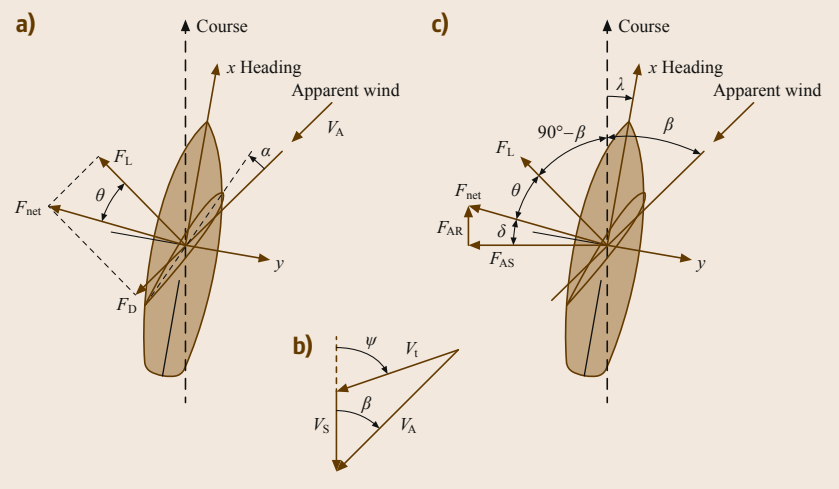


Fig. 7.15a–c Sailboat aerodynamic free body diagrams in the boat fixed frame. **(a)** The airfoil profile represents the wing and α is the angle of attack of the sail with respect to the apparent wind angle. **(b)** ψ is the true wind direction, β is the apparent wind angle. **(c)** The x- and y-axes are the boat-fixed maneuvering coordinate system, the x-axis gives the heading while the angle between course and heading is the leeway angle λ, $\delta = (\beta - \theta)$ is the angle between the net aerodynamic force F_{net} and the side force F_{AS} (after [7.25], courtesy of the Institute of Electrical and Electronics Engineers (IEEE))

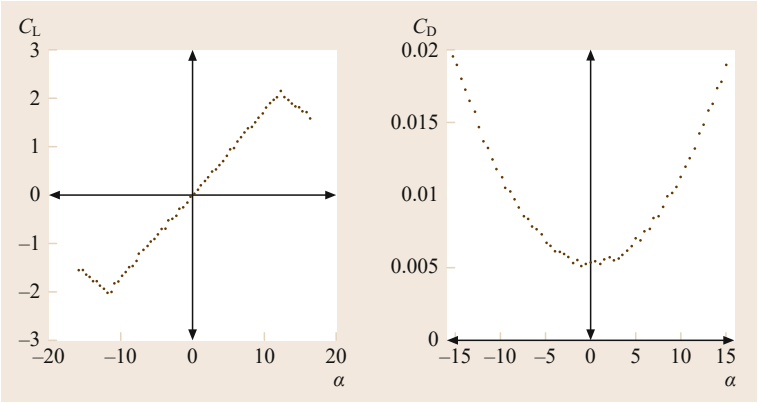


Fig. 7.16 Sailboat wing aerodynamic characteristics

3. The keel:
   Here, the keel will be a thin flat plate with a span of $S_k = 1.2\,\text{m}$ and a chord of $l_k = 1.5\,\text{m}$. In order to keep the lateral (transverse) component of the aerodynamic forces F_{AS} from pushing the yacht off course, the keel must have a small leeway angle λ (angle of attack) with respect to the boat's course (Fig. 7.17). Assuming F_{AS} is balanced entirely by the lift on the keel F_{HS} – calculate λ.

Solution 7.6

1. The hullform:
   a) The model and prototype sailboats are geometrically similar:
      i. The missing prototype values have been included in Table 7.4.
      ii. The prototype will be operated at $\text{Fr} = 0.375$, where the relevant length scale is taken as the length on the waterline, LWL (Table 7.2). What is the design speed V_S of the prototype?

$$V_S = F\sqrt{\text{gLWL}_p} = 3.51\,\text{m}\,\text{s}^{-1}$$

Table 7.4 Model and prototype dimensions

Quantity	Model	Prototype
Length overall (LOA) [m]	2.47	11
Length on waterline (LWL) [m]	2.00	$\text{LWL}_p = 8.91$
Wetted surface area [m^2]	0.965	$\text{Sw}_p = 19.1$

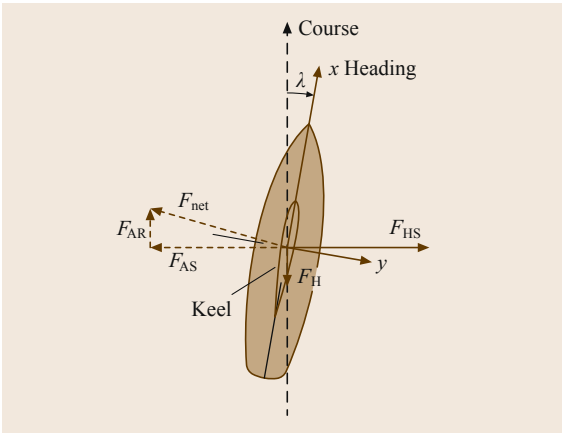


Fig. 7.17 Sailboat hydrodynamic free body diagram in the boat fixed frame. The hydrofoil profile represents the keel. At equilibrium,the hydrodynamic side force F_{HS} balances the aerodynamic side force F_{AS} and the force of hydrodynamic resistance F_{HR} balances the aerodynamic driving force F_{AR}, where F_{HS} and F_{HR} are perpendicular and parallel, respectively, to the course direction. Note that F_{HS} is generated by the keel and hull acting as an underwater wing with the leeway angle λ being the angle of attack (after [7.25]). Reproduced with permission from IEEE

b) Use Fig. 7.14 to estimate the C_D for the prototype sailboat (heel angle is not included) – calculate the weight of the prototype Δ from the volume of water displaced by the hull $\nabla = 8.0\,m^3$. Approximate the frictional drag coefficient of the vessel using the 1957 ITTC Line based on LWL (Table 7.2).

$$\Delta = \rho_{sw} g \nabla = 80.5\,kN$$

From Fig. 7.14,

$$R_R(F) = 1.18 \times 10^{-2} \Delta = 950\,N\,.$$

Using Froude's hypothesis,

$$C_D(R,F) = C_F(Re) + C_R(Fr)\,,$$

where $C_F(Re)$ is determined using the hull Reynolds number

$$Re_H = \frac{V_S LWL_p}{\upsilon_{sw}} = 2.62 \times 10^7$$

and the 1957 ITTC Line

$$C_F(Re_H) = \frac{0.075}{[\log_{10}(Re_H) - 2.0]^2} = 2.55 \times 10^{-3}$$

and

$$C_R(Fr) = \frac{R_R}{\frac{1}{2}\rho_{sw} V^2 S w_p} = 7.88 \times 10^{-3}\,,$$

so that $C_D = 10.4 \times 10^{-3}$.

c) What is the estimated hydrodynamic drag F_H on the prototype sailboat hull? We're ignoring the drag on the keel and rudder, any heel angle, the leeway angle, etc.

$$F_H = \frac{1}{2}\rho_{sw} V_s^2 S w_p C_D = 1.26\,kN\,.$$

2. The sail:
As shown in Fig. 7.12, the sailboat must be able to sail in a wind of $V_t = 20.0\,kn$ ($10.29\,m\,s^{-1}$) at an angle of $\psi = 40°$. A wing having a symmetrical airfoil cross section, a span of $s_S = 10\,m$, and a chord of $l_S = 2\,m$ will be used as the sail. Refer to Fig. 7.15 and treat the sail as a 2-D wing by ignoring the three-dimensional effects at the wingtip:

a) Determine the angle β and speed V_A of the *apparent wind*.

$$\beta = \tan^{-1}\left(\frac{V_t \sin\psi}{V_t \cos\psi + V_S}\right) = 30.1°$$

$$V_A = \frac{V_t \sin\psi}{\sin\beta} = 13.2\,m\,s^{-1}$$

b) Determine the equation of motion (for constant V_S) along the boat's course.

$$F_L \sin\beta - F_D \cos\beta - F_H = 0$$

c) Assume F_H lies along the sailboat's course. If $\alpha = 15°$, compare the component of the sail drag F_D along the boat's trajectory with F_H – can this component of F_D be ignored?
For $\alpha = 15°$, the drag coefficient of the sail $C_{DS} \approx 0.019$. The component of F_D along the boat's course is

$$F_D \cos\beta = \frac{1}{2}\rho_{air} V_A^2 C_{Ds} l_s s_s \cos\beta = 34.8\,N\,.$$

This is about 2.8% of the drag on the hull. As the sail will stall at angles of attack larger than this, we can ignore the drag on the sail, as the sail is likely to be operated at smaller angles, where the drag is lower.

d) Estimate the slope of the lift curve ($\partial C_L/\partial\alpha$) from the airfoil data provided (Fig. 7.16).

$$\frac{\partial C_L}{\partial\alpha} = \frac{2.2 - 0}{12° - 0} = 0.18\,[deg^{-1}] \rightarrow C_L = \frac{\partial C_L}{\partial\alpha}\alpha$$

e) Using the slope of the lift curve in your equation of motion, determine the angle of attack α required for the sail. Will it stall?

$$F_{\mathrm{L}} = \frac{\partial C_{\mathrm{L}}}{\partial \alpha}\alpha\frac{1}{2}\rho_{\mathrm{air}}V_{\mathrm{A}}^2 l_{\mathrm{s}}s_{\mathrm{s}}$$
$$\longrightarrow \frac{\partial C_{\mathrm{L}}}{\partial \alpha}\alpha\frac{1}{2}\rho_{\mathrm{air}}V_{\mathrm{A}}^2 l_{\mathrm{s}}s_{\mathrm{s}}\sin\beta = F_{\mathrm{H}}$$

Solving for α gives

$$\alpha = \frac{F_{\mathrm{H}}}{\frac{1}{2}\rho_{\mathrm{air}}V_{\mathrm{A}}^2 l_{\mathrm{s}}s_{\mathrm{s}}\frac{\partial C_{\mathrm{L}}}{\partial \alpha}\sin\beta} = 6.43^\circ\,.$$

From Fig. 7.16, we can see that this is smaller than the stall angle, so the sail will not stall.

3. The keel: Here, the keel will be a thin flat plate with a span of $s_{\mathrm{k}} = 1.2\,\mathrm{m}$ and a chord of $l_{\mathrm{k}} = 1.5\,\mathrm{m}$. In order to keep the lateral (transverse) component of the aerodynamic forces F_{AS} from pushing the yacht off course, the keel must have a small leeway angle λ (angle of attack) with respect to the boat's course (Fig. 7.17). Assuming F_{AS} is balanced entirely by the lift on the keel F_{HS}, calculate λ.

From thin airfoil theory for a flat plate, the lift coefficient of the keel can be calculated as $C_{\mathrm{Lk}} = 2\pi\lambda$ (where λ is in radians). Thus,

$$F_{\mathrm{HS}} = F_{\mathrm{AS}} = \frac{1}{2}\rho_{\mathrm{sw}}V_{\mathrm{S}}^2(2\pi\lambda)s_{\mathrm{k}}l_{\mathrm{k}}\,.$$

Solving for λ and converting to degrees gives $\lambda = 1.75^\circ$.

7.1.7 Screw Propellers

A typical screw propeller can be considered to be a combination of several hydrofoils arranged in a helicoidal fashion around a hub. Definitions of the basic terms are given in Fig. 7.18. The diameter of the circle traced by the tip of each blade is d and the radius is r. The tangential velocity of the blade tips is $2\pi nr$, where n is the rotational speed of the propeller in Hz. The shaft torque required to rotate the propeller is Q and the thrust generated is T. The velocity perpendicular to the plane of the rotor disk generated by the propeller is known as the advance speed and is given here by U. Thus, as shown in Fig. 7.19, the velocity of the flow approaching each blade has both a component normal to the plane of the rotor disk and parallel to the rotor disk, the resultant velocity is known as the relative velocity U_{rel}.

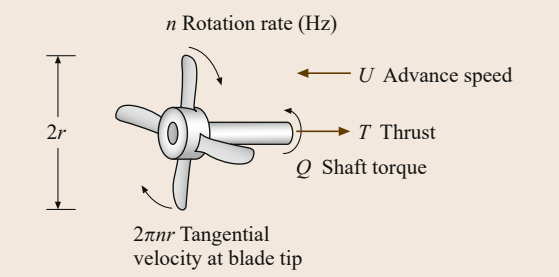


Fig. 7.18 Definition of the basic terms, here T and Q are drawn as the resultant thrust and shaft torque, respectively, of the fluid acting on the propeller

The term *pitch* can be used to refer to either an angle ϕ or a distance P (Fig. 7.20). When used in reference to angle, the pitch is the angle between the chord line of a propeller blade and the plane of the rotor disk. Alternatively, when used to reference a distance the pitch is the forward distance traveled by the rotor in one revolution when there is no slip with respect between the propeller and surrounding water.

The angle of attack of the blade α can be determined from the components of U_{rel} and the pitch angle of the blade ϕ

$$\alpha = \phi - \tan^{-1}\left(\frac{U}{2\pi nr}\right) = \phi - \tan^{-1}\left(\frac{J}{\pi}\right)\,, \quad (7.28)$$

where J is the advance ratio defined as

$$J \equiv \frac{U}{nd}\,. \quad (7.29)$$

As long as the propeller is not operated at an excessive angle of attack, flow separation on the blades can be

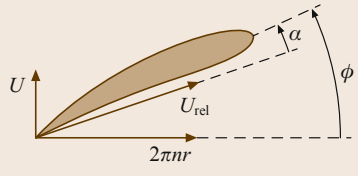


Fig. 7.19 Velocity components of the flow approaching each propeller blade

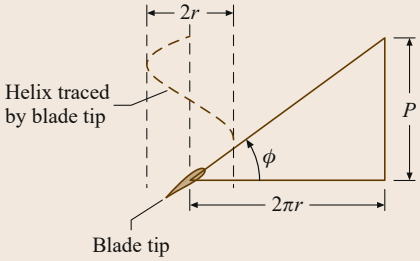


Fig. 7.20 The definition of propeller pitch

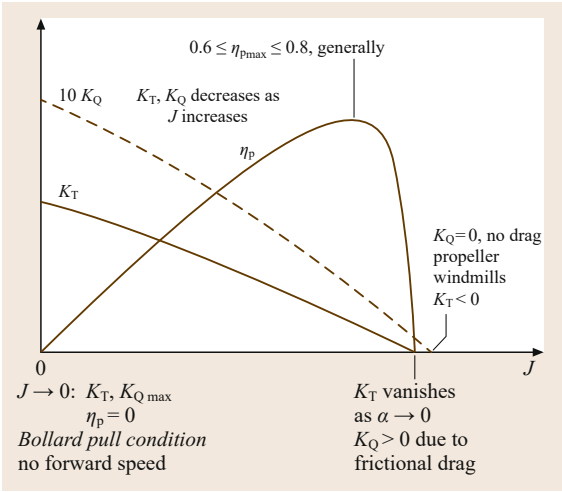


Fig. 7.21 Example of open-water propeller performance curves

avoided such that the thrust and torque on the propeller are independent of the Reynolds number. We define the thrust coefficient as

$$K_T(J) \equiv \frac{T}{\rho n^2 d^4} , \tag{7.30}$$

and the torque coefficient as

$$K_Q(J) \equiv \frac{Q}{\rho n^2 d^5} . \tag{7.31}$$

At a constant velocity, the power delivered by the propeller is UT and the power required to overcome shaft torque is $2\pi nQ$. Thus, the propeller's efficiency is the output power divided by the input power and can be written as

$$\eta_p = \frac{UT}{2\pi nQ} = \frac{U\rho n^2 d^4 K_T}{2\pi n \rho n^2 d^5 K_Q} = \frac{U}{2\pi n d}\frac{K_T}{K_Q} = \frac{J}{2\pi}\frac{K_T}{K_Q} . \tag{7.32}$$

A sketch of a typical open-water propeller performance curve is shown in Fig. 7.21. Here, the term open water indicates that the propeller was tested in the absence of any hull or other underwater surface, such that the advance velocity is not representative of any wake.

Blade Element Theory

Blade element theory can be used to estimate the output thrust and input torque on a propeller. A propeller blade is segmented into short, two-dimensional, differential elements and the lift and drag on each element are determined from the two-dimensional lift and drag coefficients, and then integrated across the span of the propeller blade. Typically, the analysis is conducted on one blade and multiplied by the number of blades N to determine the total thrust and torque on the propeller. The effects of camber and finite blade length are often omitted from the analysis for simplicity.

Let the two-dimensional drag and lift coefficients be defined as in (7.13) and (7.14), where l and d are the lift and drag per unit span of the blade, respectively, c is the blade chord length and U_{rel} (the relative velocity shown in Fig. 7.19) is used as the characteristic velocity scale. Then, the thrust and torque on a differential element are

$$\mathrm{d}T = l\cos(\phi-\alpha) - d\sin(\phi-\alpha) ,$$

and

$$\mathrm{d}Q = [l\sin(\phi-\alpha) + d\cos(\phi-\alpha)]r ,$$

respectively (r is the radial location of the blade element measured from the center of the propeller hub). Substituting the coefficients of lift and drag into these expressions gives

$$\mathrm{d}T = \frac{1}{2}\rho U_{rel}^2 c[C_l\cos(\phi-\alpha) - C_d\sin(\phi-\alpha)]\mathrm{d}r ,$$
$$\mathrm{d}Q = \frac{1}{2}\rho U_{rel}^2 c[C_l\sin(\phi-\alpha) + C_d\cos(\phi-\alpha)]r\mathrm{d}r .$$

Integrating these expressions over the span of the blade (between the radial location of the hub r_h and the tip of the blade r_0) and multiplying by the number of blades, gives the total thrust and drag on the propeller

$$T = \frac{1}{2}\rho U_{rel}^2 Nc \times \int_{r_h}^{r_0} [C_l\cos(\phi-\alpha) - C_d\sin(\phi-\alpha)]\mathrm{d}r , \tag{7.33}$$

$$Q = \frac{1}{2}\rho U_{rel}^2 Nc \times \int_{r_h}^{r_0} [C_l\sin(\phi-\alpha) + C_d\cos(\phi-\alpha)]r\mathrm{d}r . \tag{7.34}$$

Increasing the number of blades can increase the thrust output, but will also increase the torque. In theory, as the input power is the product of the torque and angular velocity and the output power is the product of the thrust and forward speed, the efficiency, which is the ratio of these products, will be independent of the number of propeller blades.

Actuator Disk Model

Here a propeller is modeled by the disk that the propeller blades trace out when spinning. The velocity upstream of the propeller is taken as U_0 and the velocity downstream is U_1. The streamlines passing through the propeller actuator disk form a continuous streamtube that contracts from the upstream side to the downstream side (Fig. 7.22). The upstream area of the streamtube is taken as A_0, the downstream area is A_1 and the area at the actuator disk itself is A_d. The velocity at the actuator disk itself is taken as the average of the upstream velocity and the downstream velocity

$$U_d = \frac{U_0 + U_1}{2} .$$

The total pressure across the actuator disk changes and energy is input into the flow at the actuator disk itself. Thus, Bernoulli's equation cannot be used across the actuator disk. However, it holds upstream and downstream of the actuator disk. The pressure on the upstream suction side of the disk p_b and on the downstream, pressure side of the disk p_f can be calculated as

$$p_b = p_\infty + \frac{1}{2}\rho U_0^2 - \frac{1}{2}\rho\left(\frac{U_1 + U_2}{2}\right)^2 ,$$

$$p_f = p_\infty + \frac{1}{2}\rho U_1^2 - \frac{1}{2}\rho\left(\frac{U_1 + U_2}{2}\right)^2 .$$

Thus, the pressure difference across the disk is

$$p_f - p_b = \frac{1}{2}\rho\left(U_1^2 - U_0^2\right) .$$

The thrust generated is

$$T = (p_f - p_b)A_d = \frac{1}{2}\rho\left(U_1^2 - U_0^2\right)A_d . \qquad (7.35)$$

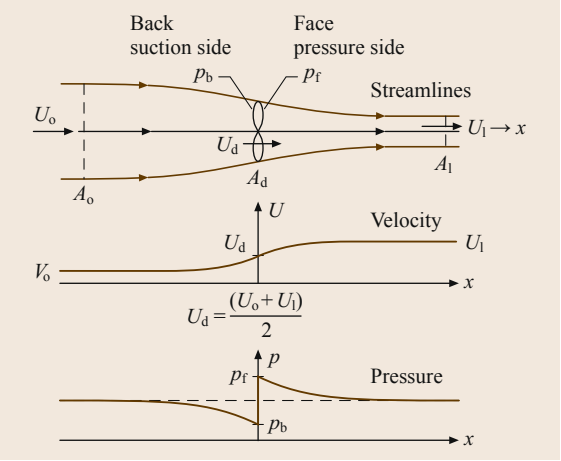


Fig. 7.22 The streamtube, velocity, and pressure across an actuator disk

Since the corresponding input power at the disk is

$$\mathbb{P}_{in} = TU_d = \frac{1}{2}\rho U_d\left(U_1^2 - U_0^2\right)A_d ,$$

and output power is

$$\mathbb{P}_{out} = TU_0 = \frac{1}{2}\rho U_0\left(U_1^2 - U_0^2\right)A_d ,$$

the efficiency is

$$\eta = \frac{\mathbb{P}_{out}}{\mathbb{P}_{in}} = \frac{2U_0}{U_0 + U_1} . \qquad (7.36)$$

Wake Effects and Propeller Slip

As explained above, the pitch is defined as the forward distance the propeller moves in one complete revolution when there is no slip (relative motion) between the propeller and surrounding water. In practice, however, slip must be present in order for a propeller to generate thrust. When no slip is present, the forward speed of the propeller would be the product of the pitch and propeller frequency, nP. Owing to hull drag, the forward speed U_S of a vessel driven by a propeller will be slightly lower. Apparent slip is the difference between nP and U_S (Fig. 7.23) and the related apparent slip ratio S_A is

$$S_A \equiv \frac{nP - U_S}{nP} . \qquad (7.37)$$

As a vessel moves through the water, it drags some water behind it, creating a wake, such that the advance velocity U at the propeller is reduced from the ship speed U_S by the wake speed. The difference between nP and U is the true slip speed and the related true slip ratio is defined as

$$S_R \equiv \frac{nP - U}{nP} . \qquad (7.38)$$

The overall propulsive efficiency is affected by the interaction of the vessel's wake and the flow into the

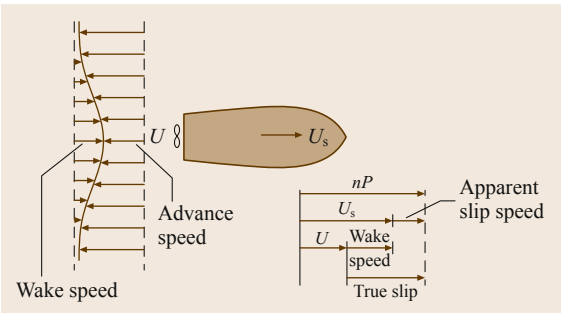


Fig. 7.23 Definition of slip

propeller. One can quantify the magnitude of the wake speed generated by a hull using the Taylor wake fraction, which is defined as

$$w \equiv \frac{U_S - U}{U_S} \,. \tag{7.39}$$

For most ships, the wake fraction lies in the range $0 < w < 0.4$, where a value of $w \approx 0.4$ would correspond to a very full (rounded) hull and is smaller for fine (slender) hulls.

The propulsion system and hull-form of a propeller-driven ship are typically designed separately. The bare-hull resistance R_T (hull drag without propulsion system, actuators, or rudders present) of the vessel is usually estimated from model test data. The performance of the propulsion system is also usually estimated from open water (no hull present) test data. However, the presence of a propeller at the stern of a ship reduces the pressure at the stern, in comparison to what it would be during bare hull resistance tests. This reduced pressure will typically result in an increase in drag. Thus, the thrust T required from the propeller is usually slightly higher than the bare hull resistance. The thrust deduction coefficient t is a measure of the difference between required thrust and bare hull resistance

$$t = \frac{T - R_T}{T} \,. \tag{7.40}$$

Generally, the thrust deduction coefficient is in the range $t < 0.2$. From (7.32), the propulsive efficiency can be expressed as the ratio of thrust power to propulsive power

$$\eta_p = \frac{UR_T}{2\pi nQ(1-t)} \,. \tag{7.41}$$

7.1.8 Air and Wind Resistance

When examining air drag on a surface-going vessel, one must account for both the velocities of the vessel through the air and the velocities of any surface winds (Fig. 7.24). The total velocity of the vessel through the air will be the vector sum of these two velocities. Of particular importance is how to model the velocity profile of each of these two components. The velocity distribution for the vessels motion will be uniform in height; the vertical profile of the wind will vary with height above the free surface owing to boundary layer effects.

For calculation of the wind resistance, the 10 min average of the wind speed at a height of 10 m (approximately 33 ft) above the water surface u_{10} is used as a reference. Generally, turbulence models are used to approximate the velocity profile of the wind. One such approximation is a modification of the *log law* for fully rough flow. If z is the height in meters above the waterline, the log law gives

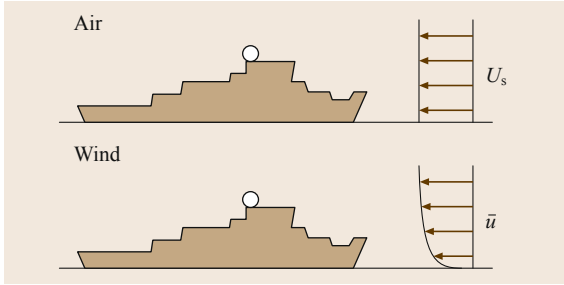


Fig. 7.24 Velocity profile of wind and air speed about a surface vessel

$$\frac{\bar{u}}{u_{10}} = \frac{\ln(z/k_s)}{\ln(10/k_s)} \,, \tag{7.42}$$

where $k_s = 6.0 \times 10^{-3}$ m. Another useful expression for modeling the wind profile [7.3] is

$$\frac{\bar{u}}{u_{10}} = \left(\frac{z}{10}\right)^{1/7} \,, \tag{7.43}$$

where z is in meters and u_{10} can be estimated from a known sea state.

Using these estimates for the speed of the wind acting on the vessel, the wind resistance can be approximated as

$$D_w = \frac{1}{2}\rho_{air} C_D A U^2 \,. \tag{7.44}$$

Here ρ_{air} is the air density and A is the frontal area of the area normal to the wind. The frontal area is used, rather than the wetted surface area (as in calculating the drag on the underwater hull) as pressure effects tend to dominate wind drag, whereas skin friction effects tend to dominate the overall drag on the underwater hull. Also, note that airflow can affect the trim and sinkage of a surface vessel, and thus the drag on the underwater hull.

7.1.9 Hydrodynamic Characterization of Marine Surface Vessels

The forces that support the weight (W) of a marine vehicle are found by integrating the pressure over the surface of the vehicle. The way in which this pressure is developed can be categorized into four broad operational regimes:

(1) Hydrostatic: The vehicle is principally supported by buoyancy forces, which are proportional to the volume of water displaced by the submerged volume

of the vessel (∇), that is, $W = \rho g \nabla$ (Archimedes' principle).

(2) Hydrodynamic: A vehicle pushes water downward as it moves across the water surface, imparting a downward momentum on fluid near the free surface. The resulting reaction force on the hull supports the vessel.

(3) Aerostatic: The pressure generated by a *lifting fan* at the top of a plenum chamber supports the vehicle's weight, even when it has no forward velocity.

(4) Aerodynamic: Some marine vehicles rely on the aerodynamic lift created by their forward velocity to support themselves near the free surface.

Hydrostatic and Hydrodynamic Regimes

The majority of marine vessels operate in either the hydrostatic or the hydrodynamic-pressure regime above. Since the hull is that part of a ship that generally supports its weight, hull design is strongly driven by the operational regime. Good hull design is critical for efficient powering, stability, controllability, and maneuverability [7.29]. The Froude number can be used to determine whether a vehicle operates in pressure regime (1), (2) or in an intermediate zone between regimes (1) and (2). Here, the waterline-length-based Froude number is defined as

$$\mathrm{Fr} \equiv \frac{U}{\sqrt{g L_{\mathrm{WL}}}} , \tag{7.45}$$

where U is vehicle speed, g is the acceleration of gravity and L_{WL} is a length-scale characteristic of the dimension of the ship parallel to the direction of the freestream flow (often taken as the design waterline LWL).

Calm Water Wave Drag and the Froude Number Limit of Displacement Hulls. As a ship or other marine vessel moves across the water's surface, it generates a distinctive pattern of waves, which create one component of the overall drag by radiating energy from the vessel to the far field (wave-making resistance). When viewed from far above the surface, the divergent and transverse waves created at the hull can be seen to interact to produce a Kelvin wave pattern (Fig. 7.19). Very close to the hull, the shape and pattern of the waves strongly depends on the form of the hull. At distances of a few hull lengths, the waves generated at all points on the hull's surface will contribute to the Kelvin wave system. However, the high-pressure regions at the bow and stern of a ship contribute most to the generation of the wave field. As the waves generated at the bow and stern move with the ship, their speed is equivalent to the speed of the ship U.

The celerity (c), or phase speed, of an attached deepwater wave, its wavelength λ, and the ship speed U are related by

$$c = \sqrt{\frac{g}{k}} = \sqrt{\frac{g\lambda}{2\pi}} = U , \tag{7.46}$$

where g is the acceleration of gravity and $k \equiv 2\pi/\lambda$ is the wavenumber. Here, we are considering the *transverse waves* (the wave fronts are perpendicular to the path of the ship), rather than divergent waves (the wave fronts lie at angles of about 19°28′ to the ship's trajectory) (Fig. 7.25).

Interference of the transverse waves generated at the bow and stern affects hull resistance. When the wave crests from the bow and the wave crests from the stern coincide, they constructively interfere to generate large waves which radiate away more energy and increase drag. If the bow wave crests coincide with the troughs in the stern waves, they destructively interfere creating attenuated waves and the drag is decreased. See (7.3), which gives 2-D wave energy/length across crest.

$$E = \frac{\rho g \lambda H^2}{8} \tag{7.47}$$

As can be seen in Fig. 7.26, when there are an integer number of wave crests spanning the hull $L_{\mathrm{WL}}/\lambda =$

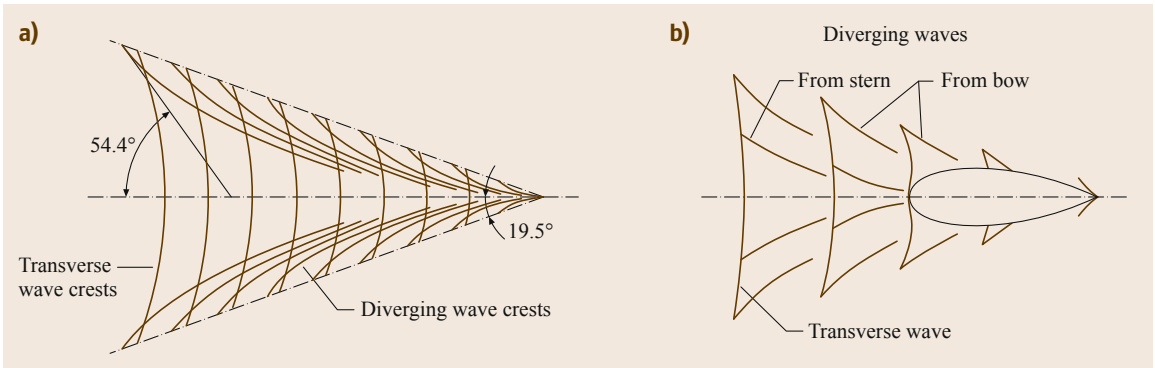


Fig. 7.25 **(a)** Far-field Kevin wave pattern; **(b)** Local bow and stern wave systems (after [7.30])

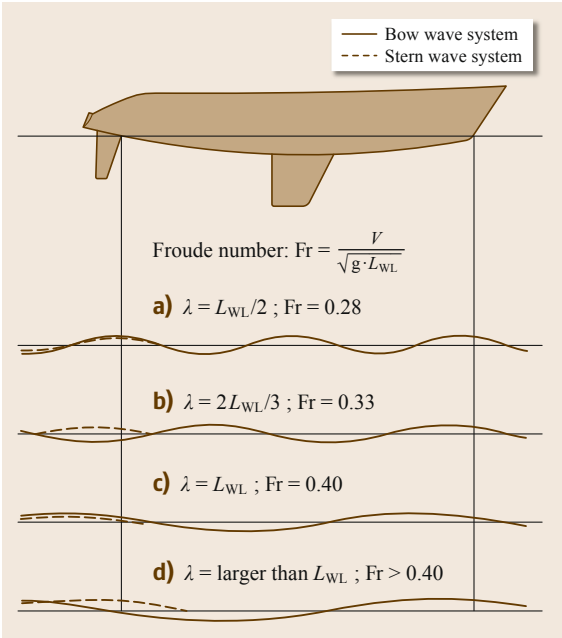


Fig. 7.26 Interference between bow and stern waves (after [7.11])

$\{1, 2, 3, \ldots\}$, constructive interference occurs; when there are an odd number of $\frac{1}{2}$ wavelengths $L_{WL}/\lambda = \{0.5, 1.5, 2.5, 3.5, \ldots\}$, the wave system behind the ship is attenuated. From (7.45)

$$U = \sqrt{\frac{g\lambda}{2\pi}} = \sqrt{\frac{gL_{WL}}{2\pi}}\sqrt{\frac{\lambda}{L_{WL}}},$$

solving for the Froude number gives

$$\mathrm{Fr} = \frac{U}{\sqrt{gL_{WL}}} = \sqrt{\frac{\lambda}{2\pi L_{WL}}}.$$

Thus, we can see that when $\mathrm{Fr} = 0.28$ and 0.4, the waves behind the ship are amplified and for $\mathrm{Fr} = 0.33$, the waves are attenuated. The corresponding effects on hull resistance are manifested as humps and hollows in the wave-resistance curve (Fig. 7.27).

When $\lambda = L_{WL}$, $\mathrm{Fr} = 0.4 \approx 1/\sqrt{2\pi}$ there are wave crests (high pressure) at both the bow and stern of the ship and a trough (low pressure) amidships. This condition corresponds to the *hull speed* of the vessel and, as explained in more detail below, gives $U_{\mathrm{hull}} = 0.4\sqrt{gL_{WL}}$ as the speed limit above which it is usually impractical to run a displacement hull vessel. When the Froude number is much higher than this, there is only a wave crest at the bow (with high pressure) and the ship is constantly *climbing a hill of water*. Ships that travel at $\mathrm{Fr} > 0.4$ are generally considered *high speed*.

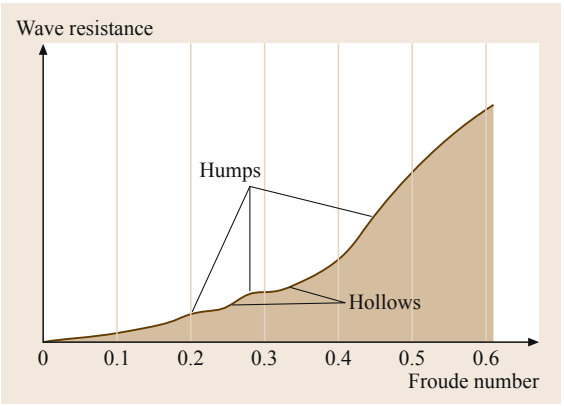


Fig. 7.27 Humps and hollows on the wave-resistance curve (after [7.11])

Fig. 7.28 RORO ferry (courtesy of J.H. Kantor)

It may be noted that:

- Another important hydrodynamic effect associated with high Fr is the generation of negative pressures along convex surfaces of the hull – these pressures have a tendency to draw a ship down into the water (squatting).
- In the design of high-speed ships, special consideration must be given to the structural weight of the ship, propeller cavitation/ventilation, seakeep-

Fig. 7.29 An ultra large crude carrier (ULCC)

ing at speed, and the effects of the wave wash created [7.3].

Basic Hull Forms. The Froude number determines the hull type (adapted from [7.29]):

1. Displacement hulls:
   $Fr \leq 0.4$ Hydrostatic buoyancy completely dominates. Examples of this type of ship include cruise ships, common carriers (bulk transports, cargo ships, reefers), roll on roll off (RORO) ships, fishing trawlers, tankers, and submarines (Figs. 7.28 and 7.29).
   The main geometric characteristics of displacement vessels (Fig. 7.30) are sharp, fine bows, midsections that are convex in both the longitudinal and transverse directions, and narrow stern sections.
   Displacement vessels are designed to operate in the narrow range of $0.3 \leq Fr < 0.4$ and are almost entirely supported by hydrostatic buoyancy forces. At their design speeds, the convex shape of their midsections prevent flow separation. When operated at higher Froude numbers ($Fr > 0.4$), their sterns tend to get drawn down into the water so that the aft part of the deck starts to become submerged, the bow starts to pitch upward, and the transverse flow along the sides clings to the hull so that it generates a sheet of spray.
2. Semidisplacement (also called semiplaning):
   $0.4 - 0.5 \leq Fr \leq 1.0 - 1.2$ both hydrostatic and hydrodynamic forces are important.
   A typical line drawing for a semidisplacement hull is shown in Fig. 7.31. These hullforms are characterized by water entry lines that are fine, a sharp wide transom that is partially submerged and straight, vee-formed transverse cross-sections near the bow. Small waterplane area (SWA) vessels, catamarans (two hulls) and trimarans (three hulls) are examples of typical semidisplacement hullforms (Figs. 7.32 and 7.33)
   Semidisplacement hulls are designed to avoid the squatting and trim angles observed in displacement hulls when $Fr \geq 0.4$. In contrast to displacement hulls, the aft part of a semidisplacement hull is almost flat and horizontal so that it terminates a sharp, submerged transom at the stern. At operational speeds, the flow separates from the transom so that

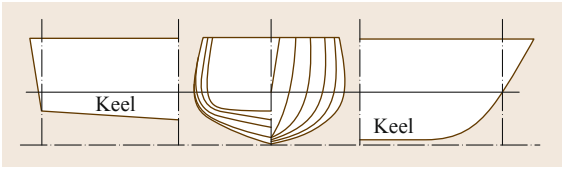


Fig. 7.31 Typical lines drawing for a semidisplacement catamarcan hull (after [7.29])

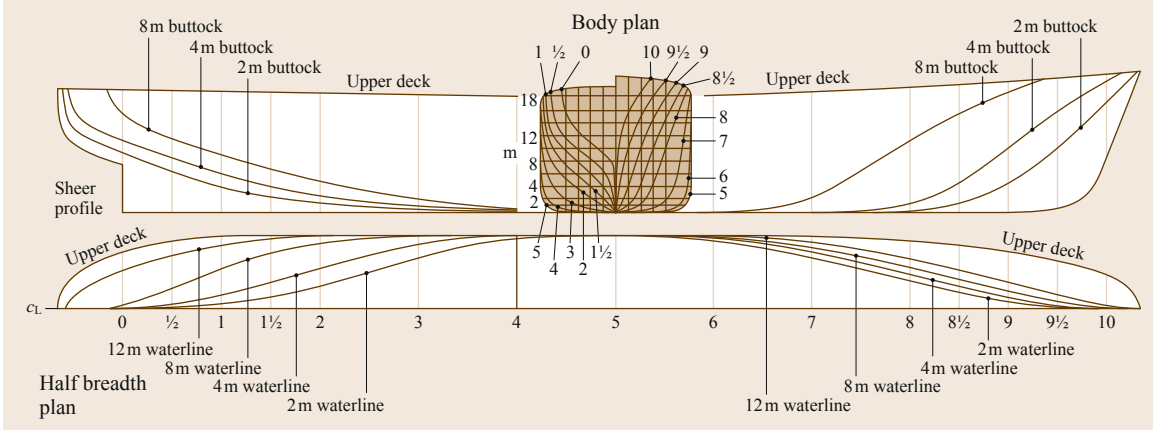


Fig. 7.30 Typical displacement hull lines drawing (after [7.31])

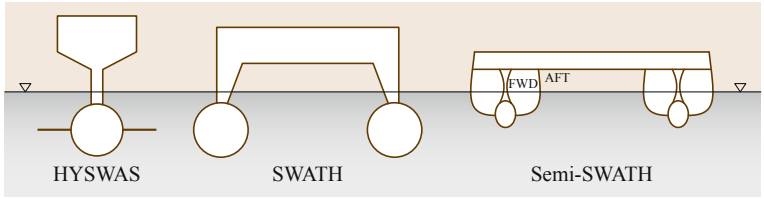


Fig. 7.32 SWATH concept – HYSWAS (hydrofoil SWA ship); SWATH (SWA twin hull) (after [7.32])

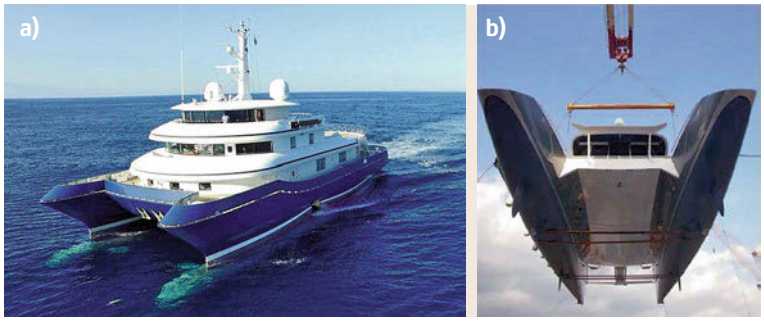


Fig. 7.33 **(a)** SWATH hull form (courtesy of Yacht Silver Cloud); **(b)** Catamaran hull form (viewed from stern)

it is ventilated and exposed to the atmosphere. With this geometry, positive dynamic pressure forces are generated in the flow at the stern, which reduce the squatting and large pitch angles seen in displacement vessels at higher Froude numbers.

3. Planing hulls:
   $1.0 - 1.2 \leq \mathrm{Fr}$ mainly hydrodynamically generated forces carry the weight of the vessel. At speed, the boat starts to ride up and over the crest of its own bow wave. A lower fraction of the hull is submerged decreasing its wetted surface area and frictional resistance. Examples of planing hull vessels include: rigid hull inflatable boats (RHIBs), sometimes simply called rigid inflatable boats (RIBs), recreational power and fishing boats, and police patrol boats (Fig. 7.34).

A lines drawing for a planing hull is shown in Fig. 7.35. The principal geometric features of planing hulls are:

- The avoidance of convex surfaces (except for the bow which is out of the water at planing speeds) to limit the development of bottom suction pressures
- Sharp-edged chines at the intersection of the bottom and sides to promote the separation of flow from the hull
- A deeply submerged wide transom that is ventilated to the atmosphere
- To reduce wave impacts, V-bottomed transverse sections with a sharp bow.

At low Froude numbers (below Fr of about 1.0–1.2 pre-planing speeds), the planing hull has more resistance than displacement or semidisplacement hulls because of the sharp chines and wide, submerged transom, which promote flow separation and increase the form drag component of the hull. The performance of a planing hull is very dependent upon the longitudinal location of the center-of-gravity, which affects the trim angle of the craft when planning and in turn influences the resistance/weight ratio. Typically, the resistance of planing hulls is reduced for trim angles of about 3–4°.

Fig. 7.34 A RIB on plane (photo public domain)

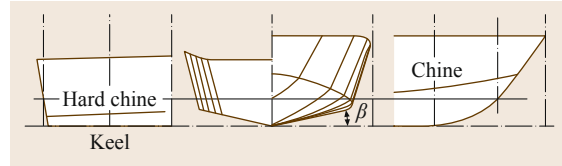


Fig. 7.35 Typical lines drawing for a hard chine planning hull (after [7.29])

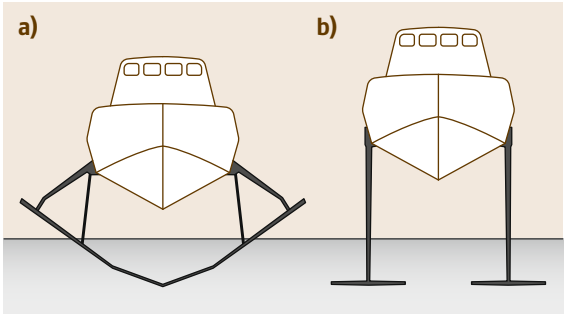


Fig. 7.36 **(a)** Surface piercing hydrofoil; **(b)** fully submerged hydrofoils (after [7.33])

Fig. 7.37 Vessel with fully submerged foils (courtesy of Sottome Jacaranda Lilan)

Transom interceptors can be used to adjust the trim angle and reduce resistance when a planing vessel is operating at design speeds [7.3].

Hydrofoil Vessels. Hydrofoil-supported vessels are a special class of vehicles that operate in the hydrodynamic pressure regime. However, they differ from conventional ships in that when they are at operating speed, the lifting function is taken away from the hull by a set of *underwater wings*. At speed, the main advantages of this approach are that the vehicle experiences lower wave drag, lower frictional drag, and the shocks caused by wave impacts are reduced. Disadvantages include a more complex propulsion drive train, a more complex attitude (trim and roll) control system, and a deeper low speed (off foil) draft. These vessels are typically operated at speeds where careful foil cross section design is required to handle cavitation and ventilation effects on the foils.

Hydrofoil vessels are typically characterized by the configuration of the foils as either surface piercing or fully submerged (Figs. 7.36 –7.38).

Fig. 7.38 A vessel with surface-piercing hydrofoils (courtesy of Martin Gimm)

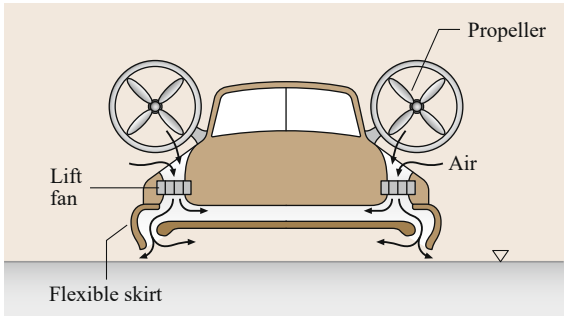


Fig. 7.39 ACV design

Fig. 7.40 ACV (photo public domain)

Aerostatic Support

Vehicles operating in the aerostatic regime use a lift fan to pressurize the air in a plenum chamber that lies below deck. Generally, the lower surface of the plenum is the free surface of the water and either the underside of the deck itself or a horizontal plate is the upper surface. When flexible skirts surround the front, back, and sides of the ship, it is known as an air cushion vehicle (ACV) or hovercraft (Figs. 7.39 and 7.40).

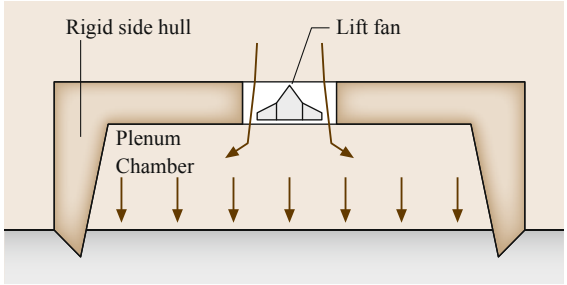


Fig. 7.41 CAB vehicle design (after [7.32])

Fig. 7.42 CAB (photo by Mark Harkin)

ACVs can rise completely above the water surface on a cushion of air and can operate amphibiously. When the front and back of a catamaran are enclosed using flexible skirts to create a plenum chamber, the vehicle is known as a captured air bubble (CAB) vehicle (Figs. 7.41 and 7.42). Since part of the weight of a CAB is supported by the cat hulls, this type of vehicle is not amphibious. Advantages of using ACV and CAB vessels are reduced wave and frictional drag (air resistance is lower than water resistance); disadvantages include poor tracking (primarily ACVs) and maintenance issues related to the use of flexible skirts and seals. Turbofans are often used to propel ACVs, while CABs are usually propelled by waterjets or surface-piercing propellers.

Aerodynamic Support

When a wing is operated near the ground (or alternatively the free surface of a lake/sea), less than about $\frac{1}{2}$

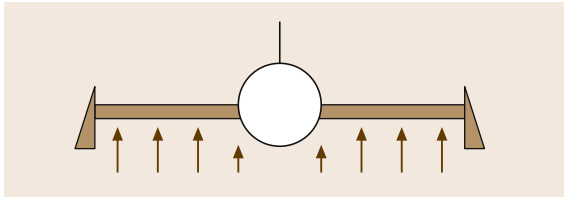


Fig. 7.43 Wing in ground-effect (WIG) concept (after [7.32])

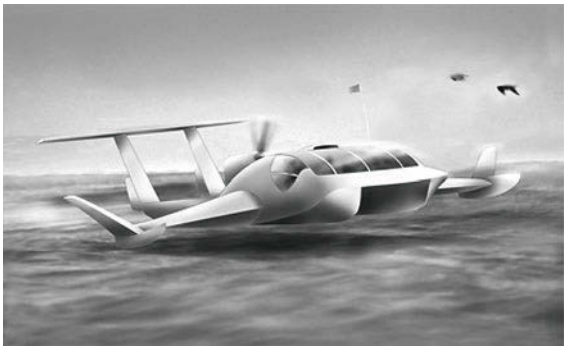


Fig. 7.44 Artist's impression of a WIG. An air-cushion effect is generated between the floats during takeoff (artist: Bjarne Stenberg; after [7.3], reproduced with permission)

a wingspan away, the lift coefficient of the wing increases and the drag due to lift is substantially reduced. A *cushion of air* is essentially trapped under the wing, which increases lift and reduces the downwash angle of the flow behind the wing (Fig. 7.43). Flying vehicles that utilize this *ground effect* can be designed with smaller wings and higher power-to-weight ratios than conventional aircraft. Vehicles of this sort are generally referred to as wing-in-ground effect (WIG) vehicles or ground effect vehicles (GEVs). The concept works best for very large vehicles since the height required to clear waves is a smaller percentage of the wingspan – most fly about 1/20-th of a wingspan above the water's surface (Fig. 7.44). Many early WIGs utilized flying boat hulls, which are quite heavy and have a substantial amount of drag. Some vehicles use extra turbofans or turbojets, which are mounted fore of the wings with their exhausts pointed downward, to provide added lift by increasing the pressure beneath the wings (power augmented ram – PAR).

7.2 Fluid Statics

In a solid, shear stress is directly proportional to shear strain. In a fluid flow viscous shear (tangential) stresses are directly proportional to the rate of change of shear strain with time. When a fluid is at rest, there are no tangential stresses and the only force between adjacent fluid surfaces is normal to the surface. In static equilibrium, the surface force per unit area (pressure) at a point is equal in all directions. However, owing to the density of a fluid, a pressure gradient exists. If we assume that gravity is acting in the $-z$ direction, the magnitude of the pressure gradient is

$$\frac{\mathrm{d}p}{\mathrm{d}z} = -\rho \mathrm{g} \,.$$

Here, ρ is the fluid density. A typical value of ρ for seawater is $\rho = 1025.9\,\mathrm{kg\,m^{-3}}$. The pressure gradient can be integrated to give

$$p = p_0 - \rho \mathrm{g} z \,,$$

where p_0 is the pressure at $z = 0$. Thus, when density and gravity can be taken as constant, we see that the pressure varies linearly with depth. When $z = 0$ is taken as the vertical location of a free surface of a body of water, such as the sea, the local pressure can be taken as atmospheric p_{atm}. A typical value of atmospheric pressure is $p_{\mathrm{atm}} = 101\,325\,\mathrm{Pa} = 1.013\,\mathrm{bar}$. Sometimes pressure is reported as *gauge pressure* p_{gauge}, which is the absolute pressure p, measured with respect to atmospheric pressure

$$p_{\mathrm{gauge}} = p - p_{\mathrm{atm}} \,.$$

7.2.1 Pressure Forces on Surfaces

Forces on a Horizontal Planar Surface

Examine the forces on a floating ship (Fig. 7.45). Take the ship to be wall-sided, such that its sides are vertical and perpendicular to the water surface.

The waterplane area A_{w} is the cross-sectional area representing the intersection of the still waterline (water surface with no waves present) and the vessels hull.

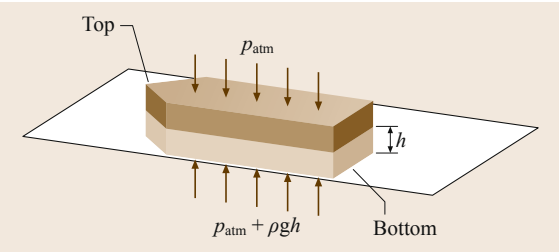


Fig. 7.45 Hydrostatic pressure on a wall-sided ship

As the pressure on the sides of the vessel is normal to the vertical axis, the buoyancy force F_{z} can be determined by integration of the pressure forces across the top and bottom surfaces. The bottom of the hull is taken to rest at a depth (draft) h below the surface (the small hydrostatic variation of air pressure is neglected).

$$\begin{aligned} F_{\mathrm{z}} = \int p \mathrm{d}A &= -\int_{\mathrm{top}} p_{\mathrm{atm}} \mathrm{d}A \\ &\quad + \int_{\mathrm{bottom}} (p_{\mathrm{atm}} + \rho g h) \mathrm{d}A \\ &= \rho g h A_{\mathrm{w}} \end{aligned}$$

We see that F_{z} is equivalent to the weight of the water displaced by the submerged volume of the hull, $\rho g h A_{\mathrm{w}}$. Note that here we used an example with a wall-sided ship that allowed us to forgo integration of the pressure forces on the sides of the hull. However, the result that the buoyancy force is equivalent to the weight of the water displaced by the submerged volume of a floating object generally holds because lateral forces will cancel due to symmetry (Archimedes' principle) [7.34].

Note that a floating object, such as a ship or boat, will trim (pivot) about its center of flotation (which is the centroid of the waterplane area) when weight is redistributed [7.35]. The center of flotation should not be confused with the center of buoyancy, which is the centroid of the submerged hull volume. The resultant buoyancy force acts vertically through the center of buoyancy and is in force and moment equilibrium with the total weight, which acts through the center of gravity.

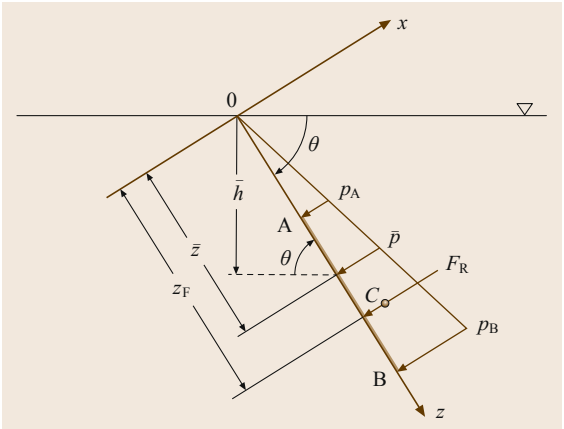


Fig. 7.46 Force on an inclined surface

Forces on an Inclined Planar Surface

Force Magnitude. Examine a submerged object with a planar surface oriented at an angle to the horizontal (Fig. 7.46).

Determine the force acting normal to a portion of the submerged surface spanning two points A and B, as shown

$$F_R = \int_A^B p\,dA = \rho g \int_A^B h\,dA\,,$$

where $h = z\sin\theta$. Then

$$F_R = \rho g \sin\theta \int_A^B z\,dA$$
$$= \rho g \sin\theta \frac{\int_A^B z\,dA}{\int_A^B dA} A = \rho g \bar{z} \sin\theta A\,.$$

Here $\bar{z}$ is the position of the area centroid, which is related to the depth of the area centroid as $\bar{h} = \bar{z}\sin\theta$ so that

$$F_R = \rho g \bar{h} A = \bar{p} A\,,$$

where $\bar{p}$ is the pressure at the area centroid. The point of the resultant force application is

$$z_F F_R = \int_A^B pz\,dA = \int_A^B \rho g h z\,dA = \int_A^B \rho g h z^2 \sin\theta\,dA$$
$$= \rho g \sin\theta \int_A^B z^2\,dA = \rho g \sin\theta I_0\,,$$

where I_0 is the area moment of inertia about the intersection between the submerged body and the water surface. Using the parallel axis theorem, I_0 can be rewritten in terms of the area moment of inertia of the submerged area about its own centroid $I_0 = I_C + \bar{z}^2 A$. Then

$$z_F F_R = \rho g \sin\theta (I_C + \bar{z}^2 A)\,,$$

and using the solution for the resultant force above, we get

$$z_F = \frac{I_C + \bar{z}^2 A}{\bar{z} A} = \bar{z} + \frac{I_C}{\bar{z} A}\,.$$

Thus, the resultant force acts below the centroid for an inclined plate.

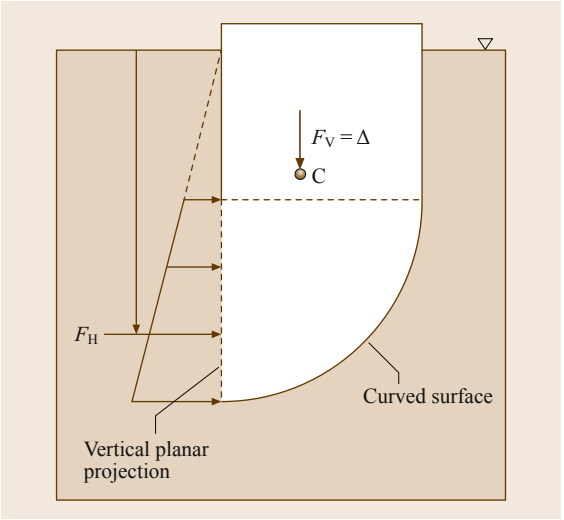


Fig. 7.47 Forces on a curved surface

Forces on Curved Surfaces

Examine the hydrostatic forces on the curved surface in Fig. 7.47. The pressure always acts normal to the surface – to find the resultant force, we can split it into a horizontal and a vertical component. The horizontal component F_H will have the same magnitude and point of application h_F as a force on the projected horizontal vertical surface. The vertical component F_V acts through the volume centroid C and corresponds to the weight of the water above the curved surface.

7.2.2 Static Stability

A floating object is considered to be hydrostatically stable when it returns to its initial state of static equilibrium when disturbed by an unbalanced force or moment. Generally, the attitude of a floating body is determined by the relative location of its center of gravity with respect to its center of buoyancy. At static equilibrium, the center of buoyancy and gravity are collinear and aligned with the vertical. Some floating objects, such as spar buoys [7.34], are designed so that their center of gravity lies below their center of buoyancy and are stable for small angular displacements. Other floating systems, such as ships are designed such that their center of gravity lies above their center of buoyancy – they are stable for small angular displacements as their hullforms are typically designed so that the center of buoyancy shifts laterally providing a restoring moment.

Consider the hull of a ship, as shown in Fig. 7.48. When an external torque is applied about its longitudinal axis it will heel to an angle ϕ as shown in Fig. 7.48a. If the ship is hydrodynamically stable the center of buoyancy will shift laterally from the point B to the

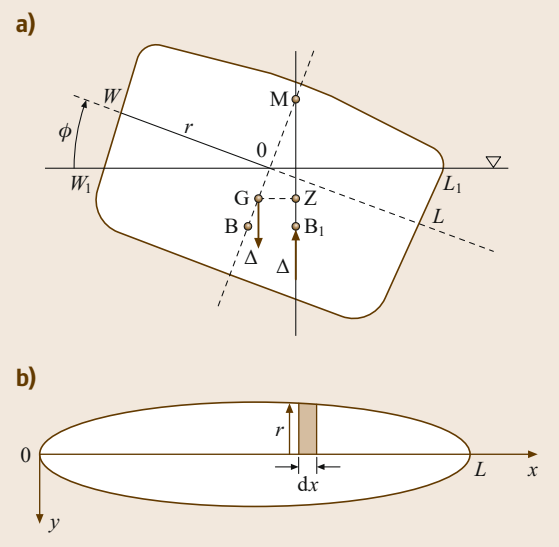


Fig. 7.48a,b Geometry for the inclined stability of a ship; **(a)** view from bow; **(b)** top view

point B_1 to provide a righting moment given by

$$RM = \Delta\overline{GZ} = \Delta\overline{GM}\sin\phi\,. \tag{7.48}$$

Here Δ is the weight (displacement) of the vessel, $\overline{GZ}$ is the righting arm and the line segment $\overline{GM}$ is the metacentric height. $\overline{GM}$ is the distance between the location of the center of gravity of the vessel G and the instantaneous metacenter M, which is the point of intersection between the vertical axis of the ship (when upright) and the line of action of gravity (when heeled).

If the ship is loaded such that M is above G, the moment will tend to right the body (positive righting moment). If M is below G, $\overline{GM} < 0$ and the body will capsize.

When an external torque is applied, perhaps through the action of wind across a deck or the placement of a weight on board, the vessel will heel. When heeled, the vessel will gain buoyancy from the newly immersed portion of the hull LOL_1, but it will also lose buoyancy owing to part of the hull which emerges from the water WOW_1. The emerged and immersed volumes will be identical for a laterally symmetric hullform. Let r be the half breadth (width) of the ship and dx be a differential element of length along the longitudinal axis of the vessel (Fig. 7.48b). The volume moment of the immersed wedge LOL_1 is given by the integral

$$\int_0^L \frac{1}{2}r(r\phi)\frac{2}{3}r\mathrm{d}x\,,$$

where L is the length of the ship. As volume moment of the emerged wedge is the same, the total volume moment of the ship when heeled is given by

$$\frac{2}{3}\int_0^L r^3\phi\mathrm{d}x\,,$$

which is $I\phi$, the product of the area moment of inertia of the waterplane about the longitudinal axis of the ship I and the heel angle ϕ. This is equivalent to the volume moment arising from the shift in the center of buoyancy, which is given by $\Delta\overline{BB_1}$, where from geometry we see that $\nabla\overline{BM}\sin\phi$, where ∇ is the volumetric displacement of the submerged hull. For small angles, $\sin\phi \approx \phi$ such that

$$\nabla\overline{BM}\phi \approx I\phi\,.$$

Often the stability of a ship or vessel is examined in terms of its initial stability, when disturbances cause small changes in ϕ. In practice, ϕ is typically taken to be less than $7{-}10^\circ$ [7.35]. Solving for $\overline{BM}$ gives

$$\overline{BM} = \frac{I}{\nabla}\,. \tag{7.49}$$

The metacentric height can then be determined from the relation

$$\overline{GM} = \overline{BM} - \overline{BG}\,, \tag{7.50}$$

where $\overline{BG}$ is determined from the geometry of the underwater hull and the known location of the center of gravity of the vessel.

7.3 Hydrodynamics

7.3.1 Flow Kinematics

Any vector field can be decomposed into a purely solenoidal (expanding/contracting) part and a rotational part. Thus, we can generally write the velocity vector as the sum of two fields. For example, in two-dimensional x–y Cartesian components we have

$$\boldsymbol{u} = \nabla\times(\psi\hat{k}) + \nabla\phi\,, \tag{7.51}$$

where ψ is the stream function, $\hat{k}$ is the unit vector perpendicular to the x–y plane, and ϕ is a scalar ve-

locity potential. It can be shown that $\nabla\cdot[\nabla\times(\psi\hat{k})]=0$ and so $\nabla\times(\psi\hat{k})$ gives the part of $\boldsymbol{u}$ with no contraction/expansion. Since $\nabla\times(\nabla\phi)=0$, $\nabla\phi$ is the part of $\boldsymbol{u}$ with no rotation. For an incompressible flow, $\nabla\cdot\boldsymbol{u}=\nabla^2\phi=0$ and for an irrotational flow $\nabla\times\boldsymbol{u}=\nabla\times[\nabla\times(\psi\hat{k})]=\nabla^2\psi=0$. Thus, for a two-dimensional incompressible, irrotational flow, both the velocity potential and stream function satisfy Laplace's equation.

7.3.2 The Navier–Stokes Equations

Material Derivative

The material derivative, sometimes also referred to as the either *total derivative* or the *substantial derivative*, is a differential operator that gives the total time rate of change of a quantity as that quantity moves through a flow field. There are essentially two contributions to the time rate of change. The first is the unsteady, time-dependent change in the quantity over time, which would happen even if the quantity were not moving with the flow. The second contribution is the change in the quantity as it moves through the flow. This second part is sometimes referred to as either the convective rate of change or the advective rate of change. For three-dimensional Cartesian coordinates, the total derivative can effectively be written as

$$\mathrm{d}f=\frac{\partial f}{\partial t}\mathrm{d}t+\frac{\partial f}{\partial x}\mathrm{d}x+\frac{\partial f}{\partial y}\mathrm{d}y+\frac{\partial f}{\partial z}\mathrm{d}z\,.$$

Taking the total derivative with respect to time gives

$$\frac{\mathrm{d}f}{\mathrm{d}t}=\frac{\partial f}{\partial t}+\frac{\partial f}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t}+\frac{\partial f}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}t}+\frac{\partial f}{\partial z}\frac{\mathrm{d}z}{\mathrm{d}t}\,,$$

which can be written compactly as

$$\frac{\mathrm{d}f}{\mathrm{d}t}=\underbrace{\frac{\partial f}{\partial t}}_{\substack{\text{unsteady, local}\\ \text{rate of change}}}+\underbrace{\boldsymbol{u}\cdot\nabla f}_{\substack{\text{rate of change}\\ \text{due to convection}}}\,, \tag{7.52}$$

where the gradient operator in three-dimensional Cartesian coordinates is

$$\nabla(\,)=\frac{\partial(\,)}{\partial x}\hat{i}+\frac{\partial(\,)}{\partial y}\hat{j}+\frac{\partial(\,)}{\partial z}\hat{k}\,,$$

and the velocity is

$$\boldsymbol{u}=\frac{\mathrm{d}x}{\mathrm{d}t}\hat{i}+\frac{\mathrm{d}y}{\mathrm{d}t}\hat{j}+\frac{\mathrm{d}z}{\mathrm{d}t}\hat{k}\,.$$

Example 7.7

In a river with a velocity field $\boldsymbol{v}(\boldsymbol{x},t)$ with respect to the ground, the temperature is given by the function $f(\boldsymbol{x},t)$. A small boat is in the river and has a velocity $\boldsymbol{U}(t)$ with respect to the ground:

1. Give the time rate of change of f experienced by the boat when forced to move at the velocity $\boldsymbol{U}$

$$\frac{\mathrm{d}f}{\mathrm{d}t}=\frac{\partial f}{\partial t}+\boldsymbol{U}\cdot\nabla f\,.$$

2. If the engine is turned off and the vessel freely drifts with the surrounding current, give the time rate of change of f experienced by the boat

$$\frac{\mathrm{d}f}{\mathrm{d}t}=\frac{\partial f}{\partial t}+\boldsymbol{v}\cdot\nabla f\,.$$

3. The boat is anchored. Give the time rate of change of f experienced by the vessel

$$\frac{\mathrm{d}f}{\mathrm{d}t}=\frac{\partial f}{\partial t}\,.$$

Conservation of Mass

The integral form of conservation of mass for a fluid volume V enclosed by surface S is

$$\underbrace{\frac{\partial}{\partial t}\int_V\rho\mathrm{d}V}_{\substack{\text{rate of change}\\ \text{of mass in } V}}=\underbrace{-\int_S\rho\boldsymbol{u}\cdot\mathrm{d}\boldsymbol{S}}_{\substack{\text{flux of mass}\\ \text{across surface of } V}}\,. \tag{7.53}$$

This is a mathematical statement that the time rate of change of mass within V is equal to the net flux of mass into or out of V, which is enclosed by the surface S (Fig. 7.49). The differential form obtained through the application of Gauss' theorem (divergence theorem) to the integral above is

$$\frac{\partial\rho}{\partial t}+\nabla\cdot(\rho\boldsymbol{u})=\frac{\mathrm{d}\rho}{\mathrm{d}t}+\rho\nabla\cdot\boldsymbol{u}=0\,. \tag{7.54}$$

For a flow of constant density, this reduces to

$$\nabla\cdot\boldsymbol{u}=0\,. \tag{7.55}$$

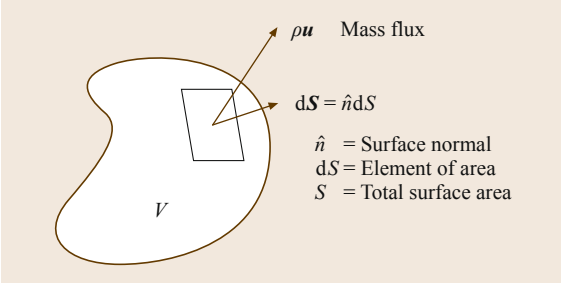


Fig. 7.49 Fluid volume V for formulation of mass conservation

Conservation of Momentum

This is basically a statement of Newton's second law in a fluid flow. Mass per unit volume multiplied by fluid acceleration (the material derivative of the velocity) is the sum of applied forces. In an incompressible fluid, the applied forces can consist of viscous shear and pressure acting at the surface of a fluid element, or body forces acting on the volume of the fluid element, giving

$$\underbrace{\rho\frac{\mathrm{d}\boldsymbol{u}}{\mathrm{d}t}}_{\text{mass}\times\text{acceleration}} = \underbrace{\underbrace{-\nabla p+\mu\nabla^2\boldsymbol{u}}_{\text{pressure and viscous forces}}+\underbrace{\rho\boldsymbol{g}}_{\text{body forces}}}_{\text{applied external forces}} . \tag{7.56}$$

The integral form of conservation of momentum in the volume V (Fig. 7.50) can be written as

$$\underbrace{\frac{\partial}{\partial t}\int\limits_V \rho\boldsymbol{u}\mathrm{d}V}_{\text{time rate of change of momentum in V}} + \underbrace{\int\limits_S \rho\boldsymbol{u}(\boldsymbol{u}\cdot\mathrm{d}\boldsymbol{S})}_{\text{flux of momentum into or out of V}}$$
$$= \underbrace{\underbrace{\int\limits_V \rho\boldsymbol{g}\mathrm{d}V}_{\text{body forces}} + \underbrace{\int\limits_S \boldsymbol{\tau}\cdot\mathrm{d}\boldsymbol{S}}_{\text{pressure and viscous stress terms}}}_{\text{applied external forces}} , \tag{7.57}$$

where $\boldsymbol{\tau}$ is the second-order stress tensor

$$\boldsymbol{\tau} \equiv p\mathbf{I}+\mu[\nabla\boldsymbol{u}+(\nabla\boldsymbol{u})^{\mathrm{T}}] , \tag{7.58}$$

$\mathbf{I}$ is the identity matrix, T denotes the transpose, and $\nabla\boldsymbol{u}$ is the gradient of the velocity vector (also a second-order tensor).

The integral form of the conservation laws is generally used when one is trying to determine the effect of a volume of flow. The differential forms are used when trying to determine the distribution of quantities, such as velocity or pressure within a flow.

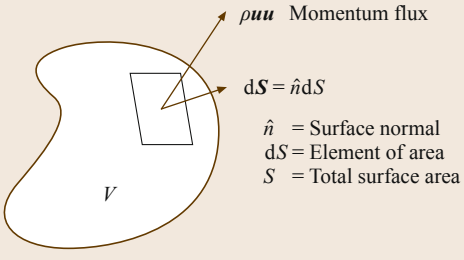


Fig. 7.50 Fluid volume V for formulation of momentum conservation

Example 7.8

An experimental underwater magnetohydrodynamic drive unit is being developed (Fig. 7.51). The drive consists of a long, deep, and narrow channel with magnets mounted in the channel walls. Actuation of the magnets drives seawater through the channel with a body force per unit mass of $\boldsymbol{f}$. Assume that $\boldsymbol{f}$ is constant throughout the channel. Find the velocity profile of the flow in the channel.

Solution 7.7

Let $\boldsymbol{f} = f\hat{i}$. For constant f (in time) the flow is steady. As the channel is long and deep, we can assume that $w = 0$ and $\partial u/\partial x = 0$. Since both ends of the channel are open to the surrounding seawater, we can also assume that there is no pressure gradient in the x direction. Apply the no slip condition on the top and bottom walls (Fig. 7.52).

- Continuity:

$$\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}=0\rightarrow\frac{\partial v}{\partial y}=0\rightarrow v = \text{constant} = 0 .$$

- X-momentum:

$$\rho\left(\frac{\partial u}{\partial t}+u\frac{\partial u}{\partial x}+v\frac{\partial u}{\partial y}\right) = -\frac{\partial p}{\partial x}+\rho f+\mu\left(\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}\right)$$
$$\rightarrow \frac{\partial^2 u}{\partial y^2} = -\frac{f}{\nu} .$$

Integrating twice with respect to y gives

$$u = -\frac{f}{2\nu}y^2+c_0 y+c_1 .$$

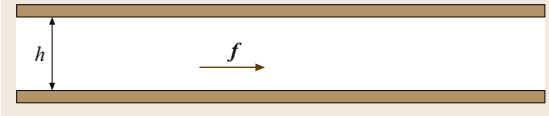


Fig. 7.51 Channel of the magnetohydrodynamic drive unit

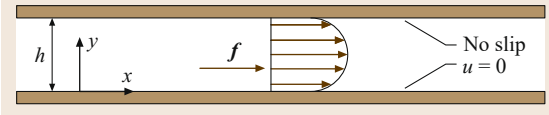


Fig. 7.52 Velocity profile in the magnetohydrodynamic drive unit

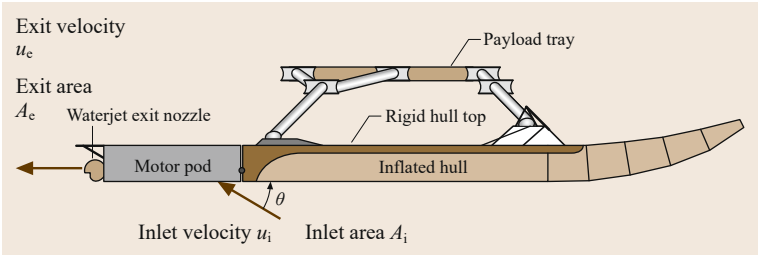


Fig. 7.53 A water-jet propelled catamaran (courtesy of Marine Advanced Research of Berkeley, USA)

Applying the boundary condition for u at $y = 0$ we get $c_1 = 0$ and using the boundary condition at $y = h$ gives $c_0 = fh/(2\nu)$. Thus, we find

$$\boxed{u = \frac{fy}{2\nu}(h - y)\,.}$$

Example 7.9

The thrust developed to propel the catamaran (twin hull boat) shown below is a result of the water pumped through the vehicle and exiting as a high-speed water jet. For the conditions shown in Fig. 7.53, find an expression for the exit velocity required to produce a total thrust T if the catamaran is moving at a constant velocity? Assume the outlet jets of water are free jets, and that the inlet jet is at an angle θ, with respect to the horizontal.

Solution 7.8

From the momentum equation (steady inviscid flow) we have

$$\int_S (\rho \boldsymbol{u})\boldsymbol{u} \cdot \mathrm{d}\boldsymbol{S} = \boldsymbol{F}_{\mathrm{ext}}\,,$$

where $\boldsymbol{F}_{\mathrm{ext}}$ represents the external forces acting on the control volume (Fig. 7.54) including the reaction force of the thrust, the weight, buoyancy, and a vertical component of the incoming momentum. The balance of forces across the control volume in the x-direction gives

$$\frac{T}{2} = \rho u_e^2 A_e - \rho u_i^2 A_i \cos\theta\,.$$

The mass flow rate is $\dot{m} = \rho u_e A_e = \rho u_i A_i$. From the continuity equation we have $\int_S \boldsymbol{u} \cdot \mathrm{d}\boldsymbol{S} = 0$ over the control surface, giving $u_e A_e = u_i A_i$. Substituting this into the equation for T, we can solve for the exit velocity

$$u_e = \sqrt{\frac{T}{2\rho A_e\left(1 - \frac{A_e}{A_i}\cos\theta\right)}}\,.$$

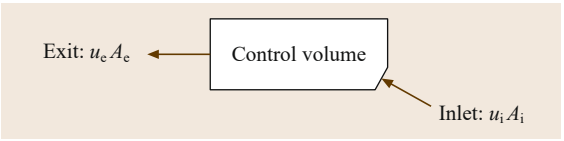


Fig. 7.54 Control volume for the waterjet

Special Forms of the N–S Equations

Rotating coordinate system [7.36]: In geophysical fluid applications positions and velocities are measured with respect to a reference frame (coordinate system) rotating with the Earth. The true acceleration, in an inertial reference frame, must account for the centripetal and Coriolis acceleration in the rotating reference frame

$$\rho\frac{\mathrm{d}\boldsymbol{u}}{\mathrm{d}t} = -\nabla p + \mu\nabla^2\boldsymbol{u} + \rho\left[\boldsymbol{g} + \underbrace{\boldsymbol{\Omega}\times\boldsymbol{\Omega}\times\boldsymbol{r}}_{\text{centripetal acceleration}} - \underbrace{2\boldsymbol{\Omega}\times\boldsymbol{u}}_{\text{Coriolis acceleration}}\right], \tag{7.59}$$

where $\boldsymbol{\Omega}$ is the angular velocity vector corresponding to the Earth's rotation, $\boldsymbol{r}$ is a position vector from the center of the Earth to a point on the surface where the acceleration is being calculated. The centripetal acceleration is largest at the equator and diminishes toward the North and South Poles. The Coriolis force tends to deflect the trajectory of a fluid particle to the right of its direction of travel in the northern hemisphere and to the left of its direction of travel in the south of the equator.

The Vortex Twisting and Tilting Equation (Vorticity Transport Equation). Under the Boussinesq approximations, which assumes that the fluid density is not a strong function of pressure (so that small changes in density are caused by small changes in temperature) and that the dynamic viscosity and thermal conductivity of the fluid are constant [7.36], the vorticity transport

equation can be written as

$$\underbrace{\frac{\mathrm{d}\boldsymbol{\omega}}{\mathrm{d}t}}_{\substack{\text{rate of change}\\ \text{of vorticity}}} = \underbrace{\boldsymbol{\omega}\cdot\nabla\boldsymbol{u}}_{1} + \underbrace{2\boldsymbol{\Omega}\cdot\nabla\boldsymbol{u}}_{2} + \underbrace{\frac{1}{\rho^2}\nabla\rho\times\nabla p}_{3} + \underbrace{\nu\nabla^2\boldsymbol{\omega}}_{4}. \tag{7.60}$$

Here:

1. This term is sometimes called the *twisting and tilting term* and gives the generation of vorticity from the stretching/compression of vortex lines, as well as, vorticity that is increased or decreased as vortex lines are tilted in a shear flow.
2. This term gives the change in vorticity that occurs as vortex lines are tilted into or away from the axis of rotation in a rotating coordinate system (as in a geostrophic flow).
3. This term gives the changes in vorticity arising from baroclinic torque (when $\nabla\rho$ and ∇p are not collinear). This term can arise, for example, when there is a disturbance along a therm in an ocean. The density gradient of the associated pycnocline can become misaligned with the pressure gradient caused by the hydrostatic pressure, causing an internal wave to propagate along the thermocline.
4. The last term is the rate of change of vorticity caused by viscous diffusion.

Euler's Equation.

$$\rho\frac{\mathrm{d}\boldsymbol{u}}{\mathrm{d}t} = -\nabla p + \rho\boldsymbol{g} \tag{7.61}$$

Bernoulli's Equation

Under certain assumptions, Bernoulli's equation relates quantities at different points in a flow. Representing the gravitational force as a conservative force in a Cartesian coordinate system such that its potential is $\Pi = gz$ and $\boldsymbol{g} = -\nabla\Pi$, the general form for a fluid of constant density is

$$\frac{\partial\boldsymbol{u}}{\partial t} + \nabla\left(\frac{\boldsymbol{u}^2}{2} + gz + \frac{p}{\rho}\right) = \boldsymbol{u}\times\boldsymbol{\omega}\,, \tag{7.62}$$

where $\boldsymbol{u}$ is the fluid velocity and $\boldsymbol{\omega}$ is the vorticity. Let

$$B = \left(\frac{\boldsymbol{u}^2}{2} + gz + \frac{p}{\rho}\right),$$

be the Bernoulli constant. Bernoulli's equation can be simplified for several different conditions:

1. For steady flow

$$\nabla\left(\frac{\boldsymbol{u}^2}{2} + gz + \frac{p}{\rho}\right) = \boldsymbol{u}\times\boldsymbol{\omega}\,.$$

The streamlines and vortex lines form surfaces. Changes in B are perpendicular to these surfaces.
2. For irrotational flows, $\boldsymbol{\omega} = 0$ and B is constant everywhere.

$$\left(\frac{\boldsymbol{u}^2}{2} + gz + \frac{p}{\rho}\right) = B = \text{constant}$$

3. If the flow is unsteady, but irrotational, such that the velocity can be represented as the gradient of a scalar potential $\boldsymbol{u} = \nabla\phi$, Bernoulli's equation reduces to

$$\frac{\partial\boldsymbol{u}}{\partial t} + \nabla B = \frac{\partial(\nabla\phi)}{\partial t} + \nabla B = \nabla\left(\frac{\partial\phi}{\partial t}\right) + \nabla B = \nabla\left(\frac{\partial\phi}{\partial t} + B\right) = 0\,.$$

As the gradient of the term in brackets above is zero, the term in brackets must either be a function of time or a constant. In general, we have

$$\frac{\partial\phi}{\partial t} + B = F(t)\,,$$

which is a form of the Bernoulli equation often used to solve problems in offshore wave mechanics [7.11].

7.3.3 Flow of an Ideal Fluid

As shown above, the vorticity transport equation (7.60) can be used to show that in a high Reynolds number flow (where viscosity becomes negligible) an initially irrotational flow will remain irrotational ($\mathrm{d}\boldsymbol{\omega}/\mathrm{d}t = 0$).

Assuming that there is no separation, any rotational flow around solid surfaces will be confined to thin boundary layers and wakes. For two-dimensional (or axisymmetric), inviscid, incompressible flow, potential flow theory is a good approximation to a real flow. Models of the flow around complex bodies can be built from adding (superposing) simple solutions for the stream function ψ and velocity potential ϕ.

As shown in Sect. 7.3.1, the stream function ensures conservation of mass

$$\nabla\cdot\boldsymbol{u} = \nabla\cdot[\nabla\times(\psi\hat{k})] = 0\,.$$

In two-dimensional Cartesian components, the velocity can be written as the curl of the stream function $\psi\hat{k}$

$$\boldsymbol{u} = u\hat{i} + v\hat{j} = \frac{\partial\psi}{\partial y}\hat{i} - \frac{\partial\psi}{\partial x}\hat{j} .$$

In two-dimensional polar coordinates, this is

$$\boldsymbol{u} = u_r\hat{r} + u_\theta\hat{\theta} = \frac{1}{r}\frac{\partial\psi}{\partial\theta}\hat{r} - \frac{\partial\psi}{\partial r}\hat{\theta} .$$

Likewise, the velocity potential ensures the irrotationality of the flow

$$\nabla\times\boldsymbol{u} = \nabla\times(\nabla\phi) = 0 .$$

In two-dimensional Cartesian components, the velocity can be written as the gradient of a scalar potential ϕ

$$\boldsymbol{u} = u\hat{i} + v\hat{j} = \frac{\partial\phi}{\partial x}\hat{i} + \frac{\partial\phi}{\partial y}\hat{j} .$$

In two-dimensional polar coordinates, this is

$$\boldsymbol{u} = u_r\hat{r} + u_\theta\hat{\theta} = \frac{\partial\phi}{\partial r}\hat{r} + \frac{1}{r}\frac{\partial\phi}{\partial\theta}\hat{\theta} .$$

For irrotational, incompressible flow, both the velocity potential and stream function satisfy Laplace's equation

$$\nabla^2\phi = 0$$

and

$$\nabla^2\psi = 0 ,$$

which is linear, such that solutions can be superposed

$$\nabla^2(\psi_1 + \psi_2) = \nabla^2\psi_1 + \nabla^2\psi_2$$

and

$$\nabla^2(\phi_1 + \phi_2) = \nabla^2\phi_1 + \nabla^2\phi_2 .$$

Solutions for ϕ and ψ can be superposed in the complex plane to find a complex velocity potential w, such that $w = \phi + \mathrm{i}\psi$, where $\mathrm{i} = \sqrt{-1}$ is the imaginary number and

$$\frac{\mathrm{d}w}{\mathrm{d}z} = u - \mathrm{i}v ,$$

gives the velocity component in two-dimensional Cartesian coordinates. Here, the complex coordinate $z = x + \mathrm{i}y = re^{\mathrm{i}\theta}$.

As Laplace's equation is second order, solutions generally require two boundary conditions in each of the independent variables (either x and y in Cartesian coordinates or r and θ in polar coordinates):

1. The velocity of the fluid at solid surfaces must match the velocity of the solid surface. Without loss of generality the problem can be formulated such that the coordinate system is in the reference frame of the surface, such that the velocity is zero. With $\hat{n}$ as a unit vector perpendicular to the surface and $\hat{s}$ as a unit vector tangent to the surface, we have

$$\frac{\partial\phi}{\partial n} = 0 \text{ and } \frac{\partial\psi}{\partial s} = 0 .$$

2. Far from the surface of the body, the velocity must approach its freestream value U, such that

$$\frac{\partial\phi}{\partial x} = U \text{ and } \frac{\partial\psi}{\partial y} = U .$$

The flow velocity can be found from solutions to Laplace's equation for the velocity potential and stream function. However, these solutions are based purely on the geometry of the flow (kinematic) and cannot be directly used to determine the forces acting on objects embedded in the flow. In order to solve for these fluidic forces, Bernoulli's equation must be used to determine the pressure around an object in the flow – by integrating the pressure around an object, one can then determine the fluidic forces acting on it.

The steady, irrotational, and incompressible form of Bernoulli's equation is often used to solve potential flow problems

$$\frac{\boldsymbol{u}^2}{2} + \mathrm{g}z + \frac{p}{\rho} = \frac{(\nabla\phi)^2}{2} + \mathrm{g}z + \frac{p}{\rho} = B , \qquad (7.63)$$

where B is constant everywhere in the flow field.

Catalog of Flow Fields

Complex flows around objects can be built from the superposition of very simple flows. Here, a reference list of some of the most basic flows is provided as a guide:

1. Simple flows:
   a) Uniform flow (Fig. 7.55)
      The velocity potential ϕ and stream function can be determined by quadrature $\partial\phi/\partial x = u = U$. We can integrate this with respect to x to find $\phi = Ux + f_1(y)$. Then, $v = \partial\phi/\partial y = \mathrm{d}f_1/\mathrm{d}y = 0$. Integrating with respect to y, we find that $f_1(y) = C$, where C is a constant. Without loss of generality we can let $C = 0$. Therefore, $\phi = Ux$ is the velocity potential for a uniform flow. Similarly, we can solve for the stream function by integrating the relation $\partial\psi/\partial y = u = U$ to

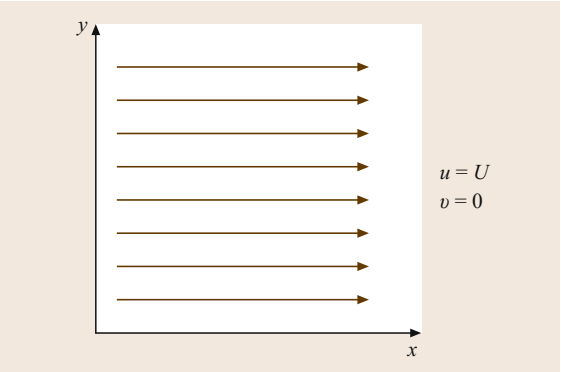


Fig. 7.55 Velocity in a uniform flow field

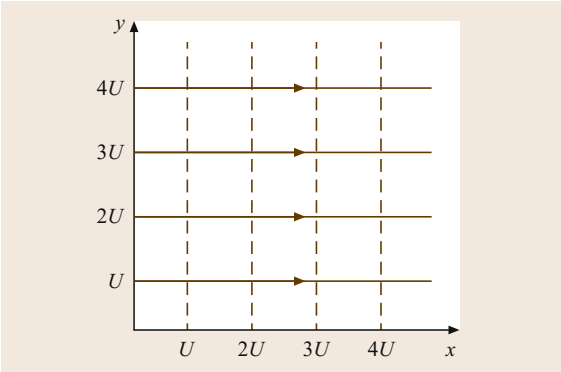


Fig. 7.56 Potential flow lines (*dashed lines*, ϕ) and streamlines (*continuous lines*, ψ) for uniform flow. The directions of the *arrows* on the streamlines are reversed when $U < 0$

get $\psi = Uy + f_2(x)$. Differentiating this expression with respect to x gives the y-component of the velocity $v = -\partial\psi/\partial x = -\mathrm{d}f_2/\mathrm{d}x = 0$. Again, upon integration we find that $f_2(y) = \mathrm{C}$ and without loss of generality we can let $C = 0$ so that $\psi = Uy$.

The complex velocity potential is then given by $w = \phi + \mathrm{i}\psi = Ux + \mathrm{i}Uy = Uz$. If we plot the velocity potential and stream function in the complex plane, we can see that they form straight lines (Fig. 7.56).

b) Source/sink flow (line source)

In this type of flow, fluid moves either radially outward (source) or inward (sink) only (Fig. 7.57). The radial velocity can be related to the strength of the source/sink through the volumetric flow rate per unit length in the out of plane direction $m = 2\pi r u_\mathrm{r}$, so that the radial velocity is given by

$$u_\mathrm{r} = \frac{m}{2\pi r} = \frac{\partial\phi}{\partial r} \,.$$

The tangential velocity is taken to be zero everywhere $u_\theta = 0$. As above, we can integrate to find the velocity potential. This operation will result in a constant of integration that can be determined by looking at the tangential velocity and again without loss of generality the constant of integration can be taken to be zero. The resulting velocity potential is found to be

$$\phi = \frac{m}{2\pi} \ln r \,.$$

The corresponding stream function and complex velocity potential can be shown to be

$$\psi = \frac{m\theta}{2\pi}$$

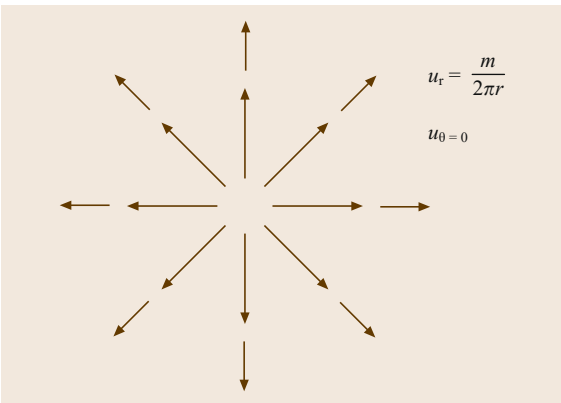


Fig. 7.57 Velocity profile in a source flow

and

$$\begin{aligned} w(z) &= \frac{m}{2\pi}(\ln r + i\theta) \\ &= \frac{m}{2\pi} \ln(r\mathrm{e}^{i\theta}) \\ &= \frac{m}{2\pi} \ln(z) \,. \end{aligned}$$

When $m > 0$, $u_\mathrm{r} > 0$ and the flow is a source flow. When $m < 0$, $u_\mathrm{r} < 0$ the flow is a sink flow. An examination of the velocity potential and stream function reveals that level curves of constant ψ (streamlines) are rays and level curves of constant ϕ are circles (Fig. 7.58).

c) Irrotational line vortex flow

The velocity field produced by a counterclockwise irrotational line vortex of circulation strength Γ (Fig. 7.59) is given by

$$u_\mathrm{r} = 0 \,, \quad u_\theta = \frac{\Gamma}{2\pi r} \,.$$

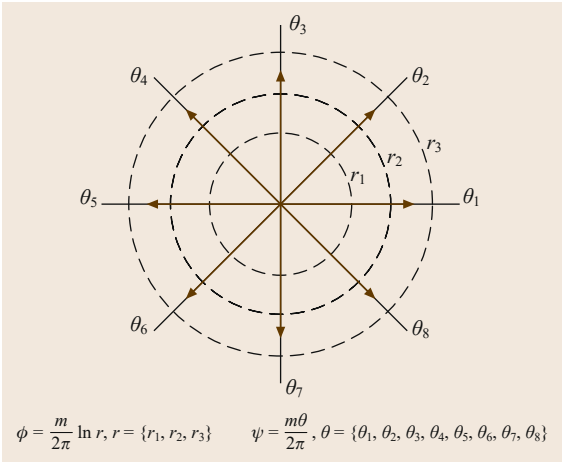


Fig. 7.58 Potential flow lines (*dashed lines*, ϕ) and streamlines (*continuous lines*, ψ) for a source flow. The directions of the *arrows* on the streamlines are reversed for a sink flow, when $m < 0$

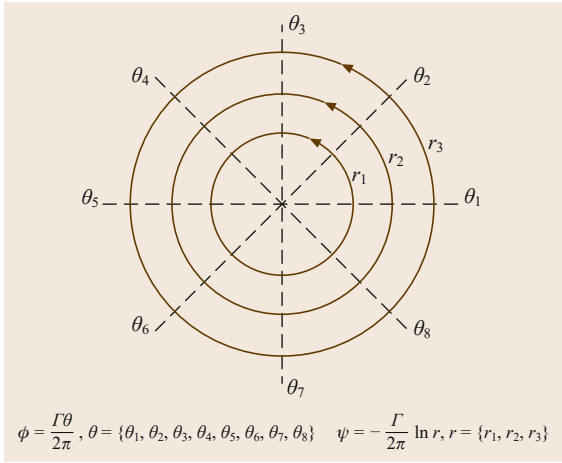


Fig. 7.60 Potential flow lines (*dashed lines*, ϕ) and streamline (*continuous lines*, ψ) for an irrotational line vortex flow. The directions of the *arrows* on the streamlines is reversed when the circulation is negative $\Gamma < 0$

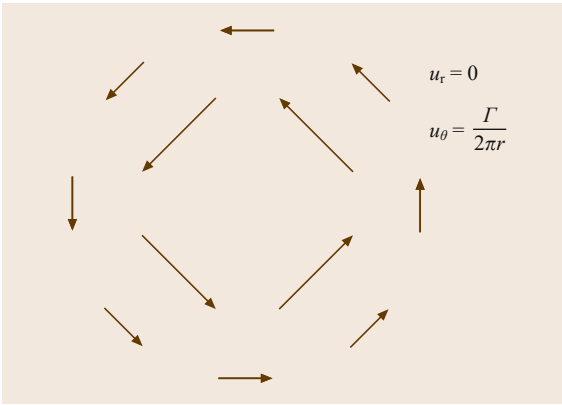


Fig. 7.59 Velocity vectors in an irrotational line vortex flow. The directions of the arrows on the streamlines is reversed when the circulation is negative $\Gamma < 0$

Integrating the velocity as above to solve for ϕ, ψ and w gives

$$\phi = \frac{\Gamma\theta}{2\pi}\,, \quad \psi = -\frac{\Gamma}{2\pi}\ln r\,, \quad w = -\frac{i\Gamma}{2\pi}\ln(z)\,.$$

An examination of the velocity potential and stream function reveals that level curves of constant ψ (streamlines) are circles and level curves of constant ϕ are rays (Fig. 7.60).

2. Superposed flows:
   a) Rankine half-body
   Add a source flow to a uniform freestream flow – for the velocity potential we get

$$\phi = Ux + \frac{m}{2\pi}\ln r = Ur\cos\theta + \frac{m}{2\pi}\ln r$$

and for the stream function we have

$$\psi = Uy + \frac{m\theta}{2\pi} = Ur\sin\theta + \frac{m\theta}{2\pi}\,. \qquad (7.64)$$

Superposed flows, such as this one, often form the outline of bodies in a flow. As stagnation points (where both components of the 2-D velocity vanish) typically lie on the surface of a body, one can determine the equation for the streamlines that pass through the body's surface by first identifying the stagnation points. The stagnation points occur at locations where $u_r = 0$ and $u_\theta = 0$. Using the expression for the stream function above, we have

$$u_r = \frac{1}{r}\frac{\partial\psi}{\partial\theta} = U\cos\theta + \frac{m}{2\pi r}$$

and

$$u_\theta = -\frac{\partial\psi}{\partial r} = -U\sin\theta\,.$$

We see that $u_\theta = 0$ when $\theta = 0$ or π. If we take $\theta = 0$, the corresponding solution for u_r would give that $r < 0$, which is nonphysical. Thus, we take the other solution giving the location of the stagnation point as

$$\theta = \pi\,, \quad r = \frac{m}{2\pi U}\,.$$

The streamline that passes through the stagnation point forms the outline of the half-body. We

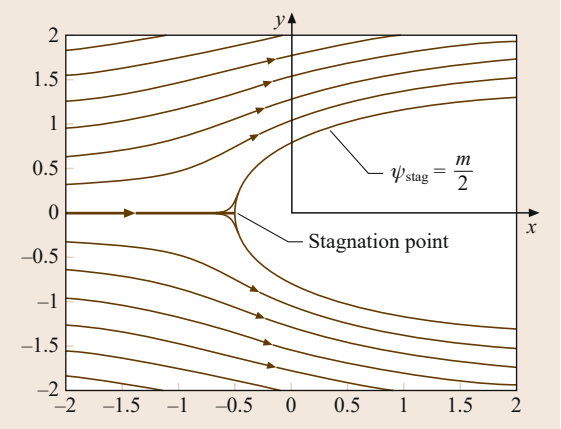


Fig. 7.61 Rankine half-body flow

can find the value of the stream function ψ_{stag} by evaluating (7.64) at the stagnation point.

$$\psi_{stag} = U\frac{m}{2\pi U}\sin\pi + \frac{m\pi}{2\pi} = \frac{m}{2}$$

The values of r and θ lying along the streamline are given by

$$Ur\sin\theta + \frac{m\theta}{2\pi} = \frac{m}{2}\,,$$

for $0 \le \theta \le \pi$. The $\frac{1}{2}$ width of the body h can be determined by examining the value of ψ_{stag} for $\theta \to 0$; using (7.64) it can be seen that $m/2 = Uh$ so that

$$h = \frac{m}{2U}\,.$$

The streamlines of the Rankine half-body flow are shown in Fig. 7.61.

b) Source and sink of equal strength
As in Fig. 7.62, let a source be placed at the location $(x, y) = (-a, 0)$ and a sink be placed at the location $(x, y) = (+a, 0)$. The combined velocity potential is

$$\phi = \frac{m}{2\pi}\ln r_1 - \frac{m}{2\pi}\ln r_2 = \frac{m}{2\pi}\ln\left(\frac{r_1}{r_2}\right)$$
$$= \frac{m}{2\pi}\ln\left(\frac{(x+a)^2+y^2}{(x-a)^2+y^2}\right)^{1/2},$$

because $r_1^2 = (x+a)^2 + y^2$ and $r_2^2 = (x-a)^2 + y^2$.
Similarly,

$$\psi = \frac{m\theta_1}{2\pi} - \frac{m\theta_2}{2\pi} = -\frac{m}{2\pi}(\theta_2 - \theta_1)\,.$$

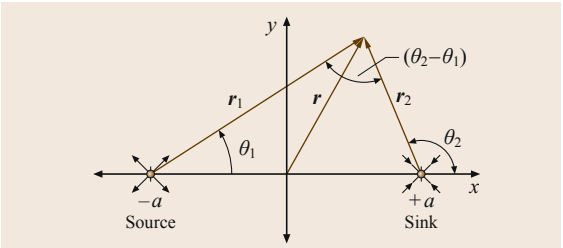


Fig. 7.62 Geometry of the source/sink flow

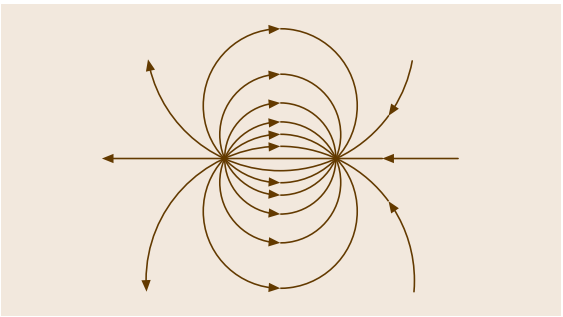


Fig. 7.63 Sketch of the flow generated by a superposed source and sink

ψ can be recast in Cartesian coordinates as

$$\psi = -\frac{m}{2\pi}\tan^{-1}\left(\frac{2ay}{x^2+y^2-a^2}\right).$$

The streamlines for the source–sink flow are shown in Fig. 7.63.

c) Rankine body
A Rankine body can be constructed by combining a source and sink of equal strength with a uniform freestream flow. The velocity potential ϕ and stream function ψ are

$$\phi = Ux + \frac{m}{2\pi}\ln\left(\frac{(x+a)^2+y^2}{(x-a)^2+y^2}\right)^{1/2},$$
$$\psi = Uy - \frac{m}{2\pi}\tan^{-1}\left(\frac{2ay}{x^2+y^2-a^2}\right).$$

As above the equation of the streamline forming the Rankine body can be determined by finding the stagnation points and then solving for the value of the stream function that passes through them. The resulting equation for the streamline is

$$\tan\left(\frac{2\pi Uy}{m}\right) = \frac{2ay}{x^2+y^2-a^2}\,.$$

The streamlines for a Rankine body flow are shown in Fig. 7.64.

d) Doublet
Let $a \to 0$ in the source/sink flow above, we have

$$\phi = \frac{m}{2\pi} \ln\left(\frac{r_1}{r_2}\right).$$

From geometry, $r_1 = 2a\cos\theta_1 + r_2\cos(\theta_2 - \theta_1)$. Therefore,

$$\phi = \frac{m}{2\pi} \ln\left(\frac{2a\cos\theta_1}{r_2} + \cos(\theta_2 - \theta_1)\right).$$

When $a \to 0$ but m is increased such that ma/π remains finite and constant, $r_2 \to r$, $\theta_2 \to \theta$, and $\cos(\theta_2 - \theta_1) \to 1$ so that

$$\lim_{a\to 0} \phi = \frac{m}{2\pi} \ln\left(\frac{2a\cos\theta}{r} + 1\right).$$

A Taylor's series expansion of the argument about 1 gives

$$\phi = \frac{ma\cos\theta}{\pi r} = \frac{\mu x}{x^2 + y^2}.$$

Similarly, for small angles $\tan\theta \approx \theta$ and

$$\psi = -\frac{m}{2\pi}\left(\frac{2ay}{x^2 + y^2}\right) = -\frac{ma}{\pi}\left(\frac{y}{x^2 + y^2}\right),$$
$$= -\frac{\mu y}{x^2 + y^2}.$$

With $r^2 = x^2 + y^2$ so that $\cos\theta = x/r$ and $\sin\theta = y/r$,

$$\phi = \frac{\mu\cos\theta}{r} \text{ and } \psi = -\frac{\mu\sin\theta}{r}.$$

The complex velocity potential is then

$$w(z) = \phi + \mathrm{i}\psi = \frac{\mu}{r}(\cos\theta - \mathrm{i}\sin\theta)$$
$$= \frac{\mu \mathrm{e}^{\mathrm{i}\theta}}{r} = \frac{\mu}{z}.$$

The streamlines for this flow are shown in Fig. 7.65.
Note that a doublet can also be produced from the superposition of a line vortex of circulation strength $-\Gamma$ and infinitesimal distance above the origin with a line vortex of circulation strength $+\Gamma$ and infinitesimal distance below the origin.

e) Uniform flow + doublet
Adding the stream functions for a uniform flow and doublet gives

$$\psi = Ur\sin\theta - \frac{\mu\sin\theta}{r},$$

where $\mu = ma/\pi$, as above. Solving for the tangential and radial velocities in the flow gives

$$u_\mathrm{r} = \frac{1}{r}\frac{\partial\psi}{\partial\theta} = \frac{1}{r}\left(Ur - \frac{\mu}{r}\right)\cos\theta,$$

and

$$u_\theta = -\frac{\partial\psi}{\partial r} = -\left(U + \frac{\mu}{r^2}\right)\sin\theta,$$

respectively. The stagnation points are located where $\theta = 0$ and $\theta = \pi$ so that $u_\theta = 0$ and

$$Ur - \frac{\mu}{r} = 0 \to r = \sqrt{\frac{\mu}{U}} \equiv R,$$

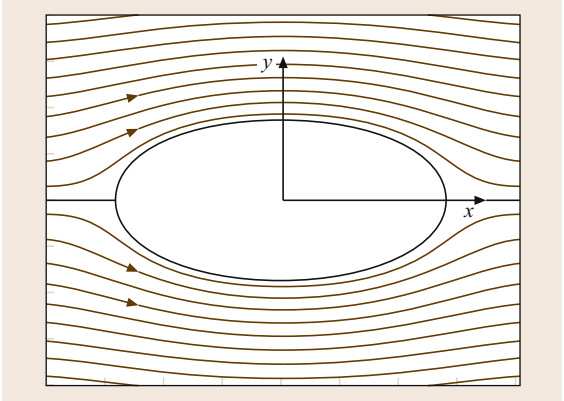


Fig. 7.64 Rankine body flow

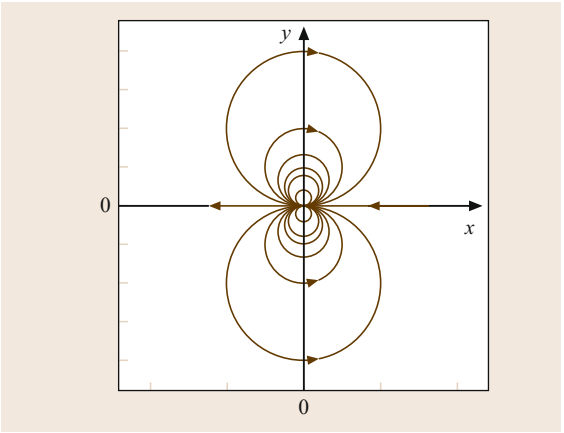


Fig. 7.65 A doublet flow

so that $u_r = 0$. We can define the stagnation streamline such that $\psi = \psi_{stag} = 0$, which gives

$$\psi_{stag}\big|_{r=R} = \frac{1}{R}(UR^2 - \mu)\sin\theta = 0\,,$$

for all θ. Thus, the body defined by ψ_{stag} is a circle of radius R. We can re-write ψ as

$$\psi = Ur\sin\theta\left(1 - \frac{R^2}{r^2}\right), \quad \text{for } r \geq R\,.$$

This gives the flow over a circular cylinder (Fig. 7.66).

Note that the forces acting on the cylinder can be determined by first using the incompressible, irrotational form of Bernoulli's equation (7.63) to determine the pressure everywhere on the surface of the cylinder and then integrating the pressure across the cylinder's surface.

The Magnus Effect

A rotating body in a freestream flow will produce a force transverse to the direction of the freestream. This can be seen by superposing a uniform flow past a circular cylinder with circulation, that is, a cylinder spinning about its axis of symmetry (Fig. 7.67).

Normally, the thin boundary layer about a spinning body would be responsible for creating the circulation; in this inviscid flow we have introduced the circulation using a line vortex.

$$\psi = Ur\sin\theta\left(1 - \frac{R^2}{r^2}\right) - \frac{\Gamma}{2\pi}\ln r + C\,,$$

where we set the constant C to $C = \Gamma/(2\pi)\ln R$ to make $\psi = \psi_s = 0$ on the cylinder surface $(r = R)$.

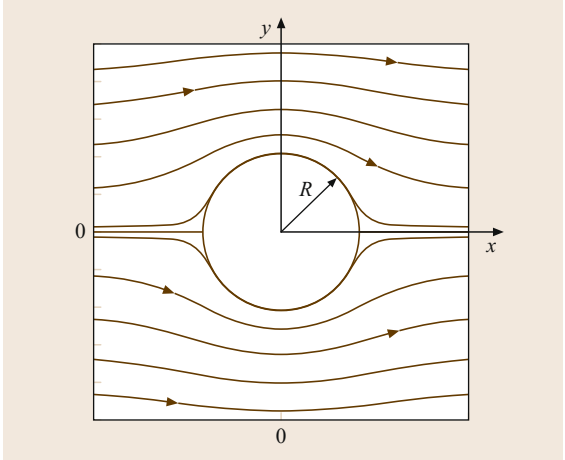


Fig. 7.66 Flow about a circular cylinder

Then

$$\psi = Ur\sin\theta\left(1 - \frac{R^2}{r^2}\right) - \frac{\Gamma}{2\pi}\ln\left(\frac{r}{R}\right).$$

The associated radial and tangential velocities are

$$u_r = \frac{1}{r}\frac{\partial\psi}{\partial\theta} = U\cos\theta\left(1 - \frac{R^2}{r^2}\right)$$

and $$u_\theta = -\frac{\partial\psi}{\partial r} = -U\sin\theta\left(1 + \frac{R^2}{r^2}\right) + \frac{\Gamma}{2\pi r}\,,$$

respectively. On $r = R$, $u_r = 0$ and $u_\theta = -2U\sin\theta + \Gamma/(2\pi R)$ (Fig. 7.68).

The pressure on the surface of the cylinder can be determined using Bernoulli's equation (7.63)

$$p_\infty + \rho\frac{U^2}{2} = p_0 = p(\theta) + \rho\frac{u_\theta^2}{2}\,,$$

where p_0 is the pressure at the stagnation point and p_∞ is the freestream pressure.

Using the value of u_θ determined above, the pressure on the cylinder is

$$\begin{aligned} p(\theta) &= p_0 - \rho\frac{u_\theta^2}{2} \\ &= p_0 - 2\rho U^2\sin^2\theta + \frac{\rho U\sin\theta\,\Gamma}{\pi R} - \frac{\rho\Gamma^2}{8\pi^2R^2}\,. \end{aligned} \tag{7.65}$$

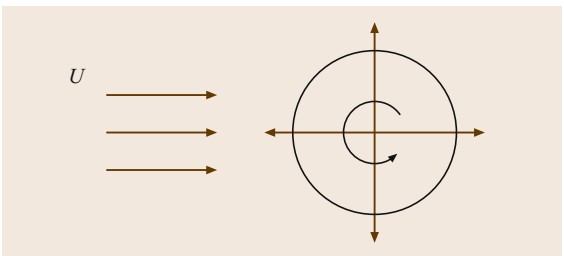


Fig. 7.67 Sketch of superposed uniform flow, cylinder, and circulation

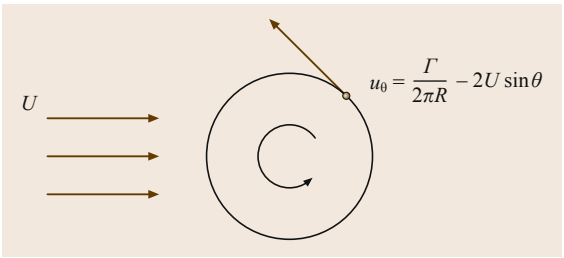


Fig. 7.68 Tangential velocity on the surface of the cylinder

The drag acting on the cylinder can be determined by integrating the streamwise component of the pressure force acting on the cylinder (Figs. 7.69 and 7.70).

$$F_x = -\int_0^{2\pi} p(\theta)\cos\theta R b \mathrm{d}\theta \ ,$$

where the element of surface area is $Rb\mathrm{d}\theta$ with b being the dimension in the z direction. Substituting the pressure around the cylinder surface into this integral, we find that the drag force per unit span is $F_x/b = 0$. This result is known as D'Alembert's paradox – according to two-dimensional, incompressible, irrotational flow theory, the drag on any body, of any shape in a uniform freestream flow is zero.

Turning now to the force component in the cross-stream direction we have

$$F_y = -\int_0^{2\pi} p(\theta)\sin\theta R b \mathrm{d}\theta \ .$$

Substituting the pressure around the cylinder surface into this integral, we find that the lift force per unit span is $F_y/b = -\rho U\Gamma$.

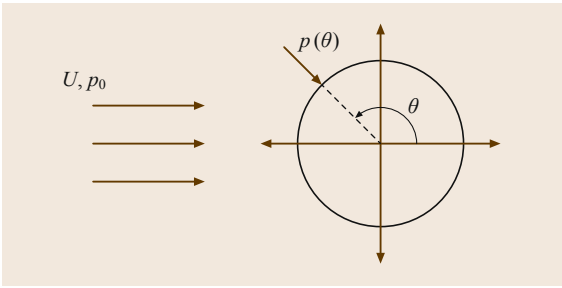


Fig. 7.69 Pressure on the cylinder surface

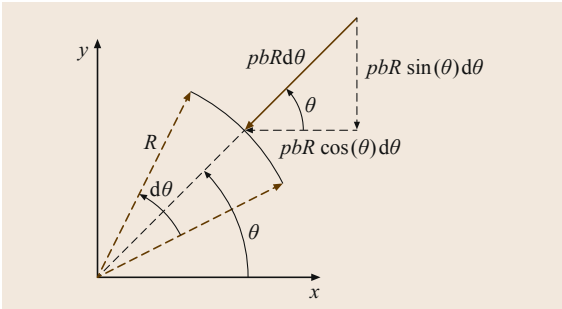


Fig. 7.70 Force components of the pressure acting on the cylinder surface. Here b is the dimension of the cylinder along the z-axis

This result is generally true and is known as the Kutta–Jukowski lift theorem. According to two-dimensional, incompressible, irrotational flow theory, the magnitude of the lift force per unit span of a cylinder of any shape, which possess circulation Γ, immersed in a freestream flow is $\rho U\Gamma$ – the direction of lift is 90° from the freestream direction.

So what does the flow look like? At the cylinder surface $r = R$, $u_r = 0$ and $u_\theta = -2U\sin\theta + \Gamma/(2\pi R)$. The stagnation points occur where $u_\theta = 0$

$$\sin\theta = \frac{\Gamma}{4\pi UR} \ .$$

For $\sin\theta \le 1$, $\Gamma < 4\pi UR$ and there are two values of θ that satisfy the equation, that is, two stagnation points (Fig. 7.71) the corresponding streamlines are shown in Fig. 7.72.

It is interesting to compare the corresponding velocity and pressure distributions about the cylinder with and without circulation. The pressure

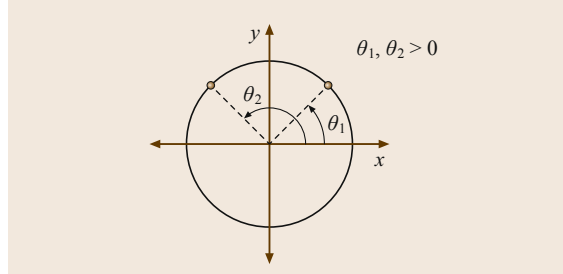


Fig. 7.71 Stagnation point locations

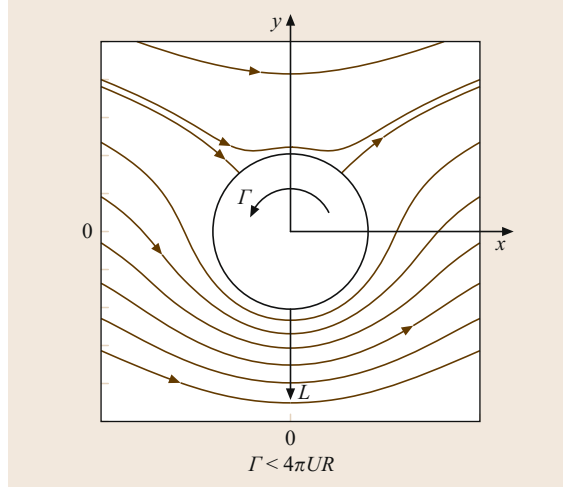


Fig. 7.72 Flow about the circular cylinder with circulation and the corresponding lift force

coefficient

$$C_p \equiv \frac{p-p_\infty}{\frac{1}{2}\rho U^2} = \left(1-\frac{u_\theta^2}{U^2}\right),$$

about the cylinder without circulation is shown in Fig. 7.73a. As can be seen from Figs. 7.66 and 7.73a the symmetrical velocity distribution produces a correspondingly symmetrical pressure coefficient distribution about both the vertical and horizontal axes of the flow. Because of this, both the lift and drag, as determined by integrating the pressure about the cylinder with no circulation, are zero. As one can imagine, the asymmetrical velocity distribution about the cylinder with circulation in a uniform flow (Fig. 7.72) will produce an asymmetrical pressure distribution about the horizontal axis. The integration of the pressure about the cylinder then gives rise to the lift force, which is acting in the cross-stream (vertical) direction.

The Blasius Integral

The Blasius integral is a convenient means of determining the lift and drag per unit span on an object in a flow when the complex velocity potential is known.

Combining the lift and drag in complex form (Fig. 7.74), we have

$$\begin{aligned} \mathrm{d}D - \mathrm{i}\mathrm{d}L &= -p\mathrm{d}y - \mathrm{i}p\mathrm{d}x \\ &= -\mathrm{i}p(\mathrm{d}x - \mathrm{i}\mathrm{d}y) = -\mathrm{i}p\mathrm{d}\bar{z}\,. \end{aligned}$$

Integrating the pressure over the surface of the body, we get the total force on the body in complex form

$$D - \mathrm{i}L = -\oint_C \mathrm{i}p\mathrm{d}\bar{z}\,, \qquad (7.66)$$

where the integral is taken in the counterclockwise direction about the closed curve C that encloses the body. Using the Bernoulli equation (7.63) to find the local

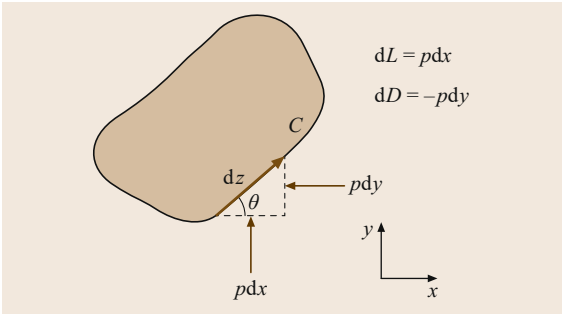


Fig. 7.74 Determination of lift dL and drag dD components from pressure on surface for Blasius integral

pressure on the body surface, we have

$$p = \left(p_\infty + \frac{1}{2}\rho U^2\right) - \frac{1}{2}\rho(u^2+v^2)\,.$$

Incorporating this expression for the pressure into the integral gives

$$D - \mathrm{i}L = -\oint_C \mathrm{i}\left[\left(p_\infty + \frac{1}{2}\rho U^2\right) - \frac{1}{2}\rho(u^2+v^2)\right]\mathrm{d}\bar{z}\,.$$

Since the integral over the closed loop of the constant term

$$\left(p_\infty + \frac{1}{2}\rho U^2\right)$$

is zero, the integral reduces to

$$D - \mathrm{i}L = \oint_C \frac{\mathrm{i}}{2}\rho(u+\mathrm{i}v)(u-\mathrm{i}v)\mathrm{d}\bar{z}\,.$$

Writing the expressions $(u+\mathrm{i}v)$ and $\mathrm{d}\bar{z}$ in polar form gives $(u+\mathrm{i}v) = \sqrt{u^2+v^2}\,\mathrm{e}^{\mathrm{i}\theta}$ and $\mathrm{d}\bar{z} = \mathrm{d}x - \mathrm{i}\mathrm{d}y =$

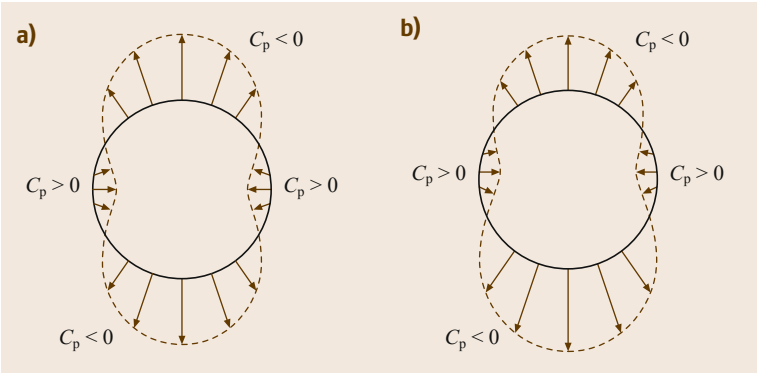


Fig. 7.73 **(a)** Symmetrical pressure distribution about the cylinder with no circulation. **(b)** Asymmetrical pressure distribution about the cylinder with circulation. The lift generated results from asymmetry of the pressure distribution

$\sqrt{dx^2+dy^2}e^{-i\theta}$, respectively (Fig. 7.75). Therefore, the product of these two terms can be rewritten as

$$\begin{aligned}(u+iv)d\bar{z} &= \sqrt{u^2+v^2}e^{i\theta}\sqrt{dx^2+dy^2}e^{-i\theta}\\ &= \sqrt{u^2+v^2}e^{-i\theta}\sqrt{dx^2+dy^2}e^{i\theta}\\ &= (u-iv)dz\,.\end{aligned}$$

Then the integral becomes

$$D-iL=\frac{i\rho}{2}\oint_C(u-iv)^2dz=\frac{i\rho}{2}\oint_C\left(\frac{dw}{dz}\right)^2dz\,. \tag{7.67}$$

Use of this integral is known as the Blasius theorem for calculating the lift and drag on a body of any cross-sectional shape in a steady, two-dimensional, incompressible, irrotational flow. The contour integral does not need to be taken at the surface of the body – so long as there are no singularities (e.g., sources/sinks, irrotational vortices, etc.) between the body and C (Fig. 7.76).

At a very large distance from the body all singularities appear to be located at $z=0$. We can use series expansions to simplify the Blasius integral in this case using the Residue Theorem.

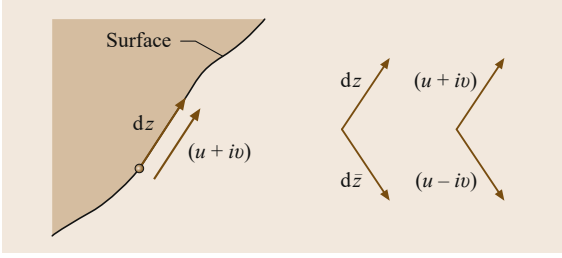


Fig. 7.75 As the flow is everywhere tangent to the surface, $(u+iv)$ and dz are in the same direction. The reflection of these complex vectors about the horizontal axis gives $(u-iv)$ and $d\bar{z}$

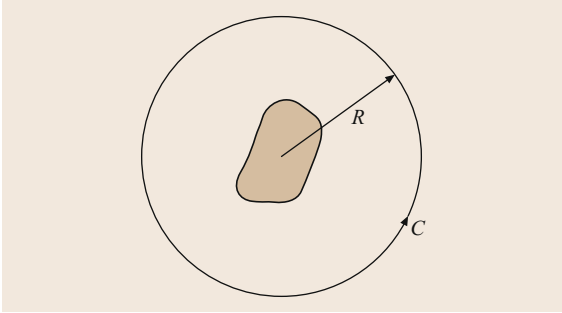


Fig. 7.76 Contour integral about a body in a flow to evaluate the Blasius integral

Example 7.10

If the stream function ψ can be expressed as the sum of two functions in the form $\psi=\psi_1+\psi_2$, the streamlines can be drawn when the curves $\psi_1=$ constant, $\psi_2=$ constant are known. Take a small constant ω and draw the curves $\psi_1=\omega,2\omega,3\omega,\ldots$ and $\psi_2=\omega,2\omega,3\omega,\ldots$ as shown in Fig. 7.77.

At the points marked 3, $\psi=\psi_1+\psi_2=3\omega$, at points marked 4, $\psi=\psi_1+\psi_2=4\omega$, etc. By joining the points with the same numeral (dashed lines in Fig. 7.77) the streamlines, along which $\psi=\psi_1+\psi_2=$ constant, can be plotted out.

Try this. By hand (using a ruler and protractor), plot the streamlines for the flow given by the complex potential

$$w(z)=(-m+ik)\ln z\,.$$

Let $m=1$, and $k=7\pi/[4\ln(10)]$, and $\omega=\pi/4$. Plot the streamlines for $\psi=n\omega$, where $n=\{0,1,2,\ldots,7\}$.

Note

It will be easier to represent z with polar coordinates r,θ.

Solution 7.9

Expanding the expression for $w(z)$ gives

$$\begin{aligned}w(z) &= (-m+ik)\ln z\\ &= -(m\ln r+k\theta)+i(k\ln r-m\theta)=\phi+i\psi\,.\end{aligned}$$

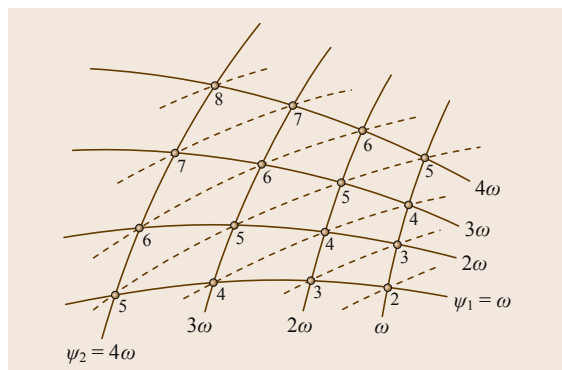


Fig. 7.77 Constructing streamlines by hand. The *solid curves* correspond to constant values of ψ_1 and ψ_2. The *dashed curves* correspond to the streamlines created by joining points with the same values of $\psi=\psi_1+\psi_2$

Therefore,

$$\psi = k \ln r - m\theta = \psi_1 + \psi_2 = \frac{7\pi}{4\ln(10)} \ln r - \theta \,.$$

Then, we have the system

$$\psi_1 = k \ln r = n\omega \,,$$
$$\psi_2 = -\theta = n\omega \,.$$

Solving the system for r gives

$$r = \mathrm{e}^{-\theta/k} = \mathrm{e}^{-n\omega/k} = \exp[\ln 10^{n/7}] = 10^{n/7} \,.$$

Values for the required n are given in Table 7.5. The resulting streamlines are plotted in Fig. 7.78.

7.3.4 Flow of a Viscous Fluid

The viscosity of a fluid, which is a measure of its resistance to deformation by shear stresses imposed on the fluid, plays a critical role in flow of the fluid, being significantly responsible for frictional resistance, flow instability, flow separation, lubrication, heat transfer, dissipation of fluid motion, and other related flow features. As discussed in Sect. 7.2.1, for Newtonian fluids, such as air and water, dynamic viscosity (units: $\mu\mathrm{N\,s\,m^{-2}}$ or $\mathrm{kg\,m^{-1}s^{-1}}$) is the coefficient of proportionality between the imposed stresses and rate of strain or deformation of the fluid. In Couette flow [7.5] between two parallel plates (Fig. 7.52), this is represented in terms of the shear stress and the velocity gradient as $\tau = \mu \partial u/\partial y$. μ is a thermodynamic quantity that varies with temperature and pressure. As shown in Sect. 7.1.2, *Reynolds number* and (7.8), the effect of the viscous forces is typically compared with the inertia forces in the fluid and expressed in terms of the Reynolds number as $\mathrm{Re} = \rho U l/\mu = U l/\nu$, where U and l are, respectively, the characteristic speed and length scales of the flow. Typically, a low Reynolds number indicates that the flow may be laminar while a large Reynolds number indicates that the flow is likely to be turbulent.

Table 7.5 Values to be plotted for the spiral flow

n	θ	r	ψ_1/ψ_2
0	0	1	0
1	$-\pi/4$	1.39	$\pi/4$
2	$-\pi/2$	1.93	$\pi/2$
3	$-3\pi/4$	2.68	$3\pi/4$
4	$-\pi$	3.73	π
5	$-5\pi/4$	5.18	$5\pi/4$
6	$-3\pi/2$	7.20	$3\pi/2$
7	$-7\pi/4$	10	$7\pi/4$

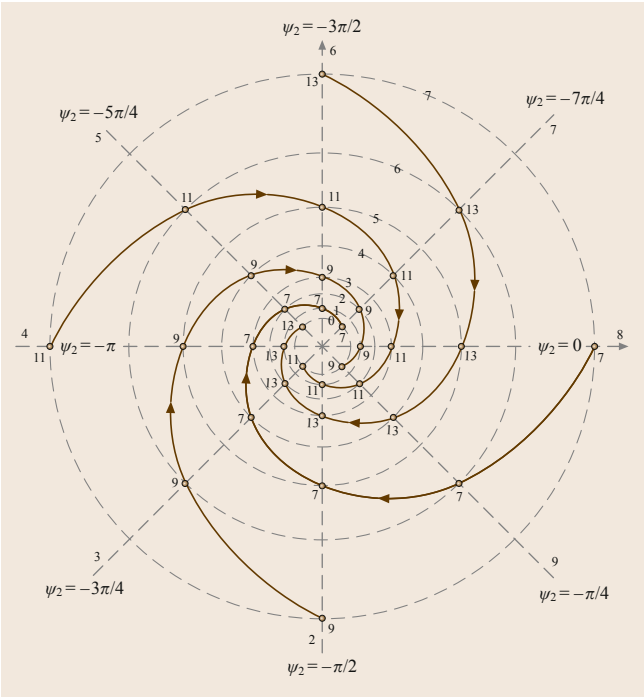


Fig. 7.78 Streamlines for a spiral flow

Laminar Flows

A laminar flow is orderly and is a solution of the Navier–Stokes equations and the associated boundary and initial conditions, so that viscous forces and pressure gradient acting on the fluid balance the inertial forces. Representing the gravitational force as a conservative force in a Cartesian coordinate system such that its potential is $\Pi = gz$ and $\boldsymbol{g} = -\nabla\Pi$, we can express the incompressible isothermal form of Navier–Stokes equations as

$$\frac{\mathrm{d}\boldsymbol{u}}{\mathrm{d}t} = -\frac{1}{\rho}\nabla p + \boldsymbol{g} + \nu\nabla^2\boldsymbol{u} \,, \tag{7.68}$$

$$\nabla \cdot \boldsymbol{u} = 0 \,. \tag{7.69}$$

If the flow is stable, small perturbations to the flow decay. If it is unstable, small perturbations grow and lead to breakdown of the orderly flow and onset of turbulence.

Exact analytical solutions of the Navier–Stokes equations include:

Hagen–Poiseuille Flow in a Circular Pipe. This is a well-known steady flow driven by a constant pressure gradient along the length of a circular pipe that is balanced by viscous forces. The flow speed at a radial distance r from the axis of symmetry in a pipe of

diameter $D = 2r_0$ is given by

$$u(r) = -\frac{1}{4\mu}\frac{\mathrm{d}(p+\rho g z)}{\mathrm{d}s}\left(r_0^2 - r^2\right) = u_{\max}\left(1 - \frac{r^2}{r_0^2}\right), \tag{7.70}$$

where s is the distance along the pipe. The mean speed is $\bar{u} = u_{\max}/2$ and the volume flow rate is $Q = \pi r_0^2 \bar{u}$. The shear stress at the pipe wall is

$$\tau_w = -\mu \frac{\mathrm{d}u}{\mathrm{d}r}|_{r=r_0} = \frac{4\mu\bar{u}}{r_0} .$$

This implies a skin-friction coefficient of

$$C_F = \frac{2\tau_w}{\rho\bar{u}^2} = \frac{16}{\mathrm{Re}_D} ,$$

where the diameter-based Reynolds number is $\mathrm{Re}_D = \rho\bar{u}D/\mu$. The friction represents a head loss of $h_f = 32\mu\bar{u}/(\rho g D^2)$ in the pipe. The laminar flow undergoes transition to turbulence when $\mathrm{Re}_D \approx 2000$ and is fully turbulent when $\mathrm{Re}_D > 3000$. See [7.32] for other internal flow solutions.

Similarity Solutions. A class of solutions of the Navier–Stokes equations obtained through introduction of a similarity variable and reduction of the set of partial differential equations to ordinary differential equations that can be solved analytically or numerically is referred to as a self-similar solution. Such solutions include:

Flow at an Axisymmetric Stagnation Point. At an axisymmetric stagnation point, with free stream flow given by $u = ar$ and $v = -2az$, in cylindrical coordinates (r, θ, z) centered on the axisymmetric stagnation point. Upon introduction of a similarity variable $\eta \equiv z\sqrt{a/\nu}$, the radial and normal components of velocity are, respectively, given by $u = arF'(\eta)$ and $v = -2F(\eta)\sqrt{a\nu}$, where $F(\eta)$ satisfies

$$F''' + 2FF'' + 1 - F'^2 = 0 ,$$
$$F(0) = F'(0) = 0 , \; F'(\infty) = 1 . \tag{7.71}$$

$F(\eta)$ and its derivatives are plotted in Fig. 7.79.

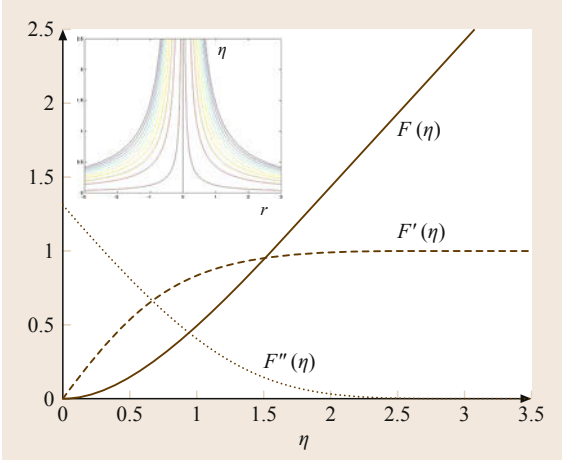


Fig. 7.79 $F(\eta)$ and $F'(\eta)$ for stagnation point flow; $F''(0) = 1.31194$

Stokes Flow Over Steadily Oscillating Wall. Over a plane wall oscillating with speed $u(0,t) = U_0 \cos\omega t$ (Fig. 7.80), upon introduction of similarity variable $\eta = y\sqrt{\omega/2\nu}$, the solution for the flow speed along the plane is given by

$$u(\eta) = U_0 \mathrm{e}^{-\eta}\cos(\omega t - \eta) . \tag{7.72}$$

This is the well-known Stokes solution; a detailed development of the solution is given in [7.5].

Flow Over a Suddenly Accelerated Plane Wall. Over a plane wall suddenly accelerated to speed U_0, using the similarity variable $\eta = y/\sqrt{4\nu t}$, the solution for the flow speed along the plane is given by

$$u(\eta) = U_0(1 - \mathrm{erf}\eta) , \tag{7.73}$$

where $\mathrm{erf}\eta$ is the error function $\mathrm{erf}(x) \equiv 2/\sqrt{\pi}\int_0^x \mathrm{e}^{-t^2}\mathrm{d}t$.

Laminar Boundary Layers. In flow past a rigid surface in air or water, the rate of convection along the surface is much greater than the rate of diffusion normal to the surface. Thus the diffusion of momentum over the surface occurs in a thin layer, which is referred to as the boundary layer. The larger the Reynolds number, the thinner the boundary layer. On this basis, the velocity $\boldsymbol{u}(\boldsymbol{x}, t) = (u, v, w)$ within the boundary layer due to an external freestream flow $\boldsymbol{U}(x, z, t) = (U_e, 0, W_e)$ can be shown to be governed by the simpler equations

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + w\frac{\partial u}{\partial z} = \frac{\partial U_e}{\partial t} + U_e\frac{\partial U_e}{\partial x} + W_e\frac{\partial U_e}{\partial z} + \nu\frac{\partial^2 u}{\partial y^2} , \tag{7.74}$$

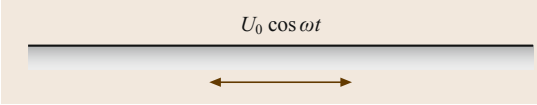


Fig. 7.80 Geometry of the oscillating wall

$$\frac{\partial w}{\partial t} + u\frac{\partial w}{\partial x} + v\frac{\partial w}{\partial y} + w\frac{\partial w}{\partial z}$$
$$= \frac{\partial W_e}{\partial t} + U_e\frac{\partial W_e}{\partial x} + W_e\frac{\partial W_e}{\partial z} + \nu\frac{\partial^2 w}{\partial y^2}\,, \qquad (7.75)$$
$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0\,. \qquad (7.76)$$

For steady two-dimensional flow over a flat plate with its upstream edge at $x = 0$ and $\boldsymbol{U}_e = (U_e, 0, 0)$, where U_e is a constant, a well-known Blasius similarity solution [7.16] for the flow in the boundary layer is obtained upon introducing the similarity variable $\eta = y\sqrt{U/(2\nu x)}$ as $u = U_e F'(\eta)$, and $v = \sqrt{\nu U/(2x)}(\eta F' - F)$, where $F(\eta)$ satisfies

$$F''' + FF'' = 0,\ F(0) = F'(0) = 0\,,$$
$$F'(\infty) = 1\,. \qquad (7.77)$$

$F(\eta)$ and its derivatives are plotted in Fig. 7.81.

The thickness of the Blasius boundary layer $\delta_{0.99} = 4.9x/\sqrt{\mathrm{Re}_x}$, where $\mathrm{Re}_x \equiv U_e x/\nu$. The momentum thickness is given by $\delta^* = 1.7208x/\sqrt{\mathrm{Re}_x}$. The friction coefficient is given by

$$c_f = \frac{2\tau_w}{\rho U^2} = \frac{0.664}{\sqrt{\mathrm{Re}_x}}\,.$$

The integrated drag coefficient on a plate of length $C_D = 2C_F(L) = 1.328/\sqrt{\mathrm{Re}_L}$. The Blasius solution is valid for $\mathrm{Re}_x > 1000$. At $\mathrm{Re}_x = O(10^6)$, the laminar boundary layer becomes unstable, leading to transition to turbulence. See [7.16] for comparison of these results with experiments as well as for other similarity solutions of boundary layer equations, including flows in a wedge and flow into a convergent channel.

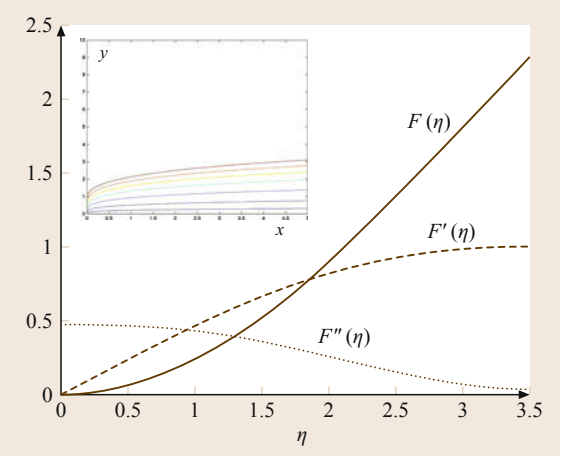


Fig. 7.81 $F(\eta)$ and its derivatives for Blasius boundary layer flow; $F''(0) = 0.469600$

Flow Separation. Flow separation occurs at the location where the surface stress at the bounding wall vanishes such that

$$\tau_w = \mu\left(\frac{\partial u}{\partial y}\right)_{y=0}\,. \qquad (7.78)$$

Close to the freestream, the curvature of the velocity profile, characterized by

$$\frac{\partial^2 u}{\partial y^2}\,, \qquad (7.79)$$

is negative since the velocity shear reduces to zero upon approach to the freestream. At the wall, however, $u = v = 0$ in (7.74) and the streamwise pressure gradient in the boundary layer, assumed constant across the boundary layer and given by

$$\frac{\mathrm{d}p}{\mathrm{d}x} = -r\left(\frac{\partial U_e}{\partial t} + U_e\frac{\partial U_e}{\partial x} + W_e\frac{\partial U_e}{\partial z}\right)\,, \qquad (7.80)$$

is balanced by the viscous force,

$$\mu\frac{\partial^2 u}{\partial y^2}|_{y=0} = \frac{\mathrm{d}p}{\mathrm{d}x}\,. \qquad (7.81)$$

Therefore, the curvature of the velocity profile has the same sign as the pressure gradient. When the pressure gradient is favorable ($\mathrm{d}p/\mathrm{d}x < 0$) or zero (as for the flat plate Blasius boundary layer) the profile curvature is less than or equal to zero throughout the boundary layer. When the pressure gradient is adverse ($\mathrm{d}p/\mathrm{d}x > 0$), such as on the aft side of a curved surface of a body, the sign of the profile curvature changes across the boundary layer and is an indication that flow separation may occur.

Laminar Wakes, Jets, and Shear Layers. Far-field behavior of laminar wakes of bodies, jets, and free shear layers can be determined as similarity solutions of the steady Navier–Stokes equations.

For an axisymmetric wake associated with a uniform flow of speed U_e past a body with related drag force D, introducing a transverse similarity variable, $\eta = r\sqrt{U/(x\nu)}$, the far-field velocity defect in the wake at a distance x (at least three times body length) downstream of the body can be shown to have a Gaussian profile [7.16] given by

$$U_e - u = \frac{D}{4\pi\mu x}\exp\left(-\frac{\eta^2}{4}\right)\,. \qquad (7.82)$$

For an axisymmetric round jet of momentum $J = \rho\int_0^a 2\pi u^2 r\mathrm{d}r$ from a circular hole of radius a, introducing a transverse similarity variable, $\eta = r/x$, the

far-field axial velocity in the jet at a distance x downstream of the hole can be shown to be given by

$$u = \frac{3J}{8\pi\mu x}\left(1 + \frac{3J\eta^2}{64\pi\rho\nu^2}\right)^{-2} . \tag{7.83}$$

Thus the speed decreases approximately linearly with downstream distance.

Finally, velocity associated with a two-dimensional laminar shear layer where the streamwise speed u transitions from U_2 to U across the layer may be represented by Blasius type profiles [7.37] upon introducing the similarity variable $\eta = y\sqrt{U/(2\nu x)}$ as $u = UF'(\eta)$, and $v = \sqrt{\nu U/(2x)}(\eta F' - F)$, where $F(\eta)$ satisfies

$$F''' + FF'' = 0 , \quad F(0) = 0 ,$$
$$F'(-\infty) = \frac{U}{U_2} , \quad F'(\infty) = 1 . \tag{7.84}$$

The case of $U/U_2 = 0.5$ is shown in Fig. 7.82.

Laminar wakes, jets, and shear layers are relatively unstable and break down into turbulent flows.

Turbulent Flows

A turbulent flow is chaotic and involves self-sustaining, random irregular fluctuations in velocity, pressure, and other dependent variables. Turbulent flows are characterized by the presence of eddies (vortices) that induce significant mixing of mass and momentum. The eddies in a turbulent flow exist over a range of spatial and temporal scales. Typically, the largest scales are respectively determined by the size of the flow domain and the time period under consideration. The smallest (Kolmogorov) length and time scales are given by $(\nu^3/\epsilon)^{1/4}$ and $(\nu/\epsilon)^{1/2}$, where ϵ is the rate of energy dissipation per unit mass. A statistical description of the turbulent flow is based on defining deviation of the instantaneous velocity, pressure, and other variables at a location from the mean, as

$$\boldsymbol{u}(\boldsymbol{x}, t) = \boldsymbol{U}(\boldsymbol{x}, t) + \boldsymbol{u}'(\boldsymbol{x}, t) , \tag{7.85}$$
$$p(\boldsymbol{x}, t) = P(\boldsymbol{x}, t) + p'(\boldsymbol{x}, t) , \tag{7.86}$$

etc., where the mean is typically considered to be a temporal average. Here, the mean terms are indicated using capitalized variables and the deviations, or fluctuations are indicated with the superscripted symbol $'$. The mean flow satisfies

$$\rho\left(\frac{\partial \boldsymbol{U}}{\partial t} + \boldsymbol{U}\cdot\nabla\boldsymbol{U}\right) = -\nabla P + \mu\nabla^2\boldsymbol{U} - \overline{\rho\boldsymbol{u}'\cdot\nabla\boldsymbol{u}'} , \tag{7.87}$$

$$\nabla\cdot\boldsymbol{U} = 0 , \tag{7.88}$$

or in tensor notation,

$$\rho\left(\frac{\partial U_i}{\partial t} + U_j\frac{\partial U_i}{\partial x_j}\right) = -\frac{\partial P}{\partial x_i} + \mu\frac{\partial^2 U_i}{\partial x_j^2} - \frac{\partial\overline{\rho u_i' u_j'}}{\partial x_j} , \tag{7.89}$$

$$\frac{\partial U_i}{\partial x_i} = 0 , \tag{7.90}$$

so that the mean flow depends on the product of fluctuations about the mean. These equations are referred to as the Reynolds averaged Navier–Stokes (RANS) equations. $\tau_{ij} = -\overline{\rho u_i' u_j'}$ is the Reynolds stress and may be computed directly, or more typically modeled using constitutive equations in terms of the mean quantities, in order to provide a closed system of equations. A simple form is given by the Boussinesq model [7.38], which for incompressible flow is given by

$$\tau_{ij} = \frac{\mu_t}{\rho_0 U}\left(\frac{\partial U_i}{\partial x_j} + \frac{\partial U_j}{\partial x_i}\right) , \tag{7.91}$$

where μ_t is known as the eddy viscosity. See [7.38] for detailed description of various models that are commonly used as well as a review of several attempts to verify the models and their implementation in numerical schemes.

Turbulence in flow over boundaries is often characterized in terms of a three-layer model: The *inner layer* close to the boundary where viscous dissipative processes dominate, the *outer layer* close to the freestream where larger scale turbulent eddies dominate the shear flow, and the *overlap layer* where both processes are equally important. Recognizing these characteristics, the flow in the inner layer is described in terms of

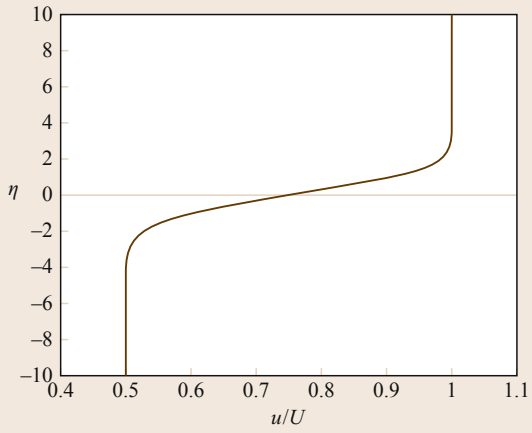


Fig. 7.82 Viscous shear layer profile

nondimensional viscous *wall units*, using the *friction velocity* $u_\tau = (\tau_w/\rho)^{1/2}$, where τ_w is the shear stress at the boundary and ρ is the fluid density, to scale the velocity and u_τ/ν to scale length. The distance normal to the boundary is then expressed as $y^+ = yu_\tau/\nu$ and the velocity in the boundary layer as $u^+ = u/u_\tau$. The velocity and length scales in the outer region are, respectively, the freestream speed U_e, and the boundary layer thickness δ. Dimensional analysis and use of the overlap region suggest that the mean speeds in the inner and outer layers, respectively, are approximately given by

$$U^+ = \begin{cases} y^+ & \text{for } y^+ \lesssim 5 \\ \frac{1}{\kappa}\ln y^+ + B - \frac{1}{\kappa}\ln(1+0.3k^+) & \text{for } y^+ \gtrsim 30 \end{cases}, \tag{7.92}$$

and

$$U_e^+ - U^+ = -\frac{1}{\kappa}\ln\frac{y}{\delta} + A\left(\frac{\delta}{\tau_w}\frac{dp_e}{dx}\right), \tag{7.93}$$

where dp_e/dx is the pressure gradient in the flow and k is average height of surface roughness [7.16]. Experimental results suggest $\kappa \approx 0.41$ and $B \approx 5$. The logarithmic dependence of the velocity in the inner layer is known as *the law of the wall*, while the outer flow dependence on U_e^+ and δ is known as the *law of the wake*.

Unlike boundary layer flows, free-stream turbulence generated through the breakdown of shear layers, wakes or jets is characterized by the maximum velocity U_{max} and the width of the shear layer, $b(x)$, where x is the downstream distance.

Turbulent Flow in a Circular Pipe. The flow in a pipe is fully turbulent when the Reynolds number based on the pipe diameter is $Re_D > 3000$. The distribution of the mean velocity in the pipe can be represented by the law of the wall across the entire pipe cross-section, the wake layer in Fig. 7.83 shows the case where $a^+ = 1000$. The mean velocity of turbulent flow is fuller than that of laminar flow as turbulent eddies mix the momentum across the pipe cross-section. The drag in the pipe is represented in terms of a friction factor $\Lambda = 4C_F$. If k is the average height of the sand-grain roughness of the pipe surface, then Λ is given by

$$\frac{1}{\Lambda^{\frac{1}{2}}} = 2\log_{10}\left(\frac{Re_D\Lambda^{\frac{1}{2}}}{1+0.1(\frac{k}{D})Re_D\Lambda^{\frac{1}{2}}}\right) - 0.8\,. \tag{7.94}$$

The equation is typically represented in terms of the Moody diagram [7.16].

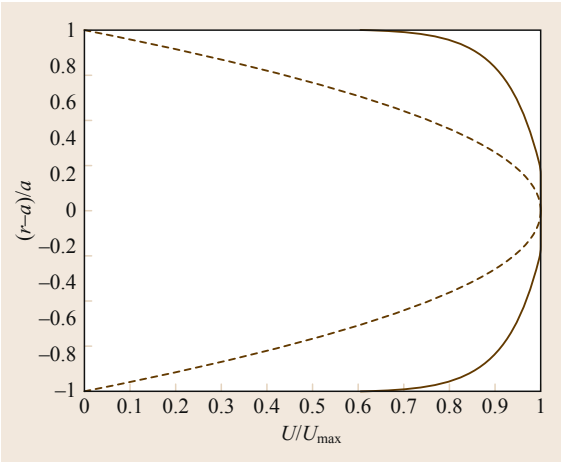


Fig. 7.83 Mean velocity profile of turbulent flow (*solid line*) in a pipe compared with corresponding laminar flow profile (*dotted line*)

Turbulent Boundary Layer Over a Flat Plate. The mean velocity in a turbulent boundary layer over a smooth flat plate is approximately given by the 1/7-th power law

$$\frac{U}{U_e} = \left(\frac{y}{\delta}\right)^{1/7}, \tag{7.95}$$

and the associated skin-friction coefficient is given by $C_F \approx 0.027/Re_x^{1/7}$.

The boundary layer thickness grows as $\delta \approx 0.16x/Re_x^{1/7}$. For fully rough surface, with sand-grain roughness $k > xRe_x/1000$, the skin-friction coefficient is given by

$$C_F \approx \left[2.87 + 1.58\log_{10}\left(\frac{x}{k}\right)\right]^{-2.5}. \tag{7.96}$$

Turbulent Mixing. Turbulent eddies tend to accelerate the mixing of mass and momentum. Mixing occurs over a range of scales and involves fluid *entrainment, dispersion*, and *diffusion* over a range of spatial and temporal time scales of the flow [7.39]. Thus, turbulent flows effectively mix fluid through entrainment and dispersion via a cascade of large to small eddies to diffusion to molecular scales that has a wide range of consequences for oceanic and engineering flows. The largest scale δ is associated with the outer eddies spanning the local region of the turbulent flow, while the smallest scale (Taylor microscale) is associated with the smallest eddies and is given by $\lambda_T = (15\nu/\epsilon)^{\frac{1}{2}}$, where $\epsilon = O(U^3/\delta)$ is the rate of dissipation per unit mass. For a jet, the range of scales varies with the flow

Reynolds number $\mathrm{Re}_\delta = U\delta/\nu$ as

$$\frac{\lambda_\mathrm{T}}{\delta} \approx 2.3\mathrm{Re}_\delta^{-1/2} \,. \tag{7.97}$$

so that the larger the Reynolds number, the greater the range of eddies and length scales involved.

Turbulent Jets and Wakes. The maximum speed in a turbulent circular jet of momentum $J = \iint_\mathrm{S} u^2 \mathrm{d}s$, where S is the cross-sectional area of the nozzle exit, decays linearly with downstream distance x, where x is measured from a virtual origin as

$$U_\mathrm{max} \approx \frac{7.4}{x}\left(\frac{J}{\rho}\right)^{1/2} . \tag{7.98}$$

The cross-sectional extent of the jet varies as $b \approx 0.4x$ and self-similar mean flow profile in the jet may be approximated [7.16, 39] as

$$U \approx U_\mathrm{max}\mathrm{sech}^2\left(\frac{10.4r}{x}\right) . \tag{7.99}$$

For an axisymmetric wake, the maximum defect in the mean velocity in the wake is given by

$$U_\mathrm{e} - U_\mathrm{min} \approx C_0 x^{-2/3} \,, \tag{7.100}$$

and the width of the wake varies as

$$b(x) \approx C_1 x^{-2/3} \,, \tag{7.101}$$

where C_0 and C_1 are constants. These results may be compared with the corresponding cases of laminar flow. The various parameters characterizing the behavior of the turbulent jets or wakes are fairly independent of flow Reynolds number.

References

7.1 B.J. Cantwell: *Introduction to Symmetry Analysis* (Cambridge Univ. Press, Cambridge 2002)
7.2 W. Valentine, K.D. von Ellenrieder: Model scaling of ocean hydrokinetic renewable energy systems, IEEE-JOE **40**(1), 27–36 (2015)
7.3 O.M. Faltinsen: *Hydrodynamics of High-Speed Marine Vehicles* (Cambridge Univ. Press, New York 2005)
7.4 M. Ferrando, M. Viviani, S. Crotti, P. Cassella, S. Caldarella: Influence of Weber number on surface piercing propellers model tests scaling, Proc 7th Int. Conf. Hydrodyn., Ischia (2006)
7.5 J.N. Newman: *Marine Hydrodynamics* (MIT Press, Cambridge 1977)
7.6 R.D. Blevins: *Flow-induced Vibration* (Van Nostrand Reinhold, New York 1990)
7.7 G.S. Triantafyllou, M.S. Triantafyllou, R. Gopalkrishnan: Wake mechanics for thrust generation in oscillating foils, Phys. Fluids A **3**(12), 2835–2837 (1991)
7.8 J.H. VanZwieten, N. Vanrietvelde, B.L. Hacker: Numerical simulation of an experimental ocean current turbine, IEEE-JOE **38**(1), 131–143 (2013)
7.9 A. Roshko: Experiments on the flow past a circular cylinder at very high Reynolds number, J. Fluid Mech. **10**(3), 345–356 (1961)
7.10 O.M. Griffin, S.E. Ramberg: Vortex shedding from a cylinder vibrating in line with an incident uniform flow, J. Fluid Mech. **75**(2), 257–271 (1976)
7.11 R.A. Dalrymple, R.G. Dean: *Water Wave Mechanics for Engineers and Scientists* (Prentice-Hall, Hackensack 1991)
7.12 M. Miche: Le pouvoir reflechissant des ouvrages maritimes exposes a l'action de la houle, Ann. Ponts Chaussées **121**, 285–319 (1951)
7.13 I.A. Hunt: Design of seawalls and breakers, J. Waterw. Harb. Coast. Eng. Div. **85**(3), 123–152 (1959)
7.14 R.E. Randall: *Elements of Ocean Engineering* (Society of Naval Architects and Marine Engineers, Jersey City 2010)
7.15 M.S. Longuet-Higgins: Longshore currents generated by obliquely incident sea waves: 1, J. Geophys. Res. **75**(33), 6778–6789 (1970)
7.16 F.M. White: *Viscous Fluids Flow* (McGraw-Hill, New York 1991)
7.17 B. Cantwell, D. Coles, P. Dimotakis: Structure and entrainment in the plane of symmetry of a turbulent spot, J. Fluid Mech. **87**(4), 641–672 (1978)
7.18 C.M. De Silva, E.P. Gnanamanickam, C. Atkinson, N.A. Buchmann, N. Hutchins, J. Soria, I. Marusic: High spatial range velcity measurements in a high Reyonolds number turbulent boundary layer, Phys. Fluids **26**(2), 025117 (2004)
7.19 J.D. van Manen, P. van Oossanen: Resistance and propulsion. In: *Principles of Naval Architecture*, Vol. 2, ed. by E.V. Lewis (Society of Naval Architects and Marine Engineers, New York 1988) pp. 228–462
7.20 H. Schlichting: *Boundary Layer Theory* (McGraw-Hill, New York 1968)
7.21 J.D. Anderson: *Fundamentals of Aerodynamics* (McGraw-Hill, New York 2001)
7.22 I.H. Abbott, A.E. von Doenhoff, L.S. Stivers: *Summary of Airfoil Data*, Rep. 824 (NACA, Langley 1945)
7.23 S.F. Hoerner: *Fluid Dynamic Drag* (Hoerner Fluid Dynamics, Alburqueque 1965)
7.24 P.F. Rynne, K.D. von Ellenrieder: Unmanned autonomous sailing: Current status and future role in sustained ocean observations, Mar. Technol. Soc. J. **43**(1), 21–30 (2009)
7.25 P.F. Rynne, K.D. von Ellenrieder: Development and preliminary experimental validation of a wind and

solar powered autonomous surface vehicle, IEEE-JOE **35**(4), 971–983 (2010)
7.26 G. Elkaim, R. Kelbley: Control architecture for segmented trajectory following of a wind-propelled autonomous catamaran, Proc. AIAA Guid. Navig. Control Conf., Keystone (2006)
7.27 N. Cruz, J. Alves: Autonomous sailboats: An emerging technology for ocean sampling and surveillance, Proc. IEEE OCEANS'08, Quebec City (2008)
7.28 M. Neal: A hardware proof of concept of a sailing robot for ocean observation, IEEE-JOE **31**(2), 462–469 (2006)
7.29 D. Savitsky: *On the Subject of High-speed Monohulls* (Society of Naval Architects and Marine Engineers, Athens 2003)
7.30 L. Larsson, R.E. Eliasson: *Principles of Yacht Design* (International Marine, Camden 2000)
7.31 E.C. Tupper: *Introduction to Naval Architecture: Formerly Muckle's Naval Architecture for Marine Engineers* (Butterworth-Heinemann, Oxford 2013)
7.32 T. Lamb: *Ship Design and Construction* (Society of Naval Architects and Marine Engineers, Jersey City 2004)
7.33 W.M. Ellsworth, W.C. O'Neill: *The U.S. Navy Hydrofoil Development Program, A Status Report*, Tech. Note SDD-OH50-62 (Naval Research and Development, 1970)
7.34 H.O. Berteaux: *Buoy Engineering* (Wiley, New York 1976)
7.35 T.C. Gillmer, B. Johnson: *Introduction to Naval Architecture* (Naval Institute Press, Annapolis 1982)
7.36 P.K. Kundu, I.M. Cohen: *Fluid Mechanics* (Academic Press, San Diego 2008)
7.37 R. Lock: The velocity distribution in the laminar boundary layer between parallel streams, Q. J. Mech. Appl. Math. **4**(1), 42–63 (1951)
7.38 NASA Langley Research Center: Turbulence Modeling Resource, http://turbmodels.larc.nasa.gov/
7.39 H. Tennekes, J.L. Lumley: *A First Course in Turbulence* (MIT, Cambridge 1972)

8. Ocean Electromagnetics

John J. Holmes

Even though acoustic waves can travel long distances in the sea with little attenuation, ocean electromagnetics has important applications in the areas of geophysical surveys and searches of the seafloor and sub-bottom, communications across the sea–air boundary, and high data transfer rate at short ranges. Unlike in-air propagation of electromagnetic fields, the finite conductivity of seawater results in a frequency-dependent phase velocity, attenuation, intrinsic impedance, and reflection and transmission coefficients at the ocean's surface. After giving a short summary of the electric and magnetic properties of the ocean, this chapter begins with Maxwell's equations and develops the mathematical descriptions of electromagnetic fields and dipole sources within a conducting media. The differences between plane wave reflection and transmission at the surface of fresh water and seawater are used to highlight how electromagnetic propagation within the electrically conducting ocean is so very different than the more familiar radio frequency transmissions in air. In addition, equations are presented that describe the fields from submerged electric and magnetic dipoles that are located both far and near the sea surface. These formulations are valid over the frequency range from 0 Hz to a few MHz. Finally, a brief discussion of ocean electromagnetics at optical wavelengths is given at the end of this chapter.

8.1 Electromagnetism in an Ocean Environment 177
8.2 Electromagnetic Field Theory 178
8.3 Plane Wave Propagation 180
8.4 Reflection and Transmission of a Plane Wave at the Surface of Fresh Water 182
8.5 Plane Wave Incident on Seawater 184
8.6 Magnetic and Electric Dipoles in an Unbounded Ocean 186
8.7 Magnetic and Electric Dipoles in a Bounded Ocean 188
8.8 Electromagnetic Propagation in the Ocean at Optical Wavelengths 193
References 195

Even though acoustic waves can travel long distances in the sea with little attenuation, ocean electromagnetics has several important applications. For example; electromagnetic surveys, searches, and communication systems can operate effectively in acoustically noisy environments, they can cross the sea–air boundary and be received by high speed airborne sensors, and they have much higher data transfer rates at short ranges. In addition, electromagnetic surveys of seafloor gas hydrate pockets can determine their methane content, a shortfall of seismic techniques. Moreover, the short range limitations of underwater electromagnetic field propagation are actually beneficial when pinpointing submerged targets during naval operations.

8.1 Electromagnetism in an Ocean Environment

The oldest known application of electromagnetism in an ocean environment is the magnetic compass used for navigation. Today, electromagnetic sources and sensors are useful tools found in many diverse areas of oceanographic study such as physical oceanography, seafloor geophysics, marine chemistry, and biology. Advances in the understanding of electromagnetism within the boundaries of these scientific disciplines have found application to engineering systems for energy generation, oil and gas exploitation and recovery, commercial fishing, weather prediction, Tsunami warning, archeology, and the military. This chapter will

present a brief description of the science behind electromagnetic fields and their interaction with an ocean environment.

Pure water is a very good dielectric. Its *permittivity* is given by $\varepsilon = \varepsilon_0 \varepsilon_w$, where ε_0 is the free-space permittivity ($\approx 8.8541878176 \times 10^{-12}$ F/m), and ε_w is the water's dielectric constant equal to approximately 80.1 at 20 °C. Due to the strong covalent bonds of the hydrogen and oxygen atoms in the water molecule, there are no electrons in the conduction band until they are pumped up to that energy level by applying high amplitude electric fields. If the field reaches a few megavolts/cm, pure water will undergo a dielectric *breakdown* forming an electric arc in the process. Although pure water is a very good insulator, seawater is a good electric conductor.

The disassociated salt ions in seawater cause it to be a good electric conductor at all field levels. Although electrons are the primary charge carriers in metallic conductors, the positively charged salt ions, called *cations* and negatively charged *anions*, are both charge carriers in seawater. If a source of electric potential gradient is introduced into the seawater, the cations will migrate toward the source's *cathode* (negative pole), while the anions travel toward the *anode* (positive pole). Electrochemical reactions occur at the anode, releasing electrons in the process that then flow in the source's metallic conductor to the cathode where they recombine with the cations.

The mobility of seawater ions is hampered by their interaction with water molecules as they migrate, resulting in loss of energy in the form of heat. At the engineering level, seawater can be treated as a simple conductor with an *electric conductivity* σ ranging from about 2.5 S/m in cold deep waters, and approaches 6 S/m in very warm waters. (A value of 4 S/m is typically used for open ocean seawater.) This finite conductivity attenuates electromagnetic energy as it propagates through seawater, which is categorized as a lossy medium.

Fortunately, seawater is nonmagnetic. Its *magnetic permeability* is the same as free-space given by $\mu_0 = 4\pi 10^{-7}$ H/m. This means the only way that seawater can be a source of a static magnetic field is by electric current flowing through it.

This chapter will focus on the science of electromagnetic field propagation and attenuation in seawater, and the reflection and transmission of the fields at the sea's surface. Emphasis will be placed on electromagnetic field theory in the ultralow frequency (ULF) band of ≈0 to 3 Hz, and the extremely low frequency (ELF) band from 3 Hz to 3 kHz. Since the attenuation of electromagnetic waves increases rapidly with frequency, radio frequency bands will not be covered here. Although remote observations of the sea surface from airborne and satellite-based microwave radiometers and radars are important tools for oceanographic research, they are beyond the scope of this discussion. (The interested reader should consult [8.1, pp. 405–510] on this topic.) A brief cursory discussion of ocean electromagnetics at optical wavelengths will be given at the end of this chapter.

8.2 Electromagnetic Field Theory

It is not possible to give an in-depth description of general electromagnetic field theory here. A comprehensive treatment of engineering electromagnetics is presented in [8.2]. In order to introduce some basic principles for electromagnetic fields in a lossy media such as seawater, the formulations for magnetic and electric fields in an unbounded ocean will be presented first.

All electromagnetic field theories are based on *Maxwell's equations*. In differential form, they are

$$\nabla \times \boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial t}\,, \tag{8.1}$$

$$\nabla \times \boldsymbol{H} = \boldsymbol{J} + \frac{\partial \boldsymbol{D}}{\partial t}\,, \tag{8.2}$$

$$\nabla \cdot \boldsymbol{D} = q_e\,, \tag{8.3}$$

$$\nabla \cdot \boldsymbol{B} = 0\,, \tag{8.4}$$

where $\boldsymbol{E}$ (v/m) and $\boldsymbol{H}$ (A/m) are the electric and magnetic field *intensity vectors*, $\boldsymbol{D}$ (c/m^2) and $\boldsymbol{B}$ (wb/m^2 or T) are the electric and magnetic *flux density vectors*, $\boldsymbol{J}$ (A/m^2) is the electric *current density*, q_e (c/m^3) represents the *electric charge density*, and t stands for time (seconds). Maxwell's equations by themselves are sufficient to solve any problem in electromagnetics.

Although magnetic charge and magnetic current densities are not realizable, they can be included in Maxwell's equations through the *generalized* current concept to aid mathematically in the solutions of certain problems. For example, a magnet can be represented mathematically as two opposite polarity magnetic charges at its ends. Adding the virtual magnetic sources to Maxwell's equations, and separating the current density into its impressed and conduction components, allows (8.1) through (8.4) to be written in a more symmetrical form as

$$\nabla \times \boldsymbol{E} = -\boldsymbol{M}_i - \frac{\partial \boldsymbol{B}}{\partial t}\,, \tag{8.5}$$

$$\nabla \times \boldsymbol{H} = \boldsymbol{J}_\mathrm{i} + \boldsymbol{J}_\mathrm{c} + \frac{\partial \boldsymbol{D}}{\partial t} , \tag{8.6}$$

$$\nabla \cdot \boldsymbol{D} = q_\mathrm{e} , \tag{8.7}$$

$$\nabla \cdot \boldsymbol{B} = q_\mathrm{m} , \tag{8.8}$$

where $\boldsymbol{M}_\mathrm{i}$ $(\mathrm{v/m^2})$ is the *impressed magnetic current density*, $\boldsymbol{J}_\mathrm{i}$ $(\mathrm{A/m^2})$ is the *impressed electric current density* produced by an active source, $\boldsymbol{J}_\mathrm{c}$ is the *conduction current* density flowing in the media, and the equivalent *magnetic charge density* is represented as q_m $(\mathrm{wb/m^3})$. The symmetric form of Maxwell's equations, (8.5) through (8.8), will be used here.

The electromagnetic properties of seawater are very different than air or free space. As a result, the interaction of electromagnetic fields with the ocean is unlike that in air. The relationships between the field intensities and flux densities are related by the *constitutive parameters* of the media as given by

$$\boldsymbol{B} = \mu \boldsymbol{H} , \tag{8.9}$$

$$\boldsymbol{D} = \varepsilon \boldsymbol{E} , \tag{8.10}$$

$$\boldsymbol{J} = \sigma \boldsymbol{E} , \tag{8.11}$$

where μ, ε, and σ for seawater have been defined previously. In air, $\mu = \mu_0$, $\varepsilon = \epsilon_0$, and $\sigma = 0$.

Continuity of electric charge is an important relationship that can be derived from Maxwell's equations. The *continuity equation* is given by

$$\nabla \cdot \boldsymbol{J} = -\frac{\partial \rho}{\partial t} . \tag{8.12}$$

This expression relates how the charge in a volume changes with the current density entering or leaving it.

The first two Maxwell's equations, (8.5) and (8.6), are a set of first-order differential equations that are coupled. The first step in their solution is to uncouple them. Using (8.9)–(8.11) to rewrite (8.5)–(8.8) gives

$$\nabla \times \boldsymbol{E} = -\boldsymbol{M}_\mathrm{i} - \mu \frac{\partial \boldsymbol{H}}{\partial t} , \tag{8.13}$$

$$\nabla \times \boldsymbol{H} = \boldsymbol{J}_\mathrm{i} + \sigma \boldsymbol{E} + \varepsilon \frac{\partial \boldsymbol{E}}{\partial t} , \tag{8.14}$$

$$\nabla \cdot \boldsymbol{E} = \frac{q_\mathrm{e}}{\varepsilon} , \tag{8.15}$$

$$\nabla \cdot \boldsymbol{H} = \frac{q_\mathrm{m}}{\mu} . \tag{8.16}$$

Taking the curl of both sides of (8.13) and (8.14) produces

$$\nabla \times \nabla \times \boldsymbol{E} = -\nabla \times \boldsymbol{M}_\mathrm{i} - \mu \frac{\partial}{\partial t} (\nabla \times \boldsymbol{H}) , \tag{8.17}$$

$$\nabla \times \nabla \times \boldsymbol{H} = \nabla \times \boldsymbol{J}_\mathrm{i} + \sigma \nabla \times \boldsymbol{E} + \varepsilon \frac{\partial}{\partial t} (\nabla \times \boldsymbol{E}) . \tag{8.18}$$

Substituting (8.14) into the right-hand side of (8.17), and using the vector identity $\nabla \times \nabla \times \boldsymbol{A} = \nabla \cdot (\nabla \cdot \boldsymbol{A}) - \nabla^2 \boldsymbol{A}$ on the left-hand side, gives

$$\nabla (\nabla \cdot \boldsymbol{E}) - \nabla^2 \boldsymbol{E} = -\nabla \times \boldsymbol{M}_\mathrm{i} - \mu \frac{\partial \boldsymbol{J}_\mathrm{i}}{\partial t} - \mu\sigma \frac{\partial \boldsymbol{E}}{\partial t} - \mu\varepsilon \frac{\partial^2 \boldsymbol{E}}{\partial t^2} . \tag{8.19}$$

Using (8.15), (8.19) can be rewritten as

$$\nabla^2 \boldsymbol{E} = \nabla \times \boldsymbol{M}_\mathrm{i} + \frac{1}{\varepsilon} \nabla q_\mathrm{e} + \mu \frac{\partial \boldsymbol{J}_\mathrm{i}}{\partial t} + \mu\sigma \frac{\partial \boldsymbol{E}}{\partial t} + \mu\varepsilon \frac{\partial^2 \boldsymbol{E}}{\partial t^2} . \tag{8.20}$$

This is an uncoupled second-order differential equation for $\boldsymbol{E}$.

The same process can be applied to obtain an uncoupled second-order differential equation for $\boldsymbol{H}$. Substituting (8.13) into the right-hand side of (8.18) and using the vector identity from the above on the left-hand side results in

$$\nabla (\nabla \cdot \boldsymbol{H}) - \nabla^2 \boldsymbol{H} = \nabla \times \boldsymbol{J}_\mathrm{i} - \sigma \boldsymbol{M}_\mathrm{i} - \varepsilon \frac{\partial \boldsymbol{M}_\mathrm{i}}{\partial t} - \mu\sigma \frac{\partial \boldsymbol{H}}{\partial t} - \mu\varepsilon \frac{\partial^2 H}{\partial t^2} . \tag{8.21}$$

Substituting (8.16) into (8.21) and reducing gives

$$\nabla^2 \boldsymbol{H} = -\nabla \times \boldsymbol{J}_\mathrm{i} + \sigma \boldsymbol{M}_\mathrm{i} + \frac{1}{\mu} \nabla q_\mathrm{m} + \varepsilon \frac{\partial \boldsymbol{M}_\mathrm{i}}{\partial t} + \mu\sigma \frac{\partial \boldsymbol{H}}{\partial t} + \mu\varepsilon \frac{\partial^2 \boldsymbol{H}}{\partial t^2} . \tag{8.22}$$

Equations (8.20) and (8.22) are vector wave equations for $\boldsymbol{E}$ and $\boldsymbol{H}$.

Only time-harmonic electromagnetic fields will be considered here. Time-harmonic fields can be represented by the real part of $\mathrm{e}^{\mathrm{i}\omega t}$, where ω is the angular frequency given by $2\pi f$, f is the linear frequency of the electromagnetic wave in hertz (Hz), and $i = \sqrt{-1}$. Since it is understood that $\mathrm{e}^{\mathrm{i}\omega t}$ appears in each term of Maxwell's equations, in can be dropped from the notation, allowing (8.5) through (8.8) to be rewritten as

$$\nabla \times \boldsymbol{E} = -\boldsymbol{M}_\mathrm{i} - \mathrm{i}\omega\mu \boldsymbol{H} , \tag{8.23}$$

$$\nabla \times \boldsymbol{H} = \boldsymbol{J}_\mathrm{i} + \boldsymbol{J}_\mathrm{c} + \mathrm{i}\omega\varepsilon \boldsymbol{E} , \tag{8.24}$$

$$\nabla \cdot \boldsymbol{E} = \frac{q_\mathrm{e}}{\varepsilon} , \tag{8.25}$$

$$\nabla \cdot \boldsymbol{H} = \frac{q_\mathrm{m}}{\mu} , \tag{8.26}$$

while the time-harmonic vector wave equations become

$$\nabla^2 \boldsymbol{E} = \nabla \times \boldsymbol{M}_\mathrm{i} + \frac{1}{\varepsilon}\nabla q_\mathrm{e} + \mathrm{i}\omega\mu \boldsymbol{J}_\mathrm{i} + \mathrm{i}\omega\mu\sigma \boldsymbol{E} - \omega^2 \mu\varepsilon \boldsymbol{E} \,, \tag{8.27}$$

$$\nabla^2 \boldsymbol{H} = -\nabla \times \boldsymbol{J}_\mathrm{i} + \sigma \boldsymbol{M}_\mathrm{i} + \frac{1}{\mu}\nabla q_\mathrm{m} + \mathrm{i}\omega\varepsilon \boldsymbol{M}_\mathrm{i} + \mathrm{i}\omega\mu\sigma \boldsymbol{H} - \omega^2 \mu\varepsilon \boldsymbol{H} \,. \tag{8.28}$$

Although $\boldsymbol{J}_\mathrm{i}$ and q_e can represent either real or virtual sources, $\boldsymbol{M}_\mathrm{i}$ and q_m are only virtual. In a source free region, $\boldsymbol{J}_\mathrm{i} = q_\mathrm{e} = 0$ and $\boldsymbol{M}_\mathrm{i} = q_\mathrm{m} = 0$, so that (8.27) and (8.28) reduce to the homogenous form of the wave equation

$$\nabla^2 \boldsymbol{E} - \gamma^2 \boldsymbol{E} = 0 \,, \tag{8.29}$$

$$\nabla^2 \boldsymbol{H} - \gamma^2 \boldsymbol{H} = 0 \,, \tag{8.30}$$

where

$$\gamma^2 = \mathrm{i}\omega\mu\sigma - \omega^2\mu\varepsilon \,. \tag{8.31}$$

Equations (8.29) and (8.30) describe electromagnetic waves traveling in a lossy medium with a propagation constant given by γ.

The last equations needed to completely define electromagnetic fields in the ocean are boundary conditions. Boundary conditions describe how the fields respond at an interface formed between two media of different constitutive properties, such as at the sea surface and seafloor. In this case, air is nonconducting with a permittivity and permeability the same as free space, while seawater is conductive with a high permittivity. Although the seafloor is conductive, its effective conductivity is one to four orders of magnitude less than the ocean's; depending on the depth the fields propagate below the bottom.

The generalized boundary conditions are well known. Their detailed derivation from Maxwell's equations can be found in [8.2]. Let the unit normal at the interface between two different media be given by $\hat{n}$ and pointing into the second region. Let the fields in media 1 and 2 be given by $\boldsymbol{E}_1$, $\boldsymbol{H}_1$, $\boldsymbol{D}_1$, $\boldsymbol{B}_1$ and $\boldsymbol{E}_2$, $\boldsymbol{H}_2$, $\boldsymbol{D}_2$, $\boldsymbol{B}_2$, respectively. Then the boundary conditions at the interface between them are

$$-\hat{n} \times (\boldsymbol{E}_2 - \boldsymbol{E}_1) = \boldsymbol{M}_\mathrm{s} \,, \tag{8.32}$$

$$\hat{n} \times (\boldsymbol{H}_2 - \boldsymbol{H}_1) = \boldsymbol{J}_\mathrm{s} \,, \tag{8.33}$$

$$\hat{n} \cdot (\boldsymbol{D}_2 - \boldsymbol{D}_1) = q_\mathrm{es} \,, \tag{8.34}$$

$$\hat{n} \cdot (\boldsymbol{B}_2 - \boldsymbol{B}_1) = q_\mathrm{ms} \,, \tag{8.35}$$

where $\boldsymbol{M}_\mathrm{s}$ and $\boldsymbol{J}_\mathrm{s}$ are the linear magnetic and electric surface currents along the boundary, and q_es and q_ms are the linear surface charge densities on the boundary. Although $\boldsymbol{M}_\mathrm{s}$ and q_ms are virtual boundary sources, they do arise in equivalent source representations of the fields within a closed region.

8.3 Plane Wave Propagation

The homogenous wave (8.29) and (8.30) can be solved in any of the separable coordinate systems. Their solution in rectangular coordinates will result in the expressions for plane waves. Using the method of separation of variables, the solution to (8.29) and (8.30) for a wave traveling along the positive z-axis is

$$\boldsymbol{E} = \boldsymbol{E}_0 \mathrm{e}^{-\gamma z} \,, \tag{8.36}$$

$$\boldsymbol{H} = \boldsymbol{H}_0 \mathrm{e}^{-\gamma z} \,, \tag{8.37}$$

where $\boldsymbol{E}_0$ and $\boldsymbol{H}_0$ are the amplitudes of the electric and magnetic fields. The propagation constant γ, defined by (8.31), can be written as

$$\gamma = \sqrt{\mathrm{i}\omega\mu(\sigma + \mathrm{i}\omega\varepsilon)} \,, \tag{8.38a}$$

$$\gamma = \alpha + \mathrm{i}\beta \,, \tag{8.38b}$$

and α is the attenuation constant in Np/m, and β is the phase constant in rad/m. If the waves were traveling in the negative z direction, then z in (8.36) and (8.37) would be replaced by $-z$. Solutions to the homogenous wave equation in cylindrical and spherical coordinates are derived in [8.2, pp. 12–18].

The electric and magnetic fields in (8.36) and (8.37) are linked through Maxwell's equations. For a source-free region, (8.24) can be written as

$$\boldsymbol{E} = \frac{1}{\sigma + \mathrm{i}\omega\varepsilon} \nabla \times \boldsymbol{H} \,. \tag{8.39}$$

If the magnetic field vector $\boldsymbol{H}$ in (8.39) has only an x component, given by H_x, taking its curl and substituting it in (8.36) gives

$$E_\mathrm{y} = \frac{-\gamma}{\sigma + \mathrm{i}\omega\varepsilon} H_\mathrm{x} \,. \tag{8.40}$$

Substituting (8.38a) into (8.40) and reducing gives

$$E_\mathrm{y} = -\eta H_\mathrm{x} \,, \tag{8.41}$$

where

$$\eta = \sqrt{\frac{\mathrm{i}\omega\mu}{\sigma + \mathrm{i}\omega\varepsilon}} \,. \tag{8.42}$$

The term η is called the *intrinsic impedance* of the medium and has the units of Ω. For a plane wave, the orthogonal electric and magnetic field components perpendicular to the propagation direction are related by the media's impedance.

In this section, only the ULF and ELF bands will be considered. Using the seawater dielectric permittivity constant from above, $\omega\varepsilon$ at the high end of the ELF band (3 kHz) is equal to $\approx 1.337\times10^{-5}$, which is much less than the water conductivity of 2.5 to 6. Therefore, the $\mathrm{i}\omega\varepsilon$ term will be dropped for electromagnetic frequencies up through the ELF band. In fact, displacement currents in seawater do not become appreciable until frequencies approach the high megahertz region.

At ELF frequencies and below, the seawater's attenuation constant is equal to its phase constant. Eliminating the $\mathrm{i}\omega\varepsilon$ in (8.38) gives

$$\gamma = \sqrt{\mathrm{i}\omega\mu\sigma}\,, \tag{8.43a}$$

$$\gamma = \sqrt{\frac{\omega\mu\sigma}{2}}(1+\mathrm{i})\,, \tag{8.43b}$$

$$\alpha = \beta = \sqrt{\frac{\omega\mu\sigma}{2}}\,. \tag{8.43c}$$

The distance at which the wave attenuates to e^{-1} is called the skin depth, and is denoted by the term δ. The skin depth is the reciprocal of the attenuation constant, $\delta = \frac{1}{\alpha}$, which for seawater is

$$\delta = \sqrt{\frac{2}{\omega\mu\sigma}}\,. \tag{8.44}$$

The field's wavelength λ is related to the propagation by $\lambda = \frac{2\pi}{\beta}$, and in this case can be written as

$$\lambda = 2\pi\sqrt{\frac{2}{\omega\mu\sigma}}\,, \tag{8.45}$$

while the speed of an electromagnetic wave v, is equal to $f\lambda$, and in seawater is given by

$$v = \sqrt{\frac{2\omega}{\mu\sigma}}\,. \tag{8.46}$$

Finally, the intrinsic impedance of seawater at ELF frequencies and below is expressed as

$$\eta = \sqrt{\frac{\mathrm{i}\omega\mu}{\sigma}}\,. \tag{8.47}$$

In the ocean, an electromagnetic wave's skin depth and wavelength is inversely proportional to the square root of its frequency, while its speed and intrinsic impedance is directly proportional to it.

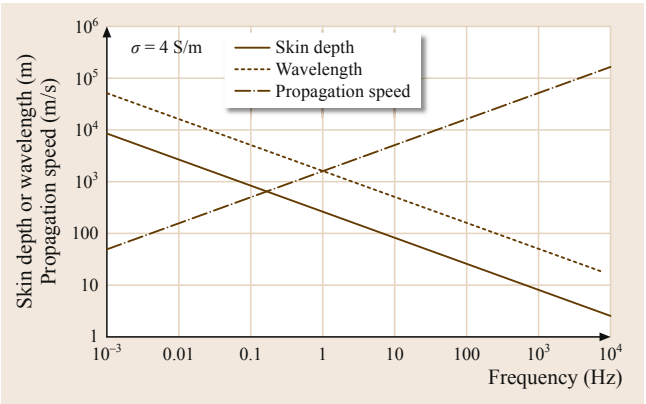


Fig. 8.1 Propagation characteristics of an electromagnetic wave in seawater

The characteristics of a propagating electromagnetic wave in the ocean are quite different than in air. In free space, the propagation speed of an electromagnetic field is always the same as that of light c (3×10^8 m/s), regardless of its frequency. This is not the case in the ocean. The skin depth, wavelength, and propagation speed given by (8.44) through (8.46) are plotted in Fig. 8.1 over the frequency range from 1 mHz to 10 kHz, using a seawater conductivity of 4 S/m. At 1 Hz, the skin depth is 252 m, while the wavelength and propagation speed are 1581 m and 1581 m/s. The propagation speed of electromagnetic fields in seawater at 1 Hz is more than 5 orders of magnitude slower than in air, and is approximately equal to the speed of sound in the ocean!

The intrinsic impedance of seawater is much smaller than free space. In air, η can be computed from (8.42) by setting $\sigma = 0$. This results in a free space impedance of 377 Ω, and is independent of frequency. Conversely, the intrinsic impedance of seawater

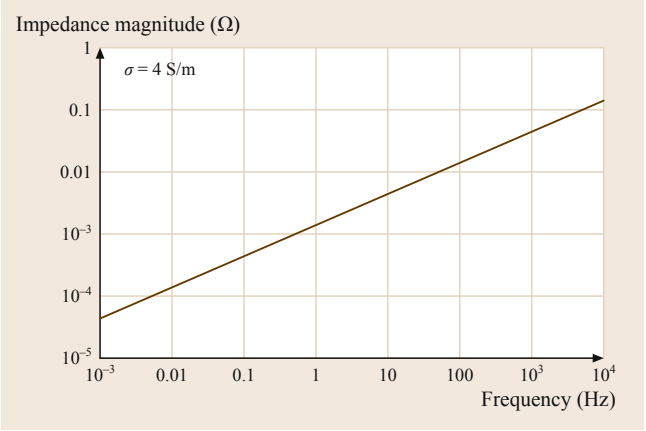


Fig. 8.2 Magnitude of the intrinsic impedance of seawater as a function of frequency

is $1.405\times10^{-3}(1+\mathrm{i})\,\Omega$ at 1 Hz. The magnitude of (8.47) has been plotted in Fig. 8.2 for the same parameters as above. As will be shown later, the difference in impedances between air and the ocean will significantly affect the reflection and transmission of electromagnetic fields at the sea surface.

As the frequency of the electromagnetic wave increases, so does its velocity. At the top end of the ELF band, an electromagnetic field propagates more than 100 times faster than sound, but the skin depth decreases to less than 3 m. For this reason, underwater electromagnetic communication systems for short distance wireless data transfers are preferred over acoustic modems. In addition, the rapid attenuation of electromagnetic fields in the ocean can be exploited to reduce interference from nearby transmitters, and for covert underwater communications systems that must avoid long range detection.

Another advantage of ULF and ELF electromagnetic fields is their ability to cross the air–ocean boundary. The description of plane wave transmission from air into conducting seawater is more complex than for fresh water. Therefore, the formulation of the *reflection and transmission coefficients* at the sea surface will begin with that of a plane wave incident on fresh water, such as a lake.

8.4 Reflection and Transmission of a Plane Wave at the Surface of Fresh Water

The reflection and transmission coefficients of a plane wave at the surface of fresh water depend on the polarization of the incident field. If the electric field vector of the incident uniform plane wave is perpendicular to the plane of incidence as shown in Fig. 8.3, then the wave is said to have a *perpendicular polarization* or *horizontal polarization*. Using the coordinate system and geometry of Fig. 8.3, the incident perpendicular polarized electric $\boldsymbol{E}^{\mathrm{i}}_{\perp}$ and magnetic $\boldsymbol{H}^{\mathrm{i}}_{\perp}$ field of a uniform plane wave can be written as

$$\boldsymbol{E}^{\mathrm{i}}_{\perp} = \hat{a}_{\mathrm{y}} E_0 \mathrm{e}^{-\mathrm{i}\beta_{\mathrm{a}}(x\sin\theta_{\mathrm{i}}+z\cos\theta_{\mathrm{i}})}\,, \tag{8.48}$$

$$\boldsymbol{H}^{\mathrm{i}}_{\perp} = (-\hat{a}_{\mathrm{x}}\cos\theta_{\mathrm{i}} + \hat{a}_{\mathrm{z}}\sin\theta_{\mathrm{i}}) \frac{E_0}{\eta_{\mathrm{a}}}\mathrm{e}^{-\mathrm{i}\beta_{\mathrm{a}}(x\sin\theta_{\mathrm{i}}+z\cos\theta_{\mathrm{i}})}\,, \tag{8.49}$$

where E_0 is the amplitude of the incident electric field, β_{a} is the propagation constant for the incident wave in air, η_{a} is the intrinsic impedance of air (377 Ω), θ_{i} is the incidence angle with respect to the vertical, and $\hat{a}_{\mathrm{x}}$, $\hat{a}_{\mathrm{y}}$, $\hat{a}_{\mathrm{z}}$ are the unit vectors in their respective directions. The relationship between the electric and magnetic field components for a uniform plane wave given by (8.41) was used to arrive at (8.49). Similarly, the reflected electric $\boldsymbol{E}^{\mathrm{r}}_{\perp}$ and magnetic $\boldsymbol{H}^{\mathrm{r}}_{\perp}$ plane waves are given by

$$\boldsymbol{E}^{\mathrm{r}}_{\perp} = \hat{a}_{\mathrm{y}}\Gamma_{\perp}E_0\mathrm{e}^{-\mathrm{i}\beta_{\mathrm{a}}(x\sin\theta_{\mathrm{r}}-z\cos\theta_{\mathrm{r}})}\,, \tag{8.50}$$

$$\boldsymbol{H}^{\mathrm{r}}_{\perp} = (\hat{a}_{\mathrm{x}}\cos\theta_{\mathrm{r}} + \hat{a}_{\mathrm{z}}\sin\theta_{\mathrm{r}}) \times \frac{\Gamma_{\perp}E_0}{\eta_{\mathrm{a}}}\mathrm{e}^{-\mathrm{i}\beta_{\mathrm{a}}(x\sin\theta_{\mathrm{r}}-z\cos\theta_{\mathrm{r}})}\,, \tag{8.51}$$

where $\Gamma_{\perp}$ is the reflection coefficient for a perpendicular polarized wave, and θ_{r} is the angle of the reflected wave with respect to the vertical. Finally, the transmitted electric $\boldsymbol{E}^{\mathrm{t}}_{\perp}$ and magnetic $\boldsymbol{H}^{\mathrm{t}}_{\perp}$ field can be expressed as

$$\boldsymbol{E}^{\mathrm{t}}_{\perp} = \hat{a}_{\mathrm{y}}T_{\perp}E_0\mathrm{e}^{-\mathrm{i}\beta_{\mathrm{f}}(x\sin\theta_{\mathrm{t}}+z\cos\theta_{\mathrm{t}})}\,, \tag{8.52}$$

$$\boldsymbol{H}^{\mathrm{t}}_{\perp} = (-\hat{a}_{\mathrm{x}}\cos\theta_{\mathrm{t}} + \hat{a}_{\mathrm{z}}\sin\theta_{\mathrm{t}}) \times \frac{T_{\perp}E_0}{\eta_{\mathrm{f}}}\mathrm{e}^{-\mathrm{i}\beta_{\mathrm{f}}(x\sin\theta_{\mathrm{t}}+z\cos\theta_{\mathrm{t}})}\,, \tag{8.53}$$

where $T_{\perp}$ is the transmission coefficient for a perpendicular polarized wave, and θ_{t} is the angle of the transmitted wave or *refraction angle* with respect to the vertical, and β_{f} and η_{f} are the propagation constant and intrinsic impedance of freshwater, respectively.

Equations (8.48)–(8.53) can be related through the boundary conditions of the continuity of the horizontal components of the electric and magnetic fields at the surface of the water. Since there are no sources at the air–water interface, $\boldsymbol{M}_{\mathrm{s}}$ and $\boldsymbol{J}_{\mathrm{s}}$ in (8.32) and (8.33)

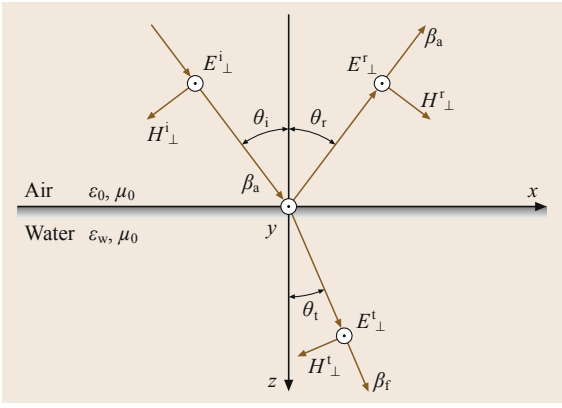


Fig. 8.3 Plane wave incident on the surface of fresh water, perpendicular polarization

are set to zero. Applying these boundary conditions to (8.48)–(8.53) at z = 0 gives

$$e^{-i\beta_a x \sin\theta_i} + \Gamma_\perp e^{-i\beta_a x \sin\theta_r} = T_\perp e^{-i\beta_f x \sin\theta_t} , \quad (8.54)$$

$$\frac{1}{\eta_a}\left(-\cos\theta_i e^{-i\beta_a x \sin\theta_i} + \cos\theta_r \Gamma_\perp e^{-i\beta_a x \sin\theta_r}\right) = -\frac{T_\perp}{\eta_f}\cos\theta_t e^{-i\beta_f x \sin\theta_t} . \quad (8.55)$$

Equations (8.54) and (8.55) are sufficient to solve for θ_r, θ_t, $\Gamma_\perp$, and $T_\perp$.

Breaking (8.54) and (8.55) into their real and imaginary parts will produce four equations with which to solve for the four unknowns. Following this procedure results in the relationships between the incident, reflected, and transmission angles given by

$$\theta_r = \theta_i , \quad (8.56a)$$

$$\beta_a \sin\theta_i = \beta_f \sin\theta_t . \quad (8.56b)$$

Equations (8.56a) and (8.56b) are called *Snell's laws* of reflection and refraction, respectively. Placing (8.56a) and (8.56b) into (8.54) and (8.55) and solving for the reflection and transmission coefficients produces the expressions

$$\Gamma_\perp = \frac{\boldsymbol{E}^r_\perp}{\boldsymbol{E}^i_\perp} = \frac{\eta_f \cos\theta_i - \eta_a \cos\theta_t}{\eta_f \cos\theta_i + \eta_a \cos\theta_t} , \quad (8.57)$$

$$T_\perp = \frac{\boldsymbol{E}^t_\perp}{\boldsymbol{E}^i_\perp} = \frac{2\eta_f \cos\theta_i}{\eta_f \cos\theta_i + \eta_a \cos\theta_t} . \quad (8.58)$$

Because air and water are both nonmagnetic (8.57) and (8.58) can be reduced to

$$\Gamma_\perp = \frac{\cos\theta_i - \sqrt{\varepsilon_w} cos\theta_t}{\cos\theta_i + \sqrt{\varepsilon_w}\cos\theta_t} , \quad (8.59)$$

$$T_\perp = \frac{2\cos\theta_i}{\cos\theta_i + \sqrt{\varepsilon_w}\cos\theta_t} , \quad (8.60)$$

where ε_w is the dielectric constant of water defined previously as 80.1.

A similar procedure is used to obtain the reflection and transmission coefficients for a uniform plane wave that has a *parallel polarization* incident on the surface of fresh water. A wave of this type also referred to as having a *vertical polarization* has an electric field that is parallel to the plane of incidence as shown in Fig. 8.4. Using the coordinate system and geometry of Fig. 8.4, the incident parallel polarized electric $\boldsymbol{E}^i_\parallel$ and magnetic $\boldsymbol{H}^i_\parallel$ field of a uniform plane wave can be written as

$$\boldsymbol{E}^i_\parallel = (\hat{a}_x \cos\theta_i - \hat{a}_z \sin\theta_i) E_0 e^{-i\beta_a(x\sin\theta_i + z\cos\theta_i)} , \quad (8.61)$$

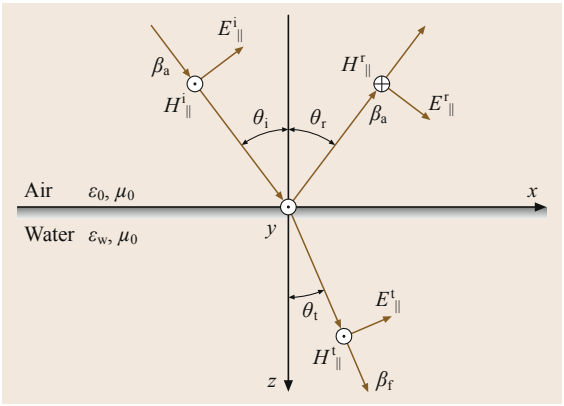


Fig. 8.4 Plane wave incident on the surface of fresh water, parallel polarization

$$\boldsymbol{H}^i_\parallel = \hat{a}_y \frac{E_0}{\eta_a} e^{-i\beta_a(x\sin\theta_i + z\cos\theta_i)} . \quad (8.62)$$

Similarly, the reflected electric $\boldsymbol{E}^r_\parallel$ and magnetic $\boldsymbol{H}^r_\parallel$ plane waves are given by

$$\boldsymbol{E}^r_\parallel = (\hat{a}_x \cos\theta_r + \hat{a}_z \sin\theta_r)\Gamma_\parallel E_0 e^{-i\beta_a(x\sin\theta_r - z\cos\theta_r)} , \quad (8.63)$$

$$\boldsymbol{H}^r_\parallel = -\hat{a}_y \frac{\Gamma_\parallel E_0}{\eta_a} e^{-i\beta_a(x\sin\theta_r - z\cos\theta_r)} , \quad (8.64)$$

and the transmitted electric $\boldsymbol{E}^t_\parallel$ and magnetic $\boldsymbol{H}^t_\parallel$ plane waves are

$$\boldsymbol{E}^t_\parallel = (\hat{a}_x \cos\theta_t - \hat{a}_z \sin\theta_t) T_\parallel E_0 e^{-i\beta_f(x\sin\theta_t + z\cos\theta_t)} , \quad (8.65)$$

$$\boldsymbol{H}^t_\parallel = \hat{a}_y \frac{T_\parallel E_0}{\eta_f} e^{-i\beta_f(x\sin\theta_t + z\cos\theta_t)} , \quad (8.66)$$

where $\Gamma_\parallel$ and $T_\parallel$ represent the reflection and transmission coefficients for a parallel polarized incident wave, respectively. All other parameters have been defined previously.

The analysis proceeds in the same manner as used for the perpendicular polarized waves. Applying the continuity boundary conditions for the tangential components of the electric and magnetic fields at the air–water interface ($z = 0$) give the equations

$$\cos\theta_i e^{-i\beta_a x \sin\theta_i} + \cos\theta_r \Gamma_\parallel e^{-i\beta_a x \sin\theta_r} = \cos\theta_t T_\parallel e^{-i\beta_f x \sin\theta_t} , \quad (8.67)$$

$$\frac{1}{\eta_a}\left(e^{-i\beta_a x \sin\theta_i} - \Gamma_\parallel e^{-i\beta_a x \sin\theta_r}\right) = \frac{T_\parallel}{\eta_f} e^{-i\beta_f x \sin\theta_t} . \quad (8.68)$$

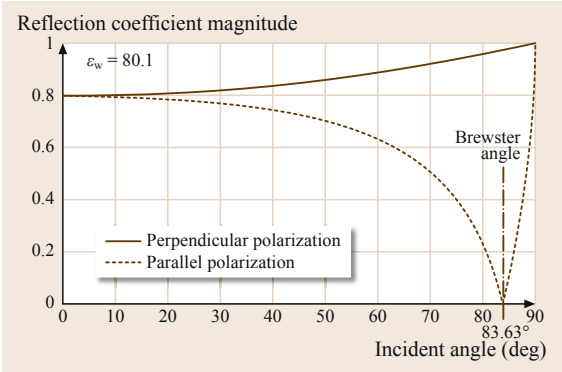


Fig. 8.5 Reflection coefficients for a perpendicular and parallel polarized uniform plane wave incident on the surface of fresh water

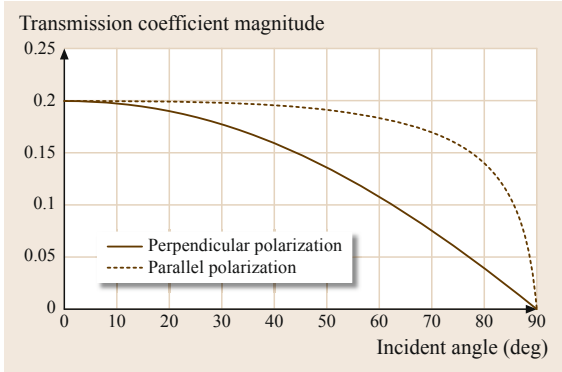


Fig. 8.6 Transmission coefficients for a perpendicular and parallel polarized uniform plane wave incident on the surface of fresh water

Once again, equating the real and imaginary parts of (8.67) and (8.68) will yield Snell's laws of reflection and refraction (8.56a) and (8.56b), along with the reflection and transmission coefficients for parallel polarization written as

$$\Gamma_{\parallel} = \frac{\boldsymbol{E}^{\mathrm{r}}_{\parallel}}{\boldsymbol{E}^{\mathrm{i}}_{\parallel}} = \frac{-\eta_{\mathrm{a}} \cos\theta_{\mathrm{i}} + \eta_{\mathrm{f}} \cos\theta_{\mathrm{t}}}{\eta_{\mathrm{a}} \cos\theta_{\mathrm{i}} + \eta_{\mathrm{f}} \cos\theta_{\mathrm{t}}} , \tag{8.69}$$

$$T_{\parallel} = \frac{\boldsymbol{E}^{\mathrm{t}}_{\parallel}}{\boldsymbol{E}^{\mathrm{i}}_{\parallel}} = \frac{2\eta_{\mathrm{f}} \cos\theta_{\mathrm{i}}}{\eta_{\mathrm{a}} \cos\theta_{\mathrm{i}} + \eta_{\mathrm{f}} \cos\theta_{\mathrm{t}}} . \tag{8.70}$$

For the interface between air and fresh water, (8.69) and (8.70) reduce to

$$\Gamma_{\parallel} = \frac{-\sqrt{\epsilon}_{\mathrm{w}} \cos\theta_{\mathrm{i}} + \cos\theta_{\mathrm{t}}}{\sqrt{\epsilon}_{\mathrm{w}} \cos\theta_{\mathrm{i}} + \cos\theta_{\mathrm{t}}} , \tag{8.71}$$

$$T_{\parallel} = \frac{2\cos\theta_{\mathrm{i}}}{\sqrt{\epsilon}_{\mathrm{w}} \cos\theta_{\mathrm{i}} + \cos\theta_{\mathrm{t}}} . \tag{8.72}$$

Although (8.59) and (8.60) are similar in form to (8.71) and (8.72), a simple example will highlight the differences.

The reflection and transmission coefficients for perpendicular and parallel polarizations were evaluated for a plane wave incident on the surface of fresh water with a dielectric constant of 80.1. Using Snell's law of refraction (8.56) at the air–water interface, the $\cos\theta_{\mathrm{t}}$ term in (8.59), (8.60), (8.71), and (8.72) can be computed from

$$\cos\theta_{\mathrm{t}} = \sqrt{1 - \frac{1}{\varepsilon_{\mathrm{w}}} \sin^2\theta_{\mathrm{i}}} . \tag{8.73}$$

The magnitudes of $\Gamma_{\perp}$ and $\Gamma_{\parallel}$ are plotted in Fig. 8.5 for all incidence angles from 0 to 90°, while $T_{\perp}$, and $T_{\parallel}$ are similarly plotted in Fig. 8.6. Interestingly, $\Gamma_{\parallel}$ goes through zero at an angle of 83.63°. This angle is called the *Brewster angle* for which there is no reflection from the water's surface. The Brewster angle only exists for incident plane waves that have parallel polarization.

8.5 Plane Wave Incident on Seawater

The general forms of the reflection and transmission coefficients computed for fresh water are also valid when the wave is incident on the surface of the electrically conducting ocean. In this case, the intrinsic impedance of fresh water η_{f} used in (8.57), (8.58), (8.69), and (8.70) is replaced with that of seawater η_{s} given by (8.47). Since η_{s} is a complex impedance with both real and imaginary parts, $\Gamma_{\perp}$, $T_{\perp}$, $\Gamma_{\parallel}$, and $T_{\parallel}$ will all be complex. This means that both the amplitude and phase of the reflected and transmitted fields will be modified by the ocean's surface.

The most dramatic effect of the sea surface on a uniform plane wave is its impact on the transmitted field's propagation characteristics. Rewriting Snell's law of refraction (8.56b) for seawater gives

$$\mathrm{i}\beta_{\mathrm{a}} \sin\theta_{\mathrm{i}} = \gamma \sin\theta_{\mathrm{t}} , \tag{8.74}$$

where γ has both a real and imaginary part given by (8.43c). In this case, the $\sin\theta_{\mathrm{t}}$ term in (8.74) can be expressed as

$$\sin\theta_{\mathrm{t}} = \frac{\mathrm{i}\beta_{\mathrm{a}}}{\alpha + \mathrm{i}\beta} \sin\theta_{\mathrm{i}} . \tag{8.75}$$

Using (8.75) and the identity $\cos\theta_t = \sqrt{1-\sin^2\theta_t}$, the term $\cos\theta_t$ becomes

$$\cos\theta_t = \sqrt{1-\left(\frac{i\beta_a}{\alpha+i\beta}\right)^2 \sin^2\theta_i}\,, \tag{8.76}$$

which can be reduced to a general form of

$$\cos\theta_t = s e^{i\xi}\,. \tag{8.77}$$

It is clear that the wave transmitted into the sea will be at an angle that is complex. The idea of a real and imaginary transmission angle needs further explanation.

The electric field for a perpendicular or parallel polarized field transmitted into seawater, (8.52) or (8.65), can be written in a general form given by

$$\boldsymbol{E}^t = \boldsymbol{E}_0^t e^{-\gamma(x\sin\theta_t + z\cos\theta_t)}\,, \tag{8.78}$$

that can be expanded to

$$\boldsymbol{E}^t = \boldsymbol{E}_0^t \exp[-(\alpha+i\beta)(x\sin\theta_t + z\cos\theta_t)]\,. \tag{8.79}$$

All the terms in front of the exponential in either (8.52) or (8.65) have been lumped into $\boldsymbol{E}_0^t$. Placing (8.75) and (8.77) into (8.79) gives

$$\boldsymbol{E}^t = \boldsymbol{E}_0^t \exp\left[-(\alpha+i\beta)\left(x\frac{i\beta_a}{\alpha+i\beta}\sin\theta_i + zs(\cos\xi + i\sin\xi)\right)\right], \tag{8.80}$$

and can be reduced to

$$\boldsymbol{E}^t = \boldsymbol{E}_0^t e^{-zp}\exp[-i(x\beta_a\sin\theta_i + zq)]\,, \tag{8.81}$$

where

$$p = s(\alpha\cos\xi - \beta\sin\xi)\,, \tag{8.82}$$

$$q = s(\alpha\sin\xi + \beta\cos\xi)\,. \tag{8.83}$$

The expression given by (8.81) describes a nonuniform wave.

A plane wave refracted into conducting seawater has planes of constant amplitude and phase that are not coincident. Equation (8.81) shows that the planes of constant amplitude (z = constant) are parallel to the ocean's surface as drawn in Fig. 8.7. The attenuation vector of the refracted wave is $\hat{\alpha}_s = \hat{a}_z p$, where p is the attenuation constant. To find the direction of propagation of the refracted wave, the phase term given in (8.81) is rewritten as

$$x\beta_a\sin\theta_i + zq = \beta_s(x\sin\Psi + z\cos\Psi)\,, \tag{8.84}$$

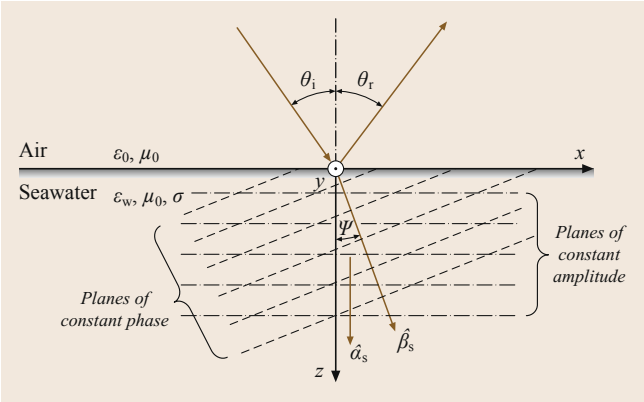


Fig. 8.7 Plane wave refracted into seawater

where

$$\beta_s = \sqrt{u^2+q^2}\,, \tag{8.85}$$

$$\sin\Psi = \frac{u}{\sqrt{u^2+q^2}}\,, \tag{8.86}$$

$$\cos\Psi = \frac{q}{\sqrt{u^2+q^2}}\,, \tag{8.87}$$

$$u = \beta_a\sin\theta_i\,. \tag{8.88}$$

The phase propagation direction for the refracted wave is $\hat{\beta}_s = \hat{a}_x\sin\Psi + \hat{a}_z\cos\Psi$, and produces planes of constant phase perpendicular to $\hat{\beta}_s$ as drawn in Fig. 8.7. The angle of refraction for the direction of propagation is determined from

$$\Psi = \tan^{-1}\left(\frac{u}{q}\right). \tag{8.89}$$

The phase velocity of the refracted wave is given by

$$v_r = \frac{\omega}{\beta_s}\,. \tag{8.90}$$

The refracted wave's attenuation, direction of propagation and velocity are all dependent on the incident angle; but to what extent? Another example is in order.

In this example, a 1 Hz plane wave is incident on the surface of the open ocean at an oblique angle θ_i. Using the intrinsic parameters for air $\sigma = 0$, μ_0, and ε_0 in (8.38) and (8.42), the free space propagation constant β_a is equal to $\frac{\omega}{c}$ (c is the speed of light in a vacuum is 3×10^8 m/s), and the intrinsic impedance of air is 377 Ω that is independent of frequency. For a seawater conductivity of 4 S/m, (8.76) can be written as

$$\cos\theta_t \approx \sqrt{1-(-i2.2\times10^{-11})\sin^2\theta_i} \approx 1$$

for all angles of incidence. This produces the values $s = 1$, $\xi = 0$, and $q = \beta$ from (8.77) and (8.83). The electromagnetic wave's angle of refraction in the ocean, given by (8.89), reduces to

$$\Psi = \tan^{-1}\left(\frac{\omega}{c\beta}\sin\theta_\mathrm{i}\right). \tag{8.91}$$

At 1 Hz, $\frac{\omega}{c\beta} = 5.27\times10^{-6}$, and the maximum angle of refraction is $\Psi_{\max} \le \tan^{-1}(5.27\times10^{-6})$ or $\Psi_{\max} \le 3\times10^{-4\circ}$ (≈ 1 s of an arc). Even at 3 kHz the refracted angle is computed to be less than 0.02°. Therefore, for frequencies up through the ELF band, a plane wave incident on the surface of the ocean at any angle will refract, propagate, and attenuate toward the bottom along a path that is nearly vertical.

The analysis presented above for a single interface can be extended to include any number of layers with different intrinsic parameters. The derivation of the iterative equation can be found in [8.2, pp. 110–120]. Plane wave reflection and transmission for multilayered ocean environments would arise when modeling depth variations of uniformly stratified seawater or seafloor conductivities.

8.6 Magnetic and Electric Dipoles in an Unbounded Ocean

Strictly speaking, plane waves do not exist in nature. They can be used to approximate an electromagnetic field that is far from its source (transmitter), or in the case where a uniform distribution of sources extend to very large distances in comparison to the receiver's range to it. An example of the latter case would be the geomagnetic fields present at the ocean's surface generated by solar-wind-induced electric currents in the ionosphere.

The two elemental sources of electromagnetic fields are the electric and magnetic point dipoles. An electric dipole source p is created by an electric current I flowing in a conductor over a linear distance l ($p = Il$), and has units of A m. A magnetic dipole m is produced if the current flows around a closed circular path that encloses an area a ($m = Ia$), and has units of A m^2. If the observation point is at a distance that is large compared to the source's dimensions, then it can be represented mathematically as an infinitesimally small or point dipole. With sources present, the time-harmonic vector wave equations, (8.27) and (8.28), are now inhomogeneous.

Although it is possible to solve for the electric and magnetic fields directly from the inhomogeneous vector wave equations, it is generally mathematically advantageous to replace them with equivalent representations in terms of the *magnetic vector potential* $\boldsymbol{A}$ and *electric vector potential* $\boldsymbol{F}$. The solutions to the inhomogeneous vector wave equations in terms of vector potentials are derived in [8.3], and can be written as

$$\boldsymbol{E} = -\mathrm{i}\omega\boldsymbol{A} + \frac{1}{\mu\sigma}\nabla(\nabla\cdot\boldsymbol{A}) - \frac{1}{\varepsilon}\nabla\times\boldsymbol{F}\,, \tag{8.92}$$

$$\boldsymbol{H} = -\mathrm{i}\omega\boldsymbol{F} + \frac{1}{\mu\sigma}\nabla(\nabla\cdot\boldsymbol{F}) + \frac{1}{\mu}\nabla\times\boldsymbol{A}\,, \tag{8.93}$$

where

$$\boldsymbol{A} = \frac{\mu}{4\pi}\iiint_V \boldsymbol{J}_\mathrm{i}\frac{\mathrm{e}^{-\gamma R}}{R}\,\mathrm{d}v'\,, \tag{8.94}$$

$$\boldsymbol{F} = \frac{\varepsilon}{4\pi}\iiint_V \boldsymbol{M}_\mathrm{i}\frac{\mathrm{e}^{-\gamma R}}{R}\,\mathrm{d}v'\,, \tag{8.95}$$

$$R = \sqrt{(x-x')^2 + (y-y')^2 + (z-z')^2}\,, \tag{8.96}$$

and the elements of the impressed electric $\boldsymbol{J}_\mathrm{i}$ and magnetic $\boldsymbol{M}_\mathrm{i}$ current densities are located at (x', y', z') with the observation point at (x, y, z). The integration is taken over the entire volume of the source.

Equations (8.92)–(8.96) are sufficient to compute the electric and magnetic fields from dipole sources. Consider first an electric point dipole of strength p located at the origin of the coordinate system shown in Fig. 8.8, and aligned along the z-axis. The impressed source's electric current density for this dipole can be represented mathematically as

$$\boldsymbol{J}_\mathrm{i} = \hat{a}_z p\delta(x')\delta(y')\delta(z')\,, \tag{8.97}$$

where the δ used here represents the Dirac delta function. Placing (8.97) into (8.94) yields the magnetic vector potential

$$A_\mathrm{z} = \frac{\mu p\mathrm{e}^{-\gamma r}}{4\pi r}\,, \tag{8.98}$$

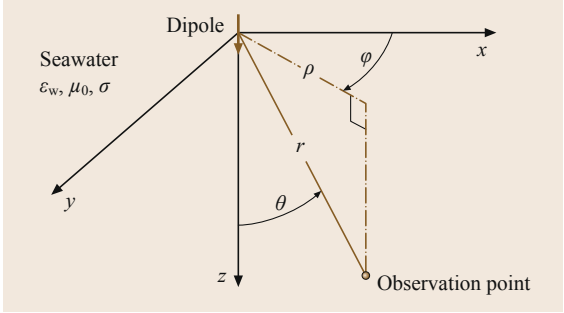


Fig. 8.8 Coordinate system for a dipole in an unbounded ocean

where r is the radial distance from the origin as shown in Fig. 8.8. Converting (8.98) to spherical coordinates and placing it into (8.92) and (8.93), the electric and magnetic fields of a point electric dipole in an unbounded ocean are written as

$$E_r = \frac{p\cos\theta}{2\pi\sigma r^3}(1+\gamma r)e^{-\gamma r}\,, \tag{8.99}$$

$$E_\theta = \frac{p\sin\theta}{4\pi\sigma r^3}(1+\gamma r+\gamma^2 r^2)e^{-\gamma r}\,, \tag{8.100}$$

$$H_\varphi = \frac{p\sin\theta}{4\pi r^2}(1+\gamma r)e^{-\gamma r}\,. \tag{8.101}$$

If the point dipole was located at coordinates (x_0, y_0, z_0) instead of the origin, then r becomes

$$\sqrt{(x-x_0)^2+(y-y_0)^2+(z-z_0)^2}\,.$$

The derivation of a magnetic point dipole follows that shown above for the electric. In this case, the magnetic point dipole has a source strength m that is also located at the origin, and is aligned along the z-axis. A dipole of this type could be produced by an electric current flowing in the φ direction of the coordinate system in Fig. 8.8. This is mathematically equivalent to a virtual magnetic current source aligned along the z-axis. Therefore, the equivalent impressed magnetic current density for this dipole can be written as

$$\boldsymbol{M}_i = \hat{a}_z m\delta(x')\delta(y')\delta(z')\,. \tag{8.102}$$

Using (8.102) in (8.95) yields the electric vector potential

$$F_z = \frac{\varepsilon m e^{-\gamma r}}{4\pi r}\,. \tag{8.103}$$

Converting (8.103) to spherical coordinates and placing it into (8.92) and (8.93), the electric and magnetic fields of a point magnetic dipole in an unbounded ocean can be written as

$$H_r = \frac{m\cos\theta}{2\pi r^3}(1+\gamma r)e^{-\gamma r}\,, \tag{8.104}$$

$$H_\theta = \frac{m\sin\theta}{4\pi r^3}(1+\gamma r+\gamma^2 r^2)e^{-\gamma r}\,, \tag{8.105}$$

$$E_\varphi = \frac{-i\omega\mu m\sin\theta}{4\pi r^2}(1+\gamma r)e^{-\gamma r}\,. \tag{8.106}$$

It should be noted that the electromagnetic fields for both the electric and magnetic dipole sources are circularly symmetric and are not a function of the φ coordinate.

The equations for the electromagnetic fields of electric and magnetic point dipoles in an unbounded ocean reduce to the standard static magnetic and electric expressions by setting the frequency in (8.99) to (8.101) and (8.104) to (8.106) to zero. Carrying this out gives the electric and magnetic fields of a static electric dipole in an unbounded ocean as

$$E_r = \frac{p\cos\theta}{2\pi\sigma r^3}\,, \tag{8.107}$$

$$E_\theta = \frac{p\sin\theta}{4\pi\sigma r^3}\,, \tag{8.108}$$

$$H_\varphi = \frac{p\sin\theta}{4\pi r^2}\,, \tag{8.109}$$

and for the static magnetic dipole (8.104) to (8.106) reduce to

$$H_r = \frac{m\cos\theta}{2\pi r^3}\,, \tag{8.110}$$

$$H_\theta = \frac{m\sin\theta}{4\pi r^3}\,, \tag{8.111}$$

$$E_\varphi = 0\,. \tag{8.112}$$

It should be noted that all electric and magnetic field components for static dipoles fall off in the radial direction as r^{-3} except for the φ component of the magnetic field of the static electric dipole, which fall off as r^{-2}, and the φ component of the electric field of the static magnetic dipole, which is zero.

Except for the zero E_φ component of the static magnetic dipole, the static and time-harmonic dipole equations differ by a factor of either $(1+\gamma r)e^{-\gamma r}$ or $(1+\gamma r+\gamma^2 r^2)e^{-\gamma r}$. The magnitude of these two multiplication factors are plotted in Fig. 8.9 as a function of the radial distance in skin depths. This data shows the range over which the electric and magnetic field equations for static sources can be used to estimate the time-harmonic fields. Depending on accuracy requirements, the static equations could be used out to distances approaching a skin depth. It is interesting that $|(1+\gamma r+\gamma^2 r^2)e^{-\gamma r}|$ is 1 at a distance equal to a half-

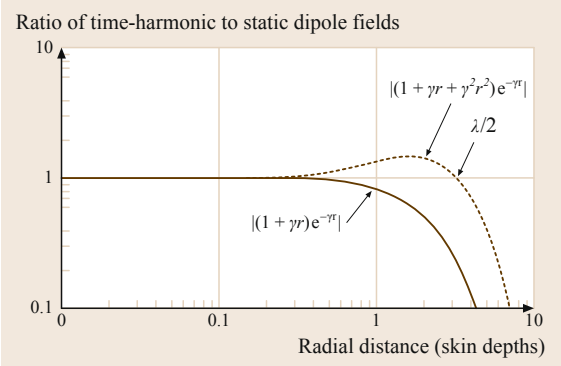


Fig. 8.9 Correction factors to convert electric and magnetic fields computed for static dipoles to those of time-harmonic sources

wavelength, as indicated in Fig. 8.9. The region in close vicinity to a time-harmonic dipole for which the static equations are sufficiently accurate is sometimes called the *quasi-DC* (direct current) range.

The equations for electric and magnetic dipoles in an unbounded ocean are sufficient if the range between the source and observation point is small in comparison to their distances from the sea surface or seafloor. If this is not the case, then reflections from one or both of the interfaces must be accounted for in the magnetic and electric field formulations. Since the dipole sources are close to an interface, simple plane wave reflection coefficients cannot be used. A more complete and, unfortunately, more complex treatment are required.

8.7 Magnetic and Electric Dipoles in a Bounded Ocean

In this section, magnetic and electric dipoles submerged in a bounded ocean will be considered. What is meant by *a bounded ocean* is that the separation between the source and observation point are comparable to their distance from either the sea surface or bottom. The change in conductivity between the seawater and air or the seafloor produces an interface that modifies the spreading and propagation of fields generated by submerged magnetic and electric sources. The effects of the interfaces are accounted for in the descriptive equations by satisfying the boundary conditions along them. Therefore, mathematical solutions to the inhomogeneous wave equations must be in a form that will easily accommodate solving for the boundary conditions when a source is near an interface.

Because of their close proximity to one of the ocean's interface, the solution to the problem of a submerged magnetic or electric dipole will also depend on their orientation with respect to the sea surface or seafloor. As a result, there are four types of dipoles that must be addressed separately when computing their electric and magnetic fields in a bounded ocean. They are; the vertical electric dipole (VED), vertical magnetic dipole (VMD), horizontal magnetic dipole (HMD), and the horizontal electric dipole (HED). The equations for the in-water fields from a subsurface time-harmonic VED in a deep ocean (only the surface interface) will be derived to demonstrate the analysis method. References will then be cited where solutions to the other source types can be found.

The geometry for a submerged vertical dipole and observation point is drawn in Fig. 8.10. As before, the air is nonconducting with free-space permittivity and permeability, while the ocean has its standard intrinsic parameters. The dipole is located on the z-axis at a depth h, with the observation point also located in the seawater at (ρ, φ, z). Although this problem can be solved in rectangular coordinates, cylindrical coordinates are employed due to the electromagnetic fields' circular symmetry about the source.

A slightly different form of the vector potential is typically used when solving the problem of dipoles in a layered medium. The alternate form is called the *Hertz* vector potential. For a VED, the electric Hertz vector potential, denoted by $\boldsymbol{\Pi}$, has only a z component and is given by

$$\Pi_z = \frac{p}{4\pi(\sigma + i\omega\epsilon)} \frac{e^{-\gamma r}}{r} , \tag{8.113}$$

where

$$r = \sqrt{\rho^2 + (z-h)^2} .$$

The general relationships between the magnetic vector potential, the electric *Hertz* vector potential, and the electric and magnetic fields are

$$\boldsymbol{A} = \mu(\sigma + i\omega\epsilon)\boldsymbol{\Pi} , \tag{8.114}$$

$$\boldsymbol{E} = -\gamma^2 \boldsymbol{\Pi} + \nabla(\nabla \cdot \boldsymbol{\Pi}) , \tag{8.115}$$

$$\boldsymbol{H} = -(\sigma + i\omega\epsilon)\nabla \times \boldsymbol{\Pi} . \tag{8.116}$$

Placing (8.113) in (8.115) and (8.116), and reducing them in cylindrical coordinates, gives

$$E_\rho = \frac{\partial^2 \Pi_z}{\partial \rho \partial z} , \tag{8.117}$$

$$E_z = -\gamma^2 \Pi_z + \frac{\partial^2 \Pi_z}{\partial z^2} , \tag{8.118}$$

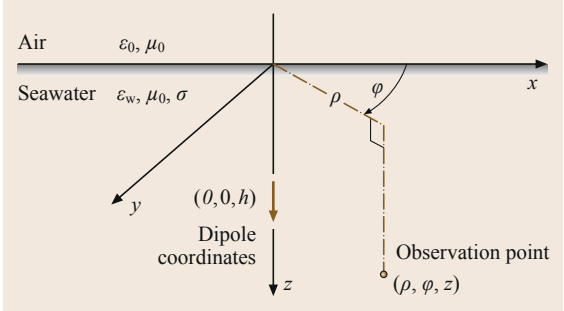


Fig. 8.10 Coordinate system for a submerged vertical dipole and submerged observation point in an electrically deep ocean

$$H_\varphi = -(\sigma + \mathrm{i}\omega\epsilon)\frac{\partial \Pi_z}{\partial \rho}\,, \tag{8.119}$$

and all other electric and magnetic field components are zero. Equations (8.113) through (8.119) apply to both the seawater and air using the appropriate constitutive parameters for each media.

The Hertz electric vector must be put in a form suitable for matching the boundary conditions at the surface interface. Using the identity

$$\frac{\mathrm{e}^{-\gamma r}}{4\pi r} = \frac{1}{4\pi}\int_0^\infty \frac{\lambda}{u}\mathrm{e}^{-u|z|}J_0(\lambda\rho)\,\mathrm{d}\lambda\,, \tag{8.120}$$

where J_0 is the Bessel function of order 0, and

$$u = (\lambda^2 + \gamma^2)^{\frac{1}{2}}\,, \tag{8.121}$$

the electric Hertz vector for a VED located at a depth h on the z-axis, (8.113), can be written as

$$\Pi_z^{(\mathrm{w})} = \frac{p}{4\pi(\sigma + \mathrm{i}\omega\varepsilon_\mathrm{w})}\int_0^\infty \frac{\lambda}{u_\mathrm{w}}\mathrm{e}^{-u_\mathrm{w}|z-h|}J_0(\lambda\rho)\,\mathrm{d}\lambda\,, \tag{8.122}$$

and the subscript and superscript, w, indicates that the constitutive parameters for seawater are to be applied. The integral representation in (8.122) can be replaced with

$$\int_0^\infty \frac{\lambda}{u_\mathrm{w}}\mathrm{e}^{-u_\mathrm{w}(z-h)}J_0(\lambda\rho)\,\mathrm{d}\lambda \quad \text{when } (z-h)\geq 0\,, \tag{8.123}$$

$$\int_0^\infty \frac{\lambda}{u_\mathrm{w}}\mathrm{e}^{u_\mathrm{w}(z-h)}J_0(\lambda\rho)\,\mathrm{d}\lambda \quad \text{when } (z-h)\leq 0\,. \tag{8.124}$$

Equations (8.120) and (8.122) are called *Hankel transforms* or *Sommerfeld integrals*.

Equation (8.122) must be modified to account for reflections from the surface interface, and an additional equation is needed to describe the transmitted fields in the air. The system of Hertz equations sufficient to solve for the electromagnetic fields in both the air $\Pi_z^{(\mathrm{a})}$ and seawater $\Pi_z^{(\mathrm{w})}$ layers are

$$\Pi_z^{(\mathrm{a})} = P_\mathrm{w}\int_0^\infty f_1(\lambda)\mathrm{e}^{u_0 z}J_0(\lambda\rho)\,\mathrm{d}\lambda \quad \text{for } z\leq 0\,, \tag{8.125}$$

$$\Pi_z^{(\mathrm{w})} = P_\mathrm{w}\int_0^\infty \frac{\lambda}{u_\mathrm{w}}\mathrm{e}^{-u_\mathrm{w}|z-h|}J_0(\lambda\rho)\,\mathrm{d}\lambda + P_\mathrm{w}\int_0^\infty f_2(\lambda)\mathrm{e}^{-u_\mathrm{w}z}J_0(\lambda\rho)\,\mathrm{d}\lambda \quad \text{for } z\geq 0\,, \tag{8.126}$$

where

$$P_\mathrm{w} = \frac{p}{4\pi\sigma}\,, \tag{8.127}$$

$$u_\mathrm{w} = (\lambda^2 + \gamma^2)^{\frac{1}{2}}\,, \tag{8.128}$$

$$u_0 = (\lambda^2 + \gamma_0^2)^{\frac{1}{2}}\,, \tag{8.129}$$

and $\gamma_0^2 = -\omega^2\mu_0\varepsilon_0$, while $\gamma^2 = \mathrm{i}\omega\mu_0\sigma$. The terms f_1 and f_2 are determined by the boundary conditions at the ocean surface.

The boundary conditions require that the tangential components of the electric and magnetic fields must be continuous across the sea surface. For this problem, E_ρ (8.117) and H_φ (8.119) are the components tangential to the interface. The continuation of these tangential field components can be expressed in terms of their electric Hertz vectors as

$$\left.\begin{aligned}\frac{\partial \Pi_z^{(\mathrm{a})}}{\partial z} &= \frac{\partial \Pi_z^{(\mathrm{w})}}{\partial z}\\ \gamma_0^2\Pi_z^{(\mathrm{a})} &= \gamma_\mathrm{w}^2\Pi_z^{(\mathrm{w})}\end{aligned}\right] \quad \text{at } z = 0\,. \tag{8.130}$$

Applying these boundary conditions results in a system of equations for f_1 and f_2 given by

$$\frac{u_0}{u_\mathrm{w}}f_1 + f_2 = \frac{\lambda}{u_\mathrm{w}}\mathrm{e}^{-u_\mathrm{w}h}\,, \tag{8.131}$$

$$\frac{\gamma_0^2}{\gamma_\mathrm{w}^2}f_1 - f_2 = \frac{\lambda}{u_\mathrm{w}}\mathrm{e}^{-u_\mathrm{w}h}\,. \tag{8.132}$$

Solving for f_1 and f_2 produces

$$f_1(\lambda) = \frac{\frac{2\lambda}{u_\mathrm{w}}\mathrm{e}^{-u_\mathrm{w}h}}{\frac{u_0}{u_\mathrm{w}} + \frac{\gamma_0^2}{\gamma_\mathrm{w}^2}}\,, \tag{8.133}$$

$$f_2(\lambda) = \frac{\lambda}{u_\mathrm{w}}\mathrm{e}^{-u_\mathrm{w}h}\left(-\frac{\frac{2u_0}{u_\mathrm{w}}}{\frac{u_0}{u_\mathrm{w}} + \frac{\gamma_0^2}{\gamma_\mathrm{w}^2}}\right). \tag{8.134}$$

Since $\gamma_0^2 \ll \gamma_\mathrm{w}^2$ for all frequencies up through the ELF band, γ_0 can be set to zero in (8.125), (8.133), and (8.134). Under these conditions, f_1 and f_2 reduce to

$$f_1(\lambda) = 2\mathrm{e}^{-u_\mathrm{w}h}\,, \tag{8.135}$$

$$f_2(\lambda) = -\frac{\lambda}{u_\mathrm{w}}\mathrm{e}^{-u_\mathrm{w}h}\,. \tag{8.136}$$

This *no-propagation* approximation for air is sometimes called the *quasi-static condition*, and should not be confused with the quasi-DC case discussed previously.

With the solutions to f_1 and f_2 in hand, (8.125) and (8.126) can be placed in (8.120) through (8.122) to compute the electric and magnetic fields above and below the sea surface. For some special cases, the field equations can be solved in the analytical form; but in general, the Sommerfeld integrals must be evaluated numerically. Unfortunately, the integrands are oscillatory and slowly damped, increasing the difficulty in their numerical integration. Special techniques have been developed to efficiently evaluate Sommerfeld-type integrals, most of which have been reviewed in [8.4].

Derivation of the more complicated equations for a VED in a shallow ocean, which includes the interface at the seafloor, follows the same procedure as shown above for the deep water analysis. In the shallow water case, a third Hertz vector is needed to describe propagation within the sea bottom, and another term must be included in (8.126) to account for reflections off the seafloor. Satisfying the boundary conditions at both the seafloor and sea surface produce four equations with four unknowns, this time f_1 through f_4. The Sommerfeld integral formulations for a VED in a three layer media can be found in [8.5], and applied to the shallow-ocean problem using the appropriate constitutive parameters.

The complementarity problem of a VMD in a deep or shallow ocean is solved in an analogous manner using a magnetic Hertz vector in place of the electric vector. The in-air fields from a time-harmonic VMD submerged in shallow water have been reported in [8.6], while the in-water fields from an airborne VMD are found in [8.7]. Because of the presence of the ocean surface and/or bottom interface, the field solutions for an HED require both horizontal and vertical electric Hertz vectors to satisfy the boundary conditions [8.8]. Although this complicates the algebra by doubling the number of boundary equations and unknowns, the Sommerfeld formulation for the in-water electric and magnetic fields from a submerged HED in shallow water have been derived, with the solution given in [8.9]. Similarly, computing the fields for an HMD in the ocean requires both the horizontal and vertical magnetic Hertz vectors. The equations for the in-air electric and magnetic fields from submerged VED, HED, and HMD sources have been complied in [8.10], while the in-water fields from these three sources located above the ocean are summarized in [8.11]. The referenced shallow water equations can be reduced to the deep-ocean case by setting the bottom conductivity equal to that of the water.

Although the electromagnetic fields from dipole sources submerged in a deep ocean have been formulated in the generalized form of Sommerfeld integrals, it is very convenient to have algebraic equations readily accessible for obtaining quick estimates; even if they are only valid for special cases. The algebraic in-water and in-air electric and magnetic field equations for VED, HED, VMD, and HMD sources submerged in a deep ocean will be presented in tabular form for a few special cases. These tabularized equations have been found to be useful as a fast and easy reference.

The coordinate system and geometry for the tabulated equations are again shown in Fig. 8.10. All sources will be located at a depth h, with the VED and VMD aligned in the positive z direction, while the HED and HMD are oriented along the positive x-axis. The planner sea surface interface is at $z = 0$. The magnetic and electric dipole source strengths will be designated as m and p, respectively.

The magnetic and electric fields for static magnetic and electric dipoles in an unbounded ocean were given previously in spherical coordinates. At times, it is more convenient to apply the formulas in the Cartesian system. For easy reference, the Cartesian field equations for a static VMD and HMD source in an unbounded ocean (no sea surface) are listed in Table 8.1. Although static magnetic sources produce no electric fields, static VED and HED dipoles create both magnetic and electric fields as given in Table 8.2 for an unbounded ocean.

If the sea surface interface is now reintroduced at $z = 0$ (Fig. 8.10), the magnetic field equations for the static VMD and HMD sources in Table 8.1 remain unchanged. These equations are valid whether the observation point is in water, air, or sea bottom. The reason is that all three media are nonmagnetic with a permeability equal to that of free-space.

The jump in conductivity at the ocean's surface does, however, affect both the electric and magnetic fields from submerged electric dipoles. Because the air is nonconducting, the electric current in the conducting seawater will stay confined within it and will not flow across the surface interface. The sea surface is equivalent to an ideal reflector of static electric current. In addition, the discontinuity in the electric current across the surface interface produces additional components of magnetic fields.

Mathematically, the effects of the sea surface on the in-water static electric fields is equivalent to placing images of the VED and HED directly above the water at a distance equal to the depth of the sources. The electric fields produced by these images are listed in Table 8.3 [8.12]. To get the total in-water static electric fields for the deep ocean case, the VED and

Table 8.1 Static VMD and HMD sources located at (0, 0, h) in an unbounded ocean

	E_x	E_y	E_z	H_x	H_y	H_z
VMD	0	0	0	$\frac{3m_z x(z-h)}{4\pi r^5}$	$\frac{3m_z y(z-h)}{4\pi r^5}$	$\frac{m_z(2(z-h)^2-x^2-y^2)}{4\pi r^5}$
HMD	0	0	0	$\frac{m_x(2x^2-y^2-(z-h)^2)}{4\pi r^5}$	$\frac{3m_x xy}{4\pi r^5}$	$\frac{3m_x x(z-h)}{4\pi r^5}$

where $r=\sqrt{x^2+y^2+(z-h)^2}$

Table 8.2 Static VED and HED sources located at (0, 0, h) in an unbounded ocean

	E_x	E_y	E_z	H_x	H_y	H_z
VED	$\frac{3p_z x(z-h)}{4\pi\sigma r^5}$	$\frac{3p_z y(z-h)}{4\pi\sigma r^5}$	$\frac{p_z(2(z-h)^2-x^2-y^2)}{4\pi\sigma r^5}$	$\frac{-p_z y}{4\pi r^3}$	$\frac{p_z x}{4\pi r^3}$	0
HED	$\frac{p_x(2x^2-y^2-(z-h)^2)}{4\pi\sigma r^5}$	$\frac{3p_x xy}{4\pi\sigma r^5}$	$\frac{3p_x x(z-h)}{4\pi\sigma r^5}$	0	$\frac{-p_x(z-h)}{4\pi r^3}$	$\frac{p_x y}{4\pi r^3}$

where $r=\sqrt{x^2+y^2+(z-h)^2}$ and σ is the seawater conductivity

Table 8.3 In-water electric fields produced by the above-water images of submerged static VED and HED sources (after [8.12])

	E_x	E_y	E_z
VED	$\frac{-3p_z x(z+h)}{4\pi\sigma r_i^5}$	$\frac{-3p_z y(z+h)}{4\pi\sigma r_i^5}$	$\frac{-p_z(2(z+h)^2-x^2-y^2)}{4\pi\sigma r_i^5}$
HED	$\frac{p_x(2x^2-y^2-(z+h)^2)}{4\pi\sigma r_i^5}$	$\frac{3p_x xy}{4\pi\sigma r_i^5}$	$\frac{3p_x x(z+h)}{4\pi\sigma r_i^5}$

where $r_i=\sqrt{x^2+y^2+(z+h)^2}$ and σ is the seawater conductivity

Table 8.4 In-water magnetic fields produced by the above-water images of submerged static VED and HED sources (after [8.12])

	H_x	H_y	H_z
VED	$\frac{p_z y}{4\pi r_i^3}$	$\frac{-p_z x}{4\pi r_i^3}$	0
HED	$\frac{-p_x xy}{4\pi\rho^4}\left[2-\frac{(z+h)(2(z+h)^2+3\rho^2)}{r_i^3}\right]$	$\frac{p_x}{4\pi}\left\{\frac{x^2}{\rho^4}\left[2-\frac{(z+h)(2(z+h)^2+3\rho^2)}{r_i^3}\right]-\frac{1}{r_i^2+(z+h)r_i}\right\}$	0

where $r_i=\sqrt{x^2+y^2+(z+h)^2}$ and $\rho=\sqrt{x^2+y^2}$

Table 8.5 In-air electric fields from static VED and HED sources submerged at a depth h in a deep ocean (after [8.12])

	E_x	E_y	E_z
VED	$\frac{3p_z x(z-h)}{2\pi\sigma r^5}$	$\frac{3p_z y(z-h)}{2\pi\sigma r^5}$	$\frac{p_z(2(z-h)^2-x^2-y^2)}{2\pi\sigma r^5}$
HED	$\frac{p_x(2x^2-y^2-(z-h)^2)}{2\pi\sigma r^5}$	$\frac{3p_x xy}{2\pi\sigma r^5}$	$\frac{3p_x x(z-h)}{2\pi\sigma r^5}$

where $r=\sqrt{x^2+y^2+(z-h)^2}$ and σ is the seawater conductivity. Note: A negative value for z is used for in-air computations

HED equations in Table 8.3 must be added to their corresponding components from Table 8.2. Likewise, the magnetic field formulations produced by the jump in the surface current density has been computed in [8.12], and are listed in Table 8.4. Adding the equations in Table 8.4 to their corresponding magnetic

Table 8.6 In-air magnetic fields from static VED and HED sources submerged at a depth h in a deep ocean (after [8.12])

	H_x	H_y	H_z
VED	0	0	0
HED	$\frac{-p_z xy}{4\pi\rho^4}\left[2-\frac{(h-z)(2(h-z)^2+3\rho^2)}{r^3}\right]$	$\frac{-p_z}{4\pi}\left\{\frac{(z-h)}{r^3}+\frac{x^2}{\rho^4}\left[2-\frac{(h-z)(2(h-z)^2+3\rho^2)}{r^3}\right]-\frac{1}{r^2+(h-z)r}\right\}$	$\frac{p_x y}{4\pi r^3}$

where $r=\sqrt{x^2+y^2+(z-h)^2}$ Note: A negative value for z is used for in-air computations

Table 8.7 Electric field subsurface-to-subsurface propagation formulas for the quasi-static range, when $|\gamma_0\rho| \ll 1 \ll |\gamma\rho|$ and $\rho \gg h$ and z

	E_x	E_y	E_z
VED	$-\frac{p_z x}{2\pi\sigma\rho^3}\frac{\gamma_0^2}{\gamma}e^{-\gamma(z+h)}$	$\frac{p_z y}{2\pi\sigma\rho^3}\frac{\gamma_0^2}{\gamma}e^{-\gamma(z+h)}$	$\frac{p_z}{2\pi\sigma\rho^3}\frac{\gamma_0^2}{\gamma^2}e^{-\gamma(z+h)}$
VMD	$\frac{3m_z y}{2\pi\sigma\rho^5}e^{-\gamma(z+h)}$	$-\frac{3m_z x}{2\pi\sigma\rho^5}e^{-\gamma(z+h)}$	0
HED	$\frac{p_x}{2\pi\sigma\rho^5}(x^2-2y^2)e^{-\gamma(z+h)}$	$\frac{3p_x xy}{2\pi\sigma\rho^5}e^{-\gamma(z+h)}$	$\frac{p_x x}{2\pi\sigma\rho^3}\frac{\gamma_0^2}{\gamma^2}e^{-\gamma(z+h)}$
HMD	$-\frac{3m_x xy\gamma}{2\pi\sigma\rho^5}e^{-\gamma(z+h)}$	$-\frac{m_x\gamma}{2\pi\sigma\rho^5}(y^2-2x^2)e^{-\gamma(z+h)}$	$\frac{m_x y\gamma_0^2}{2\pi\sigma\rho^3}e^{-\gamma(z+h)}$

where $\rho=\sqrt{x^2+y^2}$ and σ is the seawater conductivity

Table 8.8 Magnetic field subsurface-to-subsurface propagation formulas for the quasi-static range, when $|\gamma_0\rho| \ll 1 \ll |\gamma\rho|$ and $\rho \gg h$ and z

	H_x	H_y	H_z
VED	$-\frac{p_z y}{2\pi\rho^3}\frac{\gamma_0^2}{\gamma^2}e^{-\gamma(z+h)}$	$-\frac{p_z x}{2\pi\rho^3}\frac{\gamma_0^2}{\gamma^2}e^{-\gamma(z+h)}$	0
VMD	$-\frac{3m_z x}{2\pi\gamma\rho^5}e^{-\gamma(z+h)}$	$-\frac{3m_z y}{2\pi\gamma\rho^5}e^{-\gamma(z+h)}$	$-\frac{9m_z}{2\pi\gamma^2\rho^5}e^{-\gamma(z+h)}$
HED	$\frac{3p_x xy}{2\pi\gamma\rho^5}e^{-\gamma(z+h)}$	$-\frac{p_x}{2\pi\gamma\rho^5}(x^2-2y^2)e^{-\gamma(z+h)}$	$\frac{3p_x y}{2\pi\gamma^2\rho^5}e^{-\gamma(z+h)}$
HMD	$-\frac{m_x}{2\pi\rho^5}(y^2-2x^2)e^{-\gamma(z+h)}$	$\frac{3m_x xy}{2\pi\rho^5}e^{-\gamma(z+h)}$	$\frac{3m_x x}{2\pi\gamma\rho^5}e^{-\gamma(z+h)}$

where $\rho=\sqrt{x^2+y^2}$ and σ is the seawater conductivity.

components in Table 8.2 yields the total in-water magnetic field for VED and HED sources submerged in a deep ocean.

For completeness, the equations for the total in-air electric fields from static VED and HED sources submerged in a deep ocean are listed in Table 8.5. The total in-air magnetic fields for the same case are found in Table 8.6. Tables 8.5 and 8.6 are the total in-air equations and are not added to any others. It should be noted that a negative value for z is to be used when computing the fields above the ocean surface. The deep ocean equations in Tables 8.3–8.6 were reduced from those found in [8.12] for a shallow sea.

The equations in Tables 8.1 through 8.6 apply for the special case where the sources are static or quasi-DC. They are valid for frequencies that correspond to seawater skin-depths that are large compared to the distance between the source and observation point, and their depths below the surface. At the other extreme, analytic equations have been formulated for the special case where the separation between the source and observation point is much greater than a seawater skin-depth and their depths. These equations were derived by taking the asymptotic limits for the Sommerfeld formulations.

The condition that the submerged source is many seawater skin-depths from the in-water observation point is equivalent to $1 \ll |\gamma\rho|$. In addition, if $\rho \gg h$ and z and the quasi-static range applies, $|\gamma_0\rho| \ll 1$, the subsurface-to-subsurface electric field equations in

Table 8.7 can be used for the four dipole types. The corresponding subsurface-to-subsurface magnetic field equations are listed in Table 8.8. These Cartesian formulations are re-arrangements of the cylindrical forms found in Table 3–16 of [8.13].

The propagation path described by the formulations in Tables 8.7 and 8.8 is quite unique. Since $1 \ll |\gamma\rho|$, the fields that travel directly from the source to the observation point or reflected off the sea-surface are significantly attenuated. Instead, the dominated propagation path is vertical to the surface, along the water–air interface, and then directly down to the observation point. The fields are attenuated in this *up-over-and-down* path only over the vertical distances between the source and surface, and from the surface to the observation point.

There are many other special cases in which analytic propagation equations can be used instead of numerically evaluating the Sommerfeld formulations. Many of the asymptotic conditions has been tabulated in [8.13]. In addition, the region between the quasi-DC and asymptotic conditions have been subdivided into ranges for which analytic formulations are been produced using *modified image theory* [8.14].

8.8 Electromagnetic Propagation in the Ocean at Optical Wavelengths

The electromagnetic constitutive parameters of seawater are not constant at microwave frequencies and higher. Indeed, if the nominal seawater ELF constitutive parameters are used to determine the electromagnetic attenuation at an optical frequency of visible blue–green light (6×10^{14} Hz), a value of 731 dB/m is computed. And yet, experimental measurements of blue–green light attenuation in clear seawater shows only 0.15 dB/m [8.15]. Obviously, the interaction of electromagnetic waves with the ocean at optical wavelengths is quite different than at low frequencies.

Although the physics of light transmission through seawater is quite involved, the discussion here will be at the engineering level, and confined to the *apparent optical properties* (AOP) of the ocean. The AOP for seawater is applicable to many problems in ocean engineering and, with the appropriate equipment, they can be measured in situ. Here, the in air optical wavelengths of interest cover the visible range from about 400 nm for violet colors to 700 nm for red.

The primary electromagnetic parameter of interest at optical wavelengths is the *specular irradiance*. The specular irradiance, designated by the term $E_{\lambda 0}$, is the band-limited power per unit area, with units of $\mathrm{W\,m^{-2}\,nm^{-1}}$. This is a measure of the scalar power density of an electromagnetic wave. The scalar power density or specular irradiance is proportional to the square of the electric or magnetic field presented in the previous sections of this chapter.

An important ocean engineering AOP is the attenuation coefficient for the specular irradiance as the light propagates through the sea. The equation describing the attenuation of optical irradiance from point r_1 to r is given by

$$E_{\lambda 0}(r) = E_{\lambda 0}(r_1)\mathrm{e}^{-\mathrm{K}_\lambda(\mathrm{r}-\mathrm{r}_1)}\,, \qquad (8.137)$$

where K_λ is the diffuse attenuation coefficient, which is a function of wavelength. Equation (8.137) is similar to that of a plane wave given in Sect. 8.3, and in fact, does not account for any spatial spreading of the light energy. The major difference is that (8.134) describes the attenuation of power instead of field amplitude, and the attenuation coefficient K_λ is quite different from the ELF values given previously for the field attenuation constant α.

The optical diffuse attenuation coefficient is a function of wavelength, and is comprised of two major components that contribute to its overall value: absorption and scattering. Although there is absorption of light due to the intrinsic or inherent properties of clear seawater, other optical absorption mechanisms found in the ocean include organic materials such as chlorophylls in phytoplankton and dissolved organic compounds called *yellow substance* or *Gelbstoffe*. In addition, the small wavelength of visible light will cause it to scatter off organic and inorganic particulates suspended in the water, adding to its attenuation. As would be expected, the optical attenuation coefficient varies with

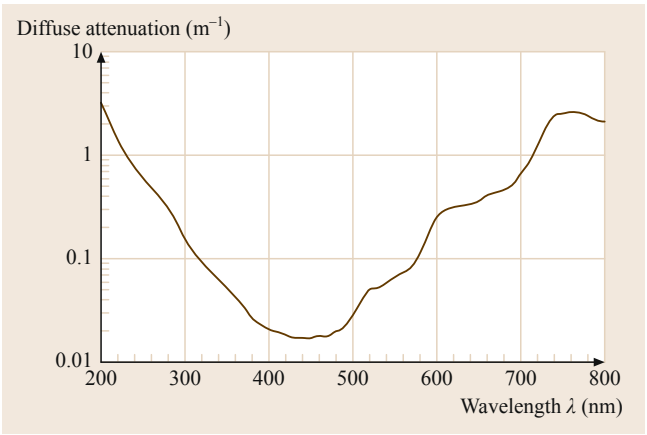


Fig. 8.11 Diffuse attenuation coefficient for clear seawater (after [8.16])

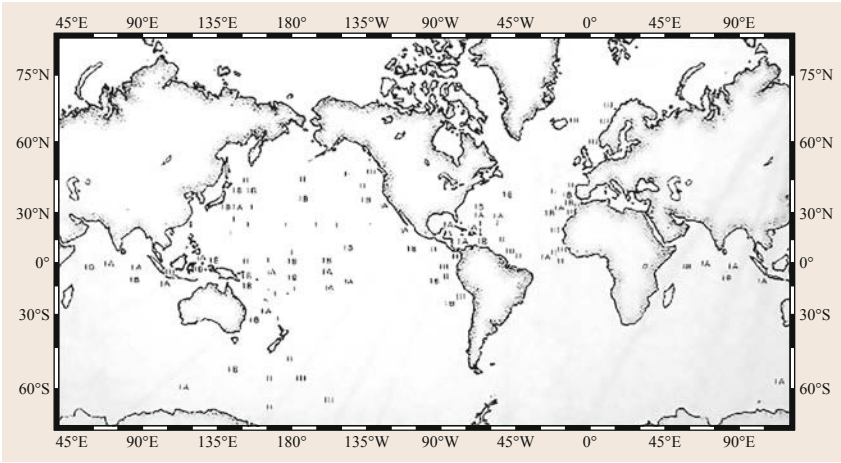


Fig. 8.12 The global distribution of Jerlov water types (after [8.1])

depth and between different regions of the world's oceans.

Laboratory measurements of the attenuation, absorption, and scattering coefficients of pure or clear seawater have been reviewed and tabulated [8.16]. The diffuse attenuation coefficients in pure seawater are plotted in Fig. 8.11, and includes water absorption and a small amount of scattering from sea salts. The estimated accuracy of this data is reported to be within +25 and −5% for wavelengths between 300 and 480 nm, and 10 and 5% from 480 to 800 nm. Below 300 nm the data is considered to be no better than an *educated estimate* [8.16]. The region of minimum attenuation between 400 and 500 nm is called the blue–green window, and has been the subject of many investigations to engineer underwater communication and detection systems [8.15, 18].

The seminal work of Jerlov produced a classification method for characterizing the optical properties of the world's oceans. Based on the measured irradiance transmissivity in the upper 10 m of the water, the oceans are categorized as Type I to III, with a subsequent subdivision of Type I into IA and IB. Coastal waters are divided into types ranging from 1 to 9. These water clarity classifications account not only for the absorption and scattering of pure seawater, but also included the effects of the suspended organic and inorganic particles as described above. A map of the global distribution of Jerlov water types is shown in Fig. 8.12 [8.1, p. 584].

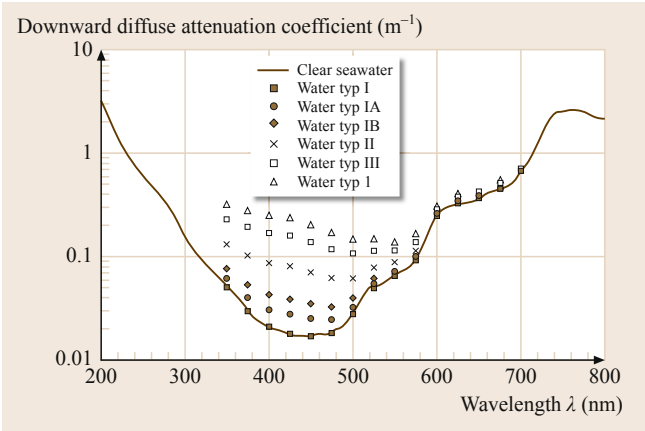


Fig. 8.13 Diffuse attenuation coefficient for Jerlov water types (after [8.17])

The downward looking diffuse attenuation coefficients for some of the Jerlov water types have been measured and tabulated [8.17]. The attenuation coefficients for ocean water Types I though III and coastal Type 1 are plotted in Fig. 8.13. As expected, the attenuation coefficient increases as water clarity decrease, with the minimum in the optical window moving toward longer wavelengths. The higher attenuation of light in natural seawaters, especially for coastal regions, limits the useful ranges of communication and detection systems that are based on blue–green lasers.

Although the dielectric constant of seawater is approximately 80 up through the ELF band, its value at an optical wavelength of 700 nm is only 1.79 at 20 °C, and increases to only 1.83 for 400 nm at 4 °C at the other extreme [8.1, pp. 532–539]. A nominal value of 1.8 gives a propagation constant of $\beta \approx 1.4 \times 10^7$ rad/m at 450 nm (20 °C), while the equivalent field attenuation constant for pure seawater is estimated from K_λ to be $\alpha \approx 0.008$ Np/m at this wavelength. Therefore, seawater can be considered a dielectric at optical wavelengths, with an average index of refraction of $n \approx 4/3$ ($n = \sqrt{\varepsilon_w}$). The index of refraction is the ratio of the speed of light in air, 3×10^8 m/s, to that in seawater, 2.24×10^8 m/s.

Since seawater can be treated as a dielectric at optical wavelengths, the expressions for the field reflection and transmission coefficients given previously for fresh

water can also be used for the ocean. The only difference is that they must be squared before applying them to irradiance, which is power. The power reflection coefficients for perpendicular $R_\perp$ and parallel polarization $R_\parallel$ can be obtained from (8.59) and (8.71), giving

$$R_\perp = \Gamma_\perp^2 = \frac{\tan^2(\theta_\mathrm{i} - \theta_\mathrm{t})}{\tan^2(\theta_\mathrm{i} + \theta_\mathrm{t})} \,, \tag{8.138}$$

$$R_\parallel = \Gamma_\parallel^2 = \frac{\sin^2(\theta_\mathrm{i} - \theta_\mathrm{t})}{\sin^2(\theta_\mathrm{i} + \theta_\mathrm{t})} \,, \tag{8.139}$$

and

$$\theta_\mathrm{r} = \theta_\mathrm{i} \,, \tag{8.140a}$$

$$\sin\theta_\mathrm{i} = n \sin\theta_\mathrm{t} \,. \tag{8.140b}$$

where all terms have been defined previously. The Brewster angle can be computed directly from $\tan\theta_\mathrm{B} = n$. These equations are valid for a smooth surface. A discussion of optical scattering from a rough ocean surface produced by wind waves can be found in [8.1, p. 528].

The maximum refraction angle of light into the ocean is limited. If the incidence angle of light entering the flat surface of the sea approaches the grazing angle ($\theta_\mathrm{i} = 90°$), then from (8.140b) the transmission angle of the refracted light is computed to be approximately 48.6°. This means that all images from the entire hemisphere above water is funneled into an upward looking cone whose sides are 48.6° from the vertical. Any image that is seen at an angle greater than this could only be coming from a submerged object.

References

8.1 J.R. Apel: *Principles of Ocean Physics* (Academic Press, San Diego 1987)

8.2 C.A. Balanis: *Advanced Engineering Electromagnetics*, 2nd edn. (Wiley, Hoboken 2012)

8.3 C.A. Balanis: *Antenna Theory*, 3rd edn. (Wiley, Hoboken 2005) pp. 133–142

8.4 J. Mosig: The weighted averages algorithm revisited, IEEE Trans. Antennas Propag. **60**(4), 2011–2018 (2012)

8.5 J.R. Wait: Electromagnetic Fields of Sources in Lossy Media. In: *Antenna Theory*, ed. by R.E. Collin, F.J. Zucker (McGraw-Hill, New York 1969) pp. 476–478

8.6 A.C. Fraser-Smith, D.M. Bubenik: ULF/ELF electromagnetic fields generated above a sea of finite depth by a submerged vertically-directed harmonic magnetic dipole, Radio Sci. **14**, 59–74 (1979)

8.7 D.M. Bubenik, A.C. Fraser-Smith: ULF/ELF electromagnetic fields generated in a sea of finite depth by a submerged vertically-directed harmonic magnetic dipole, Radio Sci. **13**, 1011–1020 (1978)

8.8 J.R. Wait: *Electromagnetic Waves in Stratified Media* (IEEE Press, Piscataway 1996) pp. 143–146

8.9 C. Liu, L.G. Zheng, Y.P. Li: Study of ELF electromagnetic fields from a submerged horizontal electric dipole positioned in a sea of finite depth, IEEE 3rd Int. Symp. Microw. Antenna Propag. EMC Technol. Wirel. Commun. (2009) pp. 152–157

8.10 A.C. Fraser-Smith, D.M. Bubenik: *Compendium of the ULF/ELF Electromagnetic Fields Generated Above a Sea of Finite Depth by Submerged Harmonic Dipoles*, Tech. Rep., Vol. E715-1 (Stanford Univ., Stanford 1980)

8.11 A.C. Fraser-Smith, D.M. Bubenik: *ULF/ELF/VLF Electromagnetic Fields Generated in a Sea of Finite Depth by Elevated Dipole Sources*, Tech. Rep., Vol. E715-2 (Stanford Univ., Stanford 1984)

8.12 D.L. Jones, C.P. Burke: The DC field components of horizontal and vertical electric dipole sources immersed in three-layer stratified media, Ann. Geophys. **15**, 503–510 (1997)

8.13 M.B. Kraichman: Electromagnetic propagation in conducting media. In: *Electromagnetics Problem Solver*, ed. by M. Fogiel (REA, Piscataway 2000) pp. 3–24, Section II

8.14 P.R. Bannister, R.L. Dube: Simple expressions for horizontal electric dipole quasi-static range subsurface-to-subsurface and subsurface-to-air propagation, Radio Sci. **13**(3), 501–507 (1978)

8.15 W.C. Cox: A 1 Mbps Underwater Communication System Using a 405 nm Laser Diode and Photomultiplier Tube, Master Thesis (North Carolina State Univ., Raleigh 2007)

8.16 R.C. Smith, K.S. Baker: Optical properties of the clearest natural waters (200–800 nm), Appl. Opt. **20**(2), 177–184 (1971)

8.17 R.W. Austin, T.J. Petzold: Spectral dependence of the diffuse attenuation coefficient of light in ocean waters, Opt. Engr. **25**, 471 (1986)

8.18 C.J. Cassidy: Airborne Laser Mine Detection System, Master Thesis (Naval Postgraduate School, Monterey 1995)

9. Digital Signal Processing

Nikolaos I. Xiros

In this chapter the concept of discrete-time signals and sampled-data systems implemented on digital hardware versus those of continuous-time signals driving analog systems and processes is introduced early on. The processes of signal sampling and analog signal reconstruction are then investigated, and the Nyquist sampling rate to avoid aliasing is explained by means of Fourier series analysis. Then the Z-transform is introduced as the tool of preference for analysis and synthesis of discrete-time linear, time-invariant systems defined by difference equations in the discrete time domain. A detailed account of the most important and practical continuous-time system mapping techniques to discrete-time ones is then presented. A brief account of digital filter structures and types is also given along with a presentation of the fast Fourier transform algorithm for the calculation of the discrete Fourier transform of discrete-time signals with finite duration admitting periodic expansion. The notions of waveform statistics as encountered in random signals and stochastic processes are then given in order to conclude with the effect on a random signal's statistics due to its propagation through a linear, time-invariant system. Finally, concepts of optimal signal estimation and the Wiener filter are presented, leading to matched filter and zero-forcing equalizer parametric designs in the discrete-time domain.

9.1 **Discrete-Time Systems** 197
9.1.1 Discrete-Time Signals and Digital Systems 197
9.1.2 Signal Sampling 198
9.1.3 Analog Signal Reconstruction Using a Discrete-Time Signal 201
9.1.4 The Z-Transform 201
9.1.5 Discrete-Time LTI Systems 202
9.1.6 Continuous-Time System Mapping 203
9.2 **Digital Filters** 206
9.2.1 Important FIR Filter Structures 206
9.2.2 Important IIR Filter Structures 208
9.3 **The Fast Fourier Transform (FFT)** 211
9.3.1 Review of Integral Transforms 211
9.3.2 The Discrete Fourier Transform (DFT) 212
9.4 **Waveform Analysis** 216
9.4.1 Definitions for Waveforms and Random Signals 216
9.4.2 Signal Power and Power Spectral Density 218
9.4.3 Waveform Propagation Through a Linear, Time-Invariant System 219
9.5 **Optimal Signal Estimation** 220
9.5.1 System Identification 220
9.5.2 Discrete-Time Wiener–Hopf Equation over a Finite-Duration Window 221
9.5.3 Signal Estimation and the Wiener Filter 222
9.6 **Concluding Remarks** 225
References 225

9.1 Discrete-Time Systems

The progress in integrated circuit manufacturing (Very Large Scale Integration (VLSI)) combined with the need for high complexity and accuracy signal analysis and processing algorithms without development cost or time increase are the basic factors that contributed to the continuously increasing use of digital systems in industry and technology [9.1, 2].

9.1.1 Discrete-Time Signals and Digital Systems

A digital processing system is a system where the basic mathematical operations and functions, including signal transformations and data series analysis, are implemented as an embedded program onboard real-

time, computer hardware [9.1, 2]. Since the computer hardware employed is digital, the signals that can be processed or generated cannot be continuous time, i. e., defined for any instant in time. Instead, the computer processes and generates tables (i. e., *vectors*) of values for the input and output signal, respectively. Such tables hold the values of the corresponding signals at discrete time instants and are referred to as discrete-time signals. The time instants for which a discrete-time signal is known or defined are in most practical cases equally spaced; this scheme is known as uniform sampling and the spacing interval between two successive samples of the signal is mentioned as sampling interval or period.

The systems, e.g., real-time digital hardware, processing discrete-time signals are known as discrete-time systems. The main advantages in using computers in marine vehicle or process instrumentation, telecommunications, acoustics etc., instead of analog (continuous-time) systems are [9.1, 3–5]:

1. (Re)use of general purpose hardware for implementing a wide variety of algorithms instead of custom-built, proprietary designs that may become obsolete causing higher acquisition and development cost.
2. Changes in the algorithm can be easily implemented as changes in the software through reprogramming without need to modify the hardware.
3. Easy implementation of adaptive or time-varying algorithms due to programming.
4. Very small to zero sensitivity to environmental conditions, e.g., ambient temperature, in contrast to, e.g., analog electric circuitry.
5. Given, well-defined accuracy in computation determined by the computer's word length.
6. User-friendly human–machine interface (HMI).
7. Network connectivity.

However, there are some disadvantages in using discrete-time systems, the most important of which is the need for converter or adapter circuits or interface systems at their input and output in order to be able to interact with the vast majority of the real-world physical or engineering systems or processes, which are most often continuous-time. Such interfaces are the analog-to-digital (A/D) and digital-to-analog (D/A) converters.

9.1.2 Signal Sampling

Modeling of Ideal Sampler

Consider a continuous-time signal $x(t)$, which is driven through a sampling circuit (sampler) operating at a constant sampling rate (frequency), $f_s = 1/T_s$, as shown in Fig. 9.1.

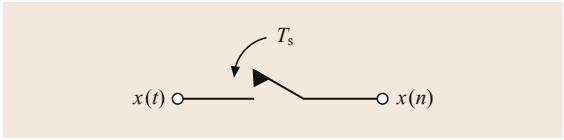


Fig. 9.1 Analog signal sampler

The sampler's output is a sequence of entries of the sampled signal, $x(nT_s)$, where n is an integer and T_s the sampling period. Sequence $x(n) = x(nT_s)$ is a discrete-time signal.

We will now represent the discrete-time signal as well as the sampling process in a continuous-time framework. Such a representation can be derived on the basis of Fig. 9.2.

A pulsed continuous-time signal $x_\Delta(t)$, which corresponds to sequence $x(nT_s)$, can be generated from the original continuous-time signal $x(t)$ (dashed line), if the latter is multiplied by the following *sampling pulse train* $P_\Delta(t)$

$$P_\Delta(t) = \sum_{n=-\infty}^{+\infty} p_\Delta(t - nT_s)\,, \tag{9.1}$$

where

$$p_\Delta(t) = \begin{cases} 1/\Delta, & t \in [0, \Delta] \\ 0 & t \notin [0, \Delta] \end{cases}, \quad 0 < \Delta \ll T_s\,. \tag{9.2}$$

In effect

$$x_\Delta(t) = x(t) \cdot P_\Delta(t)\,. \tag{9.3}$$

However, the following holds

$$\lim_{\Delta \to 0} p_\Delta(t) = \delta(t) \Rightarrow \lim_{\Delta \to 0} P_\Delta(t) = \sum_{n=-\infty}^{+\infty} \delta(t - nT_s)\,. \tag{9.4}$$

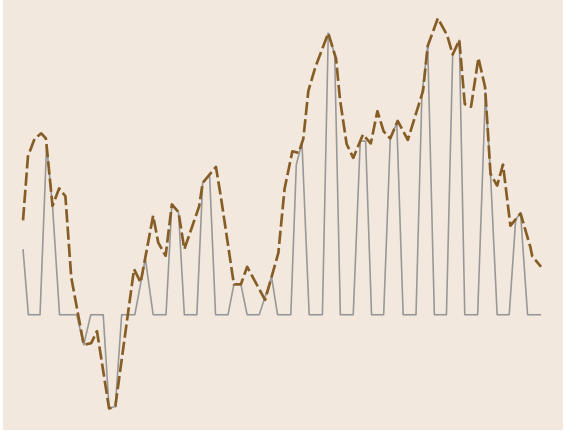


Fig. 9.2 Sampling of continuous-time waveform at uniform rate

In the above *ideal* case, the sampling pulse train is an impulse train or Dirac comb. In effect, (9.4) becomes

$$\begin{aligned} x_\delta(t) &= x(t)\cdot\sum_{n=-\infty}^{+\infty}\delta(t-nT_\mathrm{s}) \\ &= \sum_{n=-\infty}^{+\infty} x(nT_\mathrm{s})\delta(t-nT_\mathrm{s}) \\ &= \sum_{n=-\infty}^{+\infty} x(n)\delta(t-nT_\mathrm{s})\,. \end{aligned} \tag{9.5}$$

Above, the following fact for Dirac's impulse delta has been used: $\chi(t)\delta(t)=\chi(0)\delta(t)$, which implies the assumption that signal $\chi(t)$ assumes non-infinite values.

Hence, the ideal sampler operation consists of multiplicatively applying the input signal by a Dirac comb with T_s. Now we will determine the spectrum of signal $x_\delta(t)$ at the output of the sampler by employing the Fourier transform.

Fourier Series Expansion and Fourier Transform

Any periodic signal $x(t)$ with period T can be expanded to a complex Fourier series according to the following

$$x(t)=\sum_{n=0}^{+\infty} c_n \exp\left(in\frac{2\pi}{T}t\right). \tag{9.6}$$

Coefficients c_n of the series are calculated according to

$$c_n=\frac{1}{T}\int_0^T x(t)\exp\left(-in\frac{2\pi}{T}t\right)\mathrm{d}t\,. \tag{9.7}$$

The above can be extended to the case of aperiodic (non-periodic) signals assuming that period T is infinite. That is how the *Fourier transform* emerges

$$\begin{aligned} X(\omega) &= \mathcal{F}\{x(t)\} \\ &= \int_{-\infty}^{+\infty} x(t)\exp(-i\omega t)\,\mathrm{d}t, \quad \omega=\frac{2\pi}{T}\,. \end{aligned} \tag{9.8}$$

The inverse Fourier transform is defined by

$$x(t)=\frac{1}{2\pi}\int_{-\infty}^{+\infty} X(\omega)\exp(i\omega t)\,\mathrm{d}\omega\,. \tag{9.9}$$

To generalize formalism we refer to the Fourier transform of any signal, periodic or not. In the case of periodic signals, the Fourier transform is directly connected to coefficients c_n of its series expansion as follows

$$X(\omega)=2\pi\sum_{n=0}^{+\infty} c_n\delta\left(\omega-n\frac{2\pi}{T}\right). \tag{9.10}$$

The above can be derived directly by plugging it into (9.8) and generating (9.6). The Fourier transform, just like the Laplace transform, from which the former can be defined by setting $s=i\omega$, is a linear integral transform. Its properties are identical to those of the Laplace transform, and can be sought in the literature. An important property is that of multiplication-convolution duality

$$\begin{aligned} x(t) &= x_1(\xi)\otimes x_2(\xi) \\ &\triangleq \int_{-\infty}^{+\infty} x_1(\xi)\cdot x_2(t-\xi)\cdot\mathrm{d}\xi \\ &\Rightarrow X(\omega)=X_1(\omega)\cdot X_2(\omega) \\ x(t) &= x_1(t)\cdot x_2(t)\Rightarrow X(\omega) \\ &= X_1(\omega)\otimes X_2(\omega) \\ &\triangleq \int_{-\infty}^{+\infty} X_1(\xi)\cdot X_2(\omega-\xi)\cdot\mathrm{d}\xi\,. \end{aligned} \tag{9.11}$$

Here, $\otimes$ stands for the convolution operator. Note that the above can be directly applied to the Dirac comb describing the function of an ideal sampler as well as its periodicity.

Indeed, the Dirac comb in (9.4) is a periodic signal with period T_s. Its Fourier series is given below

$$\begin{aligned} c_n &= \frac{1}{T_\mathrm{s}}\int_0^{T_\mathrm{s}}\delta(t)\exp(-in\omega_\mathrm{s}t)\,\mathrm{d}t \\ &= \frac{\exp(-in\omega_\mathrm{s}0)}{T_\mathrm{s}}\int_0^{T_\mathrm{s}}\delta(t)\,\mathrm{d}t=\frac{1}{T_\mathrm{s}}, \quad \omega_\mathrm{s}=\frac{2\pi}{T_\mathrm{s}}\,. \end{aligned} \tag{9.12}$$

Hence, based on (9.10) the spectrum (Fourier transform) of the Dirac comb is given by

$$\begin{aligned} \mathcal{F}\left\{\sum_{n=-\infty}^{+\infty}\delta(t-nT_\mathrm{s})\right\} &= 2\pi\sum_{n=0}^{+\infty}\frac{1}{T_\mathrm{s}}\delta\left(\omega-n\frac{2\pi}{T_\mathrm{s}}\right) \\ &= \omega_\mathrm{s}\sum_{n=0}^{+\infty}\delta(\omega-n\omega_\mathrm{s})\,. \end{aligned} \tag{9.13}$$

Nyquist Sampling Rate – Aliasing

Combining the relationship above for the Dirac comb spectrum with (9.5), as well as the Fourier transform du-

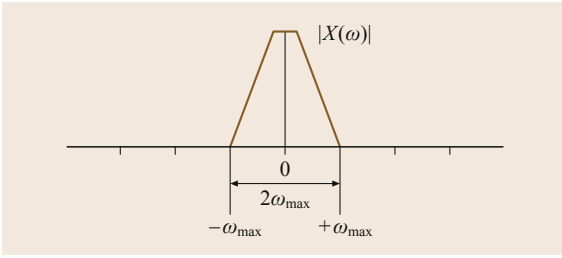


Fig. 9.3 Typical spectrum of real low-pass signal

ality between multiplication and convolution, one can determine the spectrum of signal $x_\delta(t)$ at the output of an ideal sampler

$$x_\delta(t) = x(t) \cdot \sum_{n=-\infty}^{+\infty} \delta(t - nT_s) \Rightarrow$$

$$\Rightarrow X_\delta(\omega) = X(\omega) \otimes \mathcal{F}\left\{\sum_{n=-\infty}^{+\infty} \delta(t - nT_s)\right\}$$

$$= \omega_s X(\omega) \otimes \sum_{n=0}^{+\infty} \delta(\omega - n\omega_s)$$

$$\Rightarrow X_\delta(\omega) = \omega_s \sum_{n=0}^{+\infty} X(\omega) \otimes \delta(\omega - n\omega_s)$$

$$= \omega_s \sum_{n=0}^{+\infty} \int_{-\infty}^{+\infty} X(\xi)\delta(\omega - n\omega_s - \xi)\,\mathrm{d}\xi\,.$$

The following property of the Dirac impulse delta can now be employed

$$\chi(r) \otimes \delta(r - r_0) = \int_{-\infty}^{+\infty} \chi(\xi)\delta(r - r_0 - \xi)\,\mathrm{d}\xi = \chi(r - r_0)\,. \tag{9.14}$$

Eventually one obtains

$$X_\delta(\omega) = \mathcal{F}\left\{\sum_{n=-\infty}^{+\infty} x(nT_s)\delta(t - nT_s)\right\} = \omega_s \cdot \sum_{n=0}^{+\infty} X(\omega - n\omega_s)\,. \tag{9.15}$$

Using (9.15) one can assess the effect of sampling on spectrum $X(\omega)$ of the analog, the continuous-time signal at the input. Sampling introduces, therefore, an infinite number of aliases of spectrum $X(\omega)$ of the analog signal, centered at integer multiples of the sampling circular frequency ω_s.

To visualize this, consider a real signal with all its spectral content lying within the low frequency range (low-pass or baseband signal); like, e.g., the signal with spectrum as in Fig. 9.3. Note that the amplitude spectrum $|X(\omega)|$ of the signal is an odd function of circular frequency ω due to the fact is assumed to be purely real in the time domain.

The amplitude spectrum of signal $x_\delta(t)$ is shown in the Fig. 9.4 as generated by sampling of low-pass signal $x(t)$ above at three (9.3) different sampling rates: $\omega_{s,0} = 2\omega_{max}$, $\omega_{s,1} = 4\omega_{max}$, $\omega_{s,2} = \omega_{max}$.

Using the plot, one can readily derive the following sampling rate criterion attributed to Nyquist (or the Shannon sampling theorem) for the low-pass signal like the one in Fig. 9.4

$$\omega_s \geq \omega_{s,0} = 2\omega_{max}\,. \tag{9.16}$$

If the condition above holds as seen in (9.15) (spectrum periodicity for the sampled signal), as well as Fig. 9.5, an ideal (*brick*) low-pass filter with transfer function as follows can be used to reconstruct the original continuous-time signal

$$H_{LPF}(\omega) = \begin{cases} 1, & |\omega| \leq \omega_s/2 \\ 0, & |\omega| > \omega_s/2 \end{cases}\,. \tag{9.17}$$

The original spectrum is ideally reproduced without any distortion since it holds that

$$H_{LPF}(\omega)X_\delta(\omega) = \omega_s X(\omega)\,. \tag{9.18}$$

This is made clearer next when reconstruction of a continuous-time signal based on the values of a discretized version is discussed.

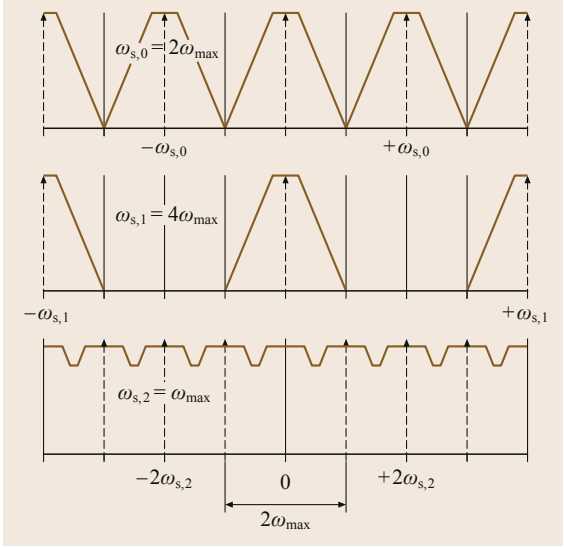


Fig. 9.4 Sampled signal spectrum at various sampling rates

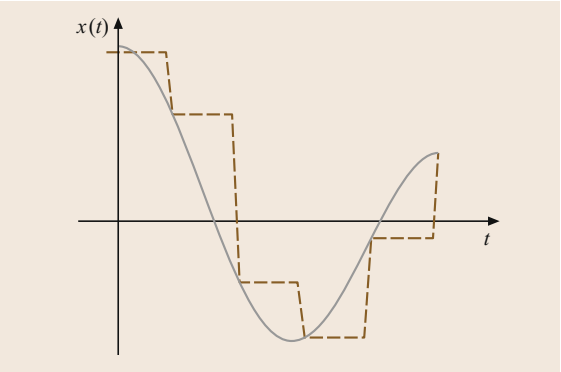


Fig. 9.5 Original continuous-time signal (*continuous line*) and reconstructed version using a non-brick filter (*dashed line*)

9.1.3 Analog Signal Reconstruction Using a Discrete-Time Signal

The simplest process to generate a continuous-time signal, $x(t)$, from a discretized sequence (discrete-time signal) $x(n)$ is to employ the zero-order hold (ZOH) network. This system *holds* the output constant and equal to the one applied to its input for an entire sampling period T_s. In effect, the output of ZOH when driven by a discretized signal is given in the Fig. 9.5.

The mathematical description of a ZOH system is achieved through the following impulse response and transfer function in the complex frequency domain of variable s

$$h_{\mathrm{ZOH}}(t) = u_{\mathrm{step}}(t) - u_{\mathrm{step}}(t - T_s)$$
$$\Leftrightarrow H_{\mathrm{ZOH}}(s) = \frac{1 - e^{-sT_s}}{s}. \tag{9.19}$$

To better understand the action of the ZOH system, its transfer function in the circular frequency domain ω (Fourier transform) is given below, as well

$$H_{\mathrm{ZOH}}(\omega) = \frac{1 - e^{-i\omega T_s}}{i\omega} = T_s e^{-i\omega T_s/2} \frac{\sin(\omega T_s/2)}{(\omega T_s/2)}$$
$$= \frac{e^{-i\pi^{f/f_s}}}{f_s} \mathrm{sinc}(f/f_s)\,. \tag{9.20}$$

In the above $\mathrm{sinc}(\chi) = \mathrm{sinc}(\pi\chi)/\pi\chi$. In Fig. 9.6 the amplitude spectra of the transfer functions of the ZOH system (normalized by setting $f_s = 1\,\mathrm{Hz}$) and the ideal reconstruction low-pass filter which was introduced with (9.17).

Indeed, in the circuit above if resistance R is small (theoretically zero), then time constant τ indicating the time needed for charging the capacitor in the hold circuit is small when compared to sampling period T_s. The

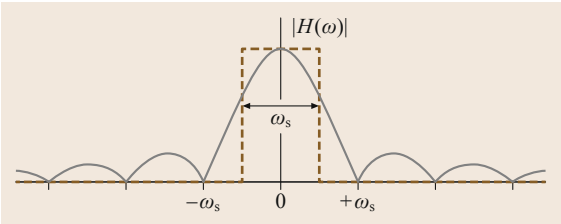


Fig. 9.6 Amplitude spectrum of transfer function $H_{\mathrm{ZOH}}(\omega)$ (*continuous line*) and $H_{\mathrm{LPF}}(\omega)$ (*dashed line*)

voltage across the terminals of the capacitor stays at the value set in the previous sampling instant, when the switch was momentarily switched on, until the next instant the switch is on, etc. The differences of the analog memory element, as the switched resistor and capacitor (RC) circuit is known, and the model ZOH system are:

a) The ohm resistance value of the circuit is non-zero and, therefore, the time to charge the capacitor is finite and not zero.
b) The duration of the conduction phase of the switch is finite and non-zero.
c) The ohm resistance applied to the terminals of the capacitor when the switch is off is not infinite (open circuit) but finite, eventually leading to discharge of the capacitor.

The use of the RC hold circuit is widespread due to its simplicity. However, if a more accurate reconstruction with less distortion, without increasing the sampling rate, then hold circuits of order higher than zero (e.g., first-order hold, FOH) so that its transfer function better approximates the ideal *brick* low-pass filter in (9.17), as shown in Fig. 9.6.

9.1.4 The Z-Transform

The single-sided (unilateral) Z-transform is a linear transform defined as follows for a discrete-time signal $x(n)$

$$X(z) = \mathcal{Z}\{x(n)\} = \sum_{n=0}^{+\infty} x(n) \cdot z^{-n}\,. \tag{9.21}$$

The Z-transform is derived from the Laplace transform of signal $x_\delta(t)$, which is generated by the ideal sampler shown in Fig. 9.7,

$$x_\delta(t) = \sum_{n=-\infty}^{+\infty} x(n)\delta(t - nT_s) \Rightarrow X_\delta(s)$$
$$= \mathcal{L}\left\{\sum_{n=-\infty}^{+\infty} x(n)\delta(t - nT_s)\right\} = \sum_{n=0}^{+\infty} x(n) e^{-n \cdot sT_s}\,. \tag{9.22}$$

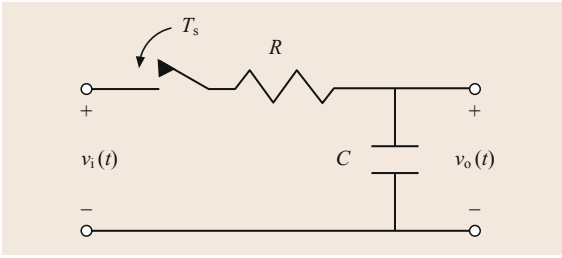


Fig. 9.7 Switched RC circuit modeled by the ZOH system

In effect,

$$X(z)|_{z^{-1}=\exp(-sT_s)} = X_\delta(s) \,. \tag{9.23}$$

Variable z^{-1} is used for describing a signal, or system as we will see later on. Its significance lies in the fact that multiplying by z^{-1} in the transform domain translates to a time shift by one sampling period, T_s, in the discrete-time domain n. Indeed,

$$y(n) = x(n-1) \Leftrightarrow Y(z) = z^{-1}X(z) \,. \tag{9.24}$$

Furthermore, here are some more of the most important properties of the Z-transform:

1. Linearity

$$\left.\begin{aligned} &X_1(z) = \mathcal{Z}\{x_1(n)\} \\ &X_2(z) = \mathcal{Z}\{x_2(n)\} \\ &k_1, k_2 \in \mathbb{R} \end{aligned}\right\} \Rightarrow k_1 \cdot X_1(z) + k_2 \cdot X_2(z) = \mathcal{Z}\{k_1 \cdot x_1(n) + k_2 \cdot x_2(n)\} \,. \tag{9.25}$$

2. Delay or advance by N discrete-time units

$$\mathcal{Z}\{x(n)\} = X(z) \Rightarrow \begin{cases} \mathcal{Z}\{x(n-N)\} = z^{-N}X(z) \\ \mathcal{Z}\{x(n+N)\} = z^{N}X(z) - \sum_{k=0}^{N-1} z^{N-k}x(k) \end{cases} . \tag{9.26}$$

3. Initial or final value theorems

$$\mathcal{Z}\{x(n)\} = X(z) \Rightarrow \begin{cases} x(n=0) = \lim_{z\to\infty}(X(z)) \\ \lim_{n\to\infty}(x(n)) = \lim_{z\to 1}((1-z^{-1})X(z)) \end{cases} . \tag{9.27}$$

Note: the final value theorem ($n \to \infty$) holds only if $X(z)$ is an analytic function of complex variable z.

4. Discrete convolution theorem

$$\left.\begin{aligned} &X_1(z) = \mathcal{Z}\{x_1(n)\} \\ &X_2(z) = \mathcal{Z}\{x_2(n)\} \\ &x(n) = x_1(k) \otimes x_2(k) \triangleq \sum_{k=-\infty}^{+\infty} x_1(k) \cdot x_2(n-k) \end{aligned}\right\} \Rightarrow X(z) = \mathcal{Z}\{x(n)\} = X_1(z) \cdot X_2(z) \,. \tag{9.28}$$

Note: because $x(n)$ needs to be *causal*, i. e., can only depend on values of signals x_1 and x_2 prior or up to instant n, the bounds of the convolutional summation are set as follows,

$$(x_1(k) \otimes x_2(k))(n) = \sum_{k=0}^{n} x_1(k) \cdot x_2(n-k) \,.$$

It is also pointed out here that the commutative and associative properties hold for the convolution operator applied on two discrete-time signals.

Finally, it is pointed out that the change of variable, $z = \exp(sT_s)$, on the basis of which the Z-transform is obtained from the Laplace transform of signal $x_\delta(t)$, defines a single-valued mapping of complex frequency s-plane to the plane of, also complex, delay operator z. Remember that the stability region on the complex plane $s = a + i\omega$, is the left-hand half-plane defined by inequality $\alpha \le 0$. In effect, the stability region on the delay operator z-plane is defined by inequality

$$|z| \le 1 \,. \tag{9.29}$$

Since

$$z = e^{sT_s} = e^{aT_s}e^{i\omega T_s} \Rightarrow \begin{cases} |z| = e^{aT_s} \\ \arg(z) = e^{i\omega T_s} \end{cases} . \tag{9.30}$$

That is, the stability region is the unit disc. Similarly, the origin of the s-plane is mapped to point $z = A + j0$ on the z-plane, while the imaginary axis of the s-plane is wrapped around the unit circle on the z-plane. Finally, the real s axis is mapped to the real z semi-axis. The Z-transform of several elementary discrete-time signals are shown in Table 9.1. In the interest of completeness and allowing for comparisons a Laplace transform table is provided in Table 9.2.

9.1.5 Discrete-Time LTI Systems

Difference Equations

A special class of dynamic systems, widely used in systems and signal theory, is linear time invariant sys-

tems (LTI) [9.1, 2, 4–6]. Such systems are described in continuous time by linear differential equations with constant coefficients. In discrete time, such a differential equation is converted to a difference equation corresponding to a discrete-time system described by a transfer function in variable z, or equivalently z^{-1}. The general form of a linear difference equation with input $u(n)$ and output $y(n)$ is the following,

$$y(n) = \sum_{k=0}^{M} b_k u(n-k) - \sum_{k=1}^{N} a_k y(n-k)\,. \qquad (9.31)$$

To obtain the iterative process started in order to obtain samples of discrete-time output signal $y(n)$ for $n \geq 0$, N initial conditions are needed: $y(-k)$, $k = 1, \ldots, N$. Using (9.31) we can conclude that the n-th sample of the output is determined as a weighed sum of the current plus M past input samples and N past output samples.

Table 9.1 Some Z-transform pairs

k-th term of the time sequence	z-transform
1 at k, 0 elsewhere (Kronecker delta sequence)	z^{-k}
1 (unit step sequence)	$\frac{z}{z-1}$
k (unit ramp sequence)	$\frac{z}{(z-1)^2}$
A^k (for complex numbers A)	$\frac{z}{z-A}$
kA^k	$\frac{Az}{(z-A)^2}$
$\frac{(k+1)(K+2)\cdots(K+n-1)}{(n-1)!}A^k$	$\frac{z^n}{(z-A)^n}$

Table 9.2 Laplace transform pairs

Time function	Laplace transform
Unit impulse $\delta(t)$	1
Unit step $\mathbb{1}(t)$	$\frac{1}{s}$
Unit ramp t	$\frac{1}{s^2}$
Polynomial t^n	$\frac{n!}{s^{n+1}}$
Exponential e^{-at}	$\frac{1}{s+a}$
Sine wave $\sin \omega t$	$\frac{\omega}{s^2+\omega^2}$
Cosine wave $\cos \omega t$	$\frac{s}{s^2+\omega^2}$
Damped sine wave $e^{-at} \sin \omega t$	$\frac{\omega}{(s+a)^2+\omega^2}$
Damped cosine wave $e^{-at} \cos \omega t$	$\frac{s+a}{(s+a)^2+\omega^2}$

Linear Time-Invariant Discrete-Time Systems

The generic difference (9.31) in the discrete-time domain defines an LTI discrete-time system. In the Z-transform domain the difference equation becomes the transfer function as follows

$$\begin{aligned} H(z^{-1}) &= \frac{Y(z)}{U(z)} \\ &= \frac{b_0 + b_1 z^{-1} + \ldots + b_M z^{-M}}{1 + a_1 z^{-1} + \ldots a_{N-1} z^{-N+1} + a_N z^{-N}} \\ &\Updownarrow \\ H(z) &= \frac{b_0 z^N + b_1 z^{N-1} + \ldots + b_M z^{N-M}}{z^N + a_1 z^{N-1} + \ldots + a_N}, \\ & M \leq N\,. \end{aligned} \qquad (9.32)$$

A special, yet very important case of discrete-time systems are finite impulse response (FIR) systems in contrast to the generic IIR (infinite impulse response) system defined in (9.31). For a FIR system it holds that $a_k = 0$, $\forall k > 0$ in (9.31), hence

$$\begin{aligned} y(n) &= \sum_{k=0}^{M} b_k u(n-k) \quad \text{and} \\ H(z^{-1}) &= b_0 + b_1 z^{-1} + \ldots + b_M z^{-M}\,. \end{aligned} \qquad (9.33)$$

In effect, the n-th output sample is *exclusively* determined as a weighed sum of the current plus M past input samples. In effect, the difference equations seizes to be recursive and that is why in the literature FIR systems are also mentioned as MA (moving average) ones.

In contrast, discrete-time LTI systems that have exclusively poles, i. e., $b_i = 0$, for every $i > 0$ in (9.32), are also known as AR (auto-regressive), while the generic IIR system is mentioned also as auto-regressive moving average (ARMA) in this framework.

9.1.6 Continuous-Time System Mapping

There are many ways (methods) to map a continuous-time system to a discrete-time equivalent [9.1, 2, 4–6]. All of these methods are based on preserving some characteristic of the continuous-time system during this procedure; this is why there is more than one way. However, *stability* (or instability) of a continuous-time system is preserved by employing any of the methods presented here.

The mapping methods considered here can be classified into the following categories:

1. Approximate differentiation or integration:
   - Backward difference
   - Forward difference
   - Trapezoidal integration (Tustin transformation).

2. Time-domain response matching at sampling instants:
   - Impulse response matching
   - Step response matching.
3. Dynamic matching, i. e., matching of all poles and zeroes appearing in the transfer function.

Approximate Differentiation or Integration

Consider a continuous-time signal $x(t)$ being sampled with sampling interval T_s and the discrete-time signal $x(n)$ is obtained. Then, the following approximations can be derived.

Approximate derivative with respect to time using backward difference

$$\dot{x}(t) \simeq \frac{x(n)-x(n-1)}{T_s} \Rightarrow s \equiv \frac{1-z^{-1}}{T_s} \,. \tag{9.34}$$

Approximate derivative with respect to time using forward difference

$$\dot{x}(t) \simeq \frac{x(n+1)-x(n)}{T_s} \Rightarrow s \equiv \frac{z-1}{T_s} = \frac{1-z^{-1}}{T_s z^{-1}} \,. \tag{9.35}$$

Approximate integral with respect to time using the trapezoidal rule

$$\begin{aligned} x_1(nT_s) &= \int_0^{nT_s} x(\xi)\,\mathrm{d}\xi = \int_0^{(n-1)T_s} x(\xi)\,\mathrm{d}\xi + \int_{(n-1)T_s}^{nT_s} x(\xi)\,\mathrm{d}\xi \\ &\triangleq x_1((n-1)T_s) + \Delta x_1(n) \,, \\ \Delta x_1(n) &\simeq T_s \cdot \frac{x(n)+x(n-1)}{2} \,. \end{aligned} \tag{9.36}$$

However,

$$\begin{aligned} X_1(s) &= \frac{1}{s}X(s) \quad \text{and} \\ X_1(z) &= z^{-1}X_1(z) + T_s\frac{1+z^{-1}}{2}X(z) \,. \end{aligned} \tag{9.37}$$

The above in effect yields

$$\frac{1}{s} \equiv \frac{T_s}{2}\cdot\frac{1+z^{-1}}{1-z^{-1}} \Leftrightarrow s \equiv \frac{2}{T_s}\cdot\frac{1-z^{-1}}{1+z^{-1}} \,. \tag{9.38}$$

Equations (9.34), (9.35), and (9.38) mean that when presented with a transfer function $H(s)$ in the complex frequency (s) domain, describing a continuous-time LTI system, it is possible to derive a discrete-time transfer function $H(z)$, or $H(z^{-1})$, defining a discrete-time LTI system by substituting variable s with the approximation desired.

Impulse Response Matching

The methods presented previously were based on approximating the integration or differentiation operator appearing in the differential (dynamic) equation of an LTI system. However, an LTI system can, also, be described in full by its response to an arbitrary input signal. Common choices include using either Dirac's delta (impulse) or the unit step signal (Heaviside signal).

In effect, if the impulse or step response of a continuous-time system is sampled, then a discrete-time description can be derived as follows.

Invariance of impulse response: the impulse delta is defined as follows in discrete time

$$\delta(n) = \begin{cases} 0, & n \neq 0 \\ 1, & n = 0 \end{cases} \Leftrightarrow \Delta(z) = 1 \,. \tag{9.39}$$

Given the transfer function, $H(z)$ of a discrete-time LTI system, it is easy to prove that, just like in the continuous time case, $H(z)$ is the Z-transform of the system's impulse response.

Indeed, the output $y(n)$ of a discrete-time LTI system is given by the following convolutional sum

$$\begin{aligned} y(n) = u(k) \otimes h(k) &\triangleq \sum_{k=-\infty}^{+\infty} u(k)\cdot h(n-k) \Rightarrow Y(z) \\ &= H(z)\cdot U(z) \,. \end{aligned} \tag{9.40}$$

In effect,

$$u(n) = \delta(n) \Rightarrow U(z) = 1 \Rightarrow Y(z) = H(z) \,.$$

Consider a continuous-time LTI system with impulse response $h(t) = \mathcal{L}^{-1}\{H(s)\}$, which is being sampled at rate f_s. This means that continuous-time signal $h_\delta(t) = \sum_{n=-\infty}^{+\infty} h(nT_s)\delta(t-nT_s)$ corresponds to discrete-time signal $h_\delta(n) \equiv h(nT_s)$, i. e., a sample set of $h(t)$.

The Z-transform of $h_\delta(n)$,

$$H_\delta(z^{-1}) = \sum_{n=0}^{+\infty} h_\delta(n) z^{-n} \,,$$

defines a discrete-time LTI system, the impulse response of which, by definition, coincides with that of continuous-time transfer function $H(s)$, at all sampling instants.

Assume, now, that for every positive M there is a positive integer ε such that

$$\sum_{n=M+1}^{+\infty} h_\delta^2(n) < \varepsilon \,. \tag{9.41}$$

Given the above, one can proceed to impulse response truncation with accuracy ε. This procedure consists of approximating the original IIR system with transfer function $H_\delta(z^{-1}) = \sum_{n=0}^{+\infty} h_\delta(n) z^{-n}$ by the following FIR one with memory M

$$\hat{H}_{\text{FIR}}(z^{-1}) = \sum_{n=0}^{M} h_\delta(n) z^{-n} = \sum_{n=0}^{M} b_n z^{-n} \,. \tag{9.42}$$

Approximating a continuous-time system with a discrete-time one, which on top might be FIR, comes with a number of advantages, especially for real-time applications.

Invariance of step response: the unit step signal in discrete time is defined as follows

$$u_{\text{step}}(n) = \begin{cases} 1, & n \geq 0 \\ 0, & n < 0 \end{cases} \quad \Leftrightarrow$$

$$U_{\text{step}}(z) = \sum_{n=0}^{+\infty} z^{-n} = \frac{1}{1 - z^{-1}} \,. \tag{9.43}$$

The Z-transform of the unit step signal is derived by applying the lemma for the infinite sum of a geometric progression with a ratio that in magnitude is less than 1.

Consider now an LTI system with impulse response as follows

$$y_{\text{step}}(t) = \mathcal{L}^{-1}\{H(s) U_{\text{step}}(s)\} = \mathcal{L}^{-1}\left\{\frac{H(s)}{s}\right\} \,. \tag{9.44}$$

We will now determine a discrete-time transfer function $H(z)$ such that the latter's unit step response coincides with that of the continuous-time system above. Mathematically, this is defined as follows

$$\mathcal{Z}^{-1}\{H(z) U_{\text{step}}(z)\} = y_{\text{step}}(n) = y_{\text{step}}(nT_{\text{s}}) \,. \tag{9.45}$$

Then, by employing (9.43), one obtains that

$$H(z) = (1 - z^{-1}) \mathcal{Z}\{y_{\text{step}}(nT_{\text{s}})\} \,. \tag{9.46}$$

Since

$$z^{-1} = \exp(-sT_{\text{s}}) \quad \text{and} \quad Y_{\text{step}}(s) = \frac{H(s)}{s} \,.$$

We can finally obtain that

$$H(z) = \mathcal{Z}\{y_{\text{ZOH}}(nT_{\text{s}})\} \,. \tag{9.47}$$

In the above, it holds that

$$y_{\text{ZOH}}(t) = h_{\text{ZOH}}(t) \otimes h(t) \Leftrightarrow$$

$$Y_{\text{ZOH}}(s) = H_{\text{ZOH}}(s) H(s) = \frac{1 - e^{-sT_{\text{s}}}}{s} H(s) \,. \tag{9.48}$$

Pole and Zero Matching of a Transfer Function

This method derives from equation

$$z^{-1} = \exp(-sT_{\text{s}})$$

connecting variable z with complex frequency s. Combining the above with the following factorized form of a continuous-time scalar transfer function

$$H(s) = K_{\text{s}} \frac{(s + z_1)(s + z_2) \ldots (s + z_m)}{(s + p_1)(s + p_2) \ldots (s + p_n)}, \quad m \leq n \,. \tag{9.49}$$

One can map every pole and every zero of the continuous-time transfer function to a corresponding pole and zero of an equivalent discrete-time transfer function

$$\begin{aligned} \zeta_{\text{i}} &= -\exp(-T_{\text{s}} z_{\text{i}}), \quad i = 1, \ldots, m \,, \\ \pi_{\text{i}} &= -\exp(-T_{\text{s}} p_{\text{i}}), \quad i = 1, \ldots, n \,. \end{aligned} \tag{9.50}$$

In effect, the following discrete-time transfer function is derived in its factorized from

$$H(z) = k_z \frac{(z+1)^{n-m}(z + \zeta_1)(z + \zeta_2) \ldots (z + \zeta_m)}{(z + \pi_1)(z + \pi_2) \ldots (z + \pi_n)} \,. \tag{9.51}$$

Factor $(z+1)^{n-m}$ is introduced in the discrete-time transfer function above in correspondence to factor $(\tau s + 1)^{n-m}$ with $\tau \to \pm j0$, which is *hidden* in the continuous-time transfer function (9.49). The intent is to equate the numerator degree to that of the denominator. This is achieved by introducing an extra zero at imaginary infinity ($\pm j\infty$) with multiplicity $(n - m)$ using factor $(\tau s + 1)^{n-m}$. In the discrete-time transfer function, factor $(\tau s + 1)^{n-m}$ is mapped to $(z + 1)^{n-m}$, since point $\pm j\infty$ on the s-plane is mapped to -1 on the z-plane.

Finally, gain k in the discrete-time transfer function is determined by equating its value at a certain z with that of the continuous-time transfer function at the equivalent point s. For example, for $z = 1$, the corresponding point is $s = 0$, and, therefore, k is calculated as follows

$$\begin{aligned} H(z)|_{z=1} &= H(s)|_{s=0} \\ &\Leftrightarrow k_z \frac{2^{n-m}(1 + \zeta_1)(1 + \zeta_2) \ldots (1 + \zeta_m)}{(1 + \pi_1)(1 + \pi_2) \ldots (1 + \pi_n)} \,, \\ &= K_{\text{s}} \frac{z_1 z_2 \ldots z_m}{p_1 p_2 \ldots p_n} \,. \end{aligned} \tag{9.52}$$

9.2 Digital Filters

In the previous section we studied different ways of describing discrete-time systems that are linear and time invariant. It was verified that the Z-transform greatly simplifies the analysis of discrete-time systems, especially those initially described by difference equations.

In this section, we study in finer detail several structures used to realize a given transfer function associated with a specific difference equation through the use of the Z-transform. The transfer functions considered here will be of the polynomial form (non-recursive filters, FIR) and of the rational-polynomial form (recursive filters, IIR). In the non-recursive case we emphasize the existence of the important subclass of linear-phase filters. Then we introduce some tools to calculate the digital network transfer function, as well as to analyze its internal behavior. We also discuss some properties of generic digital filter structures associated with practical discrete-time systems.

9.2.1 Important FIR Filter Structures

Non-recursive filters are characterized by a difference equation in the form

$$y(n) = \sum_{l=0}^{M} b_l x(n-l) \,, \tag{9.53}$$

where the b_l coefficients are directly related to the system impulse response; that is, $b_l = h(l)$. Owing to the finite length of their impulse responses, non-recursive filters are also referred to as finite-duration impulse response (FIR) filters. We can rewrite (9.53) as follows

$$y(n) = \sum_{l=0}^{M} h(l)x(n-l) \,. \tag{9.54}$$

Applying the Z-transform to the equation above, we end up with the following input–output relationship

$$H(z) = \frac{Y(z)}{X(z)} = \sum_{l=0}^{M} h(l)z^{-l} = \sum_{l=0}^{M} b_l z^{-l} \,. \tag{9.55}$$

In practical terms, (9.55) can be implemented in several distinct forms, using as basic elements the delay, the multiplier, and the adder blocks. These basic elements of digital filters and their corresponding standard symbols are depicted in Fig. 9.8. An alternative way of representing such elements is the so-called signal flow graph shown in Fig. 9.9.

These two sets of symbolisms representing the delay, multiplier, and adder elements are used throughout this text interchangeably.

Direct Form

The simplest realization of an FIR digital filter is derived from (9.55). The resulting structure, which can be seen in Fig. 9.10, is called the direct-form realization, as the multiplier coefficients are obtained directly from the filter transfer function. Such a structure is also referred to as the canonic direct form, where we understand canonic form to mean any structure that realizes a given transfer function with the minimum number of delays,

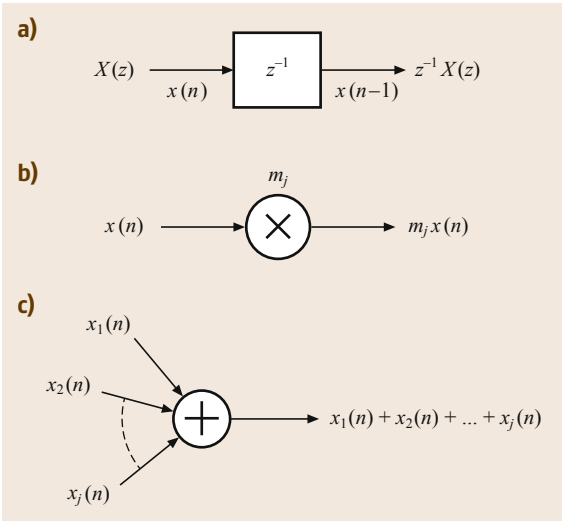


Fig. 9.8a–c Classic representation of basic elements of digital filters: **(a)** delay; **(b)** multiplier; **(c)** adder

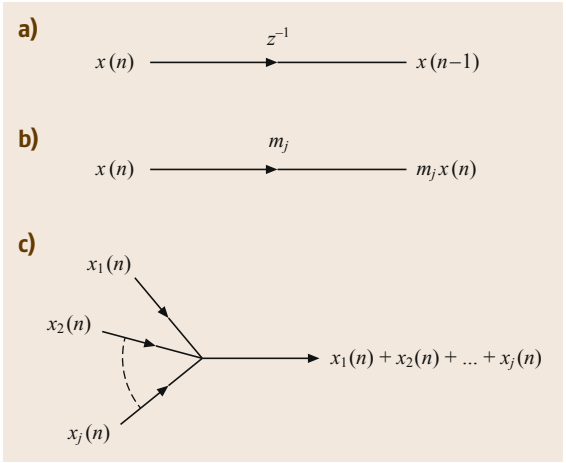


Fig. 9.9a–c Signal-flow graph representation of basic elements of digital filters: **(a)** delay; **(b)** multiplier; **(c)** adder

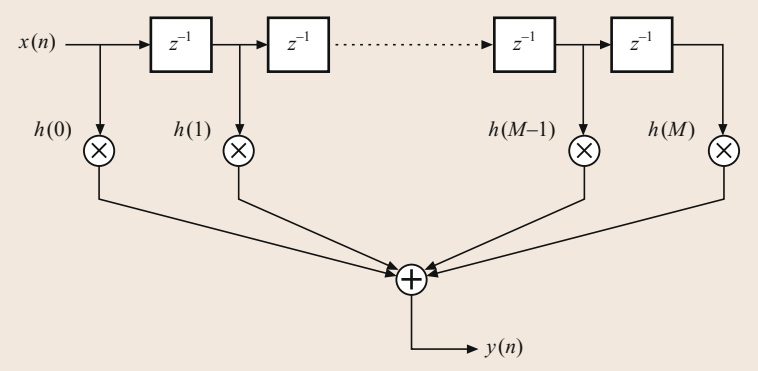


Fig. 9.10 Direct form for FIR digital filters

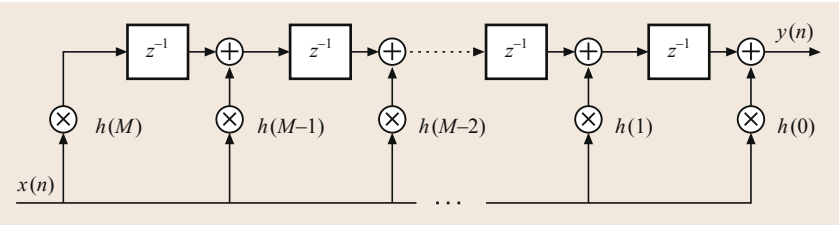


Fig. 9.11 Alternative direct form for FIR digital filters

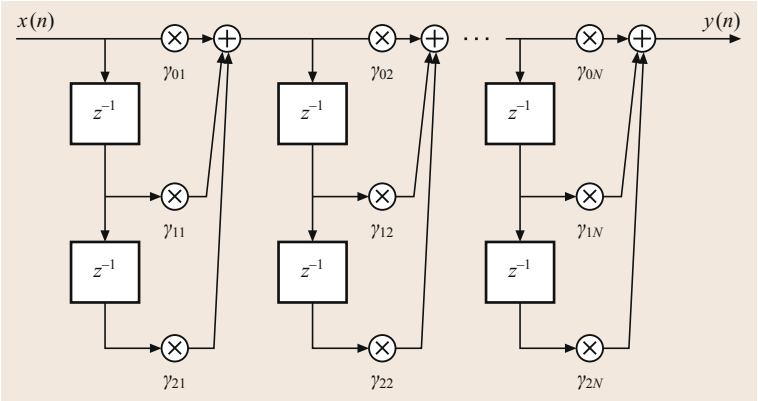


Fig. 9.12 Cascade form for FIR digital filters

multipliers, and adders. More specifically, a structure that utilizes the minimum number of delays is said to be canonical with respect to the delay element, and so on.

An alternative canonical direct form for (9.76) can be derived by expressing $H(z)$ as follows

$$(z) = \sum_{l=0}^{M} h(l)z^{-l} = h(0) + z^{-1}\left(\sum_{l=1}^{M} h(l)z^{-l}\right) . \quad (9.56)$$

The implementation of this form is shown in Fig. 9.11.

Cascade Form

Equation (9.55) can be realized through a series of equivalent structures. However, the coefficients of such distinct realizations may not be explicitly the filter impulse response or the corresponding transfer function. An important example of such a realization is the so-called cascade form, which consists of a series of second-order FIR filters connected in cascade, thus the name of the resulting structure, as seen in Fig. 9.12. The transfer function associated with such a realization is of the form

$$H(z) = \prod_{k=1}^{N} (\gamma_{0k} + \gamma_{1k}z^{-1} + \gamma_{2k}z^{-2}) . \quad (9.57)$$

In the above, if M is the filter order, then $N = M/2$ when M is even and $N = (M+1)/2$ when M is odd. In the latter case, one of the γ_{2k} vanishes.

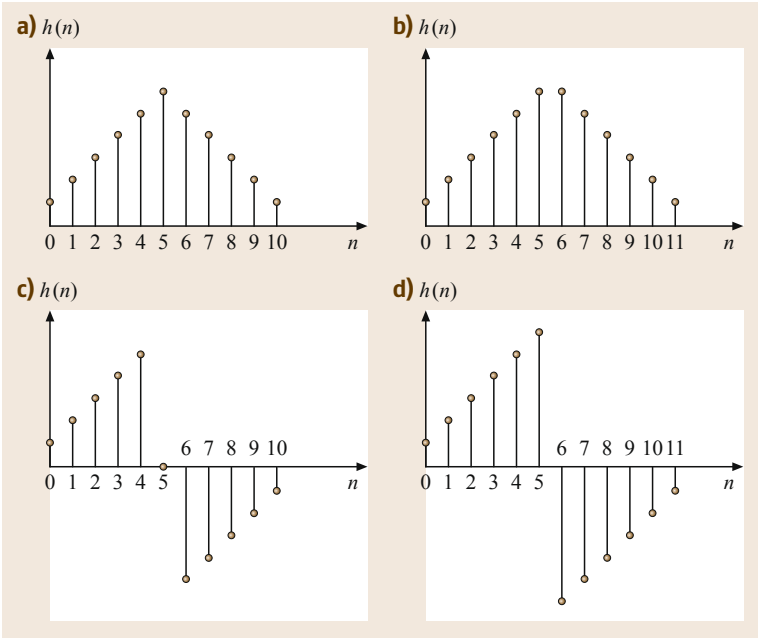


Fig. 9.13a–d Example of impulse responses of linear-phase FIR digital filters: **(a)** type I; **(b)** type II; **(c)** type III; **(d)** type IV

Linear-Phase Form

An important subclass of FIR digital filters is the one that includes linear-phase filters. Such filters are characterized by a constant group delay τ; therefore, they must present a frequency response of the following form

$$H(e^{i\omega}) = B(\omega)\exp(-i\omega\tau + i\varphi)\,. \tag{9.58}$$

In the above $B(\omega)$ is real and τ and φ are constant. We now proceed to show that linear-phase FIR filters present impulse responses of very particular forms. Specifically, if $h(n)$ is to be causal and of finite duration, for $0 \le n \le M$, we must necessarily have that

$$\tau = \frac{M}{2}\,. \tag{9.59}$$

Therefore, one obtains that

$$h(n) = e^{2i\varphi}h^*(M-n)\,. \tag{9.60}$$

This is the general equation that the coefficients of a linear-phase FIR filter must satisfy. In the common case, where all the filter coefficients are real, one finally obtains that the filter impulse response must be either *symmetric* or *antisymmetric*. In effect, the frequency response of linear-phase FIR filters with real coefficients becomes as follows

$$H(e^{i\omega}) = B(\omega)\exp\left(-i\omega\left(\frac{M}{2}\right) + i\left(\frac{k\pi}{2}\right)\right). \tag{9.61}$$

As a result, solely four distinct cases need be considered that are described by the equations above. Their types have been standardized and are referred to in literature as follows [9.2]:

- Type I: $k = 0$ and M even
- Type II: $k = 0$ and M odd
- Type III: $k = 1$ and M even
- Type IV: $k = 1$ and M odd.

Typical impulse responses of the four cases of linear-phase FIR digital filters are depicted in Fig. 9.13.

9.2.2 Important IIR Filter Structures

Direct Form

Recursive filters have transfer functions of the following form

$$H(z) = \frac{N(z)}{D(z)} = \frac{\sum_{i=0}^{M} b_i z^{-i}}{1 + \sum_{i=1}^{N} a_i z^{-i}}\,. \tag{9.62}$$

Since, in most cases, such transfer functions give rise to filters with impulse responses having infinite durations, recursive filters are also referred to as infinite-duration impulse response (IIR) filters.

We can consider that $H(z)$ as above results from the cascading of two separate filters of transfer functions $N(z)$ and $1/D(z)$. The $N(z)$ polynomial can be realized with the FIR direct form, as shown in the previous section. The realization of $1/D(z)$ can be performed as depicted in Fig. 9.14, where the FIR filter shown will

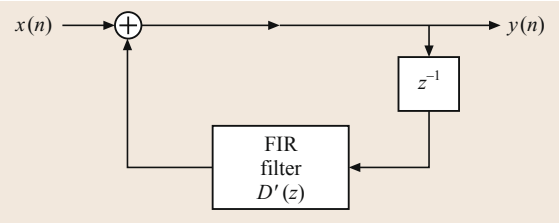


Fig. 9.14 Block diagram realization of $1/D(z)$

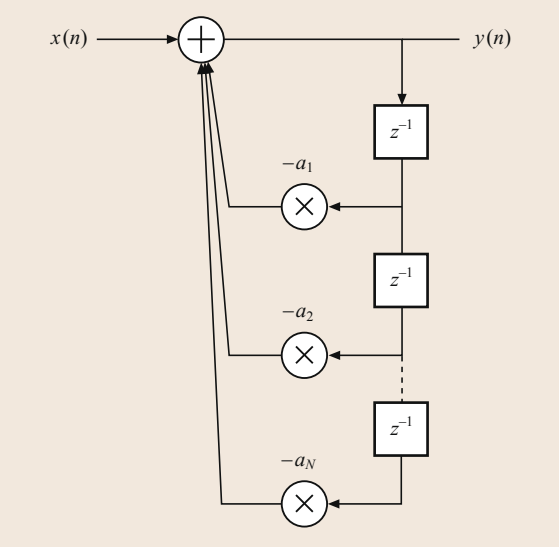


Fig. 9.15 Detailed realization of $1/D(z)$

be an $(N-1)$-th order filter with transfer function as follows

$$D'(z) = z(1-D(z)) = -z\sum_{i=1}^{N} a_i z^{-i} . \qquad (9.63)$$

The direct form of realizing $1/D(z)$ is shown in Fig. 9.15.

The complete realization of $H(z)$, as a cascade of $N(z)$ and $1/D(z)$, is shown in Fig. 9.16. Such a structure is not canonic with respect to the delays, since for an (M,N)-th order filter this realization requires $(N+M)$ delays.

Clearly, in the general case we can change the order in which we cascade the two separate filters; that is, $H(z)$ can be realized as

$$N(z) \times \frac{1}{D(z)} \quad \text{or} \quad \left(\frac{1}{D(z)}\right) \times N(z) .$$

In the second option, all delays employed start from the same node, which allows us to eliminate the consequent redundant delays. In that manner, the resulting structure, usually referred to as the Type 1 canonic direct

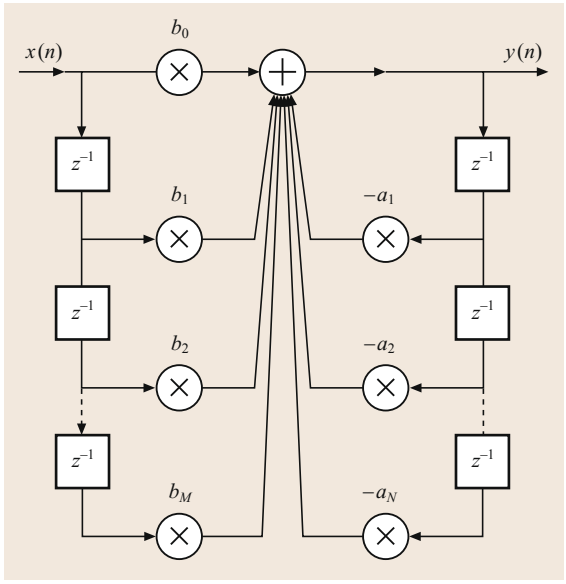


Fig. 9.16 Non-canonic IIR direct form realization

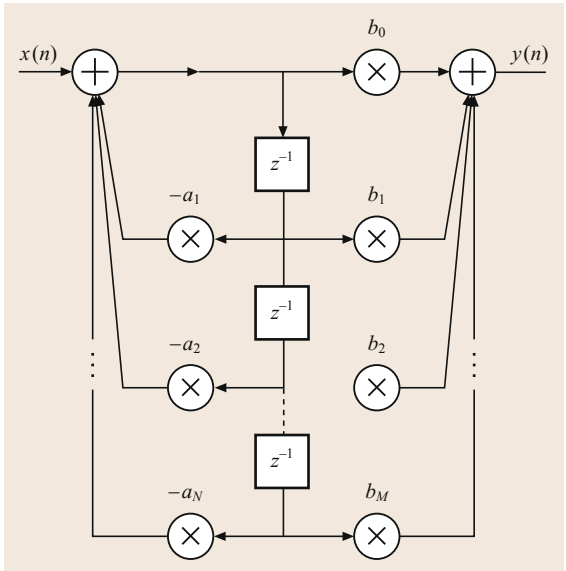


Fig. 9.17 Type 1 canonic direct form for IIR filters

form, is the one depicted in Fig. 9.17, for the special case $N = M$.

An alternative structure, the so-called Type 2 canonic direct form, is shown in Fig. 9.18. Such a realization is generated from the corresponding non-recursive form.

The majority of IIR filter transfer functions used in practice present a numerator degree M smaller than or equal to the denominator degree N. In general, one can consider, without much loss of generality, that $M = N$.

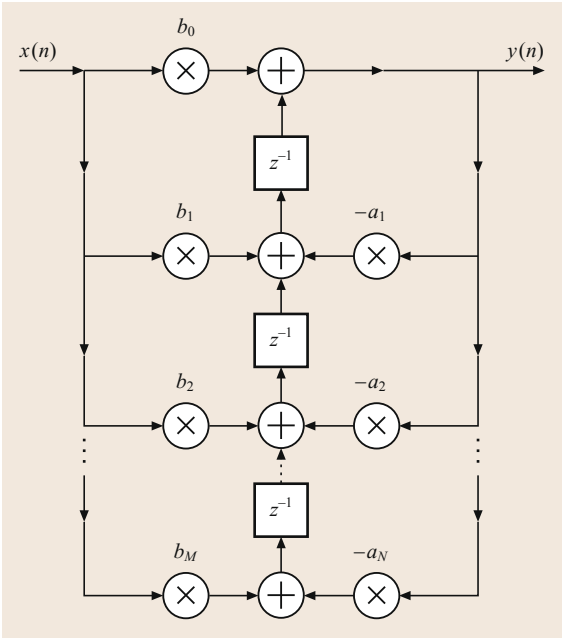


Fig. 9.18 Type 2 canonic direct form for IIR filters

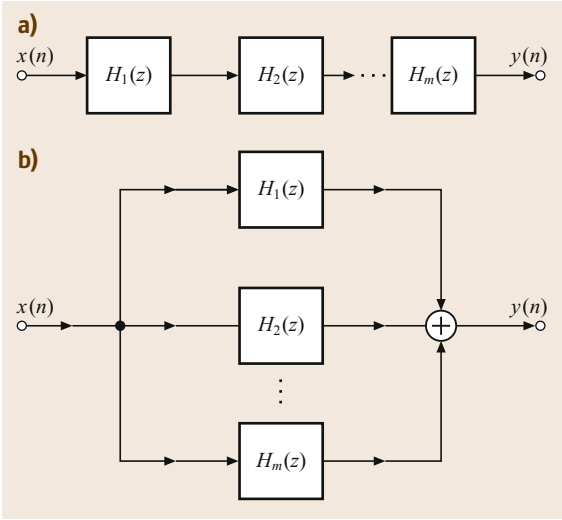


Fig. 9.19a,b Block diagrams of: **(a)** cascade form; **(b)** parallel form

In the case where $M < N$, we just make the coefficients $b_{M+1}, b_{M+2}, \ldots, b_N$ in Figs. 9.17 and 9.18 equal to zero.

Cascade Form

In the same way as their FIR counterparts, IIR digital filters present a large variety of possible alternative realizations. An important one, referred to as cascade

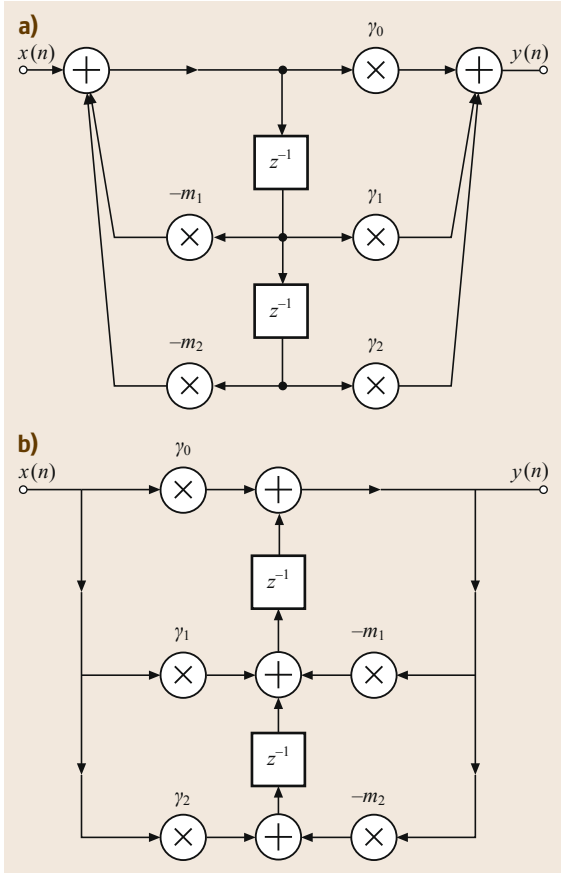


Fig. 9.20a,b Block diagrams of: **(a)** cascade form; **(b)** parallel form

realization, is depicted in Fig. 9.19a, where the basic blocks represent simple transfer functions of orders 2 or 1. In fact, the cascade form, based on second-order blocks, is associated with the following transfer function decomposition

$$H(z) = \prod_{k=1}^{m} \frac{\gamma_{0k} + \gamma_{1k} z^{-1} + \gamma_{2k} z^{-2}}{1 + m_{1k} z^{-1} + m_{2k} z^{-2}} ,$$

$$= \prod_{k=1}^{m} \frac{\gamma_{0k} z^2 + \gamma_{1k} z + \gamma_{2k}}{z^2 + m_{1k} z + m_{2k}} ,$$

$$= H_0 \prod_{k=1}^{m} \frac{z^2 + \gamma'_{1k} z + \gamma'_{2k}}{z^2 + m_{1k} z + m_{2k}} . \qquad (9.64)$$

Parallel Form

Another important realization for recursive digital filters is the parallel form represented in Fig. 9.19b. Using second-order blocks, which are the most commonly

used in practice, the parallel realization corresponds to the following transfer function decomposition

$$H(z) = \prod_{k=1}^{m} \frac{\gamma_{0k}^{\mathrm{p}} z^2 + \gamma_{1k}^{\mathrm{p}} z + \gamma_{2k}^{\mathrm{p}}}{z^2 + m_{1k} z + m_{2k}} ,$$
$$= h_0 + \prod_{k=1}^{m} \frac{\gamma_{1k}^{\mathrm{p}'} z + \gamma_{2k}^{\mathrm{p}'}}{z^2 + m_{1k} z + m_{2k}} ,$$
$$= h_0' + \prod_{k=1}^{m} \frac{\gamma_{0k}^{\mathrm{p}''} z^2 + \gamma_{1k}^{\mathrm{p}''} z}{z^2 + m_{1k} z + m_{2k}} . \qquad (9.65)$$

The above is also known as the partial-fraction decomposition. This equation indicates three alternative forms of the parallel realization, where the last two are canonic with respect to the number of multiplier elements.

It should be mentioned that each second-order block in the cascade and parallel forms can be realized by any of the existing distinct structures, as, for instance, one of the direct forms shown in Fig. 9.20.

All these digital filter realizations present different properties when one considers practical finite-precision implementations; that is, the quantization of the coefficients and the finite precision of the arithmetic operations, such as additions and multiplications. In fact, the analysis of the finite-precision effects in the distinct realizations is a fundamental step in the overall process of designing any digital filter [9.2].

9.3 The Fast Fourier Transform (FFT)

The FFT (fast Fourier transform) algorithm is a faster version of the discrete Fourier transform (DFT). FFT utilizes some clever algorithms to do the same thing as the DFT, but in much less time [9.2].

The DFT is extremely important in the area of frequency (spectral) analysis because it takes a discrete signal in the time domain and transforms that signal into its discrete frequency domain representation. Without a discrete-time to discrete-frequency transform we would not be able to compute the Fourier transform with a microprocessor or DSP-based (DSP: Digital Signal Processor) system [9.1, 2, 4–6].

It is the speed and discrete nature of the FFT that allows us to analyze a signal's spectrum, as will soon become evident.

9.3.1 Review of Integral Transforms

We first give a review of the integral transforms that have been used in the text, possibly in their unilateral (single-side) version.

The bilateral Laplace transform

$$X(s) = \mathcal{L}\{x(t)\} = \int_{-\infty}^{+\infty} x(t) e^{-st} \mathrm{d}t, x(t) \overset{\mathcal{L}}{\rightleftarrows} X(s) . \qquad (9.66)$$

The continuous-time Fourier transform

$$X(i\omega) = \mathcal{F}\{x(t)\} = \int_{-\infty}^{+\infty} x(t) e^{-i\omega t} \mathrm{d}t, x(t) \overset{\mathcal{F}}{\rightleftarrows} X(i\omega) . \qquad (9.67)$$

The bilateral Z-transform

$$X(z) = \mathcal{Z}\{x(n)\}$$
$$= \sum_{n=-\infty}^{+\infty} x(n) z^{-n}, x(n) \overset{\mathcal{Z}}{\rightleftarrows} X(z) . \qquad (9.68)$$

The Laplace transform is used to obtain a pole-zero representation of a continuous-time signal or system, $x(t)$, in the s-plane. Similarly, the Z-transform is used to find a pole-zero representation of a discrete-time signal or system, $x(n)$, in the z-plane.

The continuous-time Fourier transform can be found by evaluating the Laplace transform form at $s = i\omega$. The picture can be extended by introducing the discrete-time Fourier transform (DTFT). The DTFT can be found by evaluating the Z-transform at $z = \exp(i\Omega)$, as follows

$$X(e^{i\Omega}) = \sum_{n=-\infty}^{+\infty} x(n) e^{-i\Omega n}, x(n) \overset{\mathrm{DTFT}}{\rightleftarrows} X(e^{i\Omega}) . \qquad (9.69)$$

One needs to point out here that the frequency variable Ω is in normalized units of radians per sample rather than *absolute* units of rad/s, which apply to ω appearing in the standard Fourier transform. This can be justified by recollecting that sequences $x(n)$ are generated by sampling a continuous-time signal $x(t)$ at a certain sampling rate f_s. In this respect, the DTFT can be viewed as a discrete approximation of the Fourier

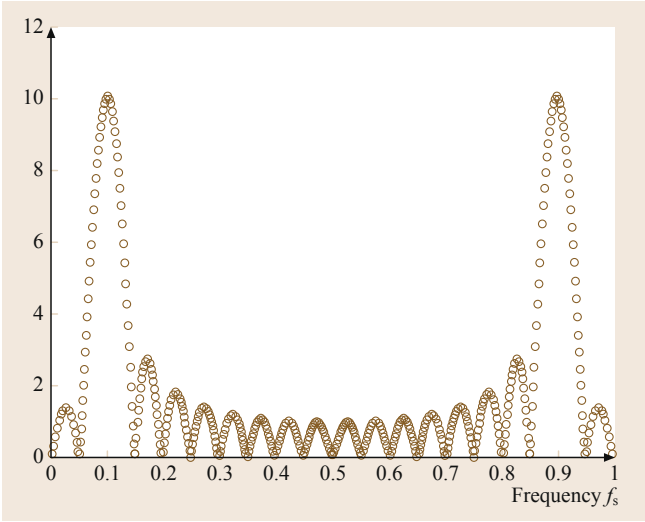


Fig. 9.21 Plot showing the symmetry of DFT

transform as explained below

$$\mathcal{F}\{x(t)\} = \int_{-\infty}^{+\infty} x(t)e^{-i\omega t}\,\mathrm{d}t$$
$$\simeq T_s \sum_{n=-\infty}^{+\infty} x(nT_s)\exp(-i\omega nT_s)$$
$$\Downarrow$$
$$\mathcal{F}\{x(t)\} \simeq T_s \cdot X(e^{i\Omega}), \Omega = \omega T_s = 2\pi\frac{f}{f_s}\,. \qquad (9.70)$$

9.3.2 The Discrete Fourier Transform (DFT)

First of all, the discrete Fourier transform (DFT) is *not* the same as the DTFT. Both start with a discrete-time signal, but DFT produces a discrete frequency domain representation while DTFT is continuous in the frequency domain. These two transforms have much in common, however. It is, therefore, helpful to have a basic understanding of the properties of the DTFT.

Periodicity

DTFT is periodic because of the fact that the signal is a discrete-time one. Indeed,

$$X(e^{i(\Omega+2\pi k)}) = \sum_{n=-\infty}^{+\infty} x(n)e^{-i(\Omega+2\pi k)n}$$
$$= \sum_{n=-\infty}^{+\infty} x(n)e^{-i\Omega n}e^{-i2\pi kn} = X(e^{i\Omega})\,. \qquad (9.71)$$

One fundamental period is, therefore, 2π, i. e., extends from $f = 0$ to f_s, where f_s is the sampling frequency. Taking advantage of this redundancy, the DFT is only defined in the region between 0 and f_s, in terms of f, or, 0 and 2π, in terms of Ω.

Symmetry

When the region between 0 and f_s is examined, it can be seen that there is even symmetry around the center point, i. e., point $f_s/2$ (half the sampling rate). Indeed,

$$\Omega_{1,2} = \pi \pm \varepsilon, 0 < \varepsilon < \pi :$$
$$X(e^{i\Omega_2})$$
$$= \sum_{n=-\infty}^{+\infty} x(n)e^{-i(\pi+\varepsilon)n} = \sum_{n=-\infty}^{+\infty} x(n)e^{-i\varepsilon n}e^{-i\pi n}$$
$$= \sum_{n=-\infty}^{+\infty} x(n)e^{-i\varepsilon n}e^{+i\pi n} = \sum_{n=-\infty}^{+\infty} x(n)e^{-i(\pi-\varepsilon)n}$$
$$= X(e^{i\Omega_1})\,. \qquad (9.72)$$

This symmetry adds redundant information. Figure 9.21 shows the DFT (implemented with Matlab's FFT function) of a cosine with a frequency one tenth of the sampling frequency. Note that the data between $0.5f_s$ and f_s is a mirror image of the data between 0 and $0.5f_s$.

Therefore, the discrete Fourier transform (DFT) can be introduced as follows

$$X_k = \sum_{n=0}^{N-1} x_n \exp\left(-\frac{i2\pi nk}{N}\right),$$
$$k = 0, 1, 2, \ldots, (N-1)\,. \qquad (9.73)$$

Note that the above is actually a transformation between a finite-length real or complex sequence x_n, corresponding to a sampled, at rate $f_s = 1/T_s$, segment of a real or complex signal with actual duration NT_s, to a generally complex sequence (discrete spectrum) of equal finite length N corresponding to frequency values in the range between 0 and $(N-1)f_s/N$ with a step equal to f_s/N; note that the corresponding range in terms of Ω is 0 to $(N-1)2\pi/N$ and the step $2\pi/N$.

Fast Fourier Transform (FFT)

FFT is simply an algorithm to speed up the DFT calculation by reducing the number of multiplications and additions required. It was popularized by J.W. Cooley and J.W. Tukey in the 1960s and, was actually a rediscovery of an idea of Runge (1903) and Danielson and Lanczos (1942), first occurring prior to the availability of computers and calculators – when numerical calculation could take many man hours. In ad-

Table 9.3 8-point DFT ($N = 8$)

$X(k) = \frac{1}{N}\sum_{n=0}^{N-1} x(n)e^{\frac{-i2\pi nk}{N}} = \frac{1}{N}\sum_{n=0}^{N-1} x(n)W_N{}^{nk} \quad W_N = e^{\frac{-i2\pi}{N}}$
$X(0) = x(0)W_8^0 + x(1)W_8^0 + x(2)W_8^0 + x(3)W_8^0 + x(4)W_8^0 + x(5)W_8^0 + x(6)W_8^0 + x(7)W_8^0$
$X(1) = x(0)W_8^0 + x(1)W_8^1 + x(2)W_8^2 + x(3)W_8^3 + x(4)W_8^4 + x(5)W_8^5 + x(6)W_8^6 + x(7)W_8^7$
$X(2) = x(0)W_8^0 + x(1)W_8^2 + x(2)W_8^4 + x(3)W_8^6 + x(4)W_8^8 + x(5)W_8^{10} + x(6)W_8^{12} + x(7)W_8^{14}$
$X(3) = x(0)W_8^0 + x(1)W_8^3 + x(2)W_8^6 + x(3)W_8^9 + x(4)W_8^{12} + x(5)W_8^{15} + x(6)W_8^{18} + x(7)W_8^{21}$
$X(4) = x(0)W_8^0 + x(1)W_8^4 + x(2)W_8^8 + x(3)W_8^{12} + x(4)W_8^{16} + x(5)W_8^{20} + x(6)W_8^{24} + x(7)W_8^{28}$
$X(5) = x(0)W_8^0 + x(1)W_8^5 + x(2)W_8^{10} + x(3)W_8^{15} + x(4)W_8^{20} + x(5)W_8^{25} + x(6)W_8^{30} + x(7)W_8^{35}$
$X(6) = x(0)W_8^0 + x(1)W_8^6 + x(2)W_8^{12} + x(3)W_8^{18} + x(4)W_8^{24} + x(5)W_8^{30} + x(6)W_8^{36} + x(7)W_8^{42}$
$X(7) = x(0)W_8^0 + x(1)W_8^7 + x(2)W_8^{14} + x(3)W_8^{21} + x(4)W_8^{28} + x(5)W_8^{35} + x(6)W_8^{42} + x(7)W_8^{49}$
N^2 Complex multiplications
$\frac{1}{N}$ Scaling factor omitted

Table 9.4 8-point DFT. Applying the properties of symmetry and periodicity to W_N^r for $N = 8$

$N = 8$	$W_8^4 = W_8^{0+4} = -W_8^0 = -1$
	$W_8^5 = W_8^{1+4} = -W_8^1$
	$W_8^6 = W_8^{2+4} = -W_8^2$
	$W_8^7 = W_8^{3+4} = -W_8^3$
	$W_8^8 = W_8^{0+8} = +W_8^0 = +1$
	$W_8^9 = W_8^{1+8} = +W_8^1$
	$W_8^{10} = W_8^{2+8} = +W_8^2$
	$W_8^{11} = W_8^{3+4} = +W_8^3$
	• • •
	• • •
	• • •
Symmetry: $W_N^{r+N/2} = -W_N^r$, Periodicity: $W_N^{r+N} = W_N^r$	

dition, the German mathematician Carl Friedrich Gauss (1777–1855) had used the method more than a century earlier.

In order to understand the basic concepts of FFT and its derivation, note that the DFT expansion shown in Table 9.3 can be greatly simplified by taking advantage of the symmetry and periodicity of the twiddle factors as shown in Table 9.4. If the equations are rearranged and factored, the result is the fast Fourier transform (FFT), which requires only $(N/2)\log_2(N)$ complex multiplications. The computational efficiency of FFT versus DFT becomes highly significant when the FFT point size increases to several thousand, as shown in Table 9.5. However, notice that FFT computes *all* the output frequency components (either all or none!). If only a few spectral points need to be calculated, DFT may actually be more efficient. Calculation of a single spectral output using the DFT requires only N complex multiplications.

Table 9.5 FFT versus DFT. FFT is simply an algorithm for efficiently calculating DFT. Computational efficiency of an N-point FFT: 1) DFT: N^2 Complex multiplications; 2) FFT: $(N/2)\log_2(N)$ Complex multiplications

N	DFT multiplications	FFT multiplications	FFT efficiency
256	65 536	1024	64 : 1
512	262 144	2304	114 : 1
1024	1 048 576	5120	205 : 1
2048	4 194 304	11 264	372 : 1
4096	16 777 216	24 576	683 : 1

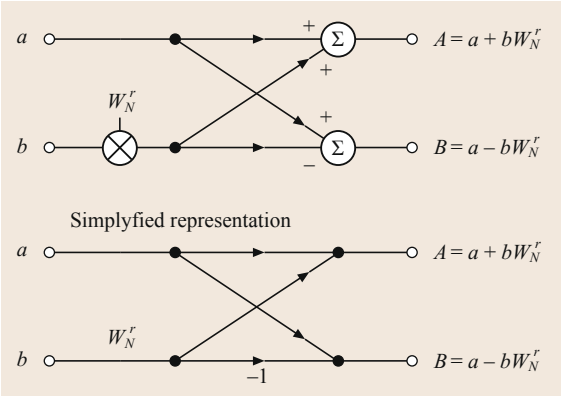


Fig. 9.22 The basic butterfly computation in the DIT-FFT algorithm

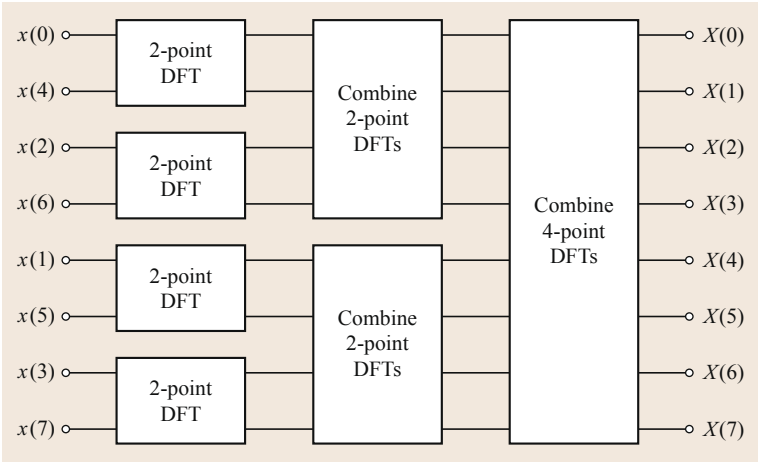


Fig. 9.23 Computation of eight-point FFT in three stages using DIT

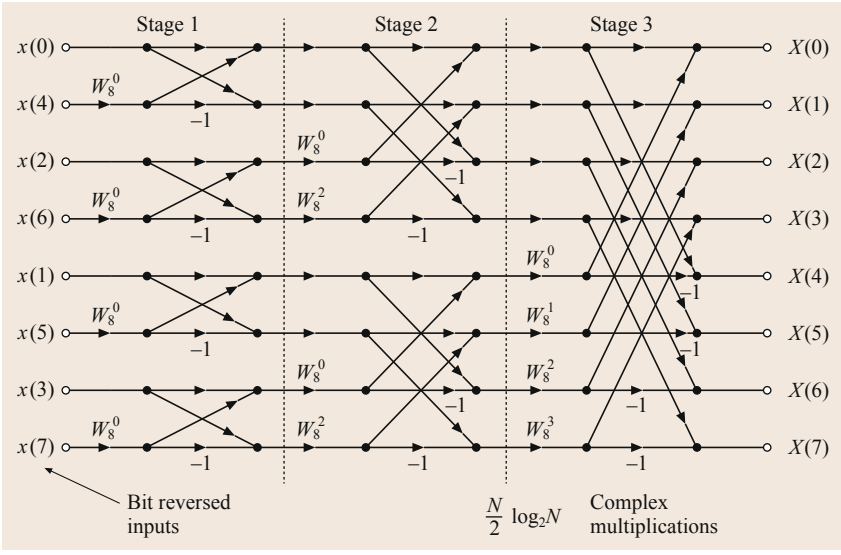


Fig. 9.24 Eight-point DIT-FFT algorithm

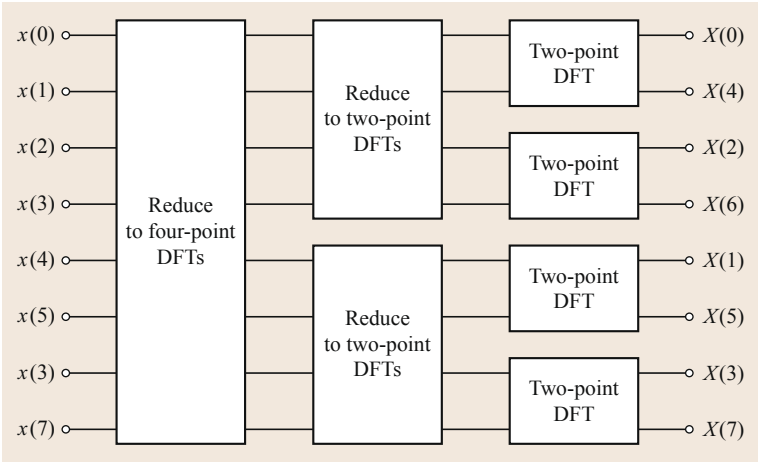


Fig. 9.25 Computation of eight-point FFT in three stages using DIF

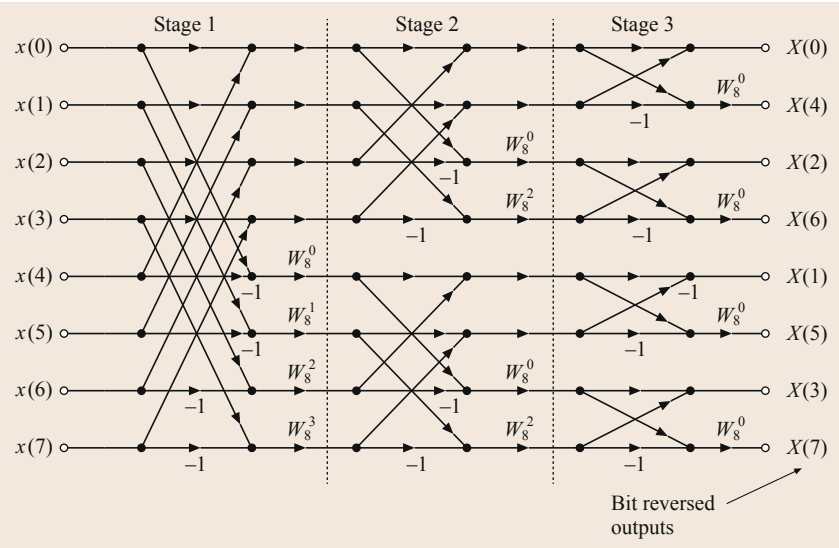


Fig. 9.26 Eight-point DIF-FFT algorithm

The radix-2 FFT algorithm breaks the entire DFT calculation down into a number of two-point DFTs. Each two-point DFT consists of a multiply-and-accumulate operation called a *butterfly*, as shown in Fig. 9.22. Two representations of the butterfly are shown in the diagram: the top diagram is the actual functional representation of the butterfly showing the digital multipliers and adders. In the simplified bottom diagram, the multiplications are indicated by placing the multiplier over an arrow, and addition is indicated whenever two arrows converge at a dot.

The eight-point decimation-in-time (DIT) FFT algorithm computes the final output in three stages as shown in Fig. 9.23. The eight input time samples are first divided (or *decimated*) into four groups of two-point DFTs. The four two-point DFTs are then combined into two four-point DFTs. The two four-point DFTs are then combined to produce the final output $X(k)$. The detailed process is shown in Fig. 9.24, where all the multiplications and additions are shown. Note that the basic two-point DFT butterfly operation forms the basis for all computations. The computation is done in three stages. After the first stage computation is complete, there is no need to store any previous results. The first stage outputs can be stored in the same registers that originally held the time samples $x(n)$. Similarly, when the second stage computation is completed, the results of the first stage computation can be deleted.

In this way, *in-place* computation proceeds to the final stage. Note that in order for the algorithm to work properly, the order of the input time samples, $x(n)$, must be properly re-ordered using a *bit reversal* algorithm.

The *bit reversal* algorithm used to perform this reordering is shown in Table 9.6. The decimal index, n, is converted to its binary equivalent. The binary bits are then placed in reverse order, and converted back to a decimal number. Bit reversing is often performed in DSP hardware in the data address generator (.*), thereby simplifying the software, reducing overhead, and speeding up the computations.

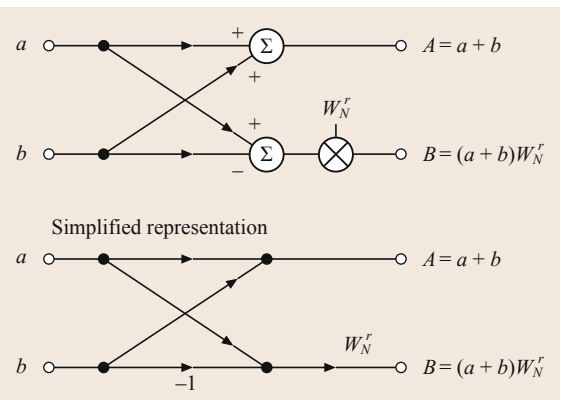


Fig. 9.27 The basic butterfly computation in the DIF-FFT algorithm

Table 9.6 Bit reversal example for $N = 8$

Decimal number:	0	1	2	3	4	5	6	7
Binary equivalent:	000	001	010	011	100	101	110	111
Bit-reversed binary:	000	100	010	110	001	101	011	111
Decimal equivalent:	0	4	2	6	1	5	3	7

The computation of FFT using *decimation-in-frequency* (DIF) is shown in Figs. 9.25 and 9.26. This method requires that the bit reversal algorithm be applied to the output $X(k)$. Note that the butterfly for the DIF algorithm differs slightly from the decimation-in-time butterfly, as shown in Fig. 9.27.

The use of decimation-in-time versus decimation-in-frequency algorithms is largely a matter of preference, as either yields the same result. System constraints may make one of the two a more optimal solution. It should be noted that the algorithms required to compute the inverse FFT are nearly identical to those required to compute the FFT, assuming complex FFTs are used. In fact, a useful method for verifying a complex FFT algorithm consists of first taking the FFT of the $x(n)$ time samples and then taking the inverse FFT of the $X(k)$. At the end of this process, the original time samples, $Re(x(n))$, should be obtained and the imaginary part, $Im(x(n))$, should be zero (within the limits of the mathematical round off errors).

The FFTs discussed up to this point are radix-2 FFTs, i. e., the computations are based on two-point butterflies. This implies that the number of points in the FFT algorithms must be a power of 2. However, non-radix-2 FFT algorithms have been developed and are available in modern computational software packages like Matlab, Mathematica, Maple, etc., to be used in a variety of applications in ocean engineering as well as beyond [9.1, 2, 4–6].

9.4 Waveform Analysis

A waveform is one recording of a deterministic signal or that of an instance of a random signal (also referred to in the literature as a stochastic process). Actually, in practice we can make a recording of finite duration, which means that a waveform is the set of signal recordings over possibly non-contiguous time intervals. Typical examples of (lumped) waveforms in ocean engineering are significant wave height at a point or average, oceanic temperature or salinity, acoustic recordings of marine life, sonar or radar signals, etc. [9.3].

9.4.1 Definitions for Waveforms and Random Signals

The time mean (or average) of a waveform is defined as follows in the continuous time domain

$$\mu_x \triangleq \lim_{T\to\infty}\left\{\frac{1}{2T}\int_{-T}^{+T} x(t)\,\mathrm{d}t\right\} . \tag{9.74}$$

The time-wise cross-correlation of two waveforms is defined as follows

$$\begin{aligned} R_{y,x}(\tau) &\triangleq \lim_{T\to\infty}\left\{\frac{1}{2T}\int_{-T}^{+T} y(t)x(t+\tau)\,\mathrm{d}t\right\} , \\ &= [y \otimes x_{-(t)}](\tau) . \end{aligned} \tag{9.75}$$

In the above, $\otimes$ stands for the convolution operation applied in the square brackets between signal y and signal $x_{(-t)}$, which is the mirrored version of signal x about origin of time ($t = 0$), with time shift τ, i. e.,

$$x_{-(t)}(t) = x(-t) .$$

Also, note that cross-correlation as defined above can be used to introduce (time-wise) autocorrelation of a single waveform if we set $y \equiv x$.

The limiting process to infinity for time variable T introduced in the equations above is required in the case when the waveform has infinite duration so that the definitions make sense and do not give rise to indeterminate or infinite sums.

On the other hand, if the waveform (or at least one of them in the cross-correlation case) is finite in duration, then the time interval ($2T$) needs to be set equal to the full finite, and not infinite, duration of the signal. In this case, the definitions need to look like the following to avoid any issues

$$\mu_x \triangleq \frac{1}{2T}\int_{-\infty}^{+\infty} x(t)\,\mathrm{d}t , \tag{9.76}$$

$$R_{y,x}(\tau) \triangleq \frac{1}{2T}\int_{-\infty}^{+\infty} y(t)x(t+\tau)\,\mathrm{d}t . \tag{9.77}$$

The above equations are well defined in the case of deterministic signals and waveforms that occur as instances of random signals. Specifically, a random signal, denoted with a capital letter, e.g., $X(t)$, in contrast to lowercase latters used for deterministic signals, is associated with a probability density function (PDF), $f_X(x)$ quantifying the probability of the signal value at any

time instant t to lie within an arbitrary real interval $[a, b]$ as follows

$$\Pr[a \le X(t) \le b] = \int_a^b f_X(x)\,\mathrm{d}x\,. \tag{9.78}$$

As well established by probability theory, the probability density function is, in turn, associated with a cumulative distribution function (CDF) for random signal $X(t)$ as follows

$$F_X(x) = \Pr[X(t) \le x] = \int_{-\infty}^{x} f_X(x_\#)\,\mathrm{d}x_\#\,. \tag{9.79}$$

One can easily verify by combining (9.78) with (9.79) that

$$\Pr[a \le X(t) \le b] = F_X(b) - F_X(a) \tag{9.80}$$

holds. Finally and assuming certain smoothness conditions hold, the PDF can be calculated as the derivative of the CDF for the same random variable, i. e.,

$$f_X(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_X(x)\,. \tag{9.81}$$

For a random signal one can, in effect, define a probabilistic mean value as follows

$$\langle X(t)\rangle = \mu_X(t) \triangleq \int_{-\infty}^{+\infty} x f_{X(t)}(x)\,\mathrm{d}x\,. \tag{9.82}$$

Along the same lines a probabilistic correlation (cross and/or autocorrelation) can be introduced

$$\begin{aligned}\langle X(t_1), Y(t_2)\rangle &= R_{X,Y}(t_1, t_2)\,,\\ &\triangleq \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xy f_{X(t_1),Y(t_2)}(x, y)\,\mathrm{d}y\,\mathrm{d}x\,.\end{aligned} \tag{9.83}$$

To properly define it, though, the joint probability density function needs to be introduced; as can be seen below, it quantifies the probability of random signal X assuming a value in an infinitesimal vicinity of x at instant t_1 and, jointly, random signal Y assuming a value in an infinitesimal vicinity of y at instant t_2

$$\begin{aligned}&f_{X(t_1),Y(t_2)}(x, y)\,\mathrm{d}x\,\mathrm{d}y\\ &= \Pr[(x \le X(t_1) \le x + \mathrm{d}x) \text{ and}\\ &\qquad (y \le Y(t_2) \le y + \mathrm{d}y)]\,.\end{aligned} \tag{9.84}$$

In effect, the probability on the right-hand side in the above is that of the joint event to find the values of signals X and Y at certain time instants within an arbitrary infinitesimal rectangle on the XY plane. Applying Bayes' theorem on the joint probability of the right-hand side of (9.84) one can derive the following equations involving conditional probabilities

$$\begin{aligned}&\Pr[x \le X(t_1) \le x + \mathrm{d}x | y \le Y(t_2) \le y + \mathrm{d}y]\\ &\quad \cdot \Pr[y \le Y(t_2) \le y + \mathrm{d}y]\\ &= f_{X(t_1),Y(t_2)}(x, y)\,\mathrm{d}x\,\mathrm{d}y\\ &= \Pr[y \le Y(t_2) \le y + \mathrm{d}y | x \le X(t_1) \le x + \mathrm{d}x]\\ &\quad \cdot \Pr[x \le X(t_1) \le x + \mathrm{d}x]\,.\end{aligned} \tag{9.85}$$

In the case that random variables X and Y are independent (not to be confused with mutually exclusive variables or events), then using (9.85) for the conditional probabilities one can see that the following holds

$$f_{X(t_1),Y(t_2)}(x, y) = f_{X(t_1)}(x) \cdot f_{Y(t_2)}(y) \tag{9.86}$$

A very important category of random signals is that of wide sense stationary (WSS). A WSS signal is one that has: (a) constant, i. e., time invariant, probabilistic mean, and, (b) autocorrelation solely depending on the time shift (delay) $\tau = t_2 - t_1$ and not on the individual time instants. In effect, for a random signal to be WSS it must hold that

$$\langle X(t)\rangle = \langle X\rangle = \mu_X\,, \tag{9.87}$$

$$\langle X(t), X(t+\tau)\rangle = R_{X,X}(\tau)\,. \tag{9.88}$$

In part, WSS signals are important because both the probabilistic mean value and autocorrelation can be obtained as the time-wise mean and autocorrelation of anyone of its instances, $x(t)$, introduced for waveforms earlier. For this to be possible though, the random signal (stochastic process) X needs to be ergodic on top of WSS. Indeed, for an ergodic process statistical (probabilistic) properties (such as its mean, correlation, and variance) can be deduced from a single, sufficiently long instance (sample realization) of the process. Therefore, for an ergodic WSS process, the following hold

$$\langle X(t)\rangle = \mu_{\mathrm{x}}\,, \tag{9.89}$$

$$\langle X(t), X(t+\tau)\rangle = R_{x,x}(\tau)\,. \tag{9.90}$$

It should be noted, however, that for the above equations to hold exactly the integrals over time in (9.74)

and (9.75) need to span the entire real line from minus to plus infinity. If not, then (9.89) and (9.90) provide only an estimate of the mean and autocorrelation values of the random signal; the accuracy of the estimate (probabilistically) increases as time parameter T in the time-wise integrals of (9.74) and (9.75) grow larger.

A final note is made regarding strictly or strongly stationary processes versus WSS ones. A strictly stationary process is a stochastic process whose joint probability distribution does not change when shifted in time. This requirement is much stronger than just mean and autocorrelation and, therefore, much harder to meet or assume in practical situations; that is why in the remaining text we will only employ WSS as well as ergodicity.

9.4.2 Signal Power and Power Spectral Density

For a waveform $x(t)$, i. e., a deterministic signal or an instance of a random one, the following definitions are given:

1. The time mean is the direct current (DC) offset (component) of the waveform and $|\mu_x|^2$ is the DC power of it.
2. Consider

$$R_{x,x}(0) = \lim_{T\to\infty}\left\{\frac{1}{2T}\int_{-T}^{+T} x^2(t)\,\mathrm{d}t\right\} \geq 0\,. \quad (9.91)$$

   This is the (total averaged) power of the waveform.
3. Then consider

$$\mathrm{cov}(x) = R_{x,x}(0) - |\mu_x|^2 \geq 0\,. \quad (9.92)$$

   This is the covariance or alternating current (AC) power of the waveform. Also, $[\mathrm{cov}(x)]^{1/2}$ is the standard deviation or RMS (root-mean-square) value of the waveform.

The use of term *power* here stands for electric power, i. e., that the waveform is considered as voltage across or electric current through a reference resistance equal to 1 ohm.

Finally, the following notation for cross-correlation (for two waveforms) or autocorrelation (for a single waveform)

$$\langle y, x\rangle = R_{y,x}(0) \triangleq \lim_{T\to\infty}\left\{\frac{1}{2T}\int_{-T}^{+T} y(t)x(t)\,\mathrm{d}t\right\}\,. \quad (9.93)$$

In the case of two distinct waveforms $\langle y, x\rangle$ is also identified as their dot product. In the case of a single waveform $\langle x, x\rangle$ is its power.

In the frequency domain and using the two-sided Fourier transform as introduced in (9.8), the (cross) power spectral density (PSD) can be defined for a couple of waveforms as follows

$$S_{y,x}(\omega) \triangleq \mathcal{F}\{R_{y,x}(\tau)\} = Y(\omega)X^*(\omega)\,. \quad (9.94)$$

Note that for a real waveform it holds that

$$X(\omega) = \mathcal{F}\{x(t)\} = \int_{-\infty}^{+\infty} x(t)\exp(-i\omega t)\,\mathrm{d}t$$

$$\Updownarrow$$

$$X^*(\omega) = \left(\int_{-\infty}^{+\infty} x(t)\exp(-i\omega t)\,\mathrm{d}t\right)^* = \int_{-\infty}^{+\infty} x^*(t)\exp(+i\omega t)\,\mathrm{d}t = X(-\omega)\,.$$

Furthermore

$$\mathcal{F}\{x(-t)\} = \int_{-\infty}^{+\infty} x(-t)\exp(-i\omega(-t))\,\mathrm{d}(-t) = \int_{-\infty}^{+\infty} x(t)\exp(i\omega t)\,\mathrm{d}t = X(-\omega)\,.$$

Therefore, for a real waveform it holds that

$$X^*(\omega) = \mathcal{F}\{x(-t)\} = X(-\omega)\,. \quad (9.95)$$

Then, using (9.94) one can directly derive the following

$$S_{x,y}(\omega) = S_{y,x}(-\omega) = S^*_{y,x}(\omega) \Leftrightarrow R_{x,y}(\tau) = R_{y,x}(-\tau)\,. \quad (9.96)$$

Moreover, in the case of autocorrelation PSD obtains the following form

$$S_{x,x}(\omega) \triangleq \mathcal{F}\{R_{x,x}(\tau)\} = |X(\omega)|^2 \geq 0 \quad \text{and} \quad S_{x,x}(\omega) = S_{x,x}(-\omega)\,. \quad (9.97)$$

In effect, the PSD of a single signal is an even, non-negative complex function of frequency. Also, for the autocorrelation it holds that it is an even function of time τ, i. e., $R_{x,x}(\tau) = R_{x,x}(-\tau)$.

A final note has to do with the extension of the PSD concept to random signals (stochastic processes in mathematical terminology). Assuming ergodicity and WSS, the Wiener–Khinchin theorem can be employed to introduce a PSD for a random signal X or at least a power spectral distribution function $F_X(f)$ as follows

$$\begin{aligned}\langle X(t), X(t+\tau)\rangle &= R_{X,X}(\tau) \\ &= \int_{-\infty}^{+\infty} e^{i\omega\tau}\,\mathrm{d}F_X(f), \omega = 2\pi f\,.\end{aligned} \tag{9.98}$$

For the above to hold, the autocorrelation of X needs to exist and be finite for any value of time shift variable τ, but its Fourier transform may not be well defined. Indeed, the Fourier transform of a random signal may not exist in general, because stationary stochastic processes may not generally be square or absolutely integrable. Nor does their autocorrelation need to be absolutely integrable, so it need not have a Fourier transform, either. However, in most practical applications the autocorrelation is integrable and even satisfies the conditions to obtain a Fourier transform. In this case, the PSD can be introduced through the Fourier transform as follows

$$\begin{aligned}S_{X,X}(\omega) &= \int_{-\infty}^{+\infty} R_{X,X}(\tau)e^{-i\omega\tau}\,\mathrm{d}\tau \Leftrightarrow R_{X,X}(\tau) \\ &= \int_{-\infty}^{+\infty} S_{X,X}(f)e^{i\omega\tau}\,\mathrm{d}f, \omega = 2\pi f\,.\end{aligned} \tag{9.99}$$

Finally, in this case the PSD is the averaged derivative of the power spectral distribution function introduced previously. This is why the latter is also referred to as the integrated spectrum of the stochastic process.

9.4.3 Waveform Propagation Through a Linear, Time-Invariant System

Consider a linear, time-invariant (LTI) system with a single input and a single output (SISO), as shown in Fig. 9.28, driven by input waveform $x(t)$ and generating as response output waveform $y(t)$.

The system is assumed to be in continuous time, as well as the waveforms at its input and its output. Then, the output can be determined as the convolution ($\otimes$) of

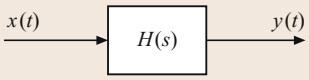


Fig. 9.28 LTI-SISO system with waveforms at its input and its output

the LTI system's scalar impulse response with the input waveform

$$\begin{aligned}y(t) = h(t)\otimes x(t) &= \int_{-\infty}^{+\infty} h(t_\#)x(t-t_\#)\,\mathrm{d}t_\# \\ &\Leftrightarrow Y(\omega) = H(\omega)\cdot X(\omega)\,.\end{aligned} \tag{9.100}$$

In the above, $h(t)$ stands for the system's impulse response that can be determined on the basis of the system's transfer function $H(s)$ in the complex frequency (Laplace transform) domain

$$h(t) = \mathcal{L}^{-1}\{H(s)\}\,. \tag{9.101}$$

It is noted here that for a causal system, the convolutional integral must be finitely bounded as follows

$$\begin{aligned}y(t) = h(t)\otimes x(t) &= \int_{0}^{t} h(t_\#)x(t-t_\#)\,\mathrm{d}t_\# \\ \text{and}\quad h(t) &= 0, \forall t < 0\,.\end{aligned} \tag{9.102}$$

For the means of the input and output waveforms the following can be derived

$$\begin{aligned}\mu_y &\triangleq \lim_{T\to\infty}\left\{\frac{1}{2T}\int_{-T}^{+T} y(t)\,\mathrm{d}t\right\} \\ &= \mu_x\cdot\int_{-\infty}^{+\infty} h(t)\,\mathrm{d}t = H(0)\cdot\mu_x\,.\end{aligned} \tag{9.103}$$

For the various correlations, as well as the PSD of the input and output, the following hold

$$\begin{aligned}R_{y,y}(t) &= R_{h,h}\otimes R_{x,x} \\ &\Leftrightarrow S_{y,y}(\omega) = |H(\omega)|^2\cdot S_{x,x}(\omega)\,,\end{aligned} \tag{9.104}$$

$$\begin{aligned}R_{y,x}(t) &= h\otimes R_{x,x} \\ &\Leftrightarrow S_{y,x}(\omega) = H(\omega)\cdot S_{x,x}(\omega)\,,\end{aligned} \tag{9.105}$$

$$R_{x,y}(t) = h_{-(t)} \otimes R_{x,x}$$
$$\Leftrightarrow S_{x,y}(\omega) = H^*(\omega) \cdot S_{x,x}(\omega) \,. \qquad (9.106)$$

All of the above can be derived on the basis of the definitions and properties given previously, as well as the commutativity and associativity of scalar convolution. For example, in the case of (9.104)

$$\begin{aligned} R_{y,y}(t) &= y \otimes y_{-(t)} \\ &= [h \otimes x] \otimes [h_{-(t)} \otimes x_{-(t)}] \\ &= R_{h,h} \otimes R_{x,x} \,. \end{aligned} \qquad (9.107)$$

It is noted here that:

1.

$$S_{h,h}(\omega) = H(\omega)H^*(\omega) = |H(\omega)|^2 \,,$$

2.

$$\begin{aligned} R_{h,h}(\tau) &= \int_{-\infty}^{+\infty} S_{h,h}(f) e^{i\omega\tau} \, \mathrm{d}f \\ &= \int_{-\infty}^{+\infty} |H(\omega)|^2 e^{i\omega\tau} \, \mathrm{d}f, \omega = 2\pi f \,, \end{aligned}$$

and

3.

$$\begin{pmatrix} x_{-(t)} \\ y_{-(t)} \\ h_{-(t)} \end{pmatrix} (t) = \begin{pmatrix} x \\ y \\ h \end{pmatrix} (-t), \forall t \in \mathbb{R} \Leftrightarrow \mathcal{F} \left\{ \begin{matrix} x_{-(t)} \\ y_{-(t)}(t) \\ h_{-(t)} \end{matrix} \right\}$$
$$= \begin{pmatrix} X^* \\ Y^* \\ H^* \end{pmatrix} (\omega \text{ or } f) \,.$$

9.5 Optimal Signal Estimation

9.5.1 System Identification

One of the most fundamental problems in science and engineering is the determination of a system's dynamics, preferably in the form of a mathematical model, when presented with the system response(s) to given deterministic or random driving inputs. This is the fundamental system identification problem [9.7, 8].

In the case of an LTI system, especially in the case of discrete-time or sampled-data systems, system identification is essentially equivalent to a problem of minimum square error approximation. Consider the problem statement as shown in Fig. 9.29.

A fundamental prerequisite for the system identification problem to have a solution in its basic form is that the additive measurement noise superimposed to the system output is of zero mean and uncorrelated to the driving forcing signal applied as input to the system. Therefore, the following two conditions must hold

$$\mu_v \triangleq \lim_{T\to\infty} \left\{ \frac{1}{2T} \int_{-T}^{+T} v(t) \, \mathrm{d}t \right\} = 0 \,, \qquad (9.108)$$

$$R_{vx}(\tau) = R_{xv}(\tau) = 0, \forall \tau$$
$$\Leftrightarrow S_{vx}(\omega) = S_{xv}(\omega) = 0, \forall \omega \,. \qquad (9.109)$$

At this point it is important to emphasize the importance of the definition of the appropriate time window T in (9.75). Indeed, it is possible that for some (commonly small) value of T condition (9.109) may not be satisfied with sufficient accuracy. However, assuming that condition (9.109) is satisfied for an infinite time window as well as appropriate stationarity and ergodicity assumptions, it is possible to determine a finite value for T such that condition (9.109) is met at arbitrarily small (*epsilon*) accuracy.

Given the property in (9.109) and by use of the linearity property of both cross-correlation and PSD, as well as (9.105), one can derive the following

$$S_{yx}(\omega) = S_{ux}(\omega) + S_{vx}(\omega) = H(\omega) \cdot S_{xx}(\omega) \,. \qquad (9.110)$$

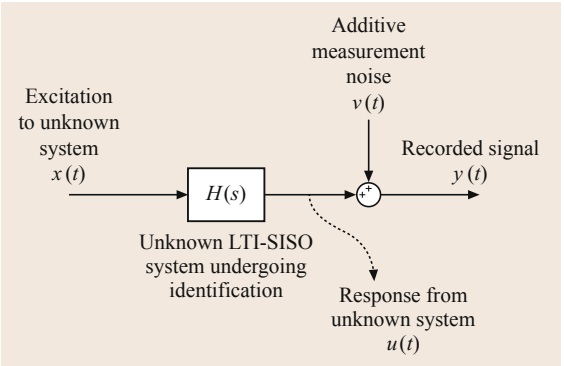


Fig. 9.29 LTI-SISO system identification problem

The above holds since $y(t) = u(t) + v(t)$ and $u(t) = h(t) \otimes x(t) \Leftrightarrow U(\omega) = H(\omega)X(\omega)$. In effect, the unknown transfer functions of the intermitted LTI system

$$H(\omega) = \frac{S_{yx}(\omega)}{S_{xx}(\omega)} \,. \tag{9.111}$$

In the time domain the equivalent equation to (9.111) is

$$R_{yx}(t) = h(t) \otimes R_{xx}(t) = \int_{-\infty}^{+\infty} h(t_\#) R_{xx}(t - t_\#)\, \mathrm{d}t_\# \,. \tag{9.112}$$

In the above equation, the unknown signal (function) is the impulse, $h(t)$. The integral equation is known as the Wiener–Hopf equation with infinite time horizon [9.1, 2, 4–6].

What is most important in both (9.111) and (9.112) is that system identification using them can be implemented with either deterministic or random signals. In effect, signals, $x(t)$, $v(t)$ and $y(t)$ may be known only probabilistically (statistically); this means that the individual instances (realizations) of the signals do not need to be known as long as their statistics are, e.g., means, correlations, or covariance coefficients.

Typically in practice, statistical characterization of signals can be effectively, yet approximately, performed by taking advantage of ergodicity and stationarity, if they apply. Assuming that these hold, then a sufficiently long, yet finite, recording of a single realization of a random signal allows for its statistical characterization in terms of its autocorrelation and cross-correlation with other signals. Such a recording, e.g., for $v(t)$ or $x(t)$ in the case of our system identification framework, can be performed once and be used along with measurements for $y(t)$ to identify the system; all due to ergodicity and stationarity.

9.5.2 Discrete-Time Wiener–Hopf Equation over a Finite-Duration Window

Integral (9.112), despite its immense theoretical value, does not really indicate how it can be applied in practice for system identification. In this passage, we will introduce a methodology allowing identification of an LTI-SISO system when presented only with sampled-data recordings of $x(t)$ and $y(t)$, with a sampling rate meeting the Nyquist requirement for both. Furthermore, it will be assumed that the observation window (time horizon) of the recordings of the forcing and the noisy response is finite in duration instead of infinite.

The outcome of the processing will be the determination of a given finite number of impulse response samples for the unknown system undergoing identification. In effect, it is possible to define a discrete-system FIR (finite impulse response) system in either the time domain, by its very impulse response, or as a transfer function in the Z-transform domain without any poles. In effect, the FIR system will approximate the behavior of the unknown LTI system undergoing identification in the sense of (truncated) impulse response matching.

We start from the integral equation for $y(t)$ given previously

$$\begin{aligned} y(t) &= h(t) \otimes x(t) + v(t) \\ &= \int_{-\infty}^{+\infty} h(t_\#) x(t - t_\#)\, \mathrm{d}t_\# + v(t) \,. \end{aligned} \tag{9.113}$$

However, to satisfy causality for the unknown system the following needs to be introduced

$$\begin{aligned} y(t) &= \int_0^t h(t_\#) x(t - t_\#)\, \mathrm{d}t_\# \\ +v(t) &= \int_0^t h(t - t_\#) x(t_\#)\, \mathrm{d}t_\# + v(t) \,. \end{aligned} \tag{9.114}$$

Converting the above to discrete time by sampling interval T_s, compliant to the Nyquist rate criterion, one can derive the following

$$y(nT_\mathrm{s}) = T_\mathrm{s} \sum_{k=0}^{n} h(kT_\mathrm{s}) x((n-k)T_\mathrm{s}) + v(nT_\mathrm{s}) \,. \tag{9.115}$$

Without loss of generality we can assume that $T_\mathrm{s} = 1$, since any other sampling interval value can be absorbed multiplicatively in the impulse response sample values.

$$y(n) = \sum_{k=0}^{n} h(k) x(n-k) + v(n) \,. \tag{9.116}$$

An important prerequisite for our identification method to be successful is that the unknown LTI system is stable. If this holds, then for an arbitrarily small ε one can find a positive integer M, such that

$$\sum_{n=M+1}^{+\infty} h^2(n) < \varepsilon \,. \tag{9.117}$$

In effect, with epsilon accuracy it is possible to truncate the impulse response of the unknown system. We note here that impulse response truncation is an approximation of the system undergoing identification by the

following FIR one with memory M

$$\hat{H}_{\mathrm{FIR}}(z^{-1}) = \sum_{n=0}^{M} h(n) z^{-n} \,. \tag{9.118}$$

Using the approximation above, (9.116) yields

$$y(n) = \sum_{k=0}^{M} h(k) x(n-k) + v(n) \,. \tag{9.119}$$

Furthermore, assuming that the recording for waveform $y(t)$ spans a finite interval $[0, T]$, the identification problem can be reduced to determining the $(M+1)$ samples of the impulse response, $h(n)$, $0 \le n \le M$, of the unknown system so that the following $(L+1)$ equations are satisfied

$$y(n) = \sum_{k=0}^{M} h(k) x(n-k) + v(n), 0 \le n \le L$$
$$\text{and} \quad L = \frac{T}{T_\mathrm{s}} \,. \tag{9.120}$$

It is noted here that containing the recording of signal $y(t)$ within interval $[0, T]$ allows limiting the need to record signal $x(t)$ also in the finite interval $[-MT_\mathrm{s}, +LT_\mathrm{s}]$.

In effect, the $((L+1)\times(M+1))$ set of algebraic (9.120) can be put in the following matrix form

$$\boldsymbol{y} = \ddot{X} \cdot \boldsymbol{h} + \boldsymbol{v} \,. \tag{9.121}$$

In the above, vectors y and v are defined as follows

$$\boldsymbol{y} = [y(0) \quad y(1) \quad \cdots \quad y(L)]^{\mathrm{T}}$$
$$\boldsymbol{v} = [v(0) \quad v(1) \quad \cdots \quad v(L)]^{\mathrm{T}} \,. \tag{9.122}$$

Vector h is defined as follows

$$\boldsymbol{h} = [h(0) \quad h(1) \quad \cdots \quad h(M)]^{\mathrm{T}} \,. \tag{9.123}$$

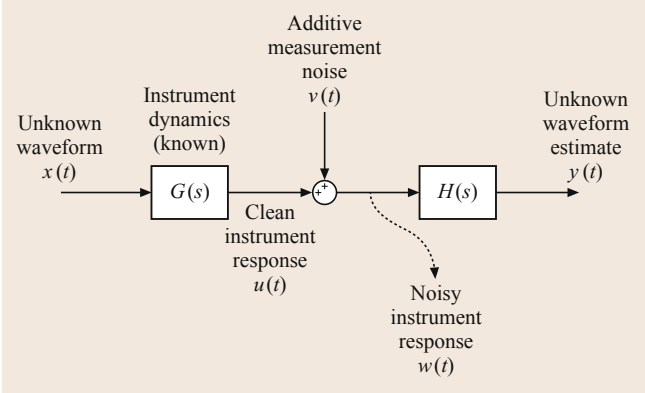


Fig. 9.30 Signal estimation framework

Matrix $\mathbf{X}$, known as observation matrix, is defined by the following

$$\mathbf{X} = \begin{bmatrix} x(0) & x(-1) & \cdots & x(-M) \\ x(1) & x(0) & \cdots & x(-M+1) \\ \vdots & \vdots & \ddots & \vdots \\ x(L) & x(L-1) & \cdots & x(-M+L) \end{bmatrix} \,. \tag{9.124}$$

The equation set in (9.121) is not square; therefore, an algebraically exact solution is not feasible. In effect, only a solution minimizing the following mean square error (MSE) vector spanning the observation window

$$\|e\|_2 = \|y - X \cdot h\|_2 \,. \tag{9.125}$$

If the above MSE vector is minimized, then the effect of the noise vector v, i. e., signal $v(t)$, is removed from the measurement vector y, i. e., signal $y(t)$. The Euclidean norm of vector e is defined as follows

$$\|e\|_2 = e^{\mathrm{T}} \cdot e = \sum_{i=0}^{L} e_i^2 \,. \tag{9.126}$$

The solution achieving minimization of $\|e\|_2$ can be proven to be that of the following square $((M+1)\times(M+1))$ set of algebraic *canonical equations*

$$(\mathbf{X}^T \cdot \mathbf{X}) \cdot h = \mathbf{X}^T \cdot y \Rightarrow h = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T y \,. \tag{9.127}$$

It is interesting to observe the equivalence between the time-domain (9.127) and (9.112); the equivalence can be straightforwardly derived if one observes that (9.112) needs to hold for any time instant t. This means that if it were converted to discrete time a (countable infinite in number) set of algebraic equations would be derived like in (9.121). Furthermore, the solution of canonical (9.127) is fully equivalent to (9.111), since there exists an one-to-one correspondence between: (a) square matrix $(\mathbf{X}^{\mathrm{T}}\mathbf{X})$ and autocorrelation $R_{xx}(t)$, and thereof PSD $S_{xx}(\omega)$; (b) vector $(\mathbf{X}^{\mathrm{T}}y)$ and cross-correlation $R_{yx}(t)$, and thereof PSD $S_{yx}(\omega)$.

On a technical note, it is pointed out that for (9.127) to hold, matrix $(\mathbf{X}^{\mathrm{T}}\mathbf{X})$ has to be invertible. In the case of zero initial conditions for x, i. e., $x(n) = 0, -M \le n < 0$, it must hold that $x(0) \neq 0$ so that $(\mathbf{X}^{\mathrm{T}}\mathbf{X})$ is, indeed, invertible.

9.5.3 Signal Estimation and the Wiener Filter

We will now look into a closely related problem: that of the estimation of a signal or waveform which is outlined in the block diagram of Fig. 9.30.

As can be seen a known, in this case, the LTI-SISO system is driven by an unknown (to be estimated) $x(t)$. or the known LTI system, the transfer function $G(s)$, in the complex frequency (Laplace) domain, or $G(\omega)$, in the Fourier (frequency) domain, or equivalently its impulse response $g(t)$ in the time domain is assumed known. Such an LTI system may model a measurement instrument, or sensor, or transducer, or telecommunications receiver.

Then, a linear filter with transfer function $H(s)$ is sought that will receive as input the noisy instrument response $w(t)$ and output waveform $y(t)$, which is expected to be the optimum, in the mean square sense, estimate of the unknown waveform $x(t)$.

The objective is to determine transfer function $H(s)$, or equivalently $H(\omega)$, of the estimation filter that minimizes the following *objective cost function*

$$\langle e, e\rangle \triangleq R_{ee}(0) = \int_{-\infty}^{+\infty} e^2(t)\,\mathrm{d}t\,. \tag{9.128}$$

In the above, the instantaneous error signal is defined as follows

$$e(t) = x(t) - y(t)\,. \tag{9.129}$$

For waveform $y(t)$ it holds that

$$\begin{aligned} y(t) = h(t) \otimes w(t) &= \int_{-\infty}^{+\infty} h(\xi)\, w(t-\xi)\,\mathrm{d}\xi \\ &\approx T_\mathrm{s} \sum_{k=-\infty}^{+\infty} h(kT_\mathrm{s})\, w(t-kT_\mathrm{s})\,. \end{aligned} \tag{9.130}$$

In the above, the convolutional integral has been approximated by the convolutional sum, assuming that sampling interval, T_s, satisfies the Nyquist criterion. Consequently, (9.128) becomes

$$\langle e, e\rangle = \int_{-\infty}^{+\infty} \left\{ x(t) - T_\mathrm{s} \sum_{k=-\infty}^{+\infty} h(kT_\mathrm{s})\, w(t-kT_\mathrm{s}) \right\}^2 \mathrm{d}t\,. \tag{9.131}$$

In effect, the impulse response of the least-squares estimator has to fulfill the following condition, so that cost $\langle e, e\rangle$ becomes minimum

$$\frac{\partial}{\partial[h(nT_\mathrm{s})]}\langle e, e\rangle = 0,\ \forall n \in Z\,. \tag{9.132}$$

However,

$$\begin{aligned} &\frac{\partial}{\partial[h(nT_\mathrm{s})]}\langle e, e\rangle \\ &= -2 \int_{-\infty}^{+\infty} \left\{ x(t) - T_\mathrm{s} \sum_{k=-\infty}^{+\infty} h(kT_\mathrm{s})\, w(t-kT_\mathrm{s}) \right\} \\ &\quad \times w(t-nT_\mathrm{s})\,\mathrm{d}t\,. \end{aligned} \tag{9.133}$$

Therefore, condition (9.132) finally yields

$$\int_{-\infty}^{+\infty} e(t)\, w(t-nT_\mathrm{s})\,\mathrm{d}t = 0,\ \forall n \in \mathbb{Z}\,. \tag{9.134}$$

Setting $\tau = nT_\mathrm{s}$ in the above and gradually diminishing T_s, we end up back in the continuous time domain, and (9.134) yields the following *orthogonality condition*

$$\begin{aligned} &\int_{-\infty}^{+\infty} e(t)\, w(t-\tau)\,\mathrm{d}t = 0,\ \forall \tau \in \mathbb{R} \\ &\qquad\qquad \Leftrightarrow R_{we}(\tau) = 0,\ \forall \tau \in \mathbb{R}\,. \end{aligned} \tag{9.135}$$

Using the above orthogonality condition

$$R_{we}(\tau) = 0 \Leftrightarrow S_{we}(\omega) = 0 \Leftrightarrow S_{xw}(\omega) = S_{yw}(\omega)\,. \tag{9.136}$$

Furthermore, since $S_{yw}(\omega) = H(\omega)S_{ww}(\omega)$, the following equation can be derived for the transfer function of the estimation filter

$$H(\omega) = \frac{S_{xw}(\omega)}{S_{ww}(\omega)}\,. \tag{9.137}$$

The above expression for the estimator is commonly known as the Wiener filter and of is of widespread use in many science and engineering application fields.

The signal power of the minimum square error, achieved by the Wiener filter in (9.137), can be determined as follows

$$\begin{aligned} S_{ee}(\omega) &= S_{xe}(\omega) - \underbrace{S_{ye}(\omega)}_{0} = S_{xx}(\omega) - S_{xy}(\omega) \\ &= S_{xx}(\omega) - \left(S_{yy}(\omega) + \underbrace{S_{ey}(\omega)}_{0} \right) \\ &\Downarrow \\ S_{ee}(\omega) &= S_{xx}(\omega) - S_{yy}(\omega)\,. \end{aligned} \tag{9.138}$$

For the derivation of the above it is reminded that

$$e = x - y \,.$$

Also, using the orthogonality condition (9.135), the following important result can be derived

$$S_{ye}(\omega) = H(\omega) S_{we}(\omega) \overset{S_{we}=0}{\Rightarrow} S_{ye}(\omega) = 0 = S_{ey}(\omega) \,.$$

It is noted here that

$$R_{ee}(\tau) = \mathcal{F}^{-1}\{S_{ee}(\omega)\} = \int_{-\infty}^{+\infty} S_{ee}(\omega) e^{\mathrm{i}\omega\tau}\, \mathrm{d}\omega \Rightarrow \langle e, e \rangle$$
$$= R_{ee}(0) = \int_{-\infty}^{+\infty} S_{ee}(\omega)\, \mathrm{d}\omega \,.$$

An alternative expression for the minimum error PSD can be derived on the basis of the following

$$S_{yy}(\omega) = |H(\omega)|^2 S_{ww}(\omega) = \frac{|S_{xw}(\omega)|^2}{S_{ww}^2(\omega)} S_{ww}(\omega)$$
$$= \frac{|S_{xw}(\omega)|^2}{S_{ww}(\omega)} \,.$$

Then

$$S_{ee}(\omega) = S_{xx}(\omega) - S_{yy}(\omega)$$
$$= \frac{S_{xx}(\omega) S_{ww}(\omega) - |S_{xw}(\omega)|^2}{S_{ww}(\omega)} \,. \qquad (9.139)$$

Further treatment of (9.137) is possible since the additive measurement noise is uncorrelated to the unknown signal $x(t)$ at the instrument input. In this case,

$$S_{xw}(\omega) = S_{xu}(\omega) + S_{xv}(\omega)$$
$$= G^*(\omega) S_{xx}(\omega) \quad \text{because} \quad S_{xv}(\omega) = 0 \,. \qquad (9.140)$$

Also

$$S_{ww}(\omega) = S_{uu}(\omega) + S_{vv}(\omega)$$
$$= |G(\omega)|^2 S_{xx}(\omega) + S_{vv}(\omega)$$
$$\text{since } u = v + w \,. \qquad (9.141)$$

Therefore, the Wiener filter in (9.137) becomes as follows

$$H(\omega) = \frac{G^*(\omega) S_{xx}(\omega)}{|G(\omega)|^2 S_{xx}(\omega) + S_{vv}(\omega)} \,. \qquad (9.142)$$

The above allows us to obtain an estimate of waveform $x(t)$ appearing at the instrument input port. As can be seen, to achieve this the signal's (waveform) statistics need to be known in advance, i. e., the signal PSD $S_{xx}(\omega)$ or autocorrelation $R_{xx}(\tau)$, as well as the dynamics of the measurement instrument, i. e., its transfer function $G(\omega)$ or its impulse response $g(t)$; last, but not least, prior information of the additive noise's statistics, ie, its PSD $S_{vv}(\omega)$ or autocorrelation $R_{vv}(\tau)$.

Two important specific subcases of (9.137) are given below:

1. *Zero-forcing equalization* [9.1, 5]:
   If $S_{vv}(\omega) \ll S_{xx}(\omega)$

$$H(\omega) = \frac{1}{G(\omega)} \,. \qquad (9.143)$$

   Evidently, in this subcase the statistics of either the unknown input signal or the additive noise need to be known in advance. Furthermore, one can easily verify that, if the inverse of the instrument's transfer function $G(\omega)$ does not introduce unstable poles, estimation error e vanishes.
2. *Matched filter* [9.1, 5]:
   If $|G(\omega)|^2 S_{xx}(\omega) \ll S_{vv}(\omega)$

$$H(\omega) = G^*(\omega) \frac{S_{xx}(\omega)}{S_{vv}(\omega)} \,. \qquad (9.144)$$

   In this subcase the major impairment toward obtaining an estimate of $x(t)$ is the additive noise rather than the instrument distortion, as was the case previously. Moreover note that if $S_{xx}(\omega) < S_{vv}(\omega)$, the instrument's output $w(t)$ will be noise-like, and yet by use of the Wiener filter it is possible to extract the unknown *information* signal $x(t)$.

Matched filter estimator design is widely used in communications engineering. At least as a first approximation, transfer function $G(\omega)$ models the telecommunication channel that introduces attenuation monotonically increasing with distance from the transmitter source. In a typical system, both the information signal $x(t)$, as well as disturbance noise $v(t)$ can be considered as *white noise*, i. e., signals with the following property

$$S_{nn}(\omega) = N_0, \forall\omega \Leftrightarrow R_{nn}(\tau) = N_0 \delta(\tau), \forall\tau \,. \qquad (9.145)$$

As can be seen, the case of a white noise signal demonstrates practically no predictability, since the value of the signal at any time instant is entirely uncorrelated

with its value at any other time instant in the past or in the future. Therefore, if both the information signal and the additive noise are modeled as white noise, (9.144) becomes

$$H(\omega) = \frac{X_0}{V_0} G^*(\omega) \,. \tag{9.146}$$

The ratio (X_0/V_0) with the PSD of the information signal $x(t)$ in the numerator and that of the additive noise in the denominator is known as signal-to-noise ratio (SNR) and is a very important parameter to characterize a system's sensitivity to external noise, as well as exogenous disturbance and crosstalk interference [9.1, 5].

9.6 Concluding Remarks

In this chapter, an account of digital signal processing concepts, methods and techniques employed in ocean engineering is given. After looking into the fundamental processes of continuous signal sampling and reconstruction, the Z-transform is introduced as the tool of convenience to analyze difference equations just like the Laplace transform is suited to analyze differential equations. Digital filters are presented then as well as the Fast Fourier Transform algorithm. Fundamentals of waveform analysis and stochastic processes are presented last as employed for system identification and signal estimation.

References

9.1 J.G. Proakis: *Digital Communications*, 4th edn. (McGraw-Hill, New York 2000)
9.2 J.G. Proakis, D. Manolakis: *Digital Signal Processing: Principles, Algorithms and Applications*, 3rd edn. (Prentice Hall, Upper Saddle River 1995)
9.3 W.A. Kuperman, J.F. Lynch: Shallow-water acoustics, Phys. Today **57**(10), 55 (2004)
9.4 L. Ljung: *System Identification: Theory for the User*, 2nd edn. (Prentice Hall, Upper Saddle River 1999)
9.5 H.V. Poor, G.W. Wornell: *Wireless Communications: Signal Processing Perspectives* (Prentice Hall, Upper Saddle River 1998)
9.6 P.S.R. Diniz: *Adaptive Filtering: Algorithms and Practical Implementation*, 2nd edn. (Springer, Berlin, Heidelberg 2002)
9.7 T. Kohonen: *Self Organization and Associative Memory*, 3rd edn. (Springer Verlag, Berlin, Heidelberg 1989)
9.8 R.E. Schapire: *The Design and Analysis of Efficient Learning Algorithms* (MIT Press, Cambridge 1992)

10. Control Theory and Applications

Nikolaos I. Xiros, Pak-Cheung Edgar An

In this chapter, a presentation of control theory and engineering as applied to ocean engineering is given. The chapter starts with the fundamentals of systems science and theory, i. e., descriptions of systems in state space and for linear, time-invariant ones in the frequency domain using the Laplace transform as well as with ordinary differential equations with respect to time. Then stability, controllability, and observability with an emphasis to linear, time-invariant systems are presented. Bode plots for sinusoidal steady-state analysis as well as the root locus technique for proportional gain feedback design are presented. For single-input, single-output systems, PID control is introduced as both a pole placement problem as well as in the framework of conventional Ziegler–Nichols methods. Pole placement design with the addition of Luenberger observers is presented for linear, time-invariant systems with any time of inputs and outputs. A brief presentation of digital controller implementations is given. Applications from ocean engineering include control of autonomous underwater vehicles and autopilots for surface vessels.

10.1 **System Theory** 227
10.1.1 Definitions and Fundamentals ... 227
10.1.2 The Laplace Transform 229
10.1.3 Linear Time-Invariant Systems.... 230
10.1.4 Multivariable Systems and State Space 233
10.1.5 Nonlinear Systems and Linearization 235
10.2 **Analysis of LTI Systems** 237
10.2.1 Block Diagrams 237
10.2.2 Stability..... 238
10.2.3 Controllability and Observability. 240
10.2.4 Sinusoidal Steady-State Response and Bode Plots..... 241
10.2.5 Analysis of Second-Order Systems 244
10.3 **SISO System Controls** 247
10.3.1 Performance Criteria 247
10.3.2 ON/OFF Control 248
10.3.3 PID Control..... 249
10.3.4 Ziegler–Nichols' Methods for PID Controller Tuning 254
10.3.5 Digital Controller Implementation..... 256
10.3.6 The Root Locus Technique 259
10.4 **Pole Placement of LTI Systems** 261
10.4.1 Input–Output Decoupling 261
10.4.2 Full-State Feedback..... 264
10.4.3 Pole Placement Design..... 264
10.5 **Course-Keeping Autopilots**..... 267
10.5.1 The Vessel in the Control Loop 267
10.5.2 Surface Vessel State-Space Model 269
10.5.3 PID Autopilots 271
10.5.4 State Observers and Use in Autopilots..... 274
References..... 275

10.1 System Theory

10.1.1 Definitions and Fundamentals

Signal

A signal is any function that contains time t as an independent variable. Typical examples of signals are the position and velocity of a moving body, for example, a ship. The symbols used for signals are commonly lower case letters of the Greek or Latin alphabet. In classical control theory signals are only functions of time. However, there has been recently a growing interest for distributed signals, that is, signals that contain more than one independent variables. For example, a beam supported at one end is described by signals that depend on both time and the distance x from the support point.

System

A confined space that accepts input signals (or excitations) and generates output signals (or responses) is

a system. Typical examples of engineering systems include electric networks, internal combustion engines, turbines, etc. Commonly, a system can be subdivided to subsystems with respect to its complexity. A marine engine, for example, includes its turbocharger, which in turn includes the turbine and the air compressor, the cylinders, the crankshaft. etc. All these subsystems may be regarded as systems with their own inputs and outputs. However, the engine is also a subsystem of a higher complexity system, that is, the propulsion plant or the vessel.

Recently the concept of system has been generalized in order to include, except technological or physical systems, biological systems, economical systems, administrative systems, etc. In correspondence the concept of signals has been expanded in order to include inputs and outputs to such systems.

Automatic Control and Feedback

The concept of control is directly related to that of *feedback* [10.1–3]. Specifically, Wiener defines automatic control as the method of controlling a system by applying as inputs the system's past responses. A similar description is given by the IEEE (Institute of Electrical and Electronics Engineers): An automatic control system, such as the cruise control system in automobiles, implements in the controller a decision process, also called the control law, that dictates the appropriate control actions to be taken for the speed to be maintained within acceptable tolerances. These decisions are taken based on how different the actual speed is from the desired, called the error, and on the knowledge of the car's response to fuel increase and decrease. This knowledge is typically captured in a mathematical model. Information about the actual speed is fed back to the controller by sensors, and the control decisions are implemented via a device, the actuator, that increases or decreases the fuel flow to the engine. This idea is also depicted in the following block diagram.

The use of block diagrams is wide in the control community and is not limited to demonstration of functionality as in Fig. 10.1. On the contrary, it is connected with mathematical properties and operations. However, coming back to Fig. 10.1, it can be seen that an automatic feedback control system includes a plant (controlled system) and a control system. The control system is composed of a controller, sensor(s), and actuator(s). When the feedback path (the line supplying information on the plant status through the sensors) is broken intentionally or unintentionally the system is called open-loop; otherwise, it is a closed-loop system.

An example of a closed-loop system is the automatic course-keeping system of a surface ocean vessel. This system includes all the components shown in Fig. 10.1. The plant in this case is the vessel. The control system consists of the autopilot unit (controller), the steering gear (actuator), and the compass and GPS (global positioning system) receiver (sensors). The controlled variable is the vessel's course and the control action is the rudder deflection from the centerline. Exogenous disturbances are present in the form of wind, wave, and current-generated forces and torques. The effect of the disturbances is demonstrated as a deviation of the actual heading of the vessel from the desired one, entered by the human operator as a waypoint (setpoint). In the case that either the waypoint is modified or excessive deviation is introduced to the controlled variable, as an effect of disturbance presence, the autopilot unit (controller) issues in response rudder position command signals (control signals) that drive the steering gear to appropriately adjust the rudder. The feedback signals are the positioning information provided by the GPS receiver and the heading indication provided by the compass. Both have additive noise which leads to more or less inaccurate readings; also, especially in the case of the GPS receiver, significant and in general variable delay between the actual occurrence of a change and its indication is experienced.

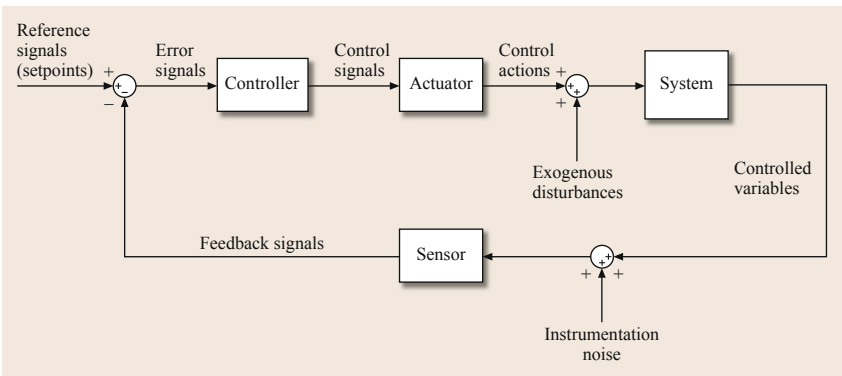


Fig. 10.1 Structure of system with automatic feedback control

Before embarking on the methods to develop a control system, a clear understanding of the behavior and analysis of the response of the plant is needed. This is commonly achieved by developing a mathematical model of the system. The type and the complexity of such a model, though, depend on various salient features of the plant in hand. One of the most important features is whether the plant is static or dynamic. A static system is one for which the output signal values at any given time instant can be determined by employing exclusively the input signal values at the exact same time instant. In effect, static systems are commonly described by algebraic equations of the following general form

$$\boldsymbol{y} = f(\boldsymbol{u}) \ . \tag{10.1}$$

In the above, $\boldsymbol{u}$ and $\boldsymbol{y}$ are the input and output signals, respectively, conveniently packed in vectors.

Another class of systems, though, which is usually the case in feedback control theory and engineering, are dynamic systems. A dynamic system is described by a number of *dynamic* equations, integro-differential equations including at least time as an independent variable. Specifically, the description of a dynamic system must encompass all the constraint equations as well as the full set of initial, or in some cases, final conditions required for the determination of a special solution of the integro-differential equations. A deeper insight in this definition as well as the fundamental description methods for linear and time-invariant dynamic systems is presented in the remainder of this chapter.

10.1.2 The Laplace Transform

The Laplace transform is a linear integral transform that is defined as follows for a signal $x(t)$

$$X(s) = \mathcal{L}\{x(t)\} = \int_0^\infty x(t)\mathrm{e}^{-st}\,\mathrm{d}t \ . \tag{10.2}$$

The (one-sided) Laplace transform is used extensively for analyses of either signals or systems in the complex frequency domain, $s = a + \mathrm{i}\omega$. The complex frequency comes as generalization of the (natural) frequency as defined especially for periodic signals, sinusoidal signals, or sums of such. This is manifested also by the fact that the Fourier transformation of any signal can be derived by substituting $s = \mathrm{i}\omega$ in its Laplace transformation or equivalently setting $a = 0$. These remarks are investigated in the following example.

Consider the signal

$$x(t) = \begin{cases} \exp(a_0 t)\sin(\omega_0 t) & t \geq 0 \ , \\ 0 & t < 0 \ . \end{cases}$$

The Laplace transform is calculated by the definition as follows

$$\begin{aligned} X(s) &= \int_0^\infty \exp(a_0 t)\sin(\omega_0 t)\mathrm{e}^{-st}\,\mathrm{d}t \\ &= \int_0^\infty \mathrm{e}^{a_0 t}\frac{\mathrm{e}^{\mathrm{i}\omega_0 t} - \mathrm{e}^{-\mathrm{i}\omega_0 t}}{2\mathrm{i}}\mathrm{e}^{-st}\,\mathrm{d}t \ . \end{aligned}$$

In the above, use of a well-known identity for the calculation of the sine signals has been made. By substituting $s_0 = a_0 + \mathrm{i}\omega_0$ and taking into account that $s_0^* = a_0 - \mathrm{i}\omega_0$, the following is obtained

$$\begin{aligned} X(s) &= \frac{1}{2\mathrm{i}}\int_0^\infty \mathrm{e}^{(s_0^* - s)t}\mathrm{d}t - \frac{1}{2\mathrm{i}}\int_0^\infty \mathrm{e}^{(s_0 - s)t}\,\mathrm{d}t \\ &= \frac{1}{2\mathrm{i}}\left(-\frac{1}{s_0 - s} + \frac{1}{s_0^* - s}\right) = \frac{\omega_0}{(s - a_0)^2 + \omega_0^2} \ . \end{aligned}$$

It is evident that signal $x(t)$ is a sine wave of angular frequency ω_0, which is either attenuating or growing, depending on the value of constant a_0. This fact is reflected to the Laplace transformation of $x(t)$, $X(s)$. Indeed, the denominator polynomial of $X(s)$, $p_c(s) = (s - a_0)^2 + \omega_0^2$ obtains the following roots

$$\begin{aligned} (s - a_0)^2 = -\omega_0^2 &\Rightarrow s - a_0 = \pm\mathrm{i}\omega_0 \\ &\Rightarrow s = s_0 \text{ or } s = s_0^* \ . \end{aligned}$$

In the case that the sine wave is shrinking it holds that $\mathrm{Re}(s_0) = \mathrm{Re}(s_0^*) = a_0 < 0$, while in the case that it is growing $a_0 > 0$. Finally, in the case that $a_0 = 0$, evidently it holds that $x(t) = \sin(\omega_0 t), t \geq 0$.

Some of the most important properties of the Laplace transform are given as follows:

1. Linearity

$$\left.\begin{aligned} &X_1(s) = \mathcal{L}\{x_1(t)\} \\ &X_2(s) = \mathcal{L}\{x_2(t)\} \\ &k_1, k_2 \in \mathbb{R} \end{aligned}\right\} \Rightarrow k_1 X_1(s) + k_2 X_2(s) = \mathcal{L}\{k_1 x_1(t) + k_2 x_2(t)\} \ . \tag{10.3}$$

2. Differentiation with respect to time

$$\mathcal{L}\{x(t)\} = X(s) \Rightarrow \mathcal{L}\{\dot{x}(t)\} = sX(s) - x(t=0)\,. \tag{10.4}$$

3. Integration with respect to time

$$\left.\begin{aligned}&\mathcal{L}\{x(t)\} = X(s)\\ &y(t) = \int x(\xi)\mathrm{d}\xi\end{aligned}\right\} \Rightarrow \mathcal{L}\left\{\int_0^t x(\xi)\mathrm{d}\xi\right\}$$
$$= \frac{X(s)}{s} \text{ and } \mathcal{L}\left\{\int_{-\infty}^t x(\xi)\mathrm{d}\xi\right\}$$
$$= \frac{X(s)}{s} + \frac{x(t=0)}{s}\,. \tag{10.5}$$

4. Initial and final value theorems

$$\mathcal{L}\{x(t)\} = X(s)$$
$$\Rightarrow \begin{cases}\lim\limits_{t\to 0}(x(t)) = \lim\limits_{s\to\infty}(sX(s))\\ \lim\limits_{t\to\infty}(x(t)) = \lim\limits_{s\to 0}(sX(s))\end{cases}\,. \tag{10.6}$$

5. Theorem of convolution

$$\left.\begin{aligned}&X_1(s) = \mathcal{L}\{x_1(t)\}\\ &X_2(s) = \mathcal{L}\{x_2(t)\}\\ &x(t) = x_1(t) * x_2(t)\\ &\quad \triangleq \int_{-\infty}^{+\infty} x_1(v)x_2(t-v)\mathrm{d}v\end{aligned}\right\} \Rightarrow$$
$$X(s) = \mathcal{L}\{x(t)\} = X_1(s)X_2(s)\,. \tag{10.7}$$

The signal $x(t)$ resulting from a convolution operation must be *causal*; it may not depend on the values of signals $x_1(t_1)$ and $x_2(t_2)$ that will occur at time instants t_1 and t_2, respectively, later than t. In effect, the limits of the convolutional integral in (10.7) must be bounded as follows

$$\left.\begin{aligned}&t_1 = v \le t\\ &t_2 = t - v \le t \Rightarrow v \ge 0\end{aligned}\right\} \Rightarrow$$
$$x_1(t) * x_2(t) = \int_0^t x_1(v)x_2(t-v)\,\mathrm{d}v\,.$$

Furthermore, the convolution operation, for which symbol $\otimes$ is also used in some texts, obtains the *commutative*, *associative*, and *distributive* with respect to addition properties. The *identity element* of convolution is the Dirac delta signal defined in the next section.

The proof of the above Laplace transform properties as well as various other lemmas of the transform can be sought after in either ordinary differential equations or introductory control systems literature [10.1, 2]. Some important Laplace transform pairs, used throughout this text, are given in Table 10.1.

10.1.3 Linear Time-Invariant Systems

Definition of LTI Systems

A category of dynamic systems of high importance in feedback control theory is *linear time-invariant* (LTI) *systems*. Such systems are described by linear differential equations with constant coefficients. The general form of such differential equations is given below for a single-input, single-output (SISO) system

$$\left(a_n\frac{\mathrm{d}^n}{\mathrm{d}t^n} + a_{n-1}\frac{\mathrm{d}^{n-1}}{\mathrm{d}t^{n-1}} + \cdots + a_1\frac{\mathrm{d}}{\mathrm{d}t} + a_0\right)y(t)$$
$$= \left(b_m\frac{\mathrm{d}^m}{\mathrm{d}t^m} + b_{m-1}\frac{\mathrm{d}^{m-1}}{\mathrm{d}t^{m-1}} + \cdots + b_1\frac{\mathrm{d}}{\mathrm{d}t} + b_0\right)u(t)\,. \tag{10.8}$$

In the above, $y(t)$ and $u(t)$ are the input and output signal, respectively, of the system. The initial conditions required are those for the up to the $(n-1)$-th derivative with respect to time of the output signal $y(t)$ at some initial time instant t_0 (commonly t_0 is set to zero by an appropriate translation of the time axis).

$$\left.\frac{\mathrm{d}^v}{\mathrm{d}t^v}y(t)\right|_{t=t_0} = y_0^{(v)}\,, \qquad v = 0, 1, \ldots, (n-1) \tag{10.9}$$

In practically realizable systems, that is, systems that can be implemented by interconnecting electrical (induction coils, capacitors, resistors, etc.), or mechanical (masses, springs, dampers, etc.), elements, it holds that $m \le n$; this assumption will be observed in the following of this text as well.

Order of an LTI System

The *order* of any LTI system is integer n, that is, the order of its differential equation.

Transfer Function of an LTI System

Considering a SISO–LTI system as in (10.8) with *zero initial conditions* in (10.9), by consecutively applying Laplace transform property (10.3) and then property (10.4), the following *algebraic* relationship between the Laplace transformations of the output and input signal

Table 10.1 Laplace and Z-transform tables

Continuous time	Laplace transform	Sampled time	Z-transform
$x(t)$	$X(s)$	$x(kT)$	$X(z)$
t	$\frac{1}{s^2}$	kT	$\frac{Tz}{(z-1)^2}$
t^2	$\frac{1}{s^2}$	$(kT)^2$	$\frac{T^2z(z+1)}{(z-1)^3}$
e^{-at}	$\frac{1}{s+a}$	e^{-akT}	$\frac{z}{z-e^{-aT}}$
$1(t)$	$\frac{1}{s}$	1	$\frac{z}{z-1}$
te^{-at}	$\frac{1}{(s+a)^2}$	$(kT)^n e^{-akT}$	$\frac{Te^{-aT}z}{(z-e^{-aT})^2}$
$\sin(wt)$	$\frac{w}{s^2+w^2}$	$\sin(wkT)$	$\frac{z\sin(wT)}{z^2-2z\cos(wT)+1}$
$\cos(wt)$	$\frac{s}{s^2+w^2}$	$\cos(wkT)$	$\frac{z(z-\cos(wT))}{z^2-2z\cos(wT)+1}$
$e^{-at}\sin wt$	$\frac{w}{(s+a)^2+w^2}$	$e^{-akT}\sin(wkT)$	$\frac{ze^{-aT}\sin(wT)}{z^2-2ze^{-aT}\cos(wT)+e^{-2aT}}$
$e^{-at}\cos(wt)$	$\frac{s+a}{(s+a)^2+w^2}$	$e^{-akT}\cos(wkT)$	$\frac{z^2-ze^{-aT}\cos(wT)}{z^2-2ze^{-aT}\cos(wT)+e^{-2aT}}$

is obtained

$$H(s)=\frac{Y(s)}{U(s)}=\frac{b_m s^m+b_{m-1}s^{m-1}+\cdots+b_1 s+b_0}{a_n s^n+a_{n-1}s^{n-1}+\cdots+a_1 s+a_0},\quad s\in\mathbb{C}\,,\quad a_n\neq 0\,. \tag{10.10}$$

The complex polynomial ratio $H(s)$ is the system's transfer function. Transfer functions are used extensively in control theory. Commonly, the polynomial coefficients are normalized with respect to $a_n \neq 0$.

Characteristic Polynomial of an LTI System

The polynomial $p_c(s)=a_n s^n+a_{n-1}s^{n-1}+\cdots+a_1 s+a_0$ appearing at the denominator of the transfer function of a SISO–LTI system, $H(s)$, is the characteristic polynomial of the system. If the transfer function is normalized with respect to $a_n \neq 0$, it follows that $a_n = 1$. In the sequel, normalization will be assumed.

System Poles

The n distinct or multiple complex roots of the characteristic polynomial $p_c(s)$ of an LTI system are the system poles. In effect, an LTI system's poles are calculated as roots of the equation

$$p_c(s)=s^n+a_{n-1}s^{n-1}+\cdots+a_1 s+a_0=0\,. \tag{10.11}$$

In the above, the polynomial coefficients are assumed normalized in order that the highest order coefficient is equal to 1.

Otherwise it is noted that if all coefficients of a polynomial are *purely real* then the roots are either also purely *real* or form pairs of *complex conjugate* numbers.

System Zeros

The m distinct or multiple complex roots of the polynomial appearing at the numerator of the transfer function, $H(s)$, of an LTI system are the system zeros.

System Impulse Response

The response of an LTI system when the input is the Dirac delta signal and the initial conditions are set to zero is the impulse response. In effect, the impulse response, $h(t)$, can be calculated as the solution of the following differential equation

$$\left(a_n\frac{\mathrm{d}^n}{\mathrm{d}t^n}+a_{n-1}\frac{\mathrm{d}^{n-1}}{\mathrm{d}t^{n-1}}+\cdots+a_1\frac{\mathrm{d}}{\mathrm{d}t}+a_0\right)h(t)=\left(b_m\frac{\mathrm{d}^m}{\mathrm{d}t^m}+b_{m-1}\frac{\mathrm{d}^{m-1}}{\mathrm{d}t^{m-1}}+\cdots+b_1\frac{\mathrm{d}}{\mathrm{d}t}+b_0\right)\delta(t)\,.$$

In the above, $\delta(t)$ stands for the Dirac delta signal, referred to also as impulse function or signal. The Dirac

signal is defined on the basis of the following properties:

1.
$$\int_{-\infty}^{+\infty} \delta(t)\mathrm{d}t = 1 \tag{10.12}$$

2.
$$\delta(t) = \begin{cases} 0\,, & t \neq 0 \\ \text{undefined} & t = 0 \end{cases}\,. \tag{10.13}$$

By using the above definition properties the following properties can be derived,

3.
$$\int_{-\infty}^{+\infty} \delta(t) f(t)\,\mathrm{d}t = f(0) \tag{10.14}$$

4.
$$\Delta(s) = \mathrm{L}\left\{\delta(t)\right\} = \int_{0}^{+\infty} \delta(t)\mathrm{e}^{-st}\mathrm{d}t = 1\,. \tag{10.15}$$

By using (10.15) it can be shown that the transfer function and the impulse response are connected through the Laplace transform.

$$\mathrm{L}\left\{h(t)\right\} = H(s)\mathrm{L}\left\{\delta(t)\right\} = H(s)\Delta(s) = H(s) \tag{10.16}$$

In effect, by employing the theorem of convolution (10.7), the time-domain input–output relationship equivalent to the one given in (10.10) in the complex frequency domain is derived

$$\begin{aligned} &Y(s) = H(s)U(s) \Leftrightarrow \\ &y(t) = h(t) * u(t) = \int_{-\infty}^{+\infty} h(v)u(t-v)\mathrm{d}v \\ &\quad = \int_{0}^{t} h(v)u(t-v)\mathrm{d}v\,. \end{aligned} \tag{10.17}$$

The interval of the convolutional integral is also reduced from $(-\infty, +\infty)$ to $(0, t)$ on the basis of *causality*. Therefore, the output signal value $y(t)$ at time t depends on input signal values $u(v)$ occurring at time instants not later than t.

Conclusively, the determination of an LTI system's transfer function can be done on the basis of (10.6), by Laplace transformation of the system's impulse response. This is why it is interesting to determine the impulse response, $h(t)$, of the system. The direct method to determine $h(t)$ is to excite the system by the Dirac signal, $\delta(t)$, and record the response. However, in many practical cases this is not feasible, especially because implementation of the Dirac signal $\delta(t)$ is not without problems or approximations.

Alternatively, the impulse response $h(t)$ and, consequently, the transfer function of a system, $H(s)$, can be calculated using its *step response*, $y_{\mathrm{step}}(t)$. The step response is the output of the system when the input signal is the *unit step signal* (also referred to as Heaviside signal or function), $u_{\mathrm{step}}(t)$, defined as follows

$$\begin{aligned} &u_{\mathrm{step}}(t) = \begin{cases} 1 & t > 0 \\ 0 & t \leq 0 \end{cases} \Rightarrow u_{\mathrm{step}}(t) = \int_{-\infty}^{t} \delta(v)\mathrm{d}v \\ &\Leftrightarrow U_{\mathrm{step}}(s) = \mathrm{L}\left\{u_{\mathrm{step}}(t)\right\} = \frac{\Delta(s)}{s} = \frac{1}{s}\,. \end{aligned} \tag{10.18}$$

In effect, on the basis of property (10.4) it can be shown that

$$\begin{aligned} &h(t) = \frac{\mathrm{d}}{\mathrm{d}t} y_{\mathrm{step}}(t) \\ &\Leftrightarrow H(s) = sY_{\mathrm{step}}(s) = s\mathrm{L}\left\{y_{\mathrm{step}}(t)\right\}\,. \end{aligned} \tag{10.19}$$

Furthermore, in the case that the transfer function of an LTI system is known and the impulse response is sought after, the technique of *partial fraction expansion* is employed. In that end, the following fact is needed

$$\frac{1}{s-\lambda} = \mathrm{L}\left\{\mathrm{e}^{\lambda t}\right\}\,, \quad \lambda \in \mathbb{C}, t \geq 0\,. \tag{10.20}$$

At first, assume that we are presented with an nth order system obtaining n distinct, purely real poles; this means that the characteristic polynomial has no multiple roots and no complex roots. Then, the characteristic polynomial can be rewritten in the following *factorized* form

$$p_c(s) = \prod_{v=1}^{n} (s - \lambda_v)\,. \tag{10.21}$$

In the above λ_v, $v = 1, \ldots, n$ are the poles of the system and as explained previously it has been assumed that $a_n = 1$. On the basis of the factorization of $p_c(s)$ the transfer function $H(s)$ may be expanded to a sum of

fractions of the form shown in (10.20)

$$H(s) = \sum_{v=1}^{n} \frac{k_v}{s-\lambda_v} \,. \tag{10.22}$$

In the above constants (coefficients) $k_v, v = 1, \ldots, n$ can be calculated by using the following relationship

$$k_v = \lim_{s\to\lambda_v} [(s-\lambda_v)H(s)] \,, \quad v = 1, \ldots, n \,. \tag{10.23}$$

Based on the fact (10.20) and the expansion (10.22) the following expression is derived for the system's impulse response

$$h(t) = \sum_{v=1}^{n} k_v \exp(\lambda_v t) \,, \quad t \geq 0 \,. \tag{10.24}$$

The partial fraction expansion can also be employed in cases for which the characteristic polynomial obtains multiple real or complex roots. First, the case of a purely real pole, λ_0, with multiplicity $l > 1$, is examined. Then, the factorized form of the characteristic polynomial of the system includes factor $(s-\lambda_0)^l$. Consequently, in the system's transfer function this factor appears in the denominator. Therefore, the partial function expansion of the following general form is needed

$$\begin{aligned} & p_c(s) = (s-\lambda_0)^l \prod_{v=1}^{n} (s-\lambda_v) \\ & \Rightarrow H(s) = \sum_{v=1}^{n} \frac{k_v}{s-\lambda_v} + \sum_{\xi=1}^{l} \frac{k_{0\xi}}{(s-\lambda_0)^\xi} \,. \end{aligned} \tag{10.25}$$

In the above, additional terms augment the expansion of $\mathbf{H}(s)$ in (10.22) in order to encompass factor $(s-\lambda_0)^l$ of the characteristic polynomial. To obtain the values of constants $k_{0\xi}$ of the expansion, the following formula is used

$$\begin{aligned} & k_{0\xi} = \lim_{s\to\lambda_0} \left\{ \frac{1}{(l-\xi)!} \frac{\mathrm{d}^{(l-\xi)}}{\mathrm{d}s^{(l-\xi)}} \left[(s-\lambda_0)^l H(s) \right] \right\} , \\ & \xi = 1, \ldots, l \,. \end{aligned} \tag{10.26}$$

The inverse Laplace transform of each additional term in the expansion (10.25) is given by the following

$$\mathrm{L}^{-1} \left\{ \frac{1}{(s-\lambda_0)^\xi} \right\} = \frac{1}{(\xi-1)!} t^{(\xi-1)} \mathrm{e}^{\lambda_0 t} \,. \tag{10.27}$$

The above comes as a generalization of (10.20). In conclusion, each additional term (10.25) corresponds to a term of the form (10.27) in the impulse response expression (10.24).

In the case that the characteristic polynomial obtains a pair of complex conjugate roots, λ_C and λ_C^*, the factorized form includes the factor

$$(s-\lambda_\mathrm{C})(s-\lambda_\mathrm{C}^*) = s^2 - 2\mathrm{Re}(\lambda_\mathrm{C})s + |\lambda_\mathrm{C}|^2 \,.$$

In this case, (10.20), (10.22), and (10.23) are still valid for each one of the poles λ_C and λ_C^*, separately. Furthermore, the following fact for the combined partial fraction expansion of this factor can be shown

$$\begin{aligned} & \frac{1}{s^2 - 2\mathrm{Re}(\lambda_\mathrm{C})s + |\lambda_\mathrm{C}|^2} = \frac{k_\mathrm{C}}{s-\lambda_\mathrm{C}} + \frac{k_{\mathrm{C}*}}{s-\lambda_\mathrm{C}^*} \\ & \Rightarrow k_{\mathrm{C}*} = (k_\mathrm{C})^* \,. \end{aligned} \tag{10.28}$$

Therefore, for calculating the constant $k_{\mathrm{C}*}$ corresponding to factor $(s-\lambda_\mathrm{C}^*)$ it suffices to calculate constant k_C corresponding to factor $(s-\lambda_\mathrm{C})$.

10.1.4 Multivariable Systems and State Space

State Equations

State equations constitute a description in the time domain encompassing, besides LTI systems, other categories of systems such as time-varying or nonlinear systems [10.1–4]. A system described by state equations may in general accept more than one scalar input signals and produce more than one scalar output signals, also referred to as measurements. If the number of scalar input signals is m and that of scalar output signals is p, the following column vectors are introduced for convenience in notation

$$\begin{aligned} & \boldsymbol{u} = \boldsymbol{u}(t) = \begin{bmatrix} u_1 & u_2 & \ldots & u_m \end{bmatrix}^\mathrm{T} \\ & \text{and} \\ & \boldsymbol{y} = \boldsymbol{y}(t) = \begin{bmatrix} y_1 & y_2 & \ldots & y_p \end{bmatrix}^\mathrm{T} \,. \end{aligned} \tag{10.29}$$

In the disciplines of dynamic systems and control, the term *state* or *state vector* is used for a set of $n \geq 1$ variables *(state variables)* conveniently packed in a column vector as follow

$$\boldsymbol{x} = \boldsymbol{x}(t) = \begin{bmatrix} x_1 & x_2 & \ldots & x_n \end{bmatrix}^\mathrm{T} \tag{10.30}$$

If the value of $\boldsymbol{x}$ is known at some initial time instant t_0, then it must be possible to determine the values of both $\boldsymbol{x}$ and $\boldsymbol{y}$ for times $t \geq t_0$ provided that (a) the values of $\boldsymbol{u}$ are given and (b) the mathematical law connecting the input, output, and state vectors is known.

Commonly, the aforementioned mathematical law is expressed as a set of n first-order differential equations with respect to time, dynamic equations. The dynamic equations connect the input vector with the state vector

$$\dot{\boldsymbol{x}}(t) \triangleq \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{x}(t) = f(\boldsymbol{x}(t), \boldsymbol{u}(t), t)\,. \tag{10.31}$$

The description of the system in state-space requires also a set of p *algebraic* output equations connecting the input and state vectors with the output one

$$\boldsymbol{y}(t) = g(\boldsymbol{x}(t), \boldsymbol{u}(t))\,. \tag{10.32}$$

By employing (10.31), it is directly concluded that if the initial state of a system is given at a time instant t_0 then the state at any future time instant can be calculated as follows

$$\boldsymbol{x}(t) = \boldsymbol{x}(t_0) + \int_{t_0}^{t} f(\boldsymbol{x}(\xi), \boldsymbol{u}(\xi), \xi)\mathrm{d}\xi\,. \tag{10.33}$$

Multivariable LTI Systems

In the case of LTI systems (10.31) and (10.32) assume the following form

$$\begin{aligned}\dot{\boldsymbol{x}} &= \mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{u}\,,\\ \boldsymbol{y} &= \mathbf{C}\boldsymbol{x} + \mathbf{D}\boldsymbol{u}\,.\end{aligned} \tag{10.34}$$

In the above $\mathbf{A}$ is an $n \times n$ matrix, $\mathbf{B}$ an $n \times m$ matrix, $\mathbf{C}$ a $p \times n$ matrix, and $\mathbf{D}$ a $p \times m$ matrix.

An input–output relationship can be deduced for an LTI system described in state-space by the state and output equations of the form (10.34). This relationship comes as a natural extension of the transfer function concept already introduced for single-input, single-output (SISO) LTI systems to the case of MIMO (multi-input, multioutput) LTI systems. To obtain it the Laplace transform is applied to the time-domain expression (10.34) by employing property (10.4)

$$\begin{aligned}s\mathbf{X}(s) - \boldsymbol{x}(0) &= \mathbf{A}\mathbf{X}(s) + \mathbf{B}\mathbf{U}(s)\,,\\ \mathbf{Y}(s) &= \mathbf{C}\mathbf{X}(s) + \mathbf{D}\mathbf{U}(s)\,.\end{aligned} \tag{10.35}$$

In the above, it has been assumed that $t_0 = 0$ for convenience and without any loss of generality. The complex frequency domain algebraic (10.35) can be solved to yield the following

$$\begin{aligned}\mathbf{X}(s) &= (s\mathbf{I}_n - \mathbf{A})^{-1}\mathbf{B}\mathbf{U}(s) + (s\mathbf{I}_n - \mathbf{A})^{-1}\boldsymbol{x}(0)\,,\\ \mathbf{Y}(s) &= \left[\mathbf{C}\,(s\mathbf{I}_n - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D}\right]\mathbf{U}(s) + \mathbf{C}\,(s\mathbf{I}_n - \mathbf{A})^{-1}\boldsymbol{x}(0)\,.\end{aligned} \tag{10.36}$$

In the above, $\mathbf{I}_n$ stands for the $n \times n$ identity matrix. Formulas (10.36) allow the analytic calculation of any LTI system's response and are referred to as full-response equations in the literature [10.1–6]. The following special cases are investigated:

1. Zero initial state response, $x(0) = 0$

$$\mathbf{Y}(s) = \left[\mathbf{C}\,(s\mathbf{I}_n - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D}\right]\mathbf{U}(s)\,. \tag{10.37}$$

2. Zero input response, $u = 0$

$$\mathbf{Y}(s) = \mathbf{C}\,(s\mathbf{I}_n - \mathbf{A})^{-1}\,x(0)\,. \tag{10.38}$$

In order to obtain the response in the time domain, the following formula of the Laplace transform theory is needed

$$(s\mathbf{I}_n - \mathbf{A})^{-1} = \mathrm{L}\left\{\mathrm{e}^{\mathbf{A}t}\right\}\,, \quad t > 0\,. \tag{10.39}$$

The above is the matrix generalization of (10.16). Expression $\mathrm{e}^{\mathbf{A}t}$, also referred to by the term system transient matrix, is defined as an $n \times n$ matrix with the following properties

$$\begin{aligned}\mathrm{e}^{\mathbf{A}t} &= \mathbf{I}_n + \mathbf{A}t + \frac{1}{2!}\mathbf{A}^2t^2 + \frac{1}{3!}\mathbf{A}^3t^3 + \cdots\\ &\Leftrightarrow \frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{\mathbf{A}t} = \mathbf{A}\cdot\mathrm{e}^{\mathbf{A}t}\,.\end{aligned} \tag{10.40}$$

By using (10.40), the full-response equation in the time domain is obtained

$$x(t) = \int_0^t \mathrm{e}^{\mathbf{A}v}\mathbf{B}u(t-v)\mathrm{d}v + \mathrm{e}^{\mathbf{A}t}x(0)\,, \quad t > 0\,. \tag{10.41}$$

The following matrix version of the theorem of convolution has been employed in the above

$$\begin{aligned}(s\mathbf{I}_n - \mathbf{A})^{-1}\mathbf{B}\mathbf{U}(s) &= \mathrm{L}\left\{\mathrm{e}^{\mathbf{A}t} * \mathbf{B}u(t)\right\}\\ &= \mathrm{L}\left\{\int_0^t \mathrm{e}^{\mathbf{A}v}\mathbf{B}u(t-v)\mathrm{d}v\right\}\,.\end{aligned} \tag{10.42}$$

Transfer Function Matrix

In the input–output relationship (10.37), the $p \times m$ matrix $\mathbf{H}(s) = \mathbf{C}\cdot(s\mathbf{I}_n - \mathbf{A})^{-1}\cdot\mathbf{B} + \mathbf{D}$ is the system's transfer matrix.

It is straightforward to show that

$$(s\mathbf{I}_n - \mathbf{A})^{-1} = |s\mathbf{I}_n - \mathbf{A}|^{-1}\,\mathrm{adj}(s\mathbf{I}_n - \mathbf{A})\,.$$

In the above, $|\bullet| = \det(\bullet)$ stands for the determinant of a matrix and adj$(\bullet)$ is the conjugate transpose of the matrix cofactors. In effect, the following conclusions are made:

- The characteristic polynomial of an LTI system with description in state-space as in (10.34) is the determinant of matrix $(s\mathbf{I}_n - \mathbf{A})$, $p_c(s) = |s\mathbf{I}_n - \mathbf{A}|$.
- The poles of a system with the general form (10.34), the roots of $p_c(s) = |s\mathbf{I}_n - \mathbf{A}|$, are the eigenvalues of matrix $\mathbf{A}$.

The above general remarks are now demonstrated in the special case of an LTI–SISO system, with a transfer function of the following format

$$\mathbf{H}(s) = \frac{\mathbf{Y}(s)}{\mathbf{U}(s)} = \frac{b_0}{s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0} \,. \quad (10.43)$$

It is noted here that although the system is SISO the state vector is of n dimension, that is, equal to the order of the system. However, there are more than one choices of the state vector's components. The problem of choosing the components of the state vector in order to formulate a state-space description of an LTI system, which is otherwise known only in the form of either a scalar transfer function or a transfer function matrix, is referred to as the *realization* problem [10.1, 2]. In general case of MIMO–LTI systems is not solvable. However, in the case of LTI–SISO systems with transfer function of the form (10.43), a *minimal* description, a description that includes no redundant components can be deduced as

$$\mathbf{H}(s) = \frac{\mathbf{Y}(s)}{\mathbf{U}(s)} = \frac{b_0}{s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0} = \frac{b_0}{s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0} \frac{\mathbf{X}_1(s)}{\mathbf{X}_1(s)} \,.$$

After introducing the auxiliary signal as above, the following differential equation is obtained in the time domain. Appropriate initial conditions of the form (10.9) for solving the differential equation must also be provided.

$$\left(\frac{\mathrm{d}^n}{\mathrm{d}t^n} + a_{n-1}\frac{\mathrm{d}^{n-1}}{\mathrm{d}t^{n-1}} + \cdots + a_1\frac{\mathrm{d}}{\mathrm{d}t} + a_0\right) x_1(t) = u(t) \Leftrightarrow x_1^{(n)}(t) = u(t) - \sum_{v=0}^{n-1} a_v x_1^{(v)}(t)$$

$$\text{and} \quad y(t) = b_0 x_1(t) \,. \quad (10.44)$$

In the above, $x_1^{(v)}(t)$ stands for the vth order time derivative of the auxiliary signal $x_1(t)$. An appropriate state vector x is

$$x = \left[x_1(t) \quad x_1^{(1)}(t) \quad \ldots \quad x_1^{(n-1)}(t)\right]^{\mathrm{T}} . \quad (10.45)$$

For the components of x the following differential equations hold,

$$\dot{x}_v = x_{v+1} \,, \quad v = 1, \ldots, n-1$$

$$\text{and} \quad \dot{x}_n = -\sum_{v=1}^{n} a_{(v-1)} x_v + u \,. \quad (10.46)$$

If the above are arranged in the matrix format the following equations are obtained

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \vdots \\ \dot{x}_n \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ 0 & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -a_0 & -a_1 & -a_2 & \ldots & -a_{n-1} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} u \,,$$

$$y = \begin{bmatrix} b_0 & 0 & 0 & \ldots & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} . \quad (10.47)$$

This form of the state equations and especially of matrix $\mathbf{A}$ is also known as the *companion canonical form*.

10.1.5 Nonlinear Systems and Linearization

Consider a nonlinear system with m inputs, p outputs, and n state variables with the following description in state space

$$\dot{x} = f(x, u) \,,$$
$$y = g(x, u) \,. \quad (10.48)$$

Significant progress has been achieved in recent years in the analysis and controller synthesis for a wide class of nonlinear systems, especially through the studies of *Isidori* [10.7, 8] and *Kokotovic* et al. [10.9]. However, the technique of linearization, the reduction to or approximation by a linear system is still extensively used.

The first step toward the linearization of a nonlinear system is to determine the equilibrium points. These points are defined by the following vector equation

$$\dot{x} = \mathbf{0} \Leftrightarrow f(x_0, u_0) = \mathbf{0} \,. \tag{10.49}$$

Linearization is then applied around some equilibrium point (x_0, u_0) by using the Taylor expansions for functions f and g

$$\begin{aligned} \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}) \approx & \boldsymbol{f}(\boldsymbol{x}_0, \boldsymbol{u}_0) + \mathbf{f}'_x(\boldsymbol{x}_0, \boldsymbol{u}_0)(\boldsymbol{x} - \boldsymbol{x}_0) \\ & + \mathbf{f}'_u(\boldsymbol{x}_0, \boldsymbol{u}_0)(\boldsymbol{u} - \boldsymbol{u}_0) \\ \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{u}) \approx & \boldsymbol{g}(\boldsymbol{x}_0, \boldsymbol{u}_0) + \mathbf{g}'_x(\boldsymbol{x}_0, \boldsymbol{u}_0)(\boldsymbol{x} - \boldsymbol{x}_0) \\ & + \mathbf{g}'_u(\boldsymbol{x}_0, \boldsymbol{u}_0)(\boldsymbol{u} - \boldsymbol{u}_0) \,. \end{aligned} \tag{10.50}$$

In the above, matrices $\mathbf{f}'_x(x_0, u_0)$, $\mathbf{f}'_u(x_0, u_0)$, $\mathbf{g}'_x(x_0, u_0)$, $\mathbf{g}'_u(x_0, u_0)$ with dimensions $n \times n$, $n \times m$, $p \times n$ and $p \times m$, respectively, are defined as follows

$$\mathbf{f}'_x(x_0, u_0) = \left[\frac{\partial}{\partial x_j} f_i(x, u) \,\middle|\, \begin{matrix} x = x_0 \\ u = u_0 \end{matrix}, i, j = 1, \ldots, n \right] ,$$

$$\mathbf{f}'_u(x_0, u_0) = \left[\frac{\partial}{\partial u_j} f_i(x, u) \,\middle|\, \begin{matrix} x = x_0 \\ u = u_0 \end{matrix}, \begin{matrix} i = 1, \ldots, n \\ j = 1, \ldots, m \end{matrix} \right] , \tag{10.51}$$

$$\mathbf{g}'_x(x_0, u_0) = \left[\frac{\partial}{\partial x_j} g_i(x, u) \,\middle|\, \begin{matrix} x = x_0 \\ u = u_0 \end{matrix}, \begin{matrix} i = 1, \ldots, p \\ j = 1, \ldots, n \end{matrix} \right] ,$$

$$\mathbf{g}'_u(x_0, u_0) = \left[\frac{\partial}{\partial u_j} g_i(x, u) \,\middle|\, \begin{matrix} x = x_0 \\ u = u_0 \end{matrix}, \begin{matrix} i = 1, \ldots, p \\ j = 1, \ldots, m \end{matrix} \right] . \tag{10.52}$$

The following change of variables is then adopted

$$\begin{aligned} & \delta x = x - x_0 \,, \quad \delta u = u - u_0 \,, \\ & \delta y = y - y_0 = y - g(x_0, u_0) \,. \end{aligned} \tag{10.53}$$

In effect, the following linearized equations are deduced for the nonlinear system (10.48) which is valid exclusively in the vicinity of the equilibrium point (x_0, u_0)

$$\begin{aligned} \delta\dot{x} &= \mathbf{f}'_x(x_0, u_0)\,\delta x + \mathbf{f}'_u(x_0, u_0)\,\delta u \,, \\ \delta y &= \mathbf{g}'_x(x_0, u_0)\,\delta x + \mathbf{g}'_u(x_0, u_0)\,\delta u \,. \end{aligned} \tag{10.54}$$

Despite its virtues, the technique of linearization has also a major drawback. This drawback originates from the fact that the elements of matrices (10.51) and (10.52) depend on the equilibrium point in the vicinity of which they are calculated. Therefore, a *family* of linear systems is obtained in order to describe a single nonlinear system. In effect, the form of the results of the analysis and control scheme synthesis must be parameterized with respect to the equilibrium point. Furthermore, linearization cannot be used for nonlinear systems that do not demonstrate equilibrium points as well as systems whose behavior is of interest when operating away from any equilibrium points.

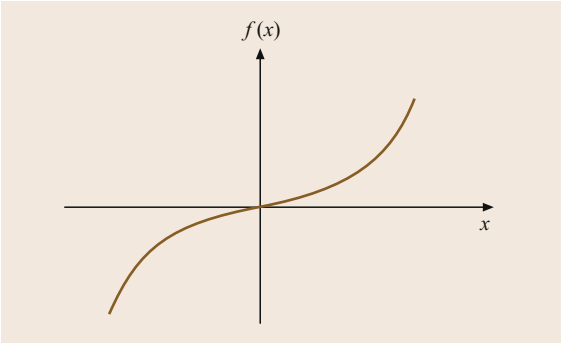


Fig. 10.2 A nonlinear function $x\,|x|$

Example 10.1 Nonlinear Function

Consider linearizing the term $x\,|x|$ shown in Fig.10.2.

One can see from the figure that the slope at any point of linearization is always non-negative and the nonlinear mapping is simply a quadratic relationship with x. Thus, the linearized term becomes

$$\begin{aligned} f(x) = x\,|x| &\approx x_0\,|x_0| + \left.\frac{\partial f}{\partial x}\right|_{x=x_0} (x - x_0) \\ &= x_0\,|x_0| + 2\,|x_0|\,(x - x_0) \,, \end{aligned} \tag{10.55}$$

where x_0 is the point of linearization. In this case, the slope and intercept are $2|x_0|$ and $x_0|x_0|$, respectively.

Example 10.2 Linearized AUV Forward Speed Model

Consider an autonomous underwater vehicle (AUV) travels at a constant heading. The dynamic modes in the longitudinal and lateral planes can be assumed uncoupled, and the forward speed model can be approximately described using only the vehicle and propeller speeds (u, Ω) as the system states, and the motor torque (τ) as the system input

$$\begin{aligned} & M\dot{u} + du\,|u| = c_T \Omega\,|\Omega| \,, \\ & \dot{\Omega} = \beta\tau - \alpha\Omega\,|\Omega| \,, \end{aligned} \tag{10.56}$$

where M is the mass of the vehicle, d, c_T, α, β are constant coefficients associated with the system. The absolute signs characterize the directional effect of the

vehicle and propeller speed. Despite the significant assumption used in the forward speed model, its mathematical equations remain highly nonlinear.

Based on the results obtained in the previous example, the forward speed model can be linearized to

$$\left.\begin{aligned} &M\dot{u} + d\{u_0|u_0| + 2|u_0|(u-u_0)\} \\ &= c_T\{\Omega_0|\Omega_0| + 2|\Omega_0|(\Omega-\Omega_0)\} \\ &\dot{\Omega} = \beta\tau - \alpha\{\Omega_0|\Omega_0| + 2|\Omega_0|(\Omega-\Omega_0)\} \end{aligned}\right\} \Rightarrow$$

$$\Rightarrow \begin{cases} M\dot{u} + 2d|u_0|u = 2c_T\Omega|\Omega_0| + du_0|u_0| \\ \quad -c_T\Omega_0|\Omega_0| \\ \dot{\Omega} + 2\alpha\Omega|\Omega_0| = \beta\tau + \alpha\Omega_0|\Omega_0| \end{cases} . \tag{10.57}$$

These linearized equations can be further simplified if the point of linearization is also the equilibrium point. That is, choose u_0, Ω_0, such that $\dot{u} = 0$, or

$$du_0|u_0| = c_T\Omega_0|\Omega_0| \ .$$

The simplified forward speed model can then be expressed as

$$\begin{aligned} &M\dot{u} + 2d|u_0|u = 2c_T\Omega|\Omega_0| \ , \\ &\dot{\Omega} + 2\alpha\Omega|\Omega_0| = \beta\tau + \alpha\Omega_0|\Omega_0| \ . \end{aligned} \tag{10.58}$$

It should be noted that the term $\alpha\Omega_0|\Omega_0|$ can be treated as a fictitious input that drives the propeller speed based on the point of linearization.

10.2 Analysis of LTI Systems

10.2.1 Block Diagrams

The block diagram of a system serves as a depiction of the interconnection structure of the various subsystems as well as the signals circulation and traffic. Each sub-block of the overall block diagram is defined by an input–output (I/O) relationship, which for LTI systems is commonly expressed in the complex frequency domain (Fig. 10.3. For example, consider the following transfer function

$$\mathbf{H}(s) = \frac{\mathbf{Y}(s)}{\mathbf{U}(s)} = \frac{1}{s+1} \ . \tag{10.59}$$

The corresponding block is shown below.

Using transfer functions for the representation of the I/O relationships of the various blocks allows to conveniently obtain the overall transfer function of a system that comprises of more than blocks interconnected in cascade (Fig. 10.4). For example, consider the system depicted in the following block diagram.

The overall transfer function of the full, interconnected system is as follows,

$$\begin{aligned} \mathbf{H}(s) &= \frac{\mathbf{Y}(s)}{\mathbf{U}(s)} = \mathbf{H}_1(s)\mathbf{H}_2(s)\mathbf{H}_3(s) \\ &= \frac{1}{s+1}\frac{s+2}{s^2+s+1}\frac{1}{s^2+5s+4} \end{aligned} \tag{10.60}$$

Except he individual transfer function blocks, another common element encountered frequently in block diagrams is the adder (or summing junction) (Fig. 10.5). An adder produces as output the algebraic sum of all the input signals.

Furthermore, a system may be defined as either open loop or closed loop on the basis of its block diagram. Indeed, if at least one output signal of a system is fed as input to at least one of its own blocks, the system is called closed-loop; otherwise, it is called open loop. A closed-loop system is shown in Fig. 10.6.

The system transfer function of the closed-loop system shown above, the I/O relationship connecting the reference signal r to the output signal y is given as

$$\mathbf{H}(s) = \frac{\mathbf{Y}(s)}{\mathbf{R}(s)} = \frac{\mathbf{G}(s)}{1+\mathbf{G}(s)\mathbf{F}(s)} \ . \tag{10.61}$$

By inspection of (10.61), it is evident that any closed loop can be reduced to a single-block system, that is, an equivalent open-loop system. The inverse is also possible; an open-loop system can also be converted to a closed loop one with unity feedback branch ($\mathbf{F}(s) \equiv 1$). Indeed, if $\mathbf{F}(s) = 1$, by solving (10.61) with respect to $\mathbf{G}(s)$, it is easily concluded that the direct (feedforward) branch must obtain the following transfer function

$$\mathbf{G}(s) = \frac{\mathbf{Y}(s)}{\mathbf{U}(s)} = \frac{\mathbf{H}(s)}{1-\mathbf{H}(s)} \tag{10.62}$$

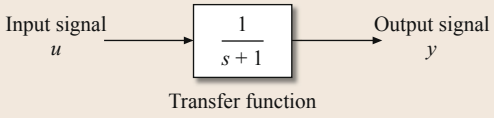


Fig. 10.3 Block representing a scalar (SISO) transfer function

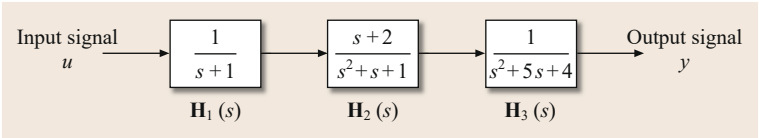


Fig. 10.4 Cascade interconnection of scalar transfer functions

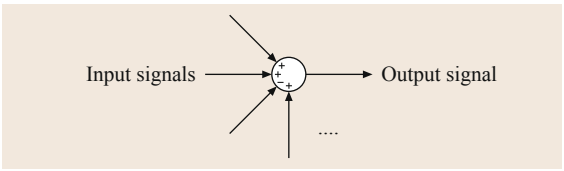


Fig. 10.5 Summing junction used in block diagrams

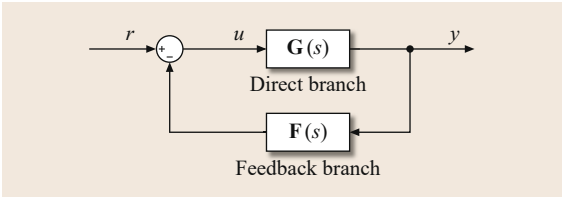


Fig. 10.6 Block diagram of a typical closed-loop system

The case that the closed-loop system in Fig. 10.6 is multivariable will also be now examined. The dimensions of the input and reference vectors, u and r, respectively, are assumed to be m and that for the output vector $\boldsymbol{y}$, p. Evidently, transfer function matrices $\mathbf{G}(s)$ and $\mathbf{F}(s)$ must be of dimensions $p \times m$ and $m \times p$, respectively. Following the same method used for deriving the scalar case (10.62), it can be shown that the $p \times m$ closed-loop, transfer function matrix $\mathbf{H}(s)$ is given by

$$\mathbf{Y}(s) = \mathbf{H}(s)\mathbf{U}(s) \,,$$
$$\text{where} \quad \mathbf{H}(s) = \left[I_p + \mathbf{G}(s)\mathbf{F}(s)\right]^{-1} \mathbf{G}(s) \,. \quad (10.63)$$

In the above, $\mathbf{I}_p$ stands for the identity $p \times p$ matrix. Equation (10.63) comes as a generalization of (10.61); it is reminded here though that commutativity in matrix multiplication does not hold.

10.2.2 Stability

Stability is a substantial structural characteristic of any dynamic system. Stability is not optional but compulsory and, therefore, must be guaranteed after the application of control to a system; otherwise, control must be employed to stabilize an unstable, open-loop system.

A working definition of stability is the following:

A system is stable if and only if it generates a bounded output when driven by any bounded input.

The above definition is generic and applicable to both linear as well as nonlinear systems [10.1]. In the case of linear, time-invariant systems (LTI), it is possible to determine stability without taking into account the class of the input signals applied; actually, the characterization can be done if either the transfer function or the state-space description is available.

Consider an LTI system with description in state space as follows

$$\dot{x} = \mathbf{A}x + \mathbf{B}u \,,$$
$$y = \mathbf{C}x + \mathbf{D}u \,. \quad (10.64)$$

The system is said to be *asymptotically* stable iff under zero input conditions (i. e., $u(t) \equiv 0$), the following criteria are met

$$\exists M(x(t=0)) \geq 0 : \forall t \geq 0 \,, \quad \|x(t)\| < M \,,$$
$$\lim_{t \to \infty} \|x(t)\| = 0 \,. \quad (10.65)$$

It is evident that the value of the bound M for the (Euclidean) norm of the state vector depends on the actual values for the initial condition. In the case that only the first of the criteria in (10.65) is met the system is said to be *critically stable*.

Some further insight is gained in these definitions of asymptotically and critically stable systems by applying them to two specific cases of LTI systems. Consider two LTI systems, (S1) and (S2), with one inputs and two outputs each. For their state-space representation, as in (10.64), assume that matrices $\mathbf{B}$ and $\mathbf{C}$ are the same for both (S1) and (S2).

$$\mathbf{B}^{\mathrm{T}} = \begin{pmatrix} 0 & 1 \end{pmatrix} \text{ and } \mathbf{C} = \mathbf{I}_2$$

Matrix $\mathbf{A}$ for both (S1) and (S2) is given below.

$$\text{(S1)}: \mathbf{A} = \begin{bmatrix} 0 & 1 \\ -1.04 & -0.4 \end{bmatrix} \,,$$
$$\text{(S2)}: \mathbf{A} = \begin{bmatrix} 0 & 1 \\ -0.09 & 0 \end{bmatrix} \quad (10.66)$$

In Fig. 10.7, the trajectories for both systems in state space are shown when the initial conditions are set to $x^{\mathrm{T}}(t=0) = \begin{bmatrix} 10.0 & 10.0 \end{bmatrix}$ for both. It can be seen that the trajectories will be similar for any other value of the

initial conditions. Both systems are stable as their trajectory remains in the interior of a disc as time elapses. For example, for the specific initial conditions value the radius of the disc can be set to $M = 40.0$. Furthermore, (S1) is asymptotically stable as the state vector's norm is converging to zero while (S2) is only critically stable as the state vector's norm, although remaining inside the disc with radius $M = 40.0$, it never decreases to zero (not even after infinite time).

Equation (10.65) is associated with the definition of stability in the time domain. The lemma given below, allows stability characterization of any LTI system (SISO or MIMO), with transfer function matrix given below, in the complex frequency domain

$$\begin{aligned}\mathbf{H}(s) &= \mathbf{C}\left(s\mathbf{I}_p - \mathbf{A}\right)^{-1}\mathbf{B} + \mathbf{D} \\ &= \mathbf{C}\frac{\operatorname{adj}\left(s\mathbf{I}_p - \mathbf{A}\right)}{\left|s\mathbf{I}_p - \mathbf{A}\right|}\mathbf{B} + \mathbf{D}\,. \end{aligned} \tag{10.67}$$

The system above is asymptotically stable iff *each and every one of its poles*, that is, the roots of the system's characteristic polynomial $p_c(s) = \left|s\mathbf{I}_p - \mathbf{A}\right|$, has *negative real parts*. In the case that at least one pair of conjugate complex poles have zero real parts (i. e., they are purely imaginary) the system is critically stable. It is noted here that the poles should be calculated as the *eigenvalues* of the system's matrix $\mathbf{A}$; this enables to avoid issues raised by a potential loss of *controllability* or *observability* (examined in the next sections) and are of concern if the poles are calculated as roots of the denominator polynomial of the transfer function.

Returning to (10.66), one may observe that matrix $\mathbf{A}$ for both systems (S1) and (S2) are in companion canonical form. In effect, their characteristic polynomial is calculated as follows

$$\begin{aligned}\text{(S1): } p_c(s) &= s^2 + 0.4s + 1.04 \\ &= (s + 0.2 + \mathrm{i})\,(s + 0.2 - \mathrm{i})\ , \\ \text{(S2): } p_c(s) &= s^2 + 0.0s + 0.09 \\ &= (s + 0.3\mathrm{i})\,(s - 0.3\mathrm{i})\ . \end{aligned}$$

In conclusion (S1) is asymptotically stable because both its poles obtain negative real part: $-0.2 < 0$. On the other hand, system (S2) is critically stable because it obtains poles with zero real part. Finally, consider the SISO system with the following transfer function

$$\begin{aligned}\mathbf{H}(s) &= \frac{1}{s^3 + 0.5s^2 - s + 3} \\ &= [(s + 1.8803)\,(s - 0.6902 + 1.0579\mathrm{i}) \\ &\quad \times (s - 0.6902 - 1.0579\mathrm{i})]^{-1}\ . \end{aligned} \tag{10.68}$$

This system is unstable because it obtains a pair of conjugate poles with positive real part: $0.6902 > 0$. The step response of the system is shown in Fig. 10.8. As can be seen, the response cannot be bounded.

There is a criterion allowing the characterization of an *unstable* system by simple inspection of its characteristic polynomial in reduced form, with no need to determine its roots. The criterion states that any polynomial, the coefficients of which are *real* and exhibit at least one *sign change*, is guaranteed to obtain at least one root in the *right-hand complex plane*. In effect, if the coefficients of the characteristic polynomial of an LTI system exhibit at least one sign change, the system is unstable. The converse, however, is not true in general; it is true in the case of first- and second- order polynomials and systems, only.

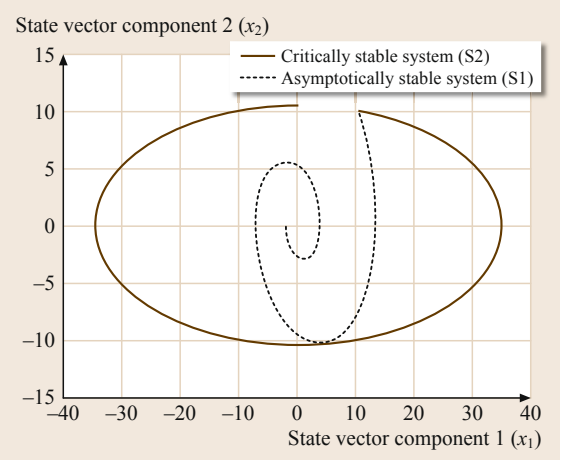


Fig. 10.7 State-space trajectories of critically and asymptotically stable systems

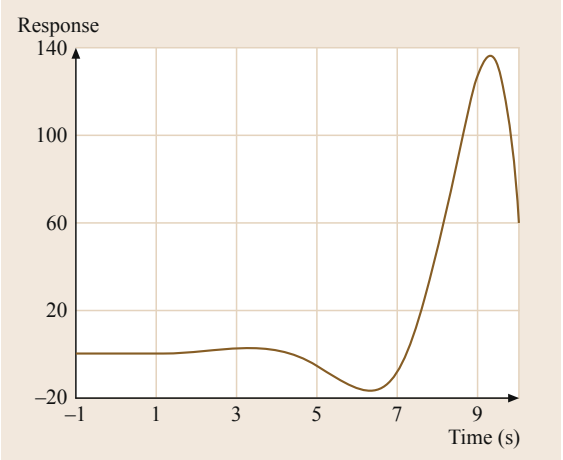


Fig. 10.8 Step response of an unstable system

Consider, for example, the characteristic polynomial of the transfer function in (10.68). This one exhibits two sign changes and therefore obtains roots in the right-hand complex plane. On the other hand, the following third-order polynomial also obtains roots in the right-hand complex plane, that is, with positive real part, although no coefficient sign changes occur

$$\begin{aligned} p(s) &= s^3 + 0.5s^2 + s + 3 \\ &= (s + 1.3694)\,(s - 0.4347 + 1.4148\mathrm{i}) \\ &\quad \times (s - 0.4347 - 1.4148\mathrm{i})\ . \end{aligned} \tag{10.69}$$

10.2.3 Controllability and Observability

The concepts of controllability and observability are fundamental in modern control system theory. For an LTI system with m inputs, p outputs and n states, controllability, and observability may be investigated by using its state-space description.

$$\begin{aligned} \dot{x} &= \mathbf{A}x + \mathbf{B}u\ , \\ y &= \mathbf{C}x + \mathbf{D}u \end{aligned} \tag{10.70}$$

The dimensions for matrix $\mathbf{A}$ are $n \times n$, for $\mathbf{B}$ $n \times m$, for $\mathbf{C}$ $p \times n$, and for $\mathbf{D}$ $p \times m$.

State Controllability

State vector controllability is the situation where there exists a control action vector that can drive the state vector to any finite value in a finite time interval. The formal definition is provided below.

The state vector $\boldsymbol{x}$ of an LTI system as in (10.70) is controllable if there exists a piecewise continuous control signal u that can drive the state vector from any initial condition x_0 to any final value x_f in a finite time interval $(t_\mathrm{f} - t_0)$.

In the case that system's matrix $\mathbf{A}$ in (10.70) is diagonal, one can easily see that, by the definition, matrix $\mathbf{B}$ must contain no rows that are identically zero in order for the LTI system to be controllable.

In the general case that $\mathbf{A}$ is not diagonal, the criterion to check controllability of the state vector for an LTI system, as in (10.70), is based on the rank of the *controllability matrix*, $\mathbf{C}^\dagger$, defined as

$$\mathbf{C}^\dagger = \begin{bmatrix} \mathbf{B} & \mathbf{AB} & \mathbf{A}^2\mathbf{B} & \cdots & \mathbf{A}^{(n-1)}\mathbf{B} \end{bmatrix}. \tag{10.71}$$

If the rank of $\mathbf{C}^\dagger$ is no less than n then the system is controllable, otherwise not. It is reminded here that the *rank* of a matrix is defined as the dimension of the vector subspace spanned by either its row or its column vectors [10.5, 6]. In effect, for an LTI system as in (10.70) to be controllable, its controllability matrix $\mathbf{C}^\dagger$ must contain n linearly independent row vectors or n linearly independent column vectors.

State Observability

State vector observability is the situation where the initial value, at any time instant, of the state vector may be determined solely on the basis of the recorded inputs (controls) and outputs (measurements). The formal definition is provided below.

The state vector $\boldsymbol{x}$ of an LTI system as in (10.70) is observable in the time interval $[t_0, t_\mathrm{f}]$ if it is possible to determine the initial value of the state vector, $x_0 = x(t_0)$, when presented with the values of the control signal u and output signal y in the finite time interval $[t_0, t_\mathrm{f}]$.

In the case that system's matrix $\mathbf{A}$ in (10.70) is diagonal, one can easily see that, by the definition, matrix $\mathbf{C}$ must contain no columns that are identically zero in order for the LTI system to be observable.

In the general case that A is not diagonal, the criterion to check observability of the state vector for an LTI system, as in (10.70), is based on the rank of the *observability matrix*, $\mathbf{O}^\dagger$, defined as

$$\mathbf{O}^\dagger = \begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \\ \mathbf{CA}^2 \\ \vdots \\ \mathbf{CA}^{(n-1)} \end{bmatrix}. \tag{10.72}$$

If the rank of $\mathbf{O}^\dagger$ is no less than n then the system is observable, otherwise not. In effect, for an LTI system as in (10.70) to be observable, its observability matrix $\mathbf{O}^\dagger$ must contain n linearly independent row vectors or n linearly independent column vectors [10.5, 6].

Zero–Pole Cancellation

An important lemma connecting the transfer function (or the transfer function matrix) with the concepts of controllability and observability is given in this section [10.1, 2].

If the *same* system pole (at least one) is cancelled by a zero in *each and every one* of the transfer function matrix elements, then the system in hand is either uncontrollable or unobservable or both.

The lemma above is now investigated in depth for the case of SISO systems. Consider an nth order, SISO system with description in state space as follows

$$\begin{aligned} \dot{x} &= \mathbf{A}x + \boldsymbol{b}u\ , \\ y &= \boldsymbol{c}^\mathrm{T}x\ . \end{aligned} \tag{10.73}$$

In the above, $\boldsymbol{b}$ and $\boldsymbol{c}$ are n-dimensioned column vectors. For simplicity, assume that $\mathbf{A}$ in (10.73) is diagonal.

$$\mathbf{A} = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} \tag{10.74}$$

The (scalar) transfer function of system is, then, given by the following

$$\mathbf{H}(s) = \frac{\mathbf{Y}(s)}{\mathbf{U}(s)} = \boldsymbol{c}^{\mathrm{T}} (s\mathbf{I}_n - \mathbf{A})^{-1} \boldsymbol{b} \,. \tag{10.75}$$

Based on the vector dynamic equation (10.73) and the diagonal form of matrix $\mathbf{A}$ it can be concluded that

$$\mathbf{X}_i(s) = \frac{b_i}{s - \lambda_i} \mathbf{U}(s) \,. \tag{10.76}$$

Furthermore, using the algebraic output equation in (10.73), the following is obtained

$$\begin{aligned} \mathbf{Y}(s) &= \sum_{i=1}^{n} c_i \mathbf{X}_i(s) = \sum_{i=1}^{n} \frac{c_i b_i}{s - \lambda_i} \mathbf{U}(s) \\ \Rightarrow \mathbf{H}(s) &= \sum_{i=1}^{n} \frac{c_i b_i}{s - \lambda_i} \,. \end{aligned} \tag{10.77}$$

If the expansion of the transfer function in (10.77) is compared against the one on in (10.22), that has been obtained through partial fraction expansion, it can be concluded that they are identical, provided that $k_i \equiv c_i b_i$. Therefore, if one of the poles, λ_i, of the system, which are determined as the eigenvalues of matrix $\mathbf{A}$, is absent from (10.22), this means that $k_i \equiv c_i b_i = 0$. In effect, this is due to either $b_i = 0$, or $c_i = 0$, or both. Taking into account that matrix $\mathbf{A}$ is diagonal, it is concluded that the system has to be either uncontrollable or unobservable or both.

In effect, the transfer function matrix of any LTI system encompasses information only for that part of the system which is both controllable and observable. As a result, it is a well-defined description of the system only if controllability and observability are guaranteed.

The importance of the above is demonstrated in the following example. Consider a 2-inputs, 2-outputs, and 3-states system with matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$ as follows

$$\mathbf{A} = \begin{bmatrix} -2 & 0 & 0 \\ 0 & -0.4 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$$\mathbf{B} = \mathbf{C}^{\mathrm{T}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{D} = \mathbf{0}_{2\times 2} \,.$$

This system is evidently both uncontrollable and unobservable. Due to the diagonal form of matrix $\mathbf{A}$, poles are easily determined to be -2, -0.4, and $+1$. Therefore, the system is unstable, too. However, the 2×2 transfer function matrix is as follows

$$\mathbf{H}(s) = \frac{1}{s^2 + 2.4s + 0.8} \begin{bmatrix} s + 0.4 & 0 \\ 0 & s + 2 \end{bmatrix} .$$

Based solely on the transfer function, one might erroneously characterize the system to be a second order and stable one, as the apparent characteristic polynomial is determined to be the following,

$$p(s) = s^2 + 2.4s + 0.8 = (s + 0.4)\,(s + 2) \,.$$

However, the correct characteristic polynomial is the following,

$$\begin{aligned} p_c(s) &= |s\mathbf{I}_n - \mathbf{A}| = s^3 + 1.4s^2 - 1.6s - 0.8 \\ &= (s + 0.4)\,(s + 2)\,(s - 1) \,. \end{aligned}$$

Therefore, the system is actually third order and unstable. This simple example shows, in effect, that in order to arrive to safe conclusions for the system, especially concerning stability, by using solely the transfer function controllability and observability must be guaranteed in advance.

10.2.4 Sinusoidal Steady-State Response and Bode Plots

In Sect. 10.1.3, the impulse and the step response of any LTI system have been introduced. Equally important is the response of an LTI system when sinusoidals as applied as inputs.

When presented with the transfer function of an LTI–SISO system, the sinusoidal steady-state response, that is, its response when driven with a pure sine wave of arbitrary frequency ω and unity amplitude, can be calculated by replacing $s = \mathrm{i}\omega$ in the transfer function. Referring to the I/O relationship in (10.10), it holds that

$$\begin{aligned} u(t) = \sin(\omega t) \Rightarrow y(t) &= y_{\mathrm{SS}}(t) \\ &= |\mathbf{H}(\mathrm{i}\omega)| \sin(\omega t + \angle H(\mathrm{i}\omega)) \,. \end{aligned} \tag{10.78}$$

In the above, $|\bullet|$ stands for a complex number's magnitude (or modulus) and $\angle\bullet$ for a complex number's phase (or argument).

The graphs of the real functions

$$A(\omega) = 20 \log |\mathbf{H}(\mathrm{i}\omega)|$$

and

$$\varphi(\omega) = \angle \mathbf{H}(\mathrm{i}\omega)$$

for real and positive ω ($\omega \geq 0$) on plots with the x-axis, where the angular frequency ω is depicted, in logarithmic scale are called magnitude and phase Bode plots, respectively.

For example, in Fig. 10.9, the Bode plots of a third-order system are shown. The transfer function of the system is as follows

$$\mathbf{H}(s) = \frac{s-1}{s^3+2.2s^2+5.4s+1} = \frac{s-1}{(s+0.2)\,(s+1+2\mathrm{i})\,(s+1-2\mathrm{i})}\,.$$

Approximate (or Asymptotic or Corner) Bode Plots

The Bode plots are widely used for frequency response characterization mainly because they can be drawn approximately on the basis of a limited set of rules of thumb. Such approximate plots, also called asymptotic or corner plots, start with bringing the transfer function of the system to the following factorized form

$$\mathbf{H}(s) = K\frac{(1+\tau_1 s)\,(1+\tau_2 s)\cdots(1+\tau_m s)}{s^q\,(1+T_1 s)\,(1+T_2 s)\cdots(1+T_n s)}\,. \tag{10.79}$$

As a working example, consider the following third-order transfer function

$$\mathbf{H}(s) = K\omega_0^2\frac{(1+\tau s)}{s\,(1+Ts)\left(s^2+2\zeta\omega_0 s+\omega_0^2\right)}\,. \tag{10.80}$$

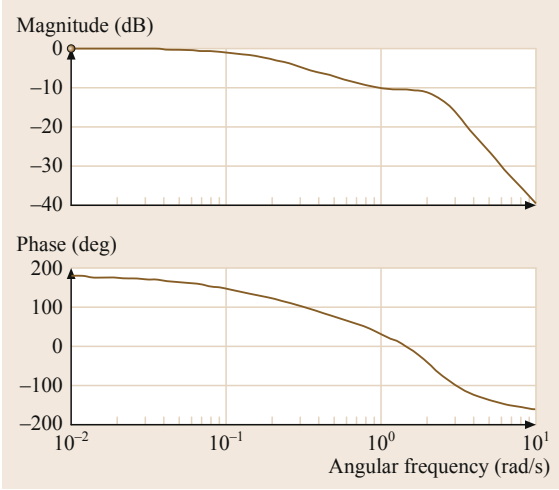


Fig. 10.9 Typical magnitude and phase Bode plots for an LTI system

The above is augmented, with respect to the transfer function in (10.79), by the quadratic factor $\left(s^2+2\zeta\omega_0 s+\omega_0^2\right)$ for which it is assumed that $\omega_0 > 0$ and $|\zeta| < 1$. This factor introduces two complex conjugate poles

$$p_{1,2} = \omega_0\left(-\zeta \pm \mathrm{i}\sqrt{1-\zeta^2}\right),$$

and therefore its investigation must be done with caution.

Substituting $s = \mathrm{i}\omega$ in (10.80) the following is obtained

$$\mathbf{H}(\mathrm{i}\omega) = K\frac{(1+\tau\omega\mathrm{i})}{(\omega\mathrm{i})\,(1+T\omega\mathrm{i})\left[1-\left(\frac{\omega}{\omega_0}\right)^2+\mathrm{i}2\zeta\left(\frac{\omega}{\omega_0}\right)\right]}\,. \tag{10.81}$$

To draw the magnitude and phase Bode plots the following manipulation is needed on the basis of (10.81),

$$\begin{aligned} A(\omega) &= 20\log|\mathbf{H}(\mathrm{i}\omega)| \\ &= 20\log|K| + 20\log|1+\tau\omega\mathrm{i}| - 20\log\omega \\ &\quad - 20\log|1+T\omega\mathbf{i}| \\ &\quad - 20\log\left|1-\left(\frac{\omega}{\omega_0}\right)^2+\mathrm{i}2\zeta\left(\frac{\omega}{\omega_0}\right)\right|\,. \end{aligned} \tag{10.82}$$

$$\begin{aligned} \varphi(\omega) &= \angle\mathbf{H}(\mathrm{i}\omega) \\ &= \angle K + \angle(1+\tau\omega\mathrm{i}) - \angle\omega \\ &\quad - \angle(1+T\omega\mathrm{i}) \\ &\quad - \angle\left[1-\left(\frac{\omega}{\omega_0}\right)^2+\mathrm{i}2\zeta\left(\frac{\omega}{\omega_0}\right)\right]\,. \end{aligned} \tag{10.83}$$

Therefore, it is concluded that in order to draw the Bode plots of a transfer function it suffices to graphically add up the various components appearing as terms in (10.82) and (10.83). The magnitude and phase contribution of each one of these components to the overall Bode plots is examined below.

Constant Gain K. The contribution of a constant to both Bode plots of a transfer function is a straight line with zero slope as follows

$$A(\omega) = 20\log|K| \text{ and } \varphi(\omega) = \begin{cases} 0^\circ\,, & K \geq 0 \\ 180^\circ\,, & K < 0\,. \end{cases} \tag{10.84}$$

Factor $(\mathrm{i}\omega)$. For the magnitude Bode plot of a transfer function that contains this factor as part of its denominator, the contribution consists of a straight line with slope $-20\,\mathrm{dB/dec}$ (decibels per decade). For the phase Bode plot of a transfer function that contains this factor as part of its denominator, the contribution consists of a horizontal straight line as follows

$$A(\omega) = -20\log\omega \quad \text{and} \quad \varphi(\omega) = -90^\circ\,. \tag{10.85}$$

In the case that the pole has multiplicity $q > 1$, as for example in (10.79), then (10.85) is modified accordingly.

$$\begin{aligned} A(\omega) &= -20q\log\omega \quad \text{and} \\ \varphi(\omega) &= -q90^\circ \end{aligned} \tag{10.86}$$

Finally, if factor $(\mathrm{i}\omega)$ appears on the numerator instead the denominator of the transfer function the above still hold but the sign has to be switched to positive.

Factor $(1+\mathrm{i}\omega T)$. Calculation of the contribution of this factor, when appearing as part of the denominator of the transfer function, is done on the basis of the following approximations

$$\begin{aligned} A(\omega) &= -20\log|1+\mathrm{i}\omega T| \\ &= -20\log\sqrt{1+\omega^2T^2} \\ &\approx \begin{cases} 0 & \omega T \ll 1 \\ -3 & \omega T \simeq 1 \\ -20\log(\omega T) & \omega T \gg 1 \end{cases}, \end{aligned} \tag{10.87}$$

$$\begin{aligned} \varphi(\omega) &= -\arctan(\omega T) \\ &\approx \begin{cases} 0^\circ & \omega T < 0.1 \\ -90^\circ & \omega T > 10.0 \end{cases}. \end{aligned} \tag{10.88}$$

As an example, consider the following first-order transfer function

$$\mathbf{H}(s) = \frac{1}{s+1} \Rightarrow \mathbf{H}(\mathrm{i}\omega) = \frac{1}{1+\mathrm{i}\omega}\,. \tag{10.89}$$

In the above, evidently $T = 1.0$.

The Bode plots of this transfer function are shown in Fig. 10.10. It is clear that these graphs can be approximated by straight lines. In this respect, the magnitude plot consists of a horizontal line until the breakpoint (or corner point) corresponding to $\omega T = 1.0$; beyond that point it can be approximated by a straight line with a slope of $-20\,\mathrm{dB/dec}$. The phase plot can be approximated by horizontal lines at the level of 0° for $\omega T < 0.1$ and -90° for $\omega T > 10.0$. These sections can then be connected with an appropriately negatively sloped straight line.

Finally, these remarks can be easily extended for the cases where factor $(1+\mathrm{i}\omega T)$ appears in the numerator of the transfer function or obtains multiplicity higher than 1.

Factor $\left[1-(\omega/\omega_0)^2+\mathrm{i}2\zeta(\omega/\omega_0)\right], |\zeta| < 1$. Calculation of the contribution of this factor, when appearing as part of the denominator of the transfer function, is done on the basis of the following approximations

$$\begin{aligned} A(\omega) &= -20\log\left|1-\left(\frac{\omega}{\omega_0}\right)^2+j2\zeta\left(\frac{\omega}{\omega_0}\right)\right| \\ &= -20\log\sqrt{\left[1-\left(\frac{\omega}{\omega_0}\right)^2\right]^2+4\zeta^2\left(\frac{\omega}{\omega_0}\right)^2} \\ \Rightarrow A(\omega) &= \begin{cases} 0\,, & \omega \ll \omega_0 \\ -20\log(2|\zeta|)\,, & \omega \simeq \omega_0 \\ -40\log\left(\frac{\omega}{\omega_0}\right), & \omega \gg \omega_0 \end{cases}, \end{aligned} \tag{10.90}$$

$$\begin{aligned} \varphi(\omega) &= -\arctan\left[\frac{2\zeta\left(\frac{\omega}{\omega_0}\right)}{1-\left(\frac{\omega}{\omega_0}\right)^2}\right] \\ \Rightarrow \varphi(\omega) &= \begin{cases} \approx 0^\circ\,, & \omega \ll \omega_0 \\ -90^\circ\,, & \omega = \omega_0 \\ \approx -180^\circ\,, & \omega \gg \omega_0 \end{cases}. \end{aligned} \tag{10.91}$$

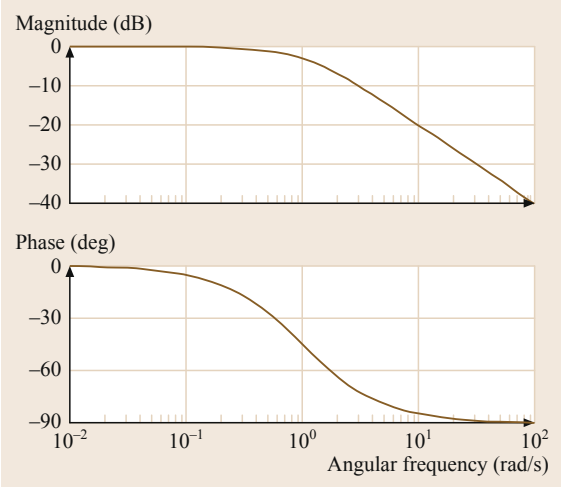


Fig. 10.10 Magnitude and phase Bode plots of a first-order factor in the denominator

The dependence of the corresponding plots on parameter ζ (known as the *damping factor* or *coefficient*) is shown in Fig. 10.11. Evidently, the smaller the damping coefficient is, the higher the intensity of resonance. By resonance, we mean the phenomenon according to which the system exhibits response of significantly amplified amplitude when driven by a sinusoidal of a specific frequency, known as the resonance frequency, $\omega = \omega_0$. In the extreme case that $\zeta = 0$, the output's amplitude is infinite. On the other hand, when $\zeta = 1$ the system obtains a double real pole located at $\omega = \omega_0$ and the analysis can be done as in the case of a first-order factor $(1 + i\omega T)$ with multiplicity equal to two.

The second-order factor Bode plots can be approximated by straight lines as in the case of factor $(1 + i\omega T)$, especially for frequencies far from the resonance point $\omega = \omega_0$. For $\omega \ll \omega_0$ a horizontal straight line is a good approximation while for $\omega \gg \omega_0$ a straight line with slope $-40\,\mathrm{dB/dec}$ represents the frequency dependence of the magnitude Bode plot. For the vicinity of the resonance point $\omega = \omega_0$, the intensity of the resonance, depending on the value of ζ, needs to be estimated. For the phase plot similar considerations with the first-order case hold; the final value, though, is now $-180°$ instead of $-90°$. Furthermore, the smaller the absolute value of the damping coefficient, the steeper the phase plot of the factor.

10.2.5 Analysis of Second-Order Systems

A large class of practical systems encountered in control engineering is either of first or second order.

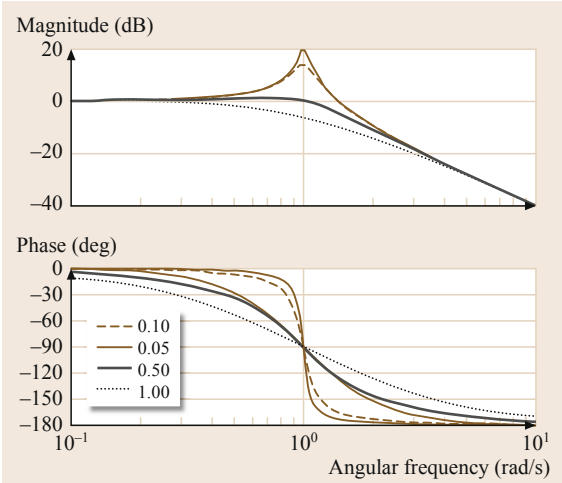


Fig. 10.11 Bode plots of a second-order factor in the denominator for various values of ζ

Furthermore, many other higher order systems can be approximated or *reduced* to first or second order ones, provided that certain conditions are valid. The importance of first- and second-order LTI systems is additionally emphasized by the fact that their poles can be conveniently calculated as functions of the coefficients of their characteristic polynomial.

In this section, the analysis of a generic second-order system without zeros, as in the transfer function below, is presented.

$$\mathbf{H}(s) = \frac{\omega_0^2}{s^2 + 2\zeta\omega_0 s + \omega_0^2} \tag{10.92}$$

The parameter $\omega_0 \geq 0$ is referred to as the eigenfrequency of the system and parameter ζ as the coefficient (or damping) coefficient.

The poles of the system are given by the expression

$$p_{1,2} = \omega_0 \left(-\zeta \pm \sqrt{\zeta^2 - 1}\right). \tag{10.93}$$

In the case that $|\zeta| < 1$, parameter $\omega_\mathrm{n} = \omega_0\sqrt{1-\zeta^2}$ is defined as the natural frequency of the system.

The step response of the generic second-order system is given by the following relationship in the complex frequency domain, if (10.19) is employed

$$\mathbf{Y}_\mathrm{step}(s) = \mathbf{H}(s)\mathbf{U}_\mathrm{step}(s) = \frac{\omega_0^2}{s\left(s^2 + 2\zeta\omega_0 s + \omega_0^2\right)}. \tag{10.94}$$

The analysis of the second-order system in hand will proceed on the basis of the following cases for the value of the damping coefficient. As shown, this coefficient plays a crucial role in determining the location of the system poles on the s-plane and effectively its stability and transient behavior:

a) $\zeta \leq -1$: Two positive purely real poles (or a single double one if $\zeta = -1$) – Unstable system
Step response (see Fig. 10.12).

b) $-1 < \zeta < 0$: Two complex conjugate poles on the right-hand s-plane – Unstable system
Step response (see Fig. 10.13).

c) $\zeta = 0$: Two conjugate purely imaginary poles – Critically stable system
Step response (Fig. 10.14)

$$y_\mathrm{step}(t) = 1 - \cos(\omega_0 t). \tag{10.95}$$

d) $0 < \zeta < 1$: Two complex conjugate poles on the left-hand s-plane – Stable system

Step response (Fig. 10.15)

$$y_{\text{step}}(t) = 1 - e^{-\sigma t}\cos(\omega_n t) - \frac{\sigma}{\omega_n} e^{-\sigma t}\sin(\omega_n t)$$
$$= 1 - \frac{e^{-\sigma t}}{\sqrt{1-\zeta^2}}\sin(\omega_n t + \theta) \ , \tag{10.96}$$

where

$$\sigma = \text{Re}\{p_{1,2}\} = -\zeta\omega_0 \quad \text{and}$$
$$\theta = \arctan\left(\frac{\sqrt{1-\zeta^2}}{\zeta}\right).$$

In Fig. 10.15, it is of interest to calculate the time (at which occurs) and the value of maximum overshoot.

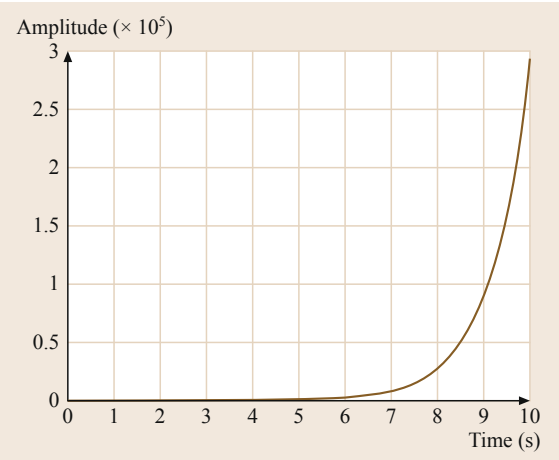


Fig. 10.12 Second-order system step response when $\zeta \leq -1$

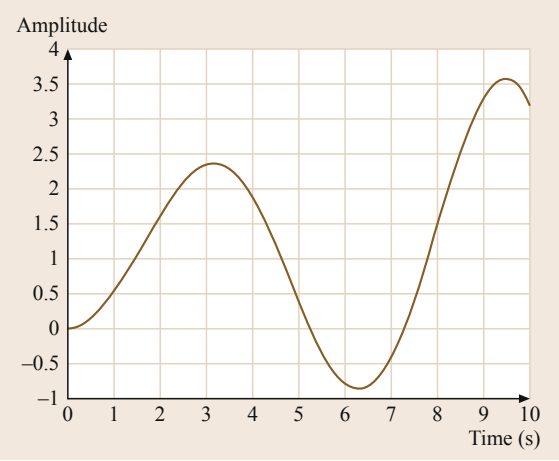


Fig. 10.13 Second-order system step response when $-1 < \zeta < 0$

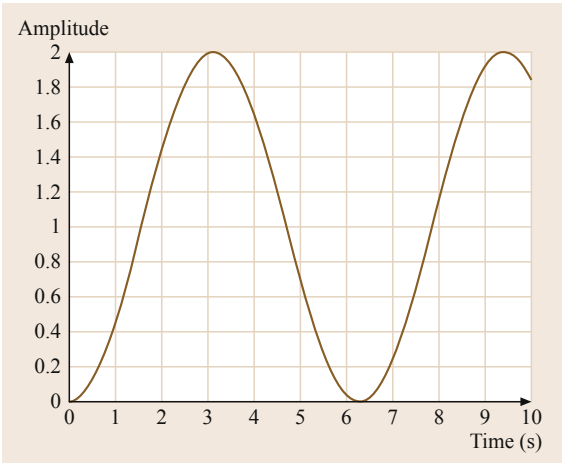


Fig. 10.14 Second-order system step response when $\zeta = 0$

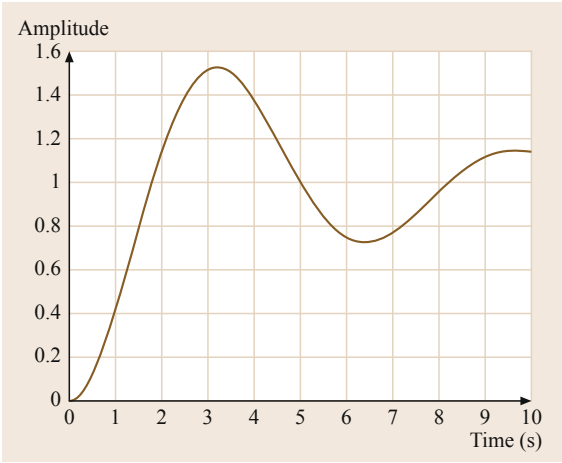


Fig. 10.15 Second-order system step response when $0 < \zeta < 1$

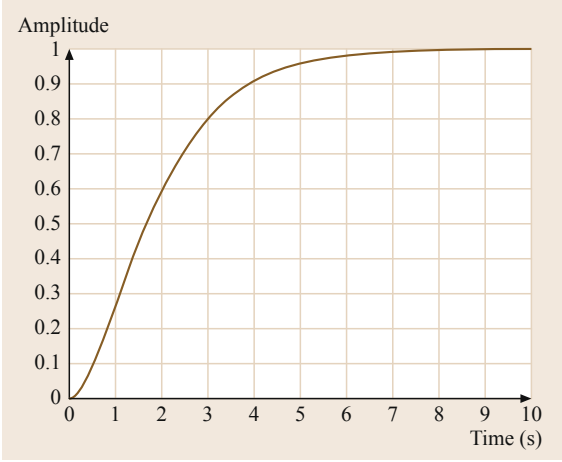


Fig. 10.16 Second-order system step response when $\zeta = 1$

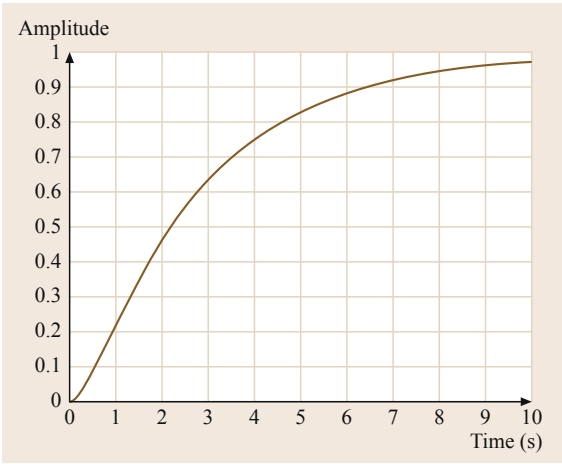


Fig. 10.17 Second-order system step response when $\zeta > 1$

This is achieved by differentiating the expression for the step response in (10.94). The derivative calculated is then equated to zero. The time of the first maximum's occurrence is the time instant of maximum overshoot t_{max} given as

$$t_{max} = \frac{\pi}{\omega_n} . \tag{10.97}$$

e) $\zeta = 1$: One double, purely real pole in the right-hand s-plane – Stable system
Step response (Fig. 10.16)

$$y_{step}(t) = 1 - e^{-\omega_0 t} - \omega_0 e^{-\omega_0 t} . \tag{10.98}$$

f) $\zeta > 1$: Two negative purely real poles – Stable system

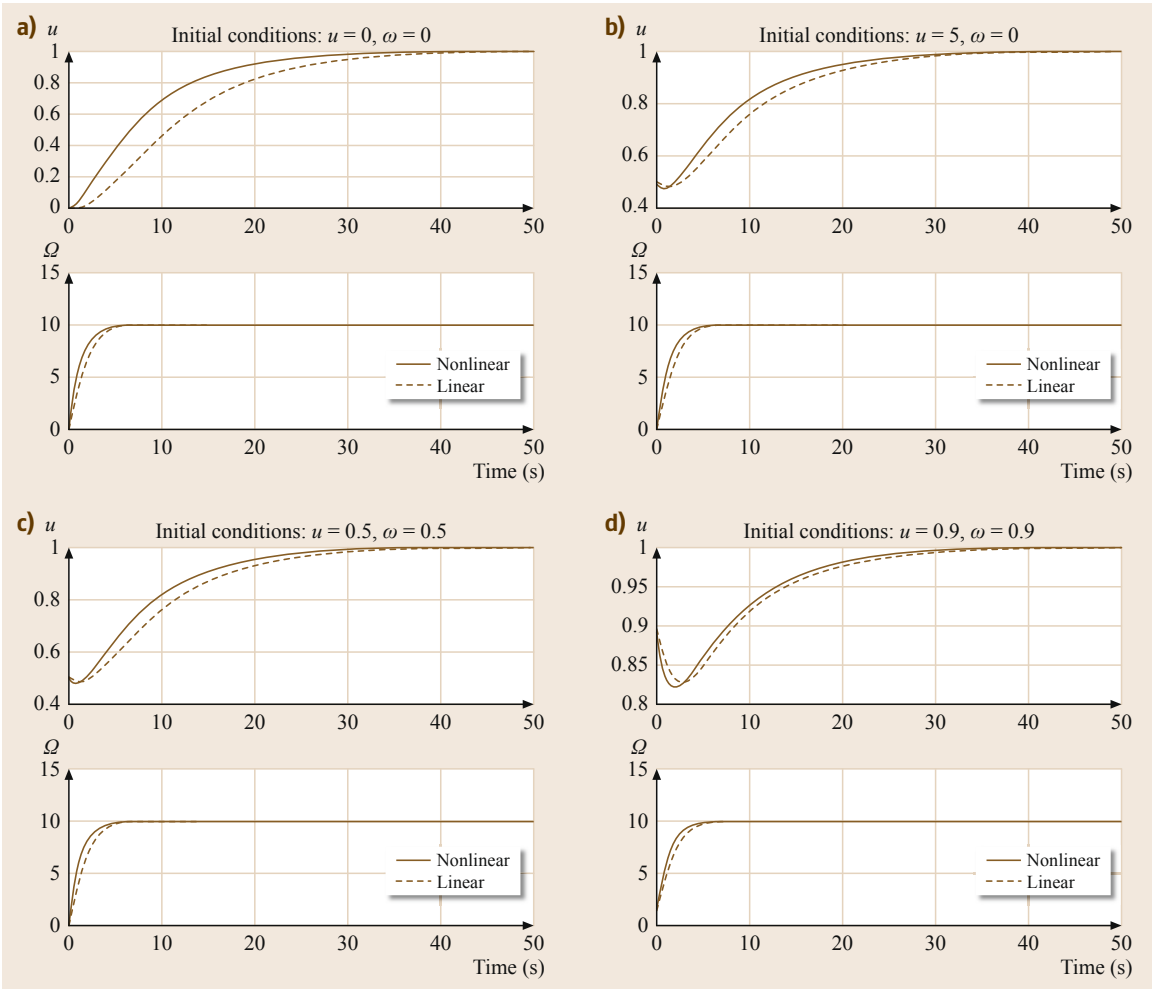


Fig. 10.18a–d Unit responses of the forward AUV speed model for four different initial conditions

Step response (Fig. 10.17)

$$y_{\text{step}}(t) = 1 - e^{-\sigma t}\cosh(at) - \frac{\sigma}{a}e^{-\sigma t}\sinh(at) \ , \tag{10.99}$$

where $\sigma = \text{Re}\{p_{1,2}\} = -\zeta\omega_0$ and $a = \omega_0\sqrt{\zeta^2 - 1}$. The Bode plots for various positive values of $\zeta > 0$ (therefore asymptotically stable cases) are given in Fig. 10.11.

Example 10.3 Linearized AUV Forward Speed Model

In this example, we will analyze the linearized AUV forward speed model introduced in Example 10.2.

Given $d = 20$, $M = 300$, $c_T = 0.2$, $\beta = 50$, $\alpha = 0.04$, $u_0 = 1$, $\Omega = 10$, the linearized model is

$$300\dot{u} + 40u = 4\Omega \ ,$$
$$\dot{\Omega} + 0.8\Omega = 50\tau + 4$$

and the transfer functions are expressed as follows using (10.67)

$$\mathbf{H}_U(s) = \frac{\mathbf{U}(s)}{\tau(s)} = \frac{200}{(300s + 40)(s + 0.8)} \ ,$$
$$\mathbf{H}_\Omega(s) = \frac{\Omega(s)}{\tau(s)} = \frac{50}{(s + 0.8)} \ ,$$

where $u(t) \leftrightarrow \mathbf{U}(s)$, $\Omega(t) \leftrightarrow \Omega(s)$, $\tau(t) \leftrightarrow \tau(s)$. These models are simulated in MATLAB/Simulink with a unit step of 0.08 Nm as the input torque. Figure 10.18 shows the simulated responses based on the nonlinear and linearized models at four different initial conditions.

One can clearly see that the linearized model becomes more accurate as the initial conditions move toward the point of linearization ($u_0 = 1$, $\Omega_0 = 10$). Also, the propeller response is visually more accurate with a shorter time constant as compared to the speed response.

10.3 SISO System Controls

This chapter presents some fundamental techniques for the control of LTI–SISO systems and basic performance criteria typically used in controller design. The techniques consist of (a) ON/OFF control; (b) PID (proportional-integral-derivative) control, and (c) controller design by use of the root locus.

The closed-loop control system given above is encountered in many practical, industrial plants and processes. The objective in introducing feedback control is to maintain the process response signal y as close as possible to the reference signal (setpoint) r. Deflection of the process output signal from the reference one is generated because either there is significant disturbance present, manifested by signal d, or the setpoint is modified or both. The first problem is referred to as the *regulation problem* while the second as the *tracking problem*.

In any of these cases, however, the controller (control system) is inserted in order to reduce the error signal, $e = r - y$, that is, the deviation of the output from the reference signal, to the lowest possible value; this can be achieved by appropriately processing the error signal in order to generate the driving control signal, that will compensate the effect of the disturbance. The block diagram in Fig. 10.19 is similar to the one shown in Fig. 10.2; for the sake of simplicity, however, the sensor and actuator blocks have been omitted, or equivalently included in the controller and process block, respectively, and the plant is assumed to be SISO.

In the complex frequency domain, the output signal is given by the relationship

$$\left.\begin{aligned}\mathbf{Y}(s) &= \mathbf{G}(s)\left[\mathbf{U}(s) + \mathbf{D}(s)\right] \\ \mathbf{U}(s) &= \mathbf{K}(s)\mathbf{E}(s) = \mathbf{K}(s)\left[\mathbf{R}(s) - \mathbf{Y}(s)\right]\end{aligned}\right\} \Rightarrow$$
$$\Rightarrow \mathbf{Y}(s) = \frac{\mathbf{G}(s)\mathbf{K}(s)}{1 + \mathbf{G}(s)\mathbf{K}(s)}\mathbf{R}(s) + \frac{\mathbf{G}(s)}{1 + \mathbf{G}(s)\mathbf{K}(s)}D(s) \ . \tag{10.100}$$

The above can be alternatively demonstrated by separately applying the general equation (10.61) for each one of the signals r and d and setting the other one identically equal to zero.

10.3.1 Performance Criteria

There are three primary criteria for which the performance of a system can be assessed. They are: transient response, steady-state response, and disturbance rejection response. These criteria are generally used for the setpoint control objective.

A *transient response* corresponds to how a system reacts to a new desired output. The associated criteria are usually defined in terms of:

- Delay time (t_d): the time it takes to reach 30% of the setpoint level

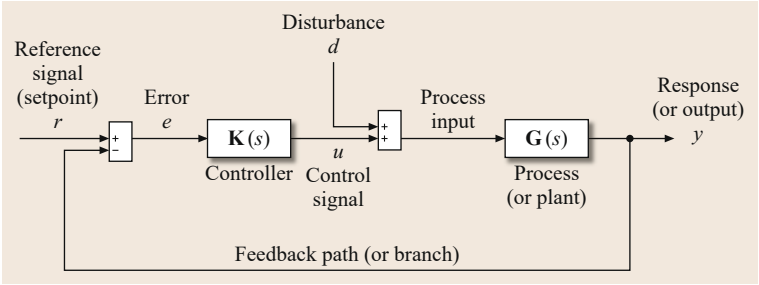


Fig. 10.19 Typical block diagram of a closed-loop system

- Rise time (t_r): the time it takes to reach 90% of the setpoint level
- Peak time (t_p): the time it takes to reach the peak value, that is, value ($A + B$)
- Maximum overshoot (M_p): a ratio of the amount of overshoot (A) to the setpoint (B)
- Settling time (t_s): the time it takes to reach within a predefined threshold around the setpoint (e.g., ± 3 or 5%).

Figure 10.20 shows a unit step response with the criteria depicted graphically. It is customary that the controller designs are based on the requirements associated with these criteria. For linear systems, their open-loop system responses can be decomposed into a combination of first-order (exponential) and second-order (oscillatory) responses, that is, factors in the transfer function as follows

$$\mathbf{H}_1(s) = \frac{1}{s+\alpha}\,,$$
$$\mathbf{H}_2(s) = \frac{1}{s^2 + 2\zeta\omega_n s + \omega_n^2}\,. \qquad (10.101)$$

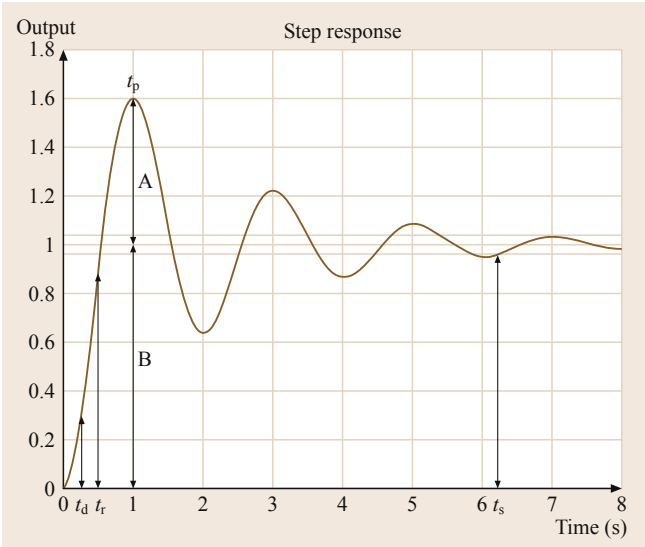


Fig. 10.20 A typical unit step response of an LTI system

In the above $\mathbf{H}_1(s)$ and $\mathbf{H}_2(s)$ correspond to the first- and second-order components. If the overall system response is dominated by its second-order component with an oscillatory response, the design requirements can be directly translated into desired natural frequency (ω_n) and damping ratio (ζ) requirements. For linear, pole-only systems, the translation can be performed using the formulas

$$t_r = \frac{\pi - \tan^{-1}\left(\frac{\sqrt{1-\zeta^2}}{\zeta}\right)}{w_d} \approx \frac{1.8}{\omega_n} \quad (\text{for } \zeta = 0.5)\,,$$
$$t_s \approx \frac{4.6}{\zeta\omega_n} \quad (\text{for } 1\% \text{ specification})\,,$$
$$M_p = \mathrm{e}^{-(\zeta/\sqrt{1-\zeta^2})\pi}\,. \qquad (10.102)$$

For an example, if the rise time must be < 1 s, maximum overshoot $< 16\%$, and the settling time less than 3 s, then the damping ratio must be > 0.5 (based on maximum overshoot) and the natural frequency > 1.8 rad/s (based on rise time) or 3.1 rad/s (based on settling time). Thus, the final design requirements are $\omega_n > 3.1$ rad/s, $\zeta > 0.5$.

A *steady-state response* corresponds to how well a system can maintain its desired output while incurring minimal offset. In many applications, the steady-state error, which is the difference between the desired output and the actual system output, is required to fall within a threshold. This can be related to the settling time requirement mentioned previously.

A *disturbance rejection response* corresponds to how well a system rejects any unexpected noise. The controller must be capable of rejecting the noise while achieving the transient and steady-state requirements.

10.3.2 ON/OFF Control

ON/OFF control is the simplest control technique and widely used in various practical systems. Typical applications include thermostats; two-positions pressure regulators; pneumatic and hydraulic actuator valves like the ones used, for example, in marine steering gears, etc.

The control law, that is, the relationship connecting the error to the control signal as implemented in the controller, is

$$u = \begin{cases} A\,\mathrm{sgn}(e)\,, & |e| \geq Z_d \\ 0\,, & |e| < Z_d \end{cases} = \begin{cases} +A\,, & e \geq +Z_d \\ 0\,, & |e| < Z_d \\ -A\,, & e \leq -Z_d \end{cases} . \tag{10.103}$$

In the above, A is a constant and $Z_d \geq 0$ a *threshold value* for the absolute error $|e|$ at which the control signal is activated. The region $|e| < Z_d$ for which the control signal is zero is called the *dead zone*; the dead zone need not be necessarily symmetrical around zero as shown in Fig. 10.21 for an ON/OFF controller with $A = 1$. Furthermore, this figure as well as (10.103) refer to controllers with *three switch positions*. Commonly, the controller switch may have only two positions, as indicated by the name ON/OFF, or four or more. However, the same features are shared by all controllers whose control law is based on switching as in (10.103); therefore, the analysis is general enough.

An important feature of all ON/OFF controllers is that the associated control law is nonlinear and, in effect, cannot be investigated with tools from linear control theory, like transfer functions (Fig. 10.22). In this end, an example is presented.

The step response of the closed-loop system for various values of parameter A and no dead zone, that is, $Z_d = 0$, is shown in Fig. 10.23. Starting with $A = 1$, one can easily see that the response is the same with the one obtained when the switch is not intermitted between the summing junction and the process. For $A = 0.1$, on the other hand, the response exhibits large steady-state error (≈ 0.9) from the setpoint value ($= 1$). For $A = 2$, one obtains a very good, from a practical viewpoint, step response as it approaches the setpoint in almost half the time needed in the case of $A = 1$. However, if one attempts to further reduce the *rise time*, by simply increasing A beyond 2, *limit cycle oscillations* of the control signal are induced. This is highly undesirable as it can cause various elements in the system (e.g., the actuator) to rapidly wear and fail.

In practical systems, dead zone cannot be exactly zero; this is mainly due to the fact that the control loop elements are not ideal and demonstrate finite sensitivity. Even a small dead zone may sever the limit cycle phenomenon. This is clearly seen in Fig. 10.24a, where $A = 2$ and the dead zone is just 3% of the absolute maximum error e observed (which is equal to 1.0 at start time). Furthermore, a relatively large dead zone is often deliberately introduced in order to reduce the frequency of the limit cycle oscillations. This is seen in Fig. 10.24b

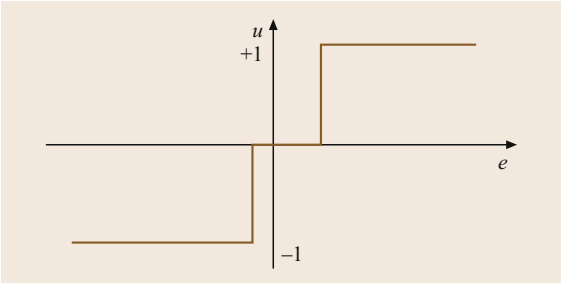


Fig. 10.21 General form of an ON/OFF controller with three positions switch

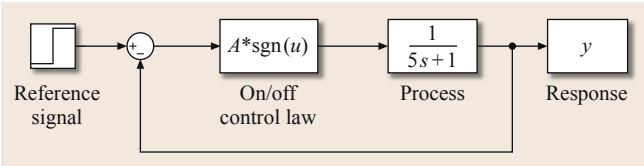


Fig. 10.22 Example for the analysis of ON/OFF controllers

where $Z_d = 0.3$ (30% of the absolute maximum error observed). However, this comes at the price of increasing the error in the response; indeed, the final value of the response as can be seen in Fig. 10.24 is ≈ 0.7, that is, 30% off the setpoint.

10.3.3 PID Control

The proportional-integral-differential (PID) control law was introduced as an alternative to ON/OFF control because it actually deals with practical control issues of SISO plants in a much improved manner.

The PID control law is commonly encountered in the following alternative forms (reference of symbols is Fig. 10.25):

(i) The analytic form which is also used in the rest of this text. The expressions for the control law in both the time and complex frequency domain are given below.

$$\begin{aligned} u(t) &= K_\mathrm{p} e(t) + K_\mathrm{i} \int_0^t e(\xi)\,\mathrm{d}\xi + K_\mathrm{d} \frac{\mathrm{d}}{\mathrm{d}t} e(t) \\ &\Updownarrow \\ \mathbf{U}(s) &= \mathbf{K}(s)\mathbf{E}(s) = \left(K_\mathrm{p} + \frac{K_\mathrm{i}}{s} + K_\mathrm{d} s \right) \mathbf{E}(s) \\ &= \frac{K_\mathrm{i} + K_\mathrm{p} s + K_\mathrm{d} s^2}{s} \mathbf{E}(s) \end{aligned} \tag{10.104}$$

In the analytic form, the control signal depends exclusively on the error signal. Furthermore, note that (a) the P-control law is a static system; (b) the PID

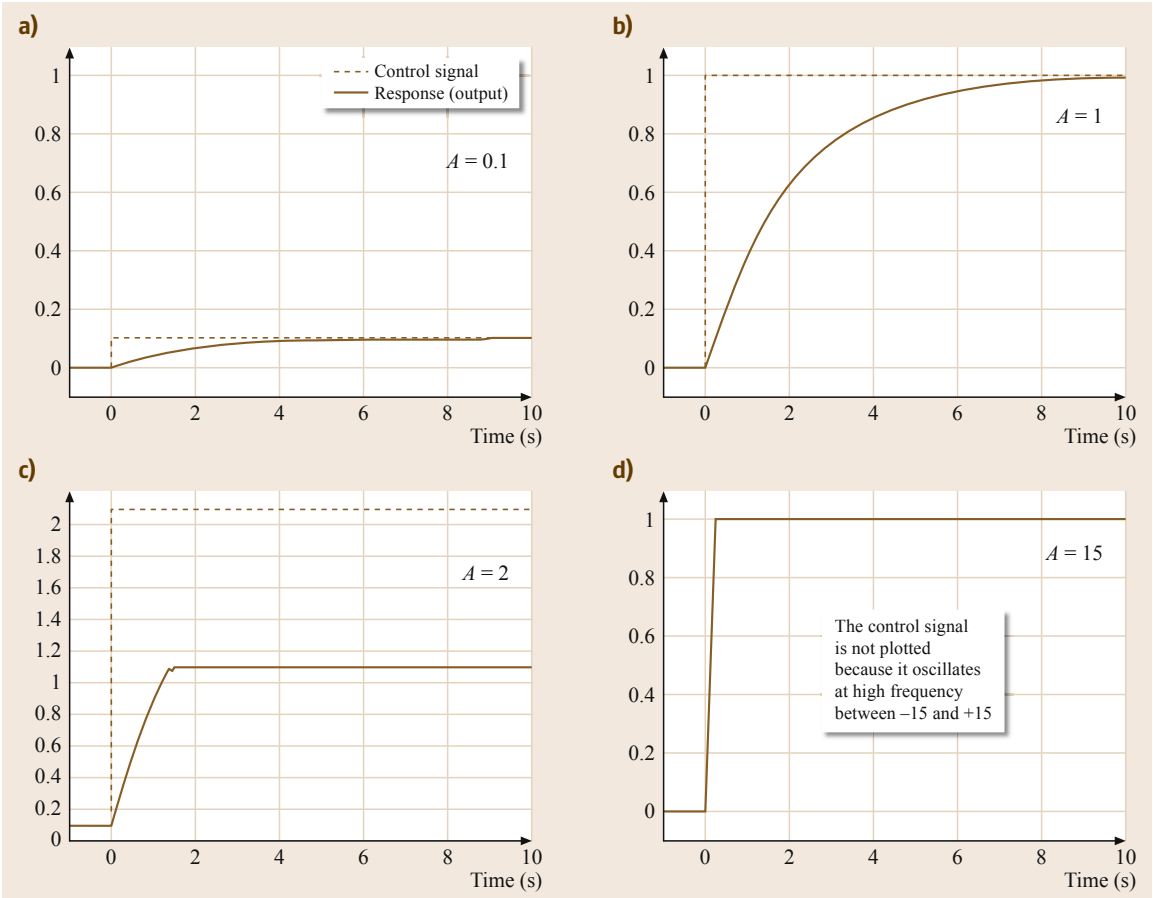


Fig. 10.23a–d Response with ON/OFF control various values of $A = 2$ and $Z_d = 0$

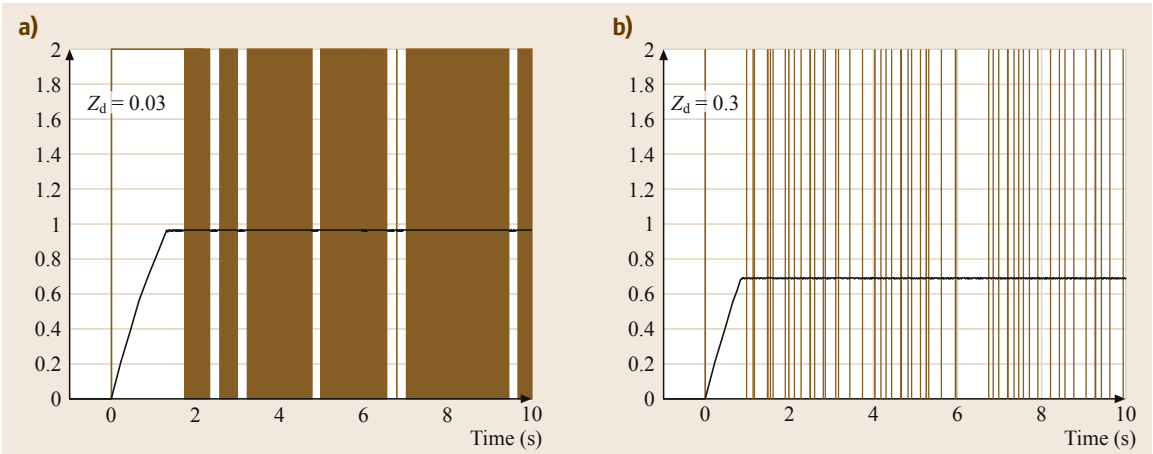


Fig. 10.24a,b Response with ON/OFF control for $A = 2$ and $Z_d = 0.03$ and 0.3

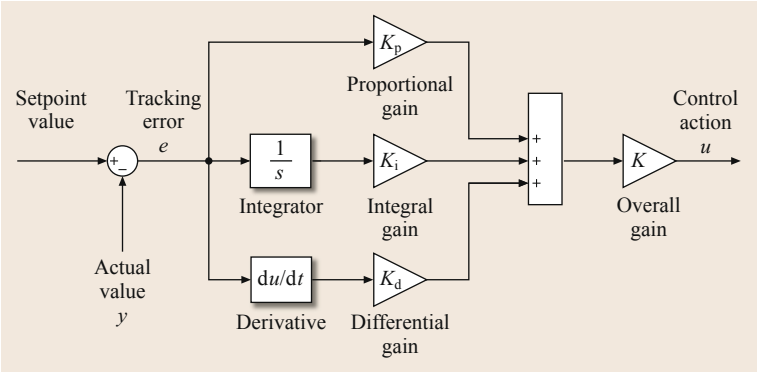


Fig. 10.25 General form of the PID controller

controller transfer function with nonzero I-term demonstrates a pole at $s = 0$; (c) the numerator polynomial of the PID controller transfer function with nonzero D-term is of higher order than its denominator one.

The block diagram of the generic PID controller is given in Fig. 10.25.

(ii) The practical form which is adopted commonly in industrial practice.

$$u(t) = K_\mathrm{p}\left(-y(t) + \frac{1}{T_\mathrm{i}}\int_0^t e(\xi)\,\mathrm{d}\xi - T_\mathrm{d}\frac{\mathrm{d}}{\mathrm{d}t}y(t)\right) \tag{10.105}$$

In this form, the error signal in both the P- and D-term is replaced by that of the process output signal (negated). In this way the closed-loop system response to abrupt changes of the setpoint, r, is significantly smoother

The problem of *tuning* a PID controller consists of determining the three P-I-D control law gains K_p, K_i and K_d so that certain transient and steady-state specifications for the closed-loop system are satisfied.

Proportional Control

For proportional control (P-control) it holds that: $K_\mathrm{i} = K_\mathrm{d} = 0$. Therefore, the tuning problem is limited to the determination of coefficient K_p in the control law. In this respect, P-control exhibits significant similarity to the ON/OFF controller presented previously. Below, a widespread version of P-control is given. In this version, the control signal is generated as the superposition of the output of a P controller plus an open-loop control action manifested as offset u_0.

$$u(t) = \begin{cases} u_0 + K_\mathrm{P}e_0\,, & e(t) > e_0 \\ u_0 + K_\mathrm{P}e(t)\,, & -e_0 \le e(t) \le e_0 \\ u_0 - K_\mathrm{P}e_0\,, & e(t) < -e_0 \end{cases} \tag{10.106}$$

In both the above relationship as well as in Fig. 10.26, one can easily assert the similarity of P-control to ON/OFF control (at least with a three-positions switch, as examined in the text).

As in the case of ON/OFF control, the control signal demonstrates saturation when the absolute value of the error signal exceeds a threshold in order to prevent excessive control signal values.

The ideal (i. e., without saturation) P controller eliminates the problem of limit cycle oscillations. On the other hand, it reduces, but does not fully eliminate, steady-state error. The effect of P-control is now investigated in two typical practical cases. At first, zero disturbance is assumed and the process block in Fig. 10.19 is assumed to be purely static as follows

$$y = K_\mathrm{SS}u = K_\mathrm{SS}(u_0 + K_\mathrm{p}e)\,. \tag{10.107}$$

In the above, K_SS stands for the steady-state gain of the process plant. After algebraic manipulation the output of the closed-loop system with P-control is obtained as follows

$$y = \frac{K_\mathrm{SS}u_0 + K_\mathrm{SS}K_\mathrm{p}r}{1 + K_\mathrm{SS}K_\mathrm{p}} \Rightarrow e = \frac{r - K_\mathrm{SS}u_0}{1 + K_\mathrm{SS}K_\mathrm{p}}\,. \tag{10.108}$$

In effect, for the steady-state error to be zero it is necessary for one of the following to hold:

- $u_0 = r/K_\mathrm{SS}$,
- $K_\mathrm{p} \to \infty$.

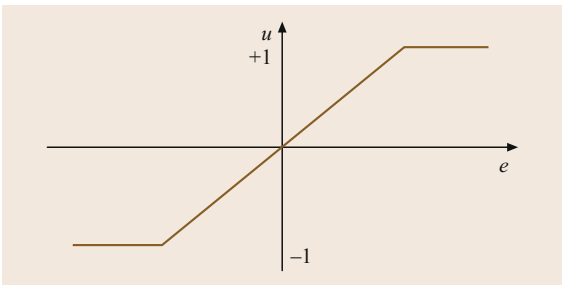


Fig. 10.26 P-control law with saturation and zero offset

The second case to be examined is when the process exhibits first-order linear dynamics and nonzero disturbance is present. Assume that $\mathbf{G}(s)$ in (10.100) is as follows

$$\mathbf{G}(s) = \frac{K_{SS}}{\tau s + 1} \,. \tag{10.109}$$

Then (10.100) with P-control ($K(s) = K_p$) yields

$$\mathbf{Y}(s) = \frac{K_{SS}K_p}{\tau s + 1 + K_{SS}K_p}\mathbf{R}(s) + \frac{K_{SS}}{\tau s + 1 + K_{SS}K_p}\mathbf{D}(s) \,. \tag{10.110}$$

By assuming that $R(s) = 0$, without any loss of generality, the final value theorem of the Laplace transform (10.6) may be employed in (10.100), to obtain the closed-loop response when there is a step change in disturbance ($d(t) = u_{step}(t) \Leftrightarrow D(s) = 1/s$).

$$\begin{aligned}
&\lim_{t\to\infty}(y(t)) = \lim_{s\to 0}(s\mathbf{Y}(s)) \\
&= \lim_{s\to 0}\left(s\frac{K_{SS}}{\tau s + 1 + K_{SS}K_p}D(s)\right) \\
&\Downarrow \\
&\lim_{t\to\infty}(y(t)) \\
&= \lim_{s\to 0}\left(s\frac{K_{SS}}{\tau s + 1 + K_{SS}K_p}\frac{1}{s}\right) = \frac{K_{SS}}{1 + K_{SS}K_p}
\end{aligned} \tag{10.111}$$

In conclusion in order to minimize the deviation of $y(t \to \infty)$ from $r = 0$ a large value for coefficient K_p is needed.

Proportional–Integral Control

For proportional-integral control (PI-control) it holds that $K_d = 0$. A control law including an integral term eliminates steady-state error which is the major drawback of P-control. It is for this reason that the PI controller finds widespread use in industry. A typical example is *speed governors* used for regulation of the rotational speed (rpm, revolutions per second) of engines, turbines, motors, etc.

PI control is defined by the transfer function

$$\begin{aligned}
&u(t) = K_p e(t) + K_i \int_0^t e(\xi)\,\mathrm{d}\xi \\
&\Updownarrow \\
&\mathbf{U}(s) = \left(K_p + \frac{K_i}{s}\right)\mathbf{E}(s) = \frac{K_p s + K_i}{s}\mathbf{E}(s) \,.
\end{aligned} \tag{10.112}$$

The effect of PI-control is now investigated in the case of an LTI–SISO process with a generic transfer function defined as a ratio of two polynomials $p_n(s)$ and $p_0(s)$ with degree m and n, respectively, and $m \le n$.

$$\mathbf{G}(s) = \frac{p_n(s)}{p_0(s)} \tag{10.113}$$

With PI control, (10.100) yields

$$\mathbf{Y}(s) = \frac{(K_i + K_p s)p_n(s)}{sp_0(s) + (K_i + K_p s)p_n(s)}\mathbf{R}(s) + \frac{sp_n(s)}{sp_0(s) + (K_i + K_p s)p_n(s)}\mathbf{D}(s) \,. \tag{10.114}$$

As an example, consider the transfer equation in (10.109). Then, (10.114) yields

$$\mathbf{Y}(s) = \frac{K_{SS}(K_i + K_p s)}{s(1 + \tau s) + K_{SS}(K_i + K_p s)}\mathbf{R}(s) + \frac{K_{SS}s}{s(1 + \tau s) + K_{SS}(K_i + K_p s)}\mathbf{D}(s) \,. \tag{10.115}$$

Assuming, as in the case of P-control, that $\mathbf{R}(s) = 0$ and applying the final value theorem for $d(t) = u_{step}(t) \Leftrightarrow \mathbf{D}(s) = 1/s$, the following is obtained

$$\begin{aligned}
y(t \to \infty) &= \lim_{s\to 0}(s\mathbf{Y}(s)) \\
&= \frac{K_{SS}\cdot 0}{0\cdot(1 + \tau\cdot 0) + K_{SS}\cdot(K_i + K_p\cdot 0)} \\
&= 0 \,.
\end{aligned} \tag{10.116}$$

As can be seen, PI control introduces a zero at $s = 0$ in the closed-loop transfer function connecting the disturbance to the output signal, resulting to elimination of the steady-state error.

A word of caution is needed here concerning closed-loop stability. The closed-loop characteristic polynomial as can be seen in (10.114) is as follows

$$p_c(s) = sp_0(s) + (K_i + K_p s)p_n(s) \,. \tag{10.117}$$

For the closed-loop system to be stable all the roots of the polynomial above must lie in the left-hand complex plane. However, the degree of this polynomial is $(n+1)$, that is, higher by 1 than the open-loop, process characteristic polynomial, $p_0(s)$. This is the price paid for the improvement in system's behavior after introducing PI-control.

In the specific case of the transfer function in (10.109), the closed-loop characteristic polynomials in

the cases when P and PI-control is employed are as follows

$$\text{P-control: } p_{c1}(s) = \tau s + K_{SS}K_p + 1\,, \tag{10.118}$$

$$\text{PI-control: } p_{c2}(s) = \tau s^2 + \left(K_{SS}K_p + 1\right) s + K_{SS}K_i\,. \tag{10.119}$$

In the case that $K_i = 0$ the polynomial in (10.119) in conjunction with the zero-pole cancellation taking place in (10.115) for $s = 0$, (10.119) coincides with the one in (10.118), that is, the one valid for P-control. Another interesting case is when $\left(K_{SS}K_p + 1\right) = 0 \Leftrightarrow K_p = -1/K_{SS}$, when polynomial $p_{c2}(s)$ has zero first-order term. In this case, the closed-loop system poles are as follows

$$p_{1,2} = \pm\sqrt{\frac{K_{SS}K_i}{\tau}}\,. \tag{10.120}$$

In conclusion, if $K_i > 0$ (and $K_{SS} > 0$) the closed-loop system becomes unstable with one pole in the right-hand complex plane, while if $K_i \le 0$ the closed-loop system is critically stable with two purely imaginary, conjugate poles. The analysis above demonstrates the importance of the proportional term in PI-control for guaranteeing closed-loop system stability.

The reason for which the PI-control law eliminates steady-state error is because the transient effect is not terminated before the output signal coincides with the setpoint. This is achieved due to the accumulation of error in the integrator, which, therefore, acts as a memory element in the closed-loop system. In effect, as long as the error integral is nonzero a persisting control action remains. The tradeoff for this advantage is that the output may be driven to damped on even sustained oscillations.

Proportional–Integral–Differential Control

This is the full form of the P-I-D control law. A differential term (D-term) is employed in order to allow us to deal with some of the problems attributed to the use of the I-term. One of the most important is the prohibitive overshoot observed in the closed-loop system response when PI-control is applied. For example, in Fig. 10.27 the response of a closed-loop system with PI control is shown exhibiting 70% overshoot above the setpoint value assumed by the output signal when steady-state is restored.

Employing the D-term in the control law incorporates prediction capabilities in the controller. Indeed, in a linear system if the error's rate of change $\mathrm{d}e/\mathrm{d}t$ is large, when a step disturbance signal is applied, it is safe to expect the output signal to demonstrate excessive overshoot. In this respect, the D-term incorporates the slope of the error signal $e(t)$ allowing to the controller to predict the imminent overshoot and to appropriately react in order to prevent it by generating a component in the control signal proportional to $\mathrm{d}e/\mathrm{d}t$.

The effect of the D-term to the response of the closed-loop system will be now investigated when the process transfer function is of the first-order as in (10.109). The controller transfer function to be used in this case is shown below.

$$\begin{aligned}\mathbf{U}(s) &= \left(K_p + \frac{K_i}{s} + K_d s\right)\mathbf{E}(s)\\ &= \frac{K_i + K_p s + K_d s^2}{s}\mathbf{E}(s)\end{aligned} \tag{10.121}$$

The general relation (10.100) for the closed-loop response yields

$$\mathbf{Y}(s) = \frac{K_{SS}(K_i + K_p s + K_d s^2)\mathbf{R}(s) + K_{SS}s\mathbf{D}(s)}{(\tau + K_{SS}K_d)\,s^2 + \left(1 + K_{SS}K_p\right)s + K_{SS}K_i}\,. \tag{10.122}$$

As can be readily seen in the above, all three coefficients of the closed-loop, second-order characteristic polynomial may be set as desired by simply adjusting the gain of each one of the terms present in the PID control law. It should be noted that the zeros and poles cannot be independently placed since they are both functions of K_p, K_i and K_d. When the stability of the closed loop takes priority, the location of zeros will be compromised, and this might result in an increased overshoot. Note that this method of determining the PID parameter values is only applicable when a system model exists.

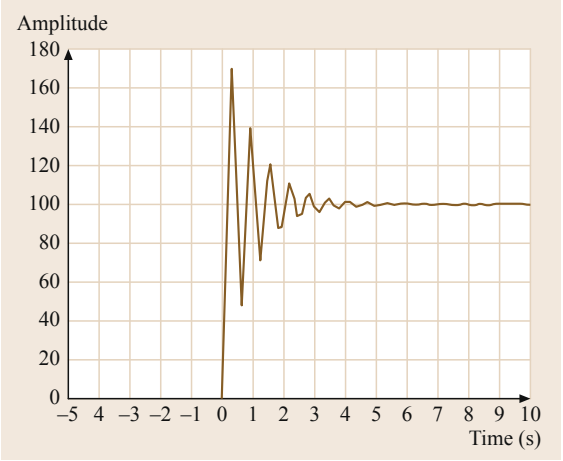


Fig. 10.27 Response of system with PI-control exhibiting prohibitive overshoot

A very interesting case of PID control, partially because of its widespread use in conventional ship course-keeping autopilots, is proportional differential control (PD-control), that is, when the I-term is absent from the control law ($K_i = 0$). Because of the missing I-term, steady-state error is exhibited in this case as well. Indeed, when $K_i = 0$ (10.122) yields

$$\mathbf{Y}(s) = \frac{K_{SS}(K_p + K_d s)}{(\tau + K_{SS}K_d)\, s + \left(1 + K_{SS}K_p\right)}\mathbf{R}(s) + \frac{K_{SS}}{(\tau + K_{SS}K_d)\, s + \left(1 + K_{SS}K_p\right)}\mathbf{D}(s)\,. \tag{10.123}$$

One can easily see that, by applying the final value theorem of the Laplace transform, for $\mathbf{R}(s) = 0$ and a step change in the disturbance signal, $d(t) = u_{step}(t)$, the steady-state response of the system, $y(t \to \infty)$, is the same with the one when simply P-control is employed

$$y(t \to \infty) = \frac{K_{SS}}{1 + K_{SS}K_p}\,.$$

It is finally noted that commonly in practice use of D-term is avoided. Although some issues in the realization of the full PID control law do exist, the main drawback is the amplification of noise due to the nonzero D-term. Indeed, consider the following relationship, demonstrating the effect of real-time signal differentiation, in the natural frequency (Fourier) domain.

$$\mathcal{L}\left\{\frac{\mathrm{d}}{\mathrm{d}t}y(t)\right\} = s\mathbf{Y}(s) \overset{s=\mathrm{i}\omega}{=} (\mathrm{i}\omega)\, Y(\mathrm{i}\omega) = \omega\,|Y(\mathrm{i}\omega)| \exp\left(\angle Y(\mathrm{i}\omega) + \frac{\pi}{2}\right)\,. \tag{10.124}$$

It is evident that high-frequency (HF, large ω), but otherwise weak in terms of power or amplitude, noise components inherent in the output signal $y(t)$ are amplified because of differentiation. This enhancement is manifested by the ω-factor in the frequency response. Therefore, the differential error signal $\mathrm{d}e/\mathrm{d}t$ is many times so noisy that cannot be practically used, at least for control applications.

10.3.4 Ziegler–Nichols' Methods for PID Controller Tuning

In many applications, open-loop system models do not exist or are highly nonlinear, and thus it is difficult to come up with accurate linear system models. To circumvent this difficulty, well-established tuning methods can be used to determine the PID parameters. The most commonly used tuning methods are those developed by Ziegler and Nichols. *Ziegler* and *Nichols* [10.10, 11] have developed an open-loop (first method) and a closed-loop tuning method (second method) for determining the PID parameters

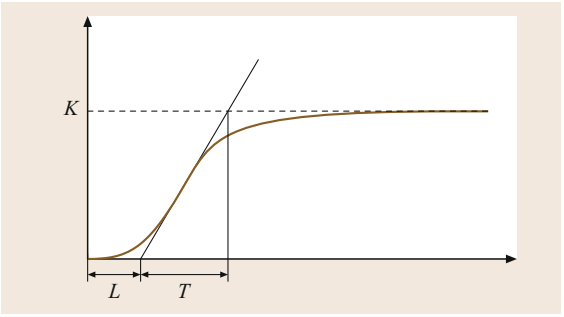


Fig. 10.28 An open-loop step response required for the first tuning method

Ziegler–Nichols' First Method

In this tuning method [10.12], the PID parameters for a system are obtained by observing its open-loop step response via experimentation. The step response must have an S-shape characterized by a time delay L, a time lag T and the static gain K as shown in Fig. 10.28.

Mathematically, this step response can be approximated by

$$\mathbf{H}(s) = \frac{K\mathrm{e}^{-Ls}}{Ts + 1}\,. \tag{10.125}$$

Mathematically, this step response can be approximated by: Once K, L, T are obtained, they can be used to design a P, PI, or PID controller using the expressions in Table 10.2. The parameters K_i, K_d can be related to T_i, T_d by the relationships

$$K_i = \frac{K_p}{T_i}\,, \quad K_d = K_p T_d\,. \tag{10.126}$$

Note that any system that has open-loop poles at the origin cannot be used with the first tuning method as its step response does not reach a steady level. In such cases, the second tuning method can be used.

Table 10.2 PID controller parameters based on the first tuning method

Type of controller	K_p	T_i	T_d
P	$\frac{T}{KL}$	∞	0
PI	$\frac{0.9T}{KL}$	$L/0.3$	0
PID	$\frac{1.2T}{KL}$	$2L$	$0.5L$

Ziegler–Nichols' Second Method

Similar to the first method, this method can be used to design a P, PI, or PID controller. Unlike the first method, the PID parameters for a system are obtained by observing its closed-loop response via experimentation. During the PID parameter identification phase, the system is driven with only a proportional controller, with its controller gain increased until the closed-loop response becomes marginally stable. The amplitude (K_{cr}) and period (P_{cr}) of oscillation are recorded and used to design a P, PI, or PID controller using the expressions in Table 10.3. As expected, this tuning method is not applicable to systems that cannot be driven to marginal stability.

Integrator Anti-Windup

In almost all physical systems, actuators saturate because they have finite range of operation, typically restricting the actuator output (control action) in terms of absolute value $u(t) = u_c(t)$, $|u_c(t)| \le u_{lim}$; $u_{lim} \times \text{sgn}(u_c(t))$, $|u_c(t)| > u_{lim}$.

If the integral control action is used, its error can be accumulated (or windup) excessively and drastically deteriorate the system performance. The integral control law thus needs to be modified

$$\dot{u}_I = \frac{k_P}{T_I} e(t) + \alpha(u(t) - u_c(t)) \,, \qquad (10.127)$$

where α is a user-defined coefficient for tracking the actuator saturation u, u_c are the control actions and control signals, respectively, (refer to Fig. 10.1). When there is actuator saturation, or $u \neq u_c$, the two terms will have opposite signs, and thus the effect of windup can be kept to a minimum. As expected, α must be chosen carefully so that it does not overcompensate the effect of windup; [10.10] suggests that $\alpha = 1/\sqrt{T_I T_D}$.

Example 10.4 AUV Forward Speed P-I-D Control

Consider the AUV forward speed model in Example 10.2 with the first tuning method. From the simulation results, we found $K = 1, L = 1.3, T = 18.7$. These are used to set up P, PI, and PID controllers for the system. Figure 10.29 shows the closed-loop responses using the P (blue), PI (red), and the PID (green) controllers. One can see that the PI controller has the worst transient performance whereas the PID controller has the best compromised performance in terms of overshoot and rise time. It should be noted that this tuning method only provides a starting point for the controller parameters, and manual fine tuning is almost always needed to further improve the system performance.

Table 10.3 PID controller parameters based on the second tuning method

Type of controller	K_p	T_i	T_d
P	$0.5K_{cr}$	∞	0
PI	$0.45K_{cr}$	$P_{cr}/1.2$	0
PID	$0.6K_{cr}$	$0.5P_{cr}$	$0.125P_{cr}$

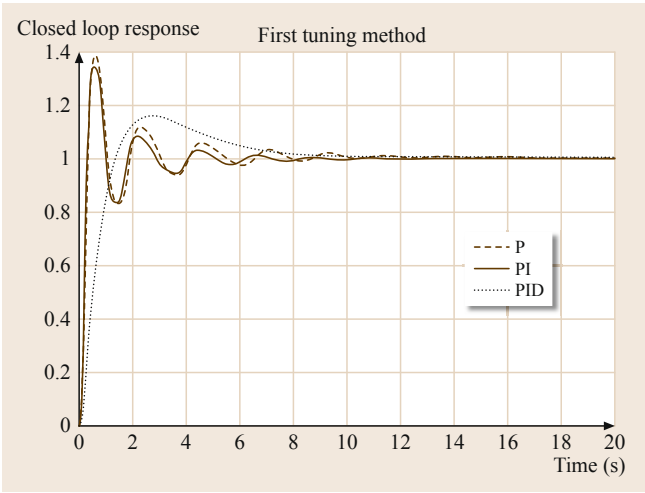


Fig. 10.29 Closed loop system responses using a P (in *blue*), PI (*red*), and a PID (in *green*) controller

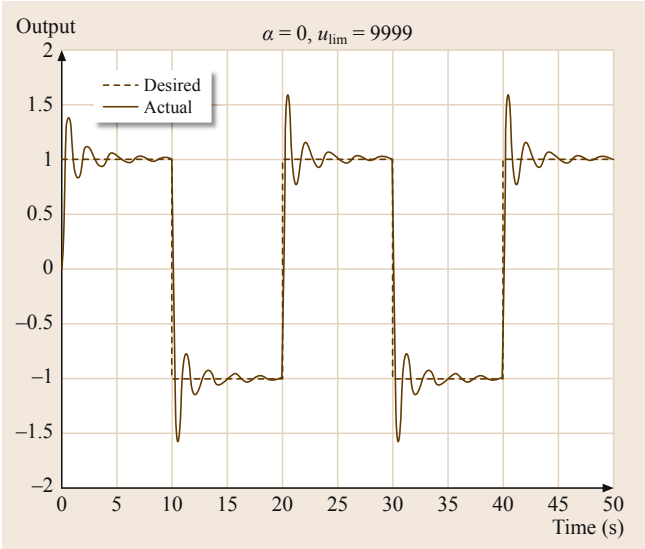


Fig. 10.30 A closed-loop AUV forward speed response using a PI controller with varying setpoints, no actuator saturation or anti-windup

To study the effect of anti-windup algorithm, the same forward speed model is controlled with a PI controller. Figure 10.30 shows the closed-loop response subject to varying setpoints over a period of 50 s. In this case, no actuator saturation is considered and $\alpha = 0$. Again, the sluggish transient response can be noticed.

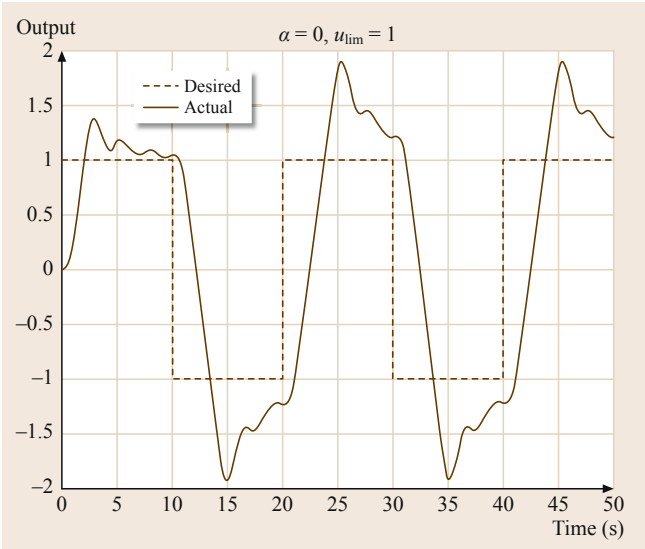


Fig. 10.31 A closed-loop AUV forward speed response using a PI controller with varying setpoints and actuator saturation but no anti-windup

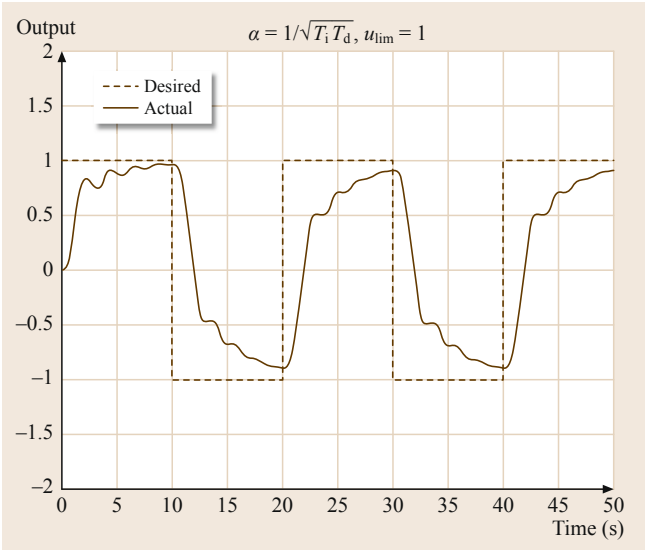


Fig. 10.32 A closed-loop AUV forward speed response using a PI controller with varying setpoints, actuator saturation and, also, anti-windup protection

In the presence of actuator saturation, $|\tau| \leq 1$, Fig. 10.31 shows the closed-loop system performance that is drastically deteriorated, especially after the first setpoint.

With the modified integral control law, Fig. 10.32 shows the much improved system performance.

10.3.5 Digital Controller Implementation

With rapid advance in computer technology, there are advantages for implementing controllers on computers because of increased flexibility, reduced cost, and more noise immunity. Controllers can be easily modified via software instead of reconfiguring or adding analog computers, thereby reducing the cost and time incurred. Since computers represent information at only low and high voltage levels, digital control systems are more immune to sensor noise and power fluctuations. In addition, many new system sensors have digital throughput, thus making the digital control choice more attractive. The only overhead in any digital control system involves two critical system components: analog-to-digital conversion (ADC) and digital-to-analog conversion (DAC).

Figure 10.33 shows a block diagram of a discrete-time closed-loop system of which the digital controller block consists of an ADC, a PC, and a DAC block. The system output is measured and digitized via the ADC and passed to the computer (PC). The digital controller implemented on the computer receives an error signal and computes a control signal, which in turn is converted back to an analog signal before it is fed to the system. Because of these additional blocks and that the signals are sampled, the closed-loop system performance and stability will have to be analyzed using different system tools from those for continuous-time systems [10.13].

Sample and Hold Function

Consider a continuous-time signal $f(t)$, shown in Fig. 10.34. If this signal is sampled at a regular time interval, the sampled values, $f^*(t)$, only exist at certain time instants.

$f^*(t)$ can be mathematically written as

$$\begin{aligned} f^*(t) &= f(0)\,[\delta(t)] + f(T)\,[\delta(t-T)] \\ &\quad + f(2T)\,[\delta(t-2T)]\cdots \\ &= \sum_{k=0}^{\infty} f(kT)\,[\delta(t-kT)]\,, \end{aligned} \tag{10.128}$$

where $\delta(t)$ is an impulse function. The Laplace transform of $f^*(t)$ can be written as [10.11]

$$F^*(s) = \sum_{k=0}^{\infty} f(kT)\mathrm{e}^{-kTs}\,. \tag{10.129}$$

In order to process the sampled signal (sampling function), its values must be kept constant between the time instants (zero-order hold function $f_\mathrm{h}(t)$), as shown

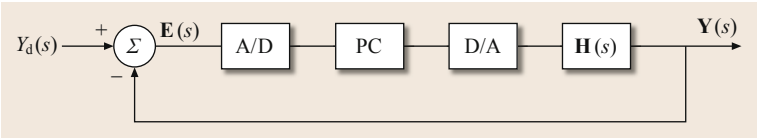


Fig. 10.33 A block diagram of a discrete-time closed-loop system

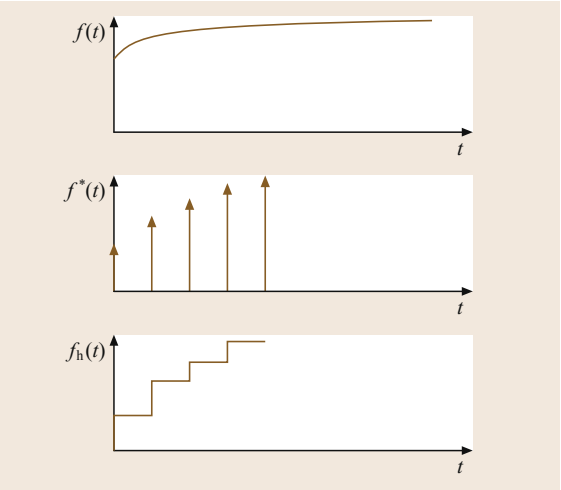


Fig. 10.34 A comparison between a continuous-time and discrete-time signal

in Fig. 10.34. An ADC process involves the sampling and holding characteristics of $f_h(t)$. Mathematically, $f_h(t)$ can be written as

$$\begin{aligned} f_h(t) &= f(0)\,[u(t)-u(t-T)] \\ &\quad + f(T)\,[u(t-T)-u(t-2T)] + \\ &\quad + f(2T)\,[u(t-2T)-u(t-3T)]\cdots \\ &= \sum_{k=0}^{\infty} f(kT)\,\{u(t-kT)-u[t-(k+1)T]\}\ , \end{aligned} \tag{10.130}$$

where $u(t)$ is an impulse function. The Laplace transform of $f_h(t)$ can be written as

$$\begin{aligned} F_h(s) &= \sum_{k=0}^{\infty} f(kT)\frac{\mathrm{e}^{-kTs}-\mathrm{e}^{-(k+1)Ts}}{s} \\ &= \frac{1-\mathrm{e}^{-Ts}}{s}\sum_{k=0}^{\infty} f(kT)\mathrm{e}^{-kTs} = \frac{1-\mathrm{e}^{-Ts}}{s}F^*(s)\ . \end{aligned} \tag{10.131}$$

It becomes apparent that the Laplace transform of a zero-order-hold function can be expressed as

$$F_{\mathrm{ZOH}}(s) = \frac{1-\mathrm{e}^{-Ts}}{s}\ , \quad F_h(s) = F_{\mathrm{ZOH}}(s)F^*(s)\ . \tag{10.132}$$

Discrete-Time System Stability

Unlike its continuous-time counterpart, the sampling time interval for a discrete-time control system needs to be chosen carefully in terms of stability and performance. The digital control system performance must be evaluated via Z-transformation instead of Laplace transformation. Some important Z-transform pairs are given in Tab. 10.1. Consider a closed-loop system in Fig. 10.37 with a sample-and-hold block.

The combined block in the feedforward path, $\mathbf{H}^*(s)$ can be written as

$$\begin{aligned} \mathbf{H}^*(s) &= \frac{1-\mathrm{e}^{-Ts}}{s}\mathbf{C}(s)\mathbf{H}(s) \\ &= \left(1-\mathrm{e}^{-Ts}\right)\frac{\mathbf{C}(s)\mathbf{H}(s)}{s}\ . \end{aligned} \tag{10.133}$$

Since $z \equiv \mathrm{e}^{sT}$ by definition, the discrete-time combined block in the feedforward path, $\mathbf{H}(z)$, can be expressed as

$$\begin{aligned} \mathbf{H}(z) &= Z\left\{\Im^{-1}\left(\mathbf{H}^*(s)\right)\right\} \\ &= Z\left\{\Im^{-1}\left(\left(1-\mathrm{e}^{-sT}\right)\frac{\mathbf{C}(s)\mathbf{H}(s)}{s}\right)\right\} \\ &= \left(1-z^{-1}\right)Z\left\{\Im^{-1}\left(\frac{\mathbf{C}(s)\mathbf{H}(s)}{s}\right)\right\}. \end{aligned} \tag{10.134}$$

Example 10.5 Discrete-Time Stability

Consider $\mathbf{C}(s) = k, \mathbf{H}(s) = a/(s(s+a))$ as shown in Fig. 10.35.

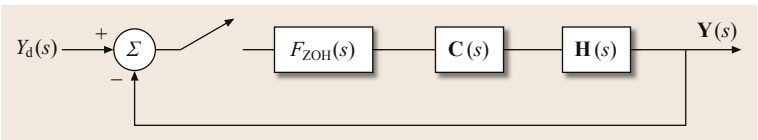


Fig. 10.35 A block diagram of a closed-loop discrete-time linear system

$$\mathbf{H}(z) = \left(1 - z^{-1}\right) Z\left\{\Im^{-1}\left[\frac{ka}{s^2(s+a)}\right]\right\} \tag{10.135}$$

is the inverse Laplace transform operator, and Z is the z-transform operator.

$Z\{\Im^{-1}[ka/(s^2(s+a))]\}$ can be computed by first evaluating the inverse Laplace transform via partial fraction techniques and then performing z-transform on the time functions via standard z-transform tables

$$\begin{aligned}
&\Im^{-1}\left\{\frac{a}{s^2(s+a)}\right\} = \Im^{-1}\left\{\frac{A}{s} + \frac{B}{s^2} + \frac{C}{s+a}\right\} \\
&= \Im^{-1}\left\{\frac{1}{s^2} - \frac{1/a}{s} + \frac{1/a}{s+a}\right\} = t - \frac{1}{a} + \frac{1}{a}\mathrm{e}^{-at} \\
&\Rightarrow \mathbf{H}(z) = k\frac{z-1}{z} \times \left\{\frac{Tz}{(z-1)^2} - \frac{z/a}{z-1} + \frac{z/a}{z-\mathrm{e}^{-aT}}\right\} \\
&= k\frac{T\left(z-\mathrm{e}^{-aT}\right) - (z-1)\,\frac{\left(1-\mathrm{e}^{-aT}\right)}{a}}{(z-1)\left(z-\mathrm{e}^{-aT}\right)} .
\end{aligned} \tag{10.136}$$

Consider $\alpha = 27$, $T = 0.1$, and we have

$$\mathbf{H}(z) = \frac{k(0.0655z + 0.027\,83)}{(z-1)(z-0.0672)} \tag{10.137}$$

Similar to the continuous-time analog, the closed-loop discrete-time transfer function, $\mathbf{H}_{\mathrm{cl}}(z)$, can be expressed as

$$\begin{aligned}
\mathbf{H}_{\mathrm{cl}}(z) &= \frac{\mathbf{H}(z)}{1 + \mathbf{H}(z)} \Rightarrow \mathbf{H}_{\mathrm{cl}}(z) \\
&= \left(k\left(0.0655z + 0.027\,83\right)\right) \\
&\quad \times \left(z^2 + \left(0.0655k - 1.0672\right)z \right. \\
&\quad \left. + 0.027\,83k + 0.0672\right)^{-1} .
\end{aligned} \tag{10.138}$$

Unlike the continuous-time stability consideration where its closed-loop poles must lie on the left-hand side of the complex s-plane, the closed-loop poles of a closed-loop discrete-time system must lie within the unit circle of the complex z-plane. At $k = 20$ the closed-loop discrete poles are located at $0.12 \pm 0.78\mathrm{i}$, and are within the unit circle. At $k = 200$ the closed-loop discrete poles are located at -0.58, -4.9, and are outside the unit circle. Thus, given a discrete-time controller, increasing its gain causes instability whereas the system stability is independent of the controller gain if a continuous-time controller is used. Similarly, when the sampling interval is increased to 0.5, the closed-loop poles become -8.17 and -0.09, and thus the closed-loop system becomes unstable.

Digital PID Controller

To implement a PID controller on a computer, its derivative and integral terms must be approximated with difference equations as signals only exist at a regular sampling time interval.

Proportional Control. The proportional term in a continuous-time system is $u(t) = k_{\mathrm{P}}(y_{\mathrm{d}}(t) - y(t))$. On a digital computer, this term is modified as

$$u(k) = k_{\mathrm{P}}(y_d(k) - y(k)) , \tag{10.139}$$

where k denotes the sampled time instants.

Integral Control. The integral term in a continuous-time system is given as

$$u(t) = \frac{k_{\mathrm{P}}}{T_{\mathrm{I}}}\int_0^t e(\tau)\mathrm{d}\tau . \tag{10.140}$$

The integral term is commonly approximated using the first-order backward difference method because of its numerical stability property

$$\begin{aligned}
&\dot{I}(t) = \frac{k_{\mathrm{P}}}{T_{\mathrm{I}}}e(t) \Rightarrow \frac{I(k) - I(k-1)}{T} = \frac{k_{\mathrm{P}}}{T_{\mathrm{I}}}e(k) \\
&\Rightarrow I(k) = I(k-1) + \frac{Tk_{\mathrm{P}}}{T_{\mathrm{I}}}e(k) .
\end{aligned} \tag{10.141}$$

Here $k = 0, 1, 2, \ldots$, and $I(-1) = 0$.

Derivative Control. The derivative (differential) term in a continuous-time system is given as

$$u(t) = k_{\mathrm{P}}T_{\mathrm{D}}\dot{e}(t) . \tag{10.142}$$

With consideration that $\dot{y}_d(t) = 0$, the derivative term is approximated using the backward difference method as follows

$$\begin{aligned}
u(t) &= k_{\mathrm{P}}T_{\mathrm{D}}\dot{e}(t) = -k_{\mathrm{P}}T_{\mathrm{D}}\dot{y}(t) \\
&\Rightarrow u(k) = -k_{\mathrm{P}}T_{\mathrm{D}}\left(\frac{y(k) - y(k-1)}{T}\right) .
\end{aligned} \tag{10.143}$$

Example 10.6 AUV Forward Speed Digital PID Control

Consider again controlling the forward speed of the AUV with a digital PID controller. Figure 10.36 shows the block diagram of the closed-loop discrete-time system generated using MATLAB/Simulink.

In the block, a continuous-time PID block is used, and its controller parameters are $K_{\mathrm{p}} = 17.26$, $T_{\mathrm{i}} = 2.6$

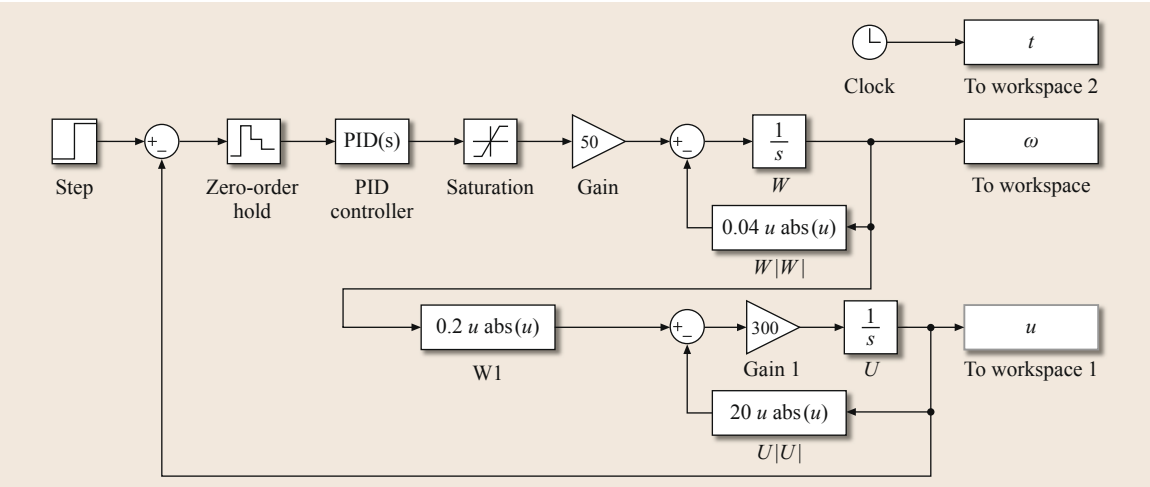


Fig. 10.36 A closed-loop discrete-time AUV speed control block diagram

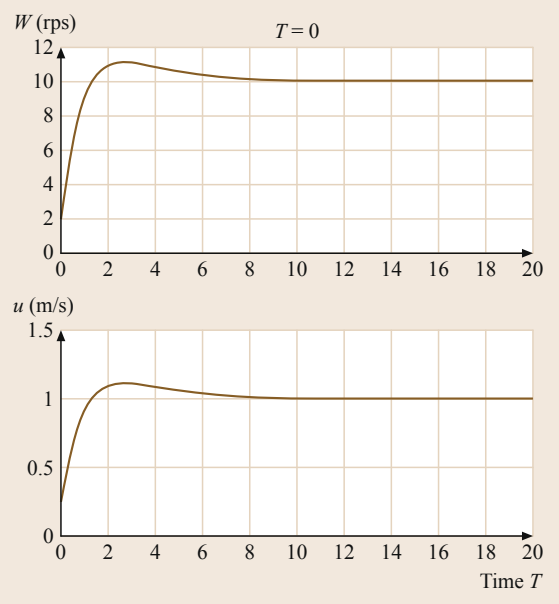


Fig. 10.37 The closed-loop vehicle and propeller speed responses without time discretization

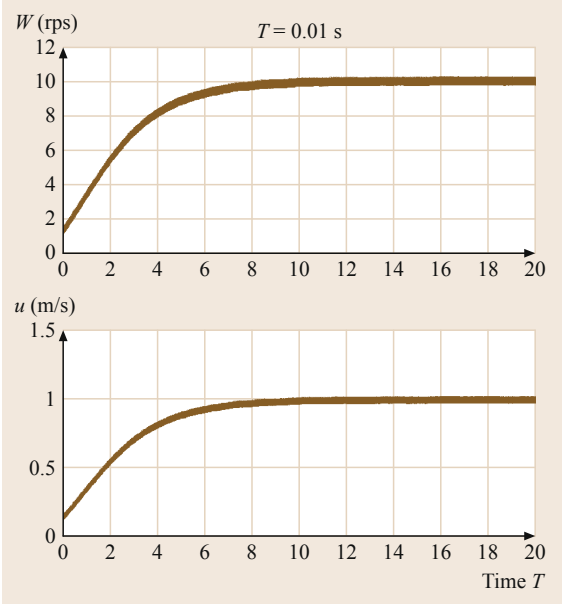


Fig. 10.38 The discrete-time closed-loop vehicle and propeller speed responses for $T = 0.01$ s

and $T_d = 0.65$. Figure 10.37 shows the closed-loop vehicle and propeller speed responses without the zero-order hold block, signifying a continuous-time control system.

In the presence of the zero-order-hold block (with $T = 0.01$ s), the closed-loop responses look significantly different, as shown in Fig. 10.38.

One can see that there is oscillation in the responses even for $T = 0.01$ s. The oscillation is worse as T is further increased to 0.1 s, as shown in Fig. 10.39. In Figs. 10.37–10.39, the x-axis represents the time in seconds.

10.3.6 The Root Locus Technique

The root locus technique was developed by W.R. Evans [10.1] for the design and comparative assessment of LTI–SISO systems. The methodology for the construction of the locus (i. e., the subset of the complex

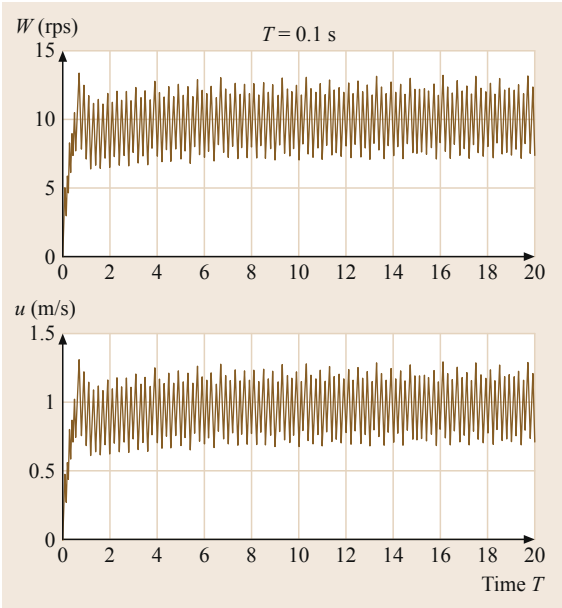


Fig. 10.39 The discrete-time closed-loop vehicle and propeller speed responses for $T = 0.1$ s

plane) of the closed-loop system poles, when a parameter (gain) of the controller varies, is shown in the present section. In specific, referring to Fig. 10.19 and forcing the disturbance d to be identically zero, a pure analog control law, that is, a gain K will be used, as shown in Fig. 10.40. The process transfer function will be presumed known in the factorized form as

$$\mathbf{G}(s) = \frac{(s+z_1)(s+z_2)\cdots(s+z_m)}{(s+p_1)(s+p_2)\cdots(s+p_n)} = \frac{s^m + b_1 s^{m-1} + \cdots + b_m}{s^n + a_1 s^{n-1} + \cdots + a_n} \qquad (10.144)$$

Note that for the root locus technique to be applied, all *poles* and *zeros* of the *open-loop*, process transfer function must be known or determined *in advance*.

In effect, the closed-loop transfer function assumes the form

$$\mathbf{H}(s) = \frac{K\mathbf{G}(s)}{1+K\mathbf{G}(s)} = [K(s+z_1)(s+z_2)\cdots(s+z_m)] \times [(s+p_1)(s+p_2)\cdots(s+p_n) + K(s+z_1)(s+z_2)\cdots(s+z_m)]^{-1} . \qquad (10.145)$$

The closed-loop characteristic polynomial, $p_c(s)$, is the denominator of the above and complying to our standard assumption that $n > m$ for the process transfer function, $\mathbf{G}(s)$, the order of $p_c(s)$ will be equal to n.

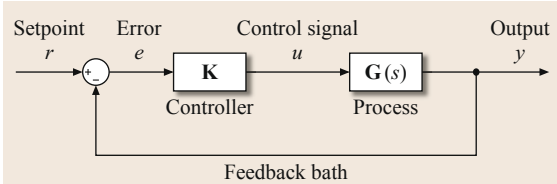


Fig. 10.40 Closed-loop block diagram for root locus analysis

The root locus (or simply the locus) of a closed-loop transfer function as in (10.145) is the location of all points on the complex plane satisfying the following for any $K \in (0, +\infty)$

$$p_c(s) = (s+p_1)(s+p_2)\cdots(s+p_n) + K(s+z_1)(s+z_2)\cdots(s+z_m) = 0 . \qquad (10.146)$$

The locus of complex plane points that satisfy (10.146) for $K \in (-\infty, 0)$ is referred to as the complementary root locus.

In the sequel, some fundamental properties of the root locus of polynomials with real coefficient are presented. These properties assist greatly in the graphical construction of the locus. Therefore, it is not necessary to determine mathematically the polynomial roots as functions of K. If this was not feasible, the locus determination would be an extremely tedious task or even impossible for polynomials of order higher than four. Proof of the lemmas are given below as it can be sought after in standard control systems literature:

1. The departure points of the locus, those for $K = 0$, are the poles of $\mathbf{G}(s)$.
2. The arrival points of the locus, those for $K \to \pm\infty$, are the zeros of $\mathbf{G}(s)$.
3. The number of discrete, component loci is n.
4. The root locus and the Complementary root locus are symmetrical with respect to the real axis.
5. For $|s| \to +\infty$ the root locus asymptotically converges toward straight lines with slopes as follows

$$\theta_i = \frac{2i+1}{n-m}\pi , \quad i = 0, 1, \ldots, (n-m)-1 . \qquad (10.147)$$

Similarly, the complementary root locus asymptotically converges toward straight lines with slopes as follows

$$\theta_i = \frac{2i}{n-m}\pi , \quad i = 0, 1, \ldots, (n-m)-1 . \qquad (10.148)$$

6. All $2(n-m)$ asymptotes in (10.147) and (10.148) for the root locus and the complementary root locus

pass through the same point on the real axis. This point is given as

$$\sigma_0 = -\frac{a_1 - b_1}{n-m} = -\frac{\sum_{i=1}^{n} p_i - \sum_{i=1}^{m} z_i}{n-m} \,. \tag{10.149}$$

7. a) A real-axis segment is part of the root locus provided that an *odd* cumulative number of poles and zeros of $\mathbf{G}(s)$ are located on its right-hand side.
   b) A real-axis segment is part of the root locus provided that an *even* cumulative number of poles and zeros of $\mathbf{G}(s)$ are located on its right-hand side.
8. The *breakpoints* (corner points) of the locus are the repeated (multiple) roots of the closed-loop characteristic polynomial $p_c(s)$; those can be determined as the roots of

$$\frac{\mathrm{d}}{\mathrm{d}s}\left[K\mathbf{G}(s)\right] = 0 \Rightarrow \frac{\mathrm{d}}{\mathrm{d}s}\mathbf{G}(s) = 0 \,. \tag{10.150}$$

9. a) The root locus angle of departure ϑ_j from the j-th pole of $\mathbf{G}(s)$, $(-p_j)$, is given by employing the following condition for $s \to (-p_j)$,

$$1 + K\mathbf{G}(s) = 0 \Rightarrow \begin{cases} |K\mathbf{G}(s)| = 1 \\ \angle K\mathbf{G}(s) = \pm(2v+1)\pi, \\ \quad v = 0, 1, 2, \ldots \end{cases} \tag{10.151}$$

The second (phase-related) of the above yields for $s \to (-p_j)$

$$\begin{aligned} &\lim_{s\to(-p_j)} \left[\angle K\mathbf{G}(s)\right] \\ &= \sum_{i=1}^{m} \angle(z_i - p_j) - \sum_{\substack{i=1 \\ i\neq j}}^{n} \angle(p_i - p_j) - \vartheta_j \\ &= \pm(2v+1)\pi \,. \end{aligned} \tag{10.152}$$

b) The root locus angle of arrival ϕ_j at the j-th zero of $\mathbf{G}(s)$, $(-z_j)$ is given by the same condition as the one used in 9(a) but applied for $s \to (-z_j)$.

$$\begin{aligned} &\lim_{s\to(-z_j)} \left[\angle K\mathbf{G}(s)\right] \\ &= \phi_j + \sum_{\substack{i=1 \\ i\neq j}}^{m} \angle(z_i - z_j) - \sum_{i=1}^{n} \angle(p_i - z_j) \\ &= \pm(2v+1)\pi \,, \quad v = 0, 1, 2, \ldots \end{aligned} \tag{10.153}$$

The properties presented previously are now applied to the construction of the root locus of various examples in Table 10.4 of process transfer function, $\mathbf{G}(s)$, as in (10.145).

10.4 Pole Placement of LTI Systems

Modern control design techniques rely heavily on the system's description in state space. The reason for this is the need to develop control systems for SISO as well as MIMO systems.

In the remaining of this chapter, LTI systems are considered. For the open-loop system (referred to as plant or process) a state-space description must be provided of the general form

$$\begin{aligned} \dot{x} &= \mathbf{A}x + \mathbf{B}u \,, \\ y &= \mathbf{C}x \,. \end{aligned} \tag{10.154}$$

In (10.154), u stands for the m-dimensional input vector (excitations), x for the n-dimensional state vector, and y for the p-dimensional output vector (measurements) of the LTI system at hand. By manipulating (10.154) the following open-loop transfer function is obtained

$$\mathbf{Y}(s) = \mathbf{H}(s)\mathbf{U}(s) = \mathbf{C}\left(s\mathbf{I}_n - \mathbf{A}\right)^{-1}\mathbf{B}\mathbf{U}(s) \tag{10.155}$$

10.4.1 Input–Output Decoupling

Consider the case below where the open-loop transfer function matrix $\mathbf{H}(s)$ is square and diagonal, that is, the process has the same number of inputs and outputs ($p = m$).

$$\mathbf{H}(s) = \begin{bmatrix} \mathbf{H}_{11}(s) & 0 & \ldots & 0 \\ 0 & \mathbf{H}_{22}(s) & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \mathbf{H}_{mm}(s) \end{bmatrix} . \tag{10.156}$$

It is straightforward that the LTI–MIMO system is equivalent to m in number LTI–SISO systems with scalar transfer functions those appearing as the diagonal, nonzero elements of the transfer function matrix, $\mathbf{H}(s)$. This case is referred to by the term *input–output decoupling*.

Table 10.4 Examples of transfer functions with their root locus

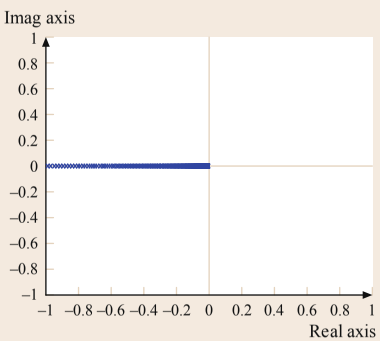
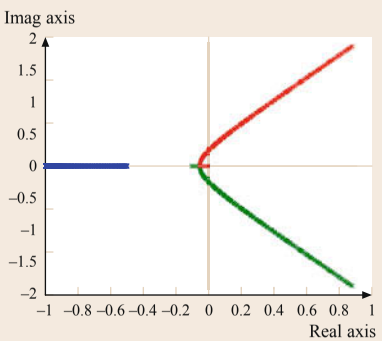
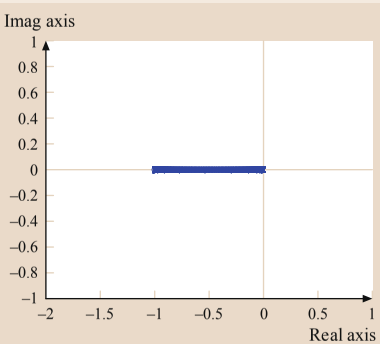
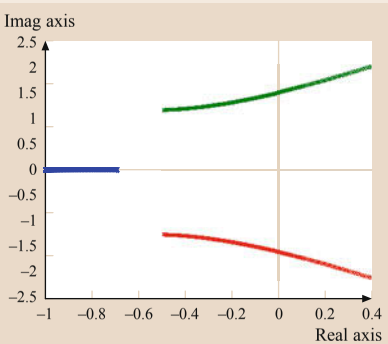
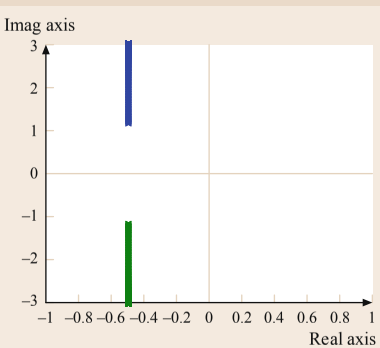
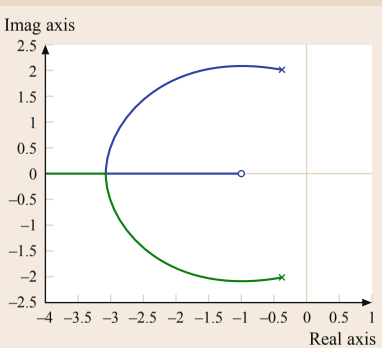
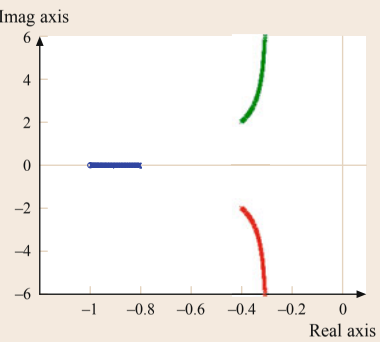
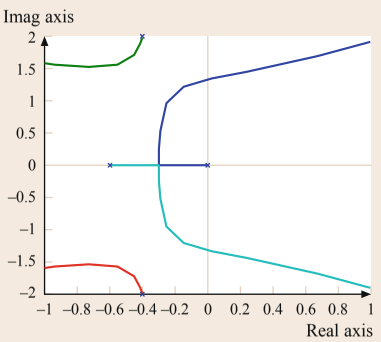
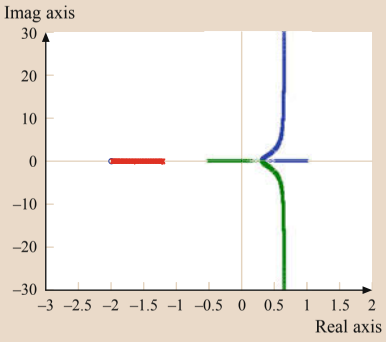
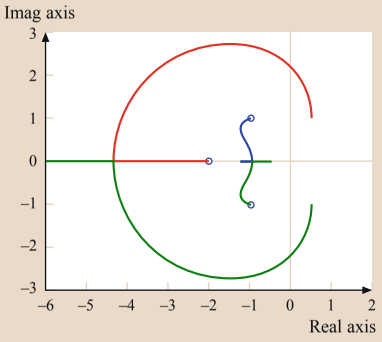
G(s)	Root Locus	G(s)	Root Locus
$\frac{1}{s}$		$\frac{1}{s\,(s+0.1)\,(s+0.5)}$	
$\frac{(s+1)}{s}$		$\frac{1}{(s+0.7)\,(s^2+s+1.69)}$	
$\frac{1}{(s^2+s+1.69)}$		$\frac{(s+1)}{(s^2+0.8s+4.16)}$	

Table 10.4 (continued)

G(s)	Root Locus	G(s)	Root Locus
$\frac{(s+1)}{(s+0.8)(s^2+0.8s+4.16)}$		$\frac{1}{s(s+0.6)(s^2+0.8s+4.16)}$	
$\frac{(s+2)}{(s+0.5)(s+1.2)(s-1)}$		$\frac{(s+2)(s^2+2s+2)}{(s+0.5)(s+1.2)(s^2-s+1.25)}$	

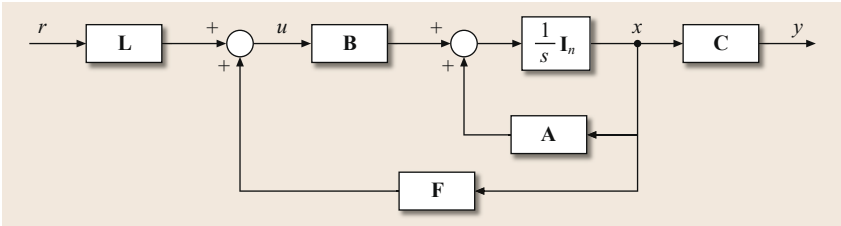


Fig. 10.41 Block diagram of a system with full-state feedback

It is evident that an input–output decoupled system as in (10.156) is only nominally MIMO because each and every one of the output signals is connected exclusively to a single input. In effect, for each and every of the m SISO systems a separate controller may be designed with techniques relevant to LTI–SISO systems.

10.4.2 Full-State Feedback

Given the description in (10.154) the control law is of the form

$$u = \mathbf{L}r + \mathbf{F}x\,. \tag{10.157}$$

In the above, r is a new input vector applied to the closed-loop system with dimension m_r. The controller design problem consists of determining the $m \times m_r$ matrix $\mathbf{L}$ and the $m \times n$ matrix $\mathbf{F}$ so that the closed-loop system demonstrates specified characteristics. The block diagram of this control scheme is shown in Fig. 10.41.

A typical example of closed-loop system specifications is input–output decoupling. In this case, the transfer function matrix, $\mathbf{H}(s)$, is square but not diagonal. Therefore, a control law of the general form in (10.157) is sought after that can render the closed-loop transfer function matrix diagonal. In effect, in the case of input–output decoupling it holds that $m = m_r = p$. By substitution of the control law in (10.157) in the state (10.154) and then application of the Laplace transform, the following is obtained

$$\left.\begin{aligned}\dot{x} &= (\mathbf{A}+\mathbf{BF})x + \mathbf{BL}r\\ y &= \mathbf{C}x\end{aligned}\right\} \Rightarrow$$
$$\Rightarrow \left\{\begin{aligned}s\mathbf{X}(s) &= (\mathbf{A}+\mathbf{BF})\mathbf{X}(s) + \mathbf{BLR}(s)\\ \mathbf{Y}(s) &= \mathbf{C}\mathbf{X}(s)\,.\end{aligned}\right\} \tag{10.158}$$

After manipulation the above yields

$$\mathbf{Y}(s) = \mathbf{G}(s)\mathbf{R}(s) = \mathbf{C}(s\mathbf{I}_n - \mathbf{A} - \mathbf{BF})^{-1}\mathbf{BLR}(s)\,. \tag{10.159}$$

Therefore, the control design problem is to determine matrices $\mathbf{F}$ and $\mathbf{L}$ in the control law so that the closed-loop transfer function matrix $\mathbf{G}(s)$ is diagonal. If this is achieved each and every component of the closed-loop system's input vector $\boldsymbol{r}$ is connected exclusively to one of the components of the output vector. Therefore, if further improvement of the system behavior is desired, m outer and decoupled SISO control loops may be employed.

10.4.3 Pole Placement Design

Generalities

As mentioned in Sect. 10.1.4, the poles of a multivariable LTI system are the eigenvalues of the system's matrix $\mathbf{A}$ and therefore can be computed as the roots of the characteristic polynomial

$$|s\mathbf{I}_n - \mathbf{A}| = 0\,. \tag{10.160}$$

The poles of a system determine its dynamics and mainly stability and transient behavior. This is the reason why in many situations full-state feedback is employed in order to relocate the poles of the open-loop system to new locations on the complex plane, specified for the closed-loop system. In effect, the related control problem is referred to as either *pole placement* or *eigenvalue assignment*. Solving the pole placement problem, allows for a closed-loop system with significantly improved characteristics. A great success story of pole placement with full-state feedback is to convert an unstable system, that is, an open-loop plant with one or more poles in the right-hand complex plane, to a stable closed-loop system.

A prerequisite for the applicability of linear, full-state feedback to a specific pole placement problem is that the open-loop plant is controllable according to the important theoretical result by *Wonham* [10.14]. In specific, if a system is state-controllable then all of its poles may be relocated arbitrarily on the complex plane. In the case that the system is uncontrollable, pole placement through state feedback is still applicable but only for a subset of the open-loop poles; the poles in this subset are the plant's controllable poles.

Given an LTI system as in (10.154) the closed-loop characteristic polynomial with full-state feedback can

be directly obtained as follows

$$p_c(s) = |s\mathbf{I}_n - \mathbf{A} - \mathbf{BF}| = s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0 \quad (10.161)$$

In (10.161), the coefficients of polynomial $p_c(s)$ are functions of the elements in the $m \times n$ feedback matrix $\mathbf{F}$. Furthermore, for the pole placement problem it can be assumed that $m = m_r$ and $L = I_m$.

The specifications in a pole placement problem consist of the desired locations of the closed-loop poles on the complex plane. As mentioned before, physically realizable systems, that is, with real parameters, have either purely real poles or pairs of complex conjugate poles. Therefore, in the factorized form of the characteristic polynomial of any physically realizable system, a purely real pole p corresponds to a factor of the form $(s-p)$ while a pair of complex conjugate poles p and p^* corresponds to the second-order factor $(s-p)(s-p^*) = s^2 - 2\mathrm{Re}(p)s + |p|^2$. By multiplicatively combining all such factors corresponding to the desired closed-loop poles and algebraically manipulating their product, a polynomial in reduced form like $\hat{p}_c(s) = s^n + \hat{a}_{n-1}s^{n-1} + \cdots + \hat{a}_1 s + \hat{a}_0$ is obtained.

Then, one has to equate the coefficients of the same powers of s of polynomials $p_c(s)$ and $\hat{p}_c(s)$. By doing that, a linear algebraic system of n equations is obtained, which if solved provides for the constants in the feedback matrix $\mathbf{F}$ of the controller

$$a_v(F_{ij}; i = 1, \ldots, m, j = 1, \ldots, n) = \hat{a}_v \,, \quad v = 1, \ldots, n \,. \quad (10.162)$$

In general, the number of equations of the algebraic system above is smaller than the number of unknowns. This means that after successfully solving the algebraic system (10.162), there remain $(m-1)n$ (if $m > 1$) free design parameters in the controller, to be used in order to meet objectives other than pole placement.

LTI–SISO Systems in Companion Canonical Form

Pole placement design is now investigated in the case of an LTI–SISO system with scalar transfer function of the form shown below.

$$\mathbf{H}(s) = \frac{\mathbf{Y}(s)}{\mathbf{U}(s)} = \frac{b_0}{s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0} \quad (10.163)$$

Further, the state-space realization in companion canonical form as in (10.47) is considered. The full-state feedback matrix $\mathbf{F}$ for this realization is, therefore, of dimensions $1 \times n$ and with the general form

$$\mathbf{F} = \begin{bmatrix} f_0 & f_1 & \ldots & f_{n-1} \end{bmatrix} . \quad (10.164)$$

The closed-loop transfer function matrix assumes the form as

$$\mathbf{A} + \mathbf{BF} = \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ 0 & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -a_0 & -a_1 & -a_2 & \ldots & -a_{n-1} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \begin{bmatrix} f_0 & f_1 & f_2 & \ldots & f_{n-1} \end{bmatrix}$$

$$\Downarrow$$

$$\mathbf{A} + \mathbf{BF} = \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ 0 & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f_0 - a_0 & f_1 - a_1 & f_2 - a_2 & \ldots & f_{n-1} - a_{n-1} \end{bmatrix} . \quad (10.165)$$

The characteristic polynomial can then be calculated as the determinant of matrix $(s\mathbf{I}_n - \mathbf{A} - \mathbf{BF}) = (s\mathbf{I}_n - (\mathbf{A} + \mathbf{BF}))$. A straightforward calculation is done if the determinant is deployed with respect to the last row, which yields the following closed-loop characteristic polynomial

$$p_c(s) = |s\mathbf{I}_n - \mathbf{A} - \mathbf{BF}| = s^n + (a_{n-1} - f_{n-1})\, s^{n-1} + \cdots + (a_1 - f_1)\, s + (a_0 - f_0) \,. \quad (10.166)$$

Then, by employing the procedure described earlier one obtains polynomial $\hat{p}_c(s)$ out of the specified pole locations of the closed-loop system. By equating the coefficients of the same powers of polynomials $p_c(s)$ and $\hat{p}_c(s)$, the following n in number design equations are obtained

$$f_i = a_i - \hat{a}_i \,, \quad i = 0, \ldots, n-1 \,. \quad (10.167)$$

Useful conclusions can be made if the pole placement controller design is compared to the root locus

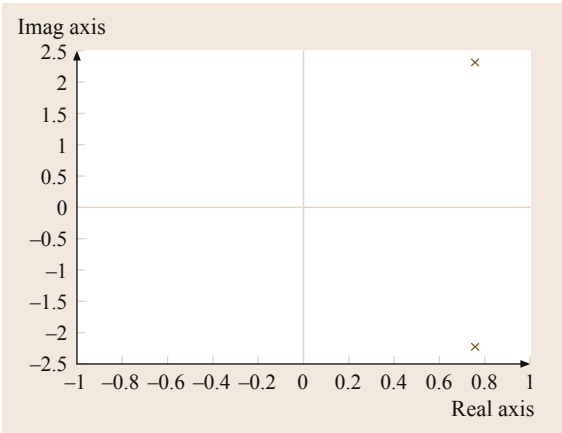


Fig. 10.42 Open-loop pole locations of plant (10.168)

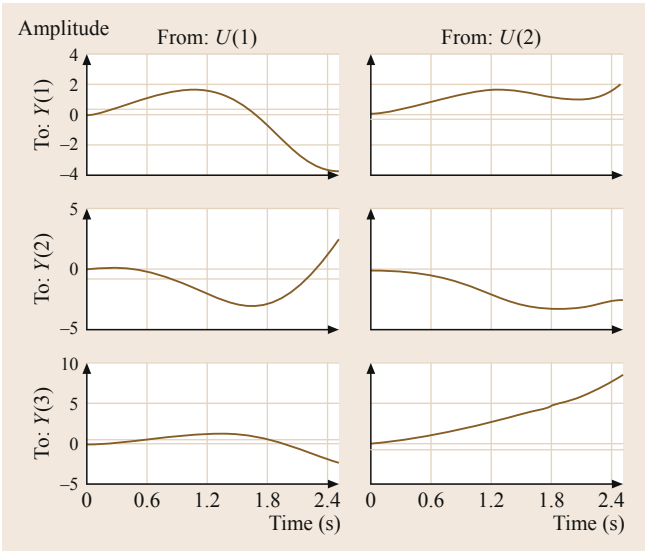


Fig. 10.43 Open-loop step response of plant (10.168)

technique where output, instead of state feedback, is employed.

LTI–MIMO Systems

It is better to investigate this case through an example. Consider the following multivariable system of the general form (10.154) with two inputs, three outputs, and three states

$$\mathbf{A} = \begin{bmatrix} 1 & 2.4 & 1 \\ -3 & 0.5 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{C} = \mathbf{I}_3 \,. \tag{10.168}$$

It can be shown that this system is controllable.

The open-loop plant transfer function matrix of the LTI system at hand is as follows

$$\mathbf{H}(s) = \begin{bmatrix} \dfrac{s^2+0.9s-1.9}{s^3-2.5s^2+7.2s-5.6} & \dfrac{s^2-0.5s+1.4}{s^3-2.5s^2+7.2s-5.6} \\ \dfrac{s^2-5s+4}{s^3-2.5s^2+7.2s-5.6} & \dfrac{-2s}{s^3-2.5s^2+7.2s-5.6} \\ \dfrac{2s-2.1}{s^3-2.5s^2+7.2s-5.6} & \dfrac{s^2-0.5s+4.2}{s^3-2.5s^2+7.2s-5.6} \end{bmatrix}_3 . \tag{10.169}$$

The open-loop characteristic polynomial of this system is, therefore

$$p_0(s) = s^3 - 2.5s^2 + 7.2s - 5.6 \,. \tag{10.170}$$

The open-loop poles, that is, the roots of the polynomial above, are

$$s_0 = 0.9807 \,, \quad s_{1,2} = 0.7596 \pm \mathrm{i}2.2656 \,. \tag{10.171}$$

The locations on the complex plane of the open-loop poles are shown in Fig. 10.42 while the open-loop plant's step response is shown in Fig. 10.43. As can be seen, the system is unstable and its step response demonstrates oscillations.

In effect, a feedback control law of the form (10.157) is sought after so that the poles of the closed-loop system are stable (demonstrate negative real part). Furthermore, in order to mitigate oscillations, the closed-loop poles are specified to be purely real, demonstrate zero imaginary part. The desired closed-loop poles are specified to be the following ones

$$p_1 = -0.5 \,, \quad p_2 = -1.0 \,, \quad p_3 = -1.5 \,. \tag{10.172}$$

These poles yield the desired characteristic polynomial

$$\begin{aligned} \hat{p}_c(s) &= (s+p_1)\,(s+p_2)\,(s+p_3) \\ &= s^3 + 3s^2 + 2.75s + 0.75 \,. \end{aligned} \tag{10.173}$$

The closed-loop matrix $(\mathbf{A}+\mathbf{BF})$, parameterized by the elements of the 3×2 matrix $\mathbf{F}$ in the control law, is given by

$$\mathbf{A}+\mathbf{BF} = \begin{bmatrix} 1+F_{11}+F_{21} & 2.4+F_{12}+F_{22} & 1+F_{13}+F_{23} \\ -3+F_{11} & 0.5+F_{12} & 1+F_{13} \\ 1+F_{21} & 1+F_{22} & 1+F_{23} \end{bmatrix} . \tag{10.174}$$

Then, the determinant $|s\mathbf{I}_n - \mathbf{A} - \mathbf{BF}|$ is calculated. This determinant is the closed-loop characteristic polynomial $p_c(s)$ of order 3. The coefficients of $p_c(s)$ are parameterized by the elements of F, F_{ij}. By equating the coefficients of polynomials $p_c(s)$ and $\hat{p}_c(s)$ and solving the resulting linear algebraic system of equations, a matrix $\mathbf{F}$ is computed as follows

$$\mathbf{F} = \begin{bmatrix} 0.3021 & 1.4785 & -0.617 \\ 2.4378 & 0.9649 & 1.2816 \end{bmatrix}. \tag{10.175}$$

The closed-loop transfer function matrix of the system with feedback matrix $\mathbf{F}$ as in (10.175) is as follows

$$\mathbf{H}(s) = \begin{bmatrix} \dfrac{s^2+1.217s+0.2183}{s^3+3s^2+2.75s+0.75} & \dfrac{s^2+1.595s+0.4767}{s^3+3s^2+2.75s+0.75} \\ \dfrac{s^2-1.281s-2.282}{s^3+3s^2+2.75s+0.75} & \dfrac{-1.685s-1.549}{s^3+3s^2+2.75s+0.75} \\ \dfrac{-1.403s-1.399}{s^3+3s^2+2.75s+0.75} & \dfrac{s^2+1.281s+0.03623}{s^3+3s^2+2.75s+0.75} \end{bmatrix}_3 . \tag{10.176}$$

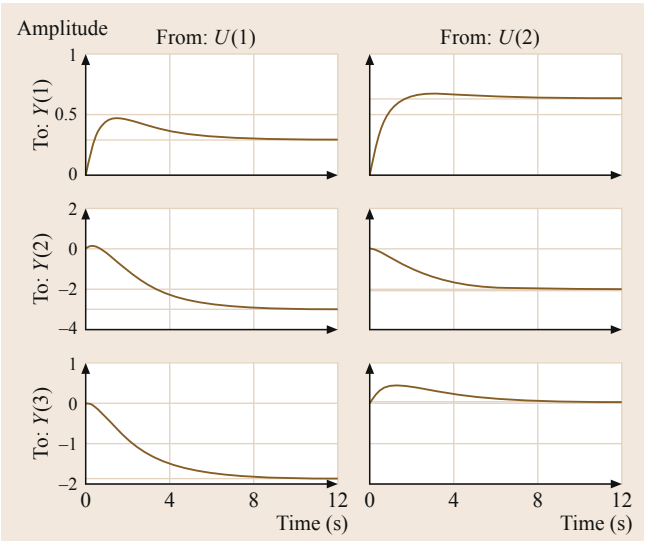


Fig. 10.44 Closed-loop step response of system with transfer function (10.176)

The step response of the closed-loop system is shown in Fig. 10.44. As can be seen, it complies with the requirements of stability and mitigated oscillations.

10.5 Course-Keeping Autopilots

10.5.1 The Vessel in the Control Loop

Surface Vessel Motions

A model describing the motions of a typical surface ship is needed. Such model should be simple, concise, as well as intuitive and fit for use in the control synthesis process.

The motions of a surface ship can be decomposed into six degrees of freedom when the vessel is considered as a rigid body (beam). In Fig. 10.45 the degrees of freedom of a ship's motion are shown and named.

The dynamic response of a vessel is modeled by a set of differential equations with respect to time. These equations concern the rate of change of linear and angular momentum of the vessel. With respect to a body-fixed (therefore noninertial in general) frame of reference, as used in ship dynamics, the equations of motion are given as

$$\frac{\mathrm{d}}{\mathrm{d}t}\left\{\int_{mS} (\boldsymbol{v}_S + \boldsymbol{\omega}_S \times \boldsymbol{r})\, \mathrm{d}m\right\} = \sum \boldsymbol{F}_{\text{ext}} ,$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\left\{\int_{mS} \boldsymbol{r} \times (\boldsymbol{v}_S + \boldsymbol{\omega}_S \times \boldsymbol{r})\, \mathrm{d}m\right\} + m_S\,(\boldsymbol{v}_S \times \boldsymbol{v}_G) = \sum \boldsymbol{M}_{\text{ext}} , \tag{10.177}$$

where m_S is the vessel's mass, $\boldsymbol{v}_S$ the linear velocity of the origin O of the body-fixed reference frame (with respect to an inertial reference frame), $\boldsymbol{\omega}_S$ the angular velocity of the origin O of the body-fixed reference frame (with respect to an inertial reference frame), and $\boldsymbol{v}_G$ the translational motion of the ship's center of mass. Vector $\boldsymbol{r}$ stands for the position vector of an arbitrary mass infinitesimal element dm shipboard with respect to the body-fixed frame. The two vector sums on the right-hand side of the differential equations above are the total force and moment applied externally to the vessel. The spatial integrals on the left-hand side are the linear and angular moment of the vessel. Finally, note the inertial terms due to the fact that a noninertial reference frame is used moving at velocity $\boldsymbol{v}_S$ with respect to an inertial one. The body-fixed and inertial reference frames used are shown in Fig. 10.46. In the same figure, the symbols used to denote the translational and

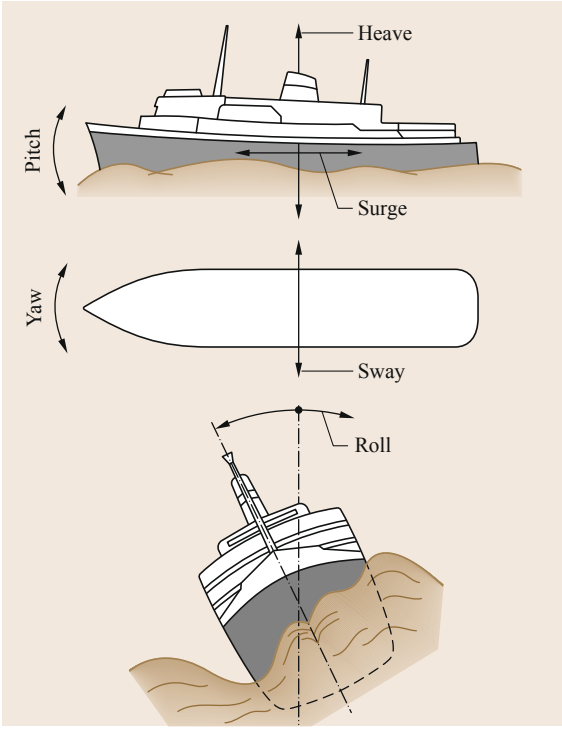


Fig. 10.45 Surface ship motions

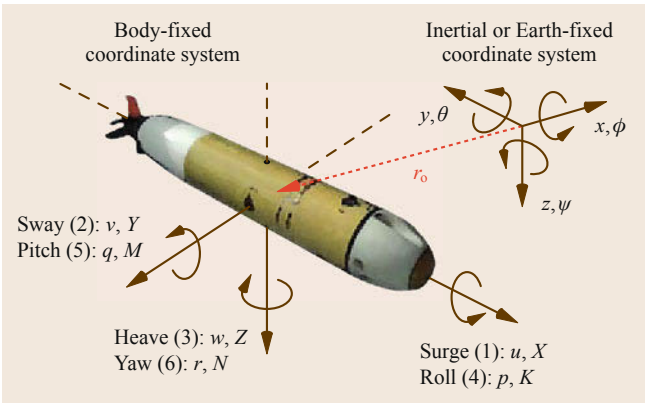


Fig. 10.46 Inertial and body-fixed reference frame used in the analysis

rotational velocities are shown as well as the ones used to denote the external forces and moments. On the axes of the inertial frame the symbols used for the infinitesimals of rotation and translation are given.

Choosing a body-fixed instead of an inertial reference frame, is based on the following reasoning: (a) acceleration measurements onboard provide with the absolute, inertial values but decomposed to the body-fixed reference frame instead of the inertial one; (b) the geometric and inertia characteristics of the vessel are time-invariant with respect to the body-fixed frame. This fact mathematically is expressed as $\dot{\boldsymbol{r}} = 0$. Also, it can be used to swap spatial integration with time differentiation in (10.177).

$$\int_{m_S} \frac{\mathrm{d}}{\mathrm{d}t} \{\boldsymbol{v}_S + \boldsymbol{\omega}_S \times \boldsymbol{r}\} \, \mathrm{d}m = \sum \boldsymbol{F}_{\text{ext}} \,,$$

$$\int_{m_S} \frac{\mathrm{d}}{\mathrm{d}t} \{\boldsymbol{r} \times (\boldsymbol{v}_S + \boldsymbol{\omega}_S \times \boldsymbol{r})\} \, \mathrm{d}m + m_S \, (\boldsymbol{v}_S \times \boldsymbol{v}_G)$$

$$= \Sigma \boldsymbol{M}_{\text{ext}} \qquad (10.178)$$

Formulation of the Linearized State Equations

Using the above, as well as the linearization procedure presented in Chap. 1, the generic surface vessel equations given previously are simplified in the sense of linearization. The linearized equations are derived by neglecting higher order, nonlinear terms in the original ship motion equations. Such terms are those including powers higher than one of $\boldsymbol{v}_S$ or $\boldsymbol{\omega}_S$ or cross-products thereof or their time derivatives. Furthermore, linearization is applied by considering perturbations around a stable equilibrium point. The derived set of linear ordinary differential equations is valid for small-amplitude variations of $\boldsymbol{v}_S$ and $\boldsymbol{\omega}_S$ around the equilibrium point. Therefore, the linearized differential equations hold for variations $\delta\boldsymbol{v}_S$ and $\delta\boldsymbol{\omega}_S$. However, in order to simplify the expressions the δ-notation is dropped, since the classical linear feedback control problem is formulated and solved on the basis of the linearized plant equations.

In the sequel the 3×3 following inertia tensor, $\overleftrightarrow{I}_S$, which is time invariant with respect to the body-fixed frame, is needed.

$$\overleftrightarrow{I}_S = \begin{bmatrix} I_{xx} & I_{xy} & I_{xz} \\ I_{xy} & I_{yy} & I_{yz} \\ I_{xz} & I_{yz} & I_{zz} \end{bmatrix} \qquad (10.179)$$

The scalar inertial entries in tensor $\overleftrightarrow{I}_S$ are given for I_{xx} and I_{xy}. For the other entries, definitions are analogous.

$$I_{xx} = \int_V \left(y^2 + z^2\right) \, \mathrm{d}m \,, \quad I_{xy} = -\int_V xy \, \mathrm{d}m \qquad (10.180)$$

For most practical autopilot designs, the following simplifying assumption that the body-fixed frame has been selected so as $\boldsymbol{r}_G = \begin{pmatrix} x_G & 0 & 0 \end{pmatrix}^{\mathrm{T}}$.

The small-signal analysis, for autopilot design in the case of ocean-going vessels moving in a time-averaged sense forward at constant speed U, can be applied if the following decomposition scheme is considered

$$\boldsymbol{v}_S = \begin{bmatrix} U+u \\ v \\ w \end{bmatrix} = \begin{bmatrix} U \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} u \\ v \\ w \end{bmatrix},$$
$$\text{with } \frac{\mathrm{d}U}{\mathrm{d}t} = \boldsymbol{\omega}_S \times U, \text{ i. e., } \dot{U} = 0\,. \tag{10.181}$$

The equilibrium point around which the linearization is applied has to be evidently $\boldsymbol{U} = \begin{bmatrix} U & 0 & 0 \end{bmatrix}^{\mathrm{T}}$.

After analyzing and linearizing the equations, the following set of first order, linear differential equations are derived

$$\begin{aligned}
&\text{Surge:} && m_S\dot{u} = X \\
&\text{Sway:} && m_S\left[\dot{v} + Ur + x_G\dot{r}\right] = Y \\
&\text{Heave:} && m_S\left[\dot{w} - Uq - x_G\dot{q}\right] = Z + Z_{\mathrm{hs}} \\
&\text{Roll:} && I_{xx}\dot{p} = K + K_{\mathrm{hs}} \\
&\text{Pitch:} && I_{yy}\dot{q} + m_S x_G\left(-\dot{w} + Uq\right) = M + M_{\mathrm{hs}} \\
&\text{Yaw:} && I_{zz}\dot{r} + m_S x_G\left(\dot{v} + Ur\right) = N\,.
\end{aligned} \tag{10.182}$$

The above are derived on the assumption of a hull symmetric with respect to plane xz. Subscript "hs" indicates a hydrostatic component in a force or moment. Hydrostatic components are in general dependent upon displacement variables like translational ones x, y, z, as well as rotational ones φ, θ, ψ; however, hydrostatic components cannot depend on translational or rotational velocities. A force or moment component that depends on the vessel's translational (u, v, w) or rotational velocities (p, q, r) or higher order time derivatives (like accelerations) is called hydrodynamic.

As seen in (10.182), after linearization surge, sway and yaw become decoupled of roll, pitch, and heave. Furthermore, note that the forces and moments affecting the surge–sway–heave motions are purely hydrodynamic. At this point, expressions for X, Y, and N are needed. These excitation signals may originate:

a) From the ocean–hull interaction.
b) From control surface actuation. The most common control surface of surface vessels is the rudder. Other control surfaces include anti-roll stabilizing fins, etc.
c) From perturbation of random nature that is either difficult or impossible or even too complicated to be modeled.

$$\begin{aligned}
X(t) &= X_{\mathrm{R}}(t) + X_{\mathrm{H}}(t) + X_0(t) \\
Y(t) &= Y_{\mathrm{R}}(t) + Y_{\mathrm{H}}(t) + Y_0(t) \\
N(t) &= N_{\mathrm{R}}(t) + N_{\mathrm{H}}(t) + N_0(t)\,,
\end{aligned} \tag{10.183}$$

where subscript R indicates the component due to rudder action, H the one due to the ocean–hull interaction, and 0 a random signal due to modeling errors and simplifications.

The rudder components evidently depend exclusively on the deflection of the rudder from the vertical xz symmetry plane of the vessel's hull

$$\begin{aligned}
&X_{\mathrm{R}}(t) = X_{\mathrm{R}}\left(\delta_{\mathrm{R}}(t)\right)\,, \quad Y_{\mathrm{R}}(t) = Y_{\mathrm{R}}\left(\delta_{\mathrm{R}}(t)\right), \\
&N_{\mathrm{R}}(t) = N_{\mathrm{R}}\left(\delta_{\mathrm{R}}(t)\right)\,.
\end{aligned} \tag{10.184}$$

For small (in the absolute sense) deflection values of angle δ_{R} the following relationships can be employed

$$X_{\mathrm{R}} = X_\delta\delta_{\mathrm{R}}\,, \quad Y_{\mathrm{R}} = Y_\delta\delta_{\mathrm{R}}\,, \quad N_{\mathrm{R}} = N_\delta\delta_{\mathrm{R}}\,, \tag{10.185}$$

where hydrodynamic derivatives X_δ, Y_δ, and N_δ are defined as

$$X_\delta = \frac{\partial X_{\mathrm{R}}}{\partial\delta_{\mathrm{R}}}\,, \quad Y_\delta = \frac{\partial Y_{\mathrm{R}}}{\partial\delta_{\mathrm{R}}}\,, \quad N_\delta = \frac{\partial N_{\mathrm{R}}}{\partial\delta_{\mathrm{R}}}\,. \tag{10.186}$$

In a similar manner, the components originating from ocean–hull interaction are dependent solely on kinematical variables of the vessel

$$\begin{aligned}
&X_{\mathrm{H}} = X_{\mathrm{H}}\left(u, \dot{u}\right)\,, \quad Y_{\mathrm{H}} = Y_{\mathrm{H}}\left(v, \dot{v}, r, \dot{r}\right), \\
&N_{\mathrm{H}} = N_{\mathrm{H}}\left(v, \dot{v}, r, \dot{r}\right)\,.
\end{aligned} \tag{10.187}$$

By using the small-signal analysis concept, the following expressions can be obtained

$$\begin{aligned}
X_{\mathrm{H}} &= X_u u + X_{\dot{u}}\dot{u}\,, \\
Y_{\mathrm{H}} &= Y_v v + Y_{\dot{v}}\dot{v} + Y_r r + Y_{\dot{r}}\dot{r}\,, \\
N_{\mathrm{H}} &= N_v v + N_{\dot{v}}\dot{v} + N_r r + N_{\dot{r}}\dot{r}\,,
\end{aligned} \tag{10.188}$$

where the hydrodynamic derivatives X_u, $X_{\dot{u}}$, Y_v, $Y_{\dot{v}}$, Y_r, $Y_{\dot{r}}$, N_v, $N_{\dot{v}}$, N_r, $N_{\dot{r}}$ are defined similarly to the ones related to the rudder in (10.186).

10.5.2 Surface Vessel State-Space Model

On the basis of the decoupled model of surge, sway and yaw in (10.182) as well as (10.183), (10.185), and (10.188) for the exogenous forces and moments applied to a vessel, the following set of differential equations can be obtained

$$\begin{aligned}
&\left(m_S - X_{\dot{u}}\right)\dot{u} = X_u u + X_\delta\delta_{\mathrm{R}} + X_0(t)\,, \\
&\left(m_S - Y_{\dot{v}}\right)\dot{v} + \left(m_S x_G - Y_{\dot{r}}\right)\dot{r} \\
&\quad = Y_v v + \left(Y_r - m_S U\right) r + Y_\delta\delta_{\mathrm{R}} + Y_0(t)\,, \\
&\left(I_{zz} - N_{\dot{r}}\right)\dot{r} + \left(m_S x_G - N_{\dot{v}}\right)\dot{v} \\
&\quad = N_v v + \left(N_r - m_S x_G U\right) r + N_\delta\delta_{\mathrm{R}} + Z_0(t)\,.
\end{aligned} \tag{10.189}$$

In matrix form and after omitting the equation for surge (which is not coupled with the ones for yaw and sway), the following is obtained

$$\begin{aligned}
&\boldsymbol{\Theta}\dot{x}_{\emptyset} = \mathbf{A}_{\emptyset}x_{\emptyset} + b_{\emptyset}\delta_{\mathrm{R}} + d_{\emptyset} \quad , \\
&\qquad\qquad \Updownarrow \\
&\begin{bmatrix} m_S - Y_{\dot{v}} & m_S x_G - Y_{\dot{r}} \\ m_S x_G - N_{\dot{v}} & I_{zz} - N_{\dot{r}} \end{bmatrix} \begin{bmatrix} \dot{v} \\ \dot{r} \end{bmatrix} = \\
&\begin{bmatrix} Y_v & Y_r - m_S U \\ N_v & N_r - m_S x_G U \end{bmatrix} \begin{bmatrix} v \\ r \end{bmatrix} + \begin{bmatrix} Y_\delta \\ N_\delta \end{bmatrix} \delta_{\mathrm{R}} + \begin{bmatrix} Y_0 \\ N_0 \end{bmatrix} .
\end{aligned} \tag{10.190}$$

The correspondence between the brief and extended versions of (10.190) is obvious.

Assuming that matrix Θ in (10.190) is invertible the following are obtained

$$\begin{aligned}
\boldsymbol{\Theta}\dot{x}_{\emptyset} &= \mathbf{A}_{\emptyset}x_{\emptyset} + b_{\emptyset}\delta_{\mathrm{R}} + d_{\emptyset} \Leftrightarrow \dot{x}_{\emptyset} \\
&= \boldsymbol{\Theta}^{-1}\mathbf{A}_{\emptyset}x_{\emptyset} + \boldsymbol{\Theta}^{-1}b_{\emptyset}\delta_{\mathrm{R}} + \boldsymbol{\Theta}^{-1}d_{\emptyset} \ , \qquad (10.191) \\
\dot{x}_{\emptyset} &= \boldsymbol{\Theta}^{-1}\mathbf{A}_{\emptyset}x_{\emptyset} + \begin{bmatrix} b & \boldsymbol{\Theta}^{-1} \end{bmatrix} \begin{bmatrix} \delta_{\mathrm{R}} \\ d_{\emptyset} \end{bmatrix} , \\
b &= \boldsymbol{\Theta}^{-1}b_{\emptyset} \ . \qquad (10.192)
\end{aligned}$$

Equation (10.192) is the dynamic equation of a system with two states and three scalar input signals one of which is the control action (rudder deflection angle δ_{R}); the other two inputs are the random disturbances Y_0 and N_0. For reasons made clear later on, the state vector $x_{\emptyset}$ in (10.192) has to be enriched by the yaw angle ψ as follows

$$\begin{aligned}
\dot{x} &= \begin{bmatrix} 0 & 0 & 1 \\ 0 & & \\ & \Theta^{-1}A_{\emptyset} & \\ 0 & & \end{bmatrix} x + \left[\begin{array}{cc|c} 0 & 0 & 0 \\ b & & \boldsymbol{\Theta}^{-1} \end{array}\right] \begin{bmatrix} \delta_{\mathrm{R}} \\ d_{\emptyset} \end{bmatrix} \\
&\Leftrightarrow \dot{x} = \mathbf{A}x + \mathbf{B}u \quad \text{with} \quad x = \begin{bmatrix} \psi \\ v \\ r \end{bmatrix} .
\end{aligned} \tag{10.193}$$

By employing the Laplace transform, the following relation is obtained

$$\begin{aligned}
x(s) &= (s\mathbf{I}_3 - \mathbf{A})^{-1} \begin{bmatrix} 0 \\ b \end{bmatrix} \delta_{\mathrm{R}}(s) \\
&\quad + (s\mathbf{I}_3 - \mathbf{A})^{-1} d(s) \ , \\
\text{where} \quad d &= \begin{bmatrix} 0 \\ \boldsymbol{\Theta}^{-1}d_{\emptyset} \end{bmatrix} .
\end{aligned} \tag{10.194}$$

In many commonplace cases of autopilot design, the single feedback signal is the yaw angle $\psi(t)$ through means of either a magnetic compass or a gyrocompass. In these cases the input to the controller (autopilot) is the yaw error, that is, the difference obtained by subtracting ψ from the heading setpoint.

In effect, the dynamical equation (10.194) is complemented by the following output, algebraic one connecting the state vector $\boldsymbol{x}$ to the scalar output signal $\psi(t)$ (or $\psi(s)$ in the complex frequency domain).

$$y = cx = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} x = \psi \tag{10.195}$$

Since x is of order 3, the transfer function from the control action δ_{R} to the output signal ψ is of order 3 as well. Because of the special form of matrix $\mathbf{A}$ (exhibiting a full zero column) one of the eigenvalues and open-loop plant poles is 0. The full characteristic polynomial is given by the following

$$p_0(s) = s\left|sI_2 - \boldsymbol{\Theta}^{-1}A_{\emptyset}\right| . \tag{10.196}$$

Combining (10.196) in the generic matrix inversion formula the following transfer function from the control action δ_{R} to yaw angle ψ is obtained

$$\begin{aligned}
\mathbf{H}_N(s) &= \frac{\boldsymbol{\psi}(s)}{\delta_{\mathrm{R}}(s)} = \frac{\kappa_0 + (s + \kappa_1)}{p_0(s)} \\
&= \frac{k + k\tau_0 s}{s[\tau_1\tau_2 s^2 + (\tau_1 + \tau_2)s + 1]} .
\end{aligned} \tag{10.197}$$

The transfer function (10.197) connecting δ_{R} with ψ is known as the Nomoto equation. For most practical cases of displacement hulls, the factor

$$\left|s\mathbf{I}_2 - \boldsymbol{\Theta}^{-1}\mathbf{A}_{\emptyset}\right| = \tau_1\tau_2 s^2 + (\tau_1 + \tau_2)s + 1$$

exhibits two purely real roots denoted as $-1/\tau_1$ and $-1/\tau_2$.

The full expression for yaw angle ψ including the disturbance signals $\mathbf{Y}_0$ and $\mathbf{N}_0$, in the complex frequency domain, has the following form

$$\boldsymbol{\psi}(s) = \mathbf{H}_N(s)\delta_{\mathrm{R}}(s) + \mathbf{T}_N(s)N_0(s) + \mathbf{T}_Y(s)\mathbf{Y}_0(s) . \tag{10.198}$$

Similarly to the Nomoto equation transfer function, the expressions for transfer functions $\mathbf{T}_N(s)$ and $\mathrm{T}_Y(s)$ can be obtained as

$$\begin{aligned}
\mathbf{T}_N(s) &= \frac{k_N + k_N\tau_N s}{s[\tau_1\tau_2 s^2 + (\tau_1 + \tau_2)s + 1]} \\
&\text{and} \\
\mathbf{T}_Y(s) &= \frac{k_Y + k_Y\tau_Y s}{s[\tau_1\tau_2 s^2 + (\tau_1 + \tau_2)s + 1]} .
\end{aligned} \tag{10.199}$$

10.5.3 PID Autopilots

Many practical, specifically low-end autopilots employ the PID control law, given as

$$\begin{aligned}\delta_{\mathrm{R}}(t) &= -K_{\mathrm{p}}\psi(t) - K_{\mathrm{i}}\int_0^t \psi(\xi)\,\mathrm{d}\xi - K_{\mathrm{d}}\frac{\mathrm{d}}{\mathrm{d}t}\psi(t)\\ &\Updownarrow\\ \delta_{\mathrm{R}}(s) &= -K(s)\psi(s)\\ &= -\left(K_{\mathrm{p}} + \frac{K_{\mathrm{i}}}{s} + K_{\mathrm{d}}s\right)\psi(s)\\ &= -\frac{K_{\mathrm{i}} + K_{\mathrm{p}}s + K_{\mathrm{d}}s^2}{s}\psi\,.\end{aligned} \tag{10.200}$$

Substituting the above in (10.198) the following closed-loop expressions can be obtained

$$\begin{aligned}\psi(s) &= \mathbf{G}_N(s)\mathbf{N}_0(s) + \mathbf{G}_Y(s)\mathbf{Y}_0(s)\\ &= \frac{T_N(s)}{1+\mathbf{H}_N(s)\mathbf{K}(s)}\mathbf{N}_0(s)\\ &\quad + \frac{T_Y(s)}{1+\mathbf{H}_N(s)\mathbf{K}(s)}\mathbf{Y}_0(s)\,.\end{aligned} \tag{10.201}$$

In effect, transfer functions $\mathbf{G}_N(s)$ and $\mathbf{G}_Y(s)$ from the yaw moment and sway force disturbance have to be of the fourth order according to (10.197) and (10.199). However, for many practical applications the following first-order open-loop transfer functions consist a good approximation

$$\begin{aligned}\tilde{\mathbf{H}}_N(s) &= \frac{k}{s(\tau_1 s+1)}\,,\\ \tilde{\mathbf{T}}_N(s) &= \frac{k_N}{s(\tau_1 s+1)}\,,\\ \tilde{\mathbf{T}}_Y(s) &= \frac{k_Y}{s(\tau_1 s+1)}\,.\end{aligned} \tag{10.202}$$

For the above approximants to be valid, it must hold that $\tau_1 \gg \tau_2$ and in effect for $s \to 0$

$$\tau_1\tau_2 s^2 + (\tau_1+\tau_2)s + 1 \approx (\tau_1 s+1)\,.$$

Similar assumptions are needed for the time constants in the numerator of the full-order transfer functions. For most practical displacement hull forms, $\tau_1 \gg \tau_2$ and therefore the so-called dominant pole technique is applicable.

In effect

$$\begin{aligned}\tilde{\mathbf{G}}_N(s) &= \frac{\dfrac{k_N}{s(\tau_1 s+1)}}{1+\dfrac{K_{\mathrm{i}}+K_{\mathrm{p}}s+K_{\mathrm{d}}s^2}{s}\dfrac{k_N}{s(\tau_1 s+1)}}\\ &= \frac{k_N s}{\tau_1 s^3 + (k_N K_{\mathrm{d}}+1)s^2 + k_N K_{\mathrm{p}} s + k_N K_{\mathrm{i}}}\,.\end{aligned} \tag{10.203}$$

The expression for $\tilde{\mathbf{G}}_Y(s)$ is similar and as seen the reduced order transfer functions will be of the third order. Further reduction (simplification) can be achieved by intervening in the controller structure.

Such intervention is the design assumption that $K_{\mathrm{i}} \ll K_{\mathrm{p}}, K_{\mathrm{d}}$. This assumption is based on the fact that the integral term is included in the PID control law only for eliminating the long-term, near-steady-state effects of the **Y** and **N** disturbance signals. That is why the I-term constant can assume relatively small values. Furthermore as seen in (10.203) the I-term constant appears exclusively in the constant (zero-order) term of the closed-loop characteristic polynomial. Therefore, small I-term constant values will give a closed-loop pole in the vicinity of 0. Such a pole, however, will cancel out with the closed-loop zero at 0. In effect, the following simplified closed-loop transfer function can be obtained for small values of K_{i}

$$\tilde{\mathbf{G}}'_N(s) = \frac{k_N}{\tau_1 s^2 + (k_N K_{\mathrm{d}}+1)s + k_N K_{\mathrm{p}}}\,. \tag{10.204}$$

Finally, note that for small, purely positive values of K_{i} the pole $s=0$ moves inside the left-hand complex plane and the closed-loop system becomes asymptotically stable.

In conclusion, two of the poles of $\tilde{\mathbf{G}}_N(s)$ can be determined as ones of $\tilde{\mathbf{G}}'_N(s)$, while the third one, which happens to be the most dominant since it is closer to the imaginary axis at least for small values of K_{i}, is the one that coincides with the origin for $K_{\mathrm{i}}=0$.

In effect, the poles of $\tilde{\mathbf{G}}'_N(s)$ define whether the response of the vessel to yaw moment disturbance exhibits oscillations or not. Indeed if the value of the damping constant ζ of the second-order characteristic polynomial $\tilde{p}'_N(s) = \tau_1 s^2 + (k_N K_{\mathrm{d}}+1)s + k_N K_{\mathrm{p}}$ is below 0.7 then oscillations in the response are to be expected. In effect, proper selection of constants K_{p}, K_{d} can be made in order to eliminate excessive oscillatory behavior in closed-loop response. On the other hand, all controller constants, but predominantly K_{i}, contribute to obtain a closed-loop system that is adequately stable.

Similar conclusions hold for transfer function $\tilde{\mathbf{G}}_Y(s)$, too.

Example 10.7 Autopilot Steering System for Disturbance Rejection

A 6200 TEU container carrier will be examined as typical application of the analysis and methodology developed for autopilot pole placement design earlier. The principal particulars of the ship are given in Table 10.5.

Considering the linearized ship model introduced by (10.190), it is easily seen that the values of a number of parameters are required to be known. The geometric and physical data for a certain ship, for example, deadweight, length, breadth, etc., are usually given. On the other hand, the values of the involved hydrodynamic derivatives have to be estimated. A variety of estimation methods exists including direct measurement campaigns onboard ships and prediction of their values through similitude-based methods based on the experimental data of towing tank. In this work, strip theory results for the nondimensional hydrodynamic derivatives will be used, which will then be dimensionalized according to standard naval architecture practice.

For a symmetrical hull, the following equations proposed by Clarke and deduced through semianalytical methods, provide an estimate of the nondimensional hydrodynamic derivatives, involved in the ship motion dynamic equations [10.15]

$$Y'_{\dot{v}} = -\pi \left(\frac{T}{L}\right)^2 \left[1 + 0.16\mathbf{C}_b \frac{B}{T} - 5.1\left(\frac{B}{L}\right)^2\right], \tag{10.205}$$

$$Y'_{\dot{r}} = -\pi \left(\frac{T}{L}\right)^2 \left[0.67\frac{B}{L} - 0.0033\left(\frac{B}{T}\right)^2\right], \tag{10.206}$$

$$N'_{\dot{v}} = -\pi \left(\frac{T}{L}\right)^2 \left[1.1\frac{B}{L} - 0.041\frac{B}{T}\right], \tag{10.207}$$

$$N'_{\dot{r}} = -\pi \left(\frac{T}{L}\right)^2 \left[\frac{1}{12} + 0.017\mathbf{C}_b \frac{B}{T} - 0.33\frac{B}{L}\right], \tag{10.208}$$

Table 10.5 Ship principal particulars, deadweight, service speed and rudder area

$L = 292\,\text{m}$
$B = 40\,\text{m}$
$T = 12\,\text{m}$
Design load draft: 87 905 MT
$C_b = 0.6097$
$A_\delta = 66.314\,\text{m}^2$
$U = 24.4\,\text{kn}$

$$Y'_v = -\pi \left(\frac{T}{L}\right)^2 \left[1 + 0.40\mathbf{C}_b \frac{B}{T}\right], \tag{10.209}$$

$$Y'_r = -\pi \left(\frac{T}{L}\right)^2 \left[-\frac{1}{2} + 2.2\frac{B}{L} - 0.080\frac{B}{T}\right], \tag{10.210}$$

$$N'_v = -\pi \left(\frac{T}{L}\right)^2 \left[\frac{1}{2} + 2.4\frac{T}{L}\right], \tag{10.211}$$

$$N'_r = -\pi \left(\frac{T}{L}\right)^2 \left[\frac{1}{4} + 0.036\frac{B}{T} - 0.56\frac{B}{L}\right], \tag{10.212}$$

$$Y'_\delta = -\frac{3A_\delta}{L^2}, \tag{10.213}$$

$$N'_\delta = -\frac{Y'_\delta}{2}. \tag{10.214}$$

In the above it must hold that

$$\frac{A_\delta}{LT} \geq 0.01\left[1 + 25\left(\frac{B}{L}\right)^2\right], \tag{10.215}$$

where L (m) and B (m) are the hull length and breadth, respectively, T (m) is the ship's draft and A_δ is the rudder area (m^2). Moreover, the moment of inertia of the ship, I_Z (kg m^2), can be approximately calculated as the one of the circumscribed parallelepiped, where m (kg) is the mass of the ship.

$$I_Z = \frac{1}{12} m(L^2 + B^2) \tag{10.216}$$

The standard dimensionalization equations, used extensively in naval architecture practice, are [10.15]

$$Y'_v = \frac{Y_v}{\frac{1}{2}\rho L^2 U}, \tag{10.217}$$

$$Y'_r = \frac{Y_r}{\frac{1}{2}\rho L^3 U}, \tag{10.218}$$

$$N'_v = \frac{N_v}{\frac{1}{2}\rho L^3 U}, \tag{10.219}$$

$$N'_r = \frac{N_r}{\frac{1}{2}\rho L^4 U}, \tag{10.220}$$

$$Y'_\delta = \frac{Y_\delta}{\frac{1}{2}\rho L^2 U^2}, \tag{10.221}$$

$$N'_\delta = \frac{N_\delta}{\frac{1}{2}\rho L^3 U^2}. \tag{10.222}$$

In the above ρ (kg/m^3) is the seawater density and U (m/s) is the forward advance speed of the ship, for

Table 10.6 A model for the test case ship maneuvering dynamics

$$\begin{bmatrix}\dot\psi\\ \dot v\\ \dot r\end{bmatrix}=\begin{bmatrix}0 & 0 & 1\\ 0 & -0.0279 & -3.7140\\ 0 & -0.0005 & -0.0694\end{bmatrix}\begin{bmatrix}\psi\\ v\\ r\end{bmatrix}+\begin{bmatrix}0 & 0 & 0\\ -0.1077 & -0.5877\times10^{-8} & -0.0006\times10^{-8}\\ 0.0023 & -0.0002\times10^{-8} & +0.0001\times10^{-8}\end{bmatrix}\begin{bmatrix}\delta_R\\ Y_0\\ N_0\end{bmatrix}$$

$$\psi(s)=\frac{0.002305s+0.0001155}{s(s^2+0.09734s+0.0001754)}\delta_R(s)+\frac{-1.518\times10^{-12}s-2.832\times10^{-12}}{s(s^2+0.09734s+0.0001754)}Y_0(s)+\frac{9.43\times10^{-13}s+2.916\times10^{-14}}{s(s^2+0.09734s+0.0001754)}N_0(s)$$

$\tau_1=555.56s$, $\tau_2=10.47s$

$\tau_0=19.96s$, $\tau_Y=0.54s$, $\tau_N=32.34s$

example, service speed. The above formulas give for conventional hulls an a priori (before construction of the full-scale hull) estimation of the ship's maneuverability.

By using these values the major hydrodynamic derivatives have been calculated as in Table 10.6.

Finally, the state-space model of the ship for service speed of 24.4 kn is obtained and thereof the Nomoto transfer functions are calculated. The results are shown in Table 10.7. The values for the time constants in (10.197) and (10.199) are shown in the same table, as well.

From the values shown in Table 10.7, it is seen that the value of τ_1 is one order of magnitude larger than the values of all other time constants. This is typical for the hull forms used in the vast majority of merchant vessels. In specific, the hydrodynamic derivatives obtain such values that instead of the third order Nomoto transfer functions, the reduced (second) order ones in (10.202) may be used for control design. Then, as explained for practical automatic pilot applications, a PD control law can be, under general assumptions, considered, especially for disturbance rejection.

Eventually, a closed-loop transfer function, $\tilde{\mathbf{G}}'_N(s)$, connecting yaw disturbance to heading deflection is derived as in (10.204); a similar governs the effect of sway disturbance on heading.

The objective of disturbance rejection is to make sure that for the worst case scenario disturbance the induced heading fluctuation lies below specified levels. In this case, this can be assured, if transfer function $\tilde{G}'_N(s)$ is matched to the following form

$$\mathbf{G}(s)=\frac{k_0\omega_0^2}{s^2+2\zeta\omega_0 s+\omega_0^2} \tag{10.223}$$

with

$$k_0=0.0175\times10^{-9}\,\mathrm{rad/(N\,m)}\,,$$
$$\omega_0=0.1308\,\mathrm{rad}\,,\quad s,\zeta=1.0\,. \tag{10.224}$$

Then a sinusoidal turning torque N_0 at any frequency and with amplitude not exceeding 10^9 N m will result to a sinusoidal heading error of the same frequency and with amplitude not exceeding $1^\circ=0.0175$ rad. To

Table 10.7 Estimated values for the ship's hydrodynamic derivatives

$Y_{\dot v}$	$=-8.2534\times10^7$ kg
$Y_{\dot r}$	$=-1.0752\times10^9$ kg m
$N_{\dot v}$	$=-2.7437\times10^8$ kg m
$N_{\dot r}$	$=-4.2223\times10^{11}$ kg m^2
Y_v	$=-5.2709\times10^6$ kg/s
Y_r	$=3.9997\times10^8$ kg m/s
N_v	$=-5.1196\times10^8$ kg m/s
N_r	$=-7.4747\times10^{10}$ kg m^2/s
Y_δ	$=-1.5878\times10^7$ N
N_δ	$=2.4190\times10^9$ N m

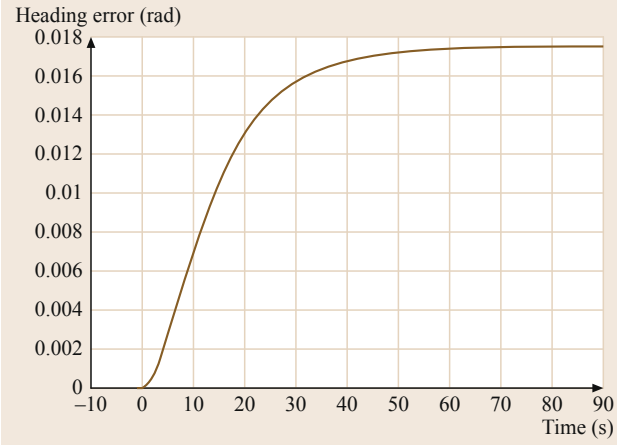


Fig. 10.47 Reduced-order model response to step yaw disturbance

match $\tilde{\mathbf{G}}'_N(s)$ with $\mathbf{G}(s)$ in (10.223) the PD controller gains need to be set as follows

$$k_\mathrm{p}=14.4311\,,\quad k_\mathrm{d}=219.1623s\,. \tag{10.225}$$

The response of the reduced order system (transfer functions of (10.202) in terms of heading error to a step disturbance with magnitude 10^9 N m is shown in Fig. 10.47. In Fig. 10.48, the step response of the full order system, with open-loop transfer functions as in (10.198) are given. As can be seen, in both cases, steady-state error of 1° is demonstrated, which is compliant to specification.

Note that the introduction of a small I-gain can remove the steady-state error, but will not significantly

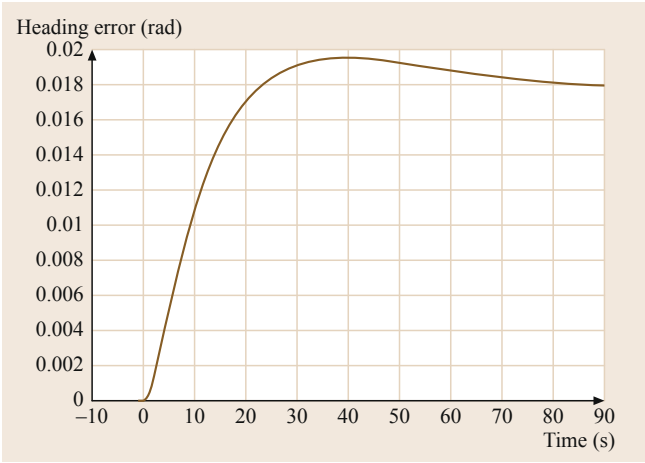


Fig. 10.48 Full-order model response to step yaw disturbance

affect the response of the ship in the first 1min, as during this interval the influence of the PD part of the controller is dominant. Furthermore, note the small overshoot demonstrated in the response of the full-order system is attributed to the presence of the zeros in the numerator of the transfer functions and not to the increase of the order of the characteristic polynomial.

In Fig. 10.49 top the full-order closed-loop system of the ship with the autopilot is shown, while in Fig. 10.49 bottom the reduced order system is shown. Note that the only feedback signal assumed to be available is heading error.

10.5.4 State Observers and Use in Autopilots

A rather important problem to apply the P(I)D control law arising in the common marine autopilot system, is the implementation of the D-term. In the marine autopilot, the D-term corresponds to the yaw rate.

Indeed, the only signal available for feedback is the heading $\psi(t)$ and its differentiation is not recommended in order to avoid amplifying the noise in the control loop. A solution to this problem in practice is to employ a system called a *state-vector observer* (or *estimator*).

In specific, a state observer allows reconstruction on the basis of the measurement (output) vector value. In the case of marine autopilots, the state vector is defined within the framework of the sway-yaw as introduced in the notes. The state observer then obtains the form as

$$\dot{\hat{x}} = \hat{\mathbf{A}}\hat{x} + \hat{B}\delta_{\mathrm{R}} + \mathbf{L}\psi \,, \tag{10.226}$$

where $\hat{x}$ is the state vector estimate (output of the state observer) and $\hat{\mathbf{A}}$, $\hat{\mathbf{B}}$, and $\mathbf{L}$ matrices of appropriate dimensions. Since they define the state observer equations, they need to be determined.

In the 1960s, Luenberger proposed a method to establish the observer matrices based on the pole placement principle introduced in the notes. Prerequisites for Luenberger's method are:

1. The system under investigation upon which state observation is to be applied has to be *observable*.
2. The measurement signals (in the case of the autopilot this is the heading error) need to be reasonably free of any noise or interference.

In the case that the second of the prerequisites is not met, it is possible to construct a state observer by employing optimal filtering techniques. The observer derived in this technique is known from signal theory as Kalman filter.

The Luenberger approach that will be developed here is based on obtaining a dynamic equation for the state prediction error $\hat{e}$, defined as follows

$$\hat{e} = x - \hat{x} \Rightarrow \dot{\hat{e}} = \dot{x} - \dot{\hat{x}} \,. \tag{10.227}$$

By substituting in the above $\dot{\hat{x}}$ as defined previously, $\dot{x}$ by the state equation and ψ from the output equation the following is readily obtained

$$\dot{\hat{e}} = \mathbf{A}x + \begin{bmatrix} 0 \\ \overline{b} \end{bmatrix} \delta_{\mathrm{R}} + d - \left[\hat{\mathbf{A}}\,(x - \hat{e}) + \hat{\mathbf{B}}\delta_{\mathrm{R}} + \mathbf{L}cx\right],$$

$$\Downarrow$$

$$\dot{\hat{e}} = \hat{\mathbf{A}}\hat{e} + \left(\mathbf{A} - \mathbf{L}c - \hat{\mathbf{A}}\right)x + \left(\begin{bmatrix} 0 \\ \overline{b} \end{bmatrix} - \hat{\mathbf{B}}\right)\delta_{\mathrm{R}} + d \,. \tag{10.228}$$

By setting

$$\hat{A} = \mathbf{A} - \mathbf{L}c \quad \text{and} \quad \hat{B} = \begin{bmatrix} 0 \\ \overline{b} \end{bmatrix} . \tag{10.229}$$

Then the differential equation for the state estimation error becomes

$$\dot{\hat{e}} = \hat{\mathbf{A}}\hat{\mathbf{e}} + d = (\mathbf{A} - \mathbf{L}c)\hat{e} + d \,. \tag{10.230}$$

Using the Laplace transform

$$\hat{e}(s) = (s\mathbf{I}_3 - \mathbf{A} + \mathbf{L}c)^{-1} d(s) \,, \tag{10.231}$$

where the above equation is essentially a dynamic equation whose poles can be placed ad hoc by appropriately selecting the values of matrix $\mathbf{L}$. Therefore, the state estimate $\hat{x}$ is finally given by

$$\dot{\hat{x}} = \mathbf{A}\hat{x} + \begin{bmatrix} 0 \\ \overline{b} \end{bmatrix} \delta_{\mathrm{R}} + \mathbf{L}\,(\psi - c\hat{x}) \,. \tag{10.232}$$

Moreover, a good choice for the entries $\mathbf{L}$ the estimation error $\hat{e}$ tends to coincide with random vector $\boldsymbol{d}$. It is noted here that the size of d decides the accuracy of the

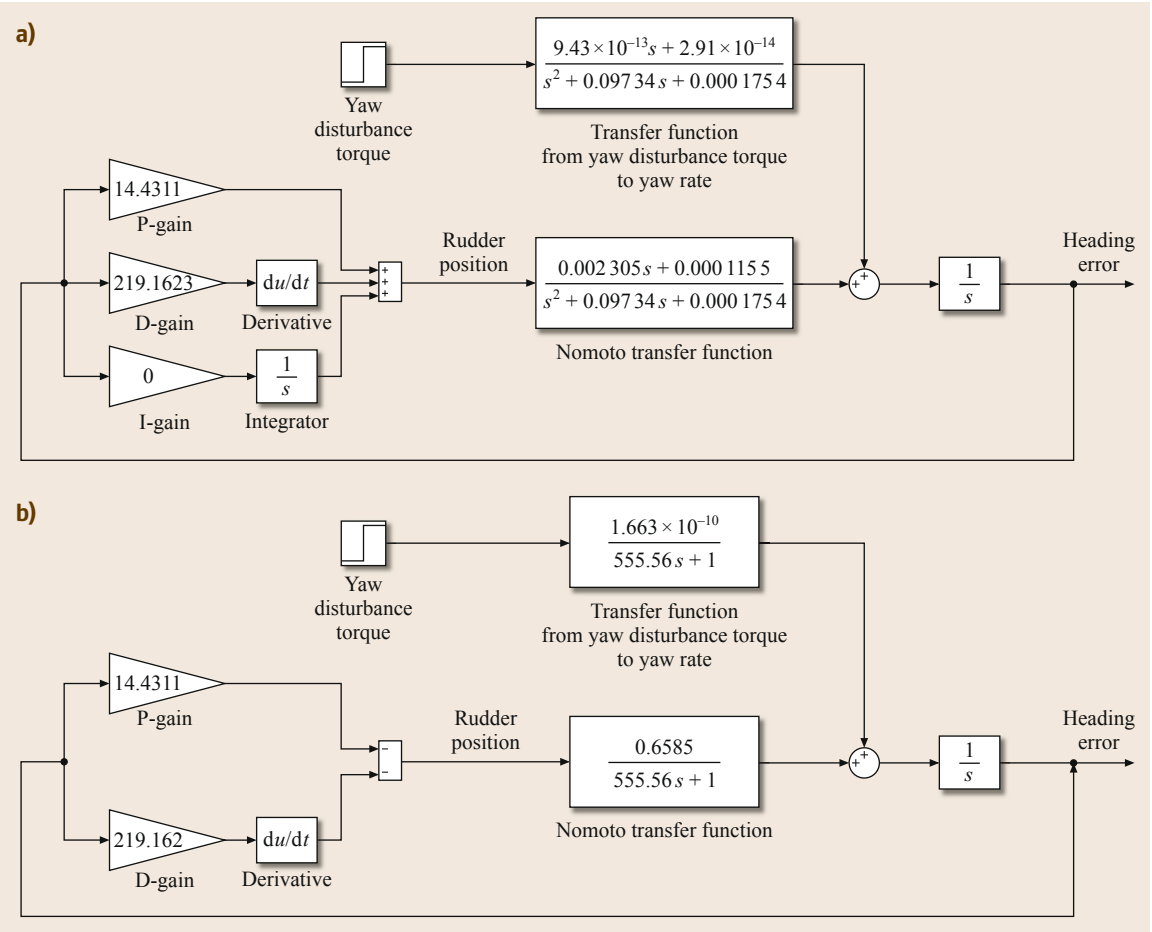


Fig. 10.49a,b Full-order (a) and reduced-order (b) model for ship autopilot design

estimation for x since the estimate $\hat{x}$ cannot approximate x any better.

Also note that (10.232) is essentially the state-space dynamic equation of the system augmented by term $\mathbf{L}\times(\psi - c\hat{x})$. This term is a correction since with a properly set $\mathbf{L}$ allows generating estimates $\hat{x}$ so that model (10.232) of the system approximates the actual system dynamics.

A Luenberger state observer obviously generates not just yaw rate r needed in PID autopilots but for the entire state vector including the sway velocity v. Therefore it can be used to implement advanced control schemes as well like pole placement with full-state feedback as presented earlier in the present chapter.

References

10.1 F. Golnaraghi, B.C. Kuo: *Automatic Control Systems* (Wiley, Hoboken 2010)

10.2 K. Ogata: *Modern Control Engineering* (Pearson Higher, Upper Saddle River 2011)

10.3 S. Skogestad, I. Postlethwaite: *Multivariable Feedback Control; Analysis and Design*, 2nd edn. (Wiley-Interscience, Chichester 2005)

10.4 K. Zhou, J.C. Doyle: *Essentials of Robust Control* (Prentice Hall, Upper Saddle River 1999)

10.5 S. Axler: *Linear Algebra Done Right*, 2nd edn. (Springer, New York 1997)

10.6 R.A. Horn, C.R. Johnson: *Matrix Analysis*, 2nd edn. (Cambridge Univ. Press, Cambridge 2013)

10.7 A. Astolfi, L. Marconi: *Analysis and Design of Nonlinear Control Systems* (Springer, Berlin, Heidelberg 2008)

10.8 A. Isidori: *Nonlinear Control Systems* (Springer, London 1995)

10.9 M. Krstic, I. Kanellakopoulos, P.V. Kokotovic: *Nonlinear and Adaptive Control Design* (Wiley-Interscience, Chichester 1995)
10.10 K.J. Astrom, R.M. Murray: *Feedback Systems: An Introduction for Scientists and Engineers* (Princeton Univ. Press, Princeton 2008)
10.11 B.T. Kulakowski, J.F. Gardner, J.L. Shearer: *Dynamic Modeling and Control of Engineering Systems* (Cambridge Univ. Press, Cambridge 2007)
10.12 J.G. Ziegler, N.B. Nichols: Optimum settings for automatic controllers, Trans. ASME **64**, 759–768 (1942)
10.13 O.J. Smith: *Feedback Control Systems* (McGraw-Hill Publishing, New York 1958)
10.14 W.M. Wonham: On pole assignment in multi-input controllable linear systems, IEEE Trans. Autom. Control **12**, 660–665 (1967)
10.15 N.I. Xiros, E. Logis, G. Charitos: An automatic steering system for robust disturbance rejection, Trans. IASME **1**(2), 324–329 (2004)

Part B Autonom

Part B Autonomous Ocean Vehicles, Subsystems and Control

Ed. by Thomas B. Curtin

11 **Highly Maneuverable Biorobotic Underwater Vehicles**
Promode R. Bandyopadhyay, Newport, USA

12 **Autonomous Underwater Gliders**
Scott A. Jenkins, San Diego, USA
Gerald D'Spain, San Diego, USA

13 **Autonomous Sea Surface Vehicles**
Stefano Brizzolara, Cambridge, USA
Robert A. Brizzolara, Arlington, USA

14 **Autonomous Underwater Vehicle Navigation**
John J. Leonard, Cambridge, USA
Alexander Bahr, Lausanne, Switzerland

15 **Acoustic Communication**
Milica Stojanovic, Boston, USA
Pierre-Philippe J. Beaujean, Boca Raton, USA

16 **Autonomous Underwater Vehicle Docking**
James G. Bellingham, Woods Hole, USA

17 **Underwater Vehicle Manipulators**
Tae Won Kim, Davis, USA
Giacomo Marani, Fairmont, USA
Junku Yuh, Seoul, Korea

18 **Non-Acoustic Sensors**
David Fries, Tampa, USA
William Kirkwood, Moss Landing, USA

19 **Cooperative Vehicle Environmental Monitoring**
Naomi Ehrich Leonard, Princeton, USA

20 **Nested Autonomy for Distributed Ocean Sensing**
Henrik Schmidt, Cambridge, USA
Michael R. Benjamin, Cambridge, USA
Stephanie M. Petillo, Woods Hole, USA
Raymond Lum, Singapore, Singapore

21 **Science of Autonomy: Time-Optimal Path Planning and Adaptive Sampling for Swarms of Ocean Vehicles**
Pierre F.J. Lermusiaux, Cambridge, USA
Tapovan Lolla, Cambridge, USA
Patrick J. Haley Jr., Cambridge, USA
Konuralp Yigit, Urbana, USA
Mattheus P. Ueckermann, Hanover, USA
Thomas Sondergaard, Vedbæk, Denmark
Wayne G. Leslie, Boston, USA

22 **Cooperative Vehicle Target Tracking**
Frank Ehlers, Kiel, Germany

23 **Rules of the Road for Unmanned Marine Vehicles**
Joseph A. Curcio, Gray, USA

24 **Autonomy: Risk Assessment**
Mario P. Brito, Southampton, UK
Gwyn Griffiths, Southampton, UK

An autonomous ocean vehicle is a mobile system whose ability to accomplish a specific mission depends on the design and interplay among its seven principal subsystems: propulsion, sensors, energy, navigation, communication, control, and autonomy. The subsystems are coupled; a change in one impacts all the others in some way. The operating environment for an ocean vehicle is spatially complex, transient, and uncertain. The need for long duration measurements and the constraints of low bandwidth and intermittent communication dictate autonomous capability. Large-scale, three-dimensional, simultaneous coverage of the ocean requires concurrent deployment of multiple vehicles. Under such conditions, cooperative, adaptive vehicle behavior provides a competitive advantage in accomplishing most missions. Of the hundreds of types of unmanned ocean vehicles that have been built, only a few are viable, commercial products. Recent advances in component technologies and subsystem performance have combined to establish new thresholds of capability that are attractive and practical for many missions. This trend is likely to continue. The goal of Part B is to document the current state-of-the-art in some key areas and chart directions for future research and development.

Propulsion is clearly a distinguishing feature of mobile platforms. For subsurface vehicles, mechanical forces and buoyancy forces are the two prime movers. Propellers have dominated mechanical propulsion and are a mature technology, which is not reviewed here. There have been significant advances in biomimetic propulsors, which have great advantages for maneuverability. Chapter 11 reviews biology-inspired systems within a unifying framework of self-regulating nonlinear dynamics. Subsurface vehicles driven by buoyancy forces (gliders) have seen explosive growth in recent years. Chapter 12 reviews the underlying principles of ocean gliders, in which the net hydrodynamic force of lift and drag produced by a wing balances the net buoyancy. Various classes of gliders are categorized and new innovative designs are described.

Ocean surface vehicles play an important and unique role in an autonomous ocean sampling network by providing connectivity at the air-sea interface. Such connectivity includes near-surface sensing, communication links, energy re-supply, and carrying aircraft and subsurface vehicles. In the evolution of autonomous surface vehicles, there has been a tendency to simply adapt existing vessels, designed for other purposes, by removing the crew. Chapter 13 examines the characteristics of surface vehicles specifically designed for autonomous operation from the keel up, and discusses the potential of and need for such multi-functional platforms in an autonomous system of systems.

Precise navigation is essential for mobile platforms and is particularly challenging underwater. Improvements in performance in recent years are related to reduction in the cost and size of component sensors, and advances in algorithms such as simultaneous localization and mapping and cooperative positioning. Chapter 14 reviews the state-of-the-art in estimating the precise position of an autonomous ocean vehicle. The degree of autonomy of an ocean vehicle is inversely related to the continuity and handshaking of its communication link. Communication with an autonomous ocean vehicle is necessary to both acquire data and/or modify its behavior. Radio, sound, and light waves each have advantages and disadvantages in specific applications. Chapter 15 focuses on acoustic channel characteristics and modulation schemes used to provide wireless connection between nodes operating underwater. Networking protocols and peer-peer routing are important enablers for cooperative, adaptive behavior of multiple vehicles.

Onboard energy is always a limitation for autonomous ocean vehicles, as it is for all mobile vehicles. Efficient energy management is a function of subsystem duty cycles, payload demand, and operating technique. One approach is to design a vehicle to carry all the energy it needs to complete its intended mission (e.g., a nuclear submarine). This approach puts a premium on high-energy density power sources and tends toward large displacement hull forms for long duration missions. An alternative approach is to treat energy as a distributed resource in the operational domain (caches, renewable sources, replenishment vehicles, ambient kinetic energy advection), and plan missions to include re-supply and/or ambient exploitation. Most mature transportation systems use a distributed approach with some type of supporting infrastructure. Common features of re-supply are localization of the mobile vehicle relative to an energy source and physical connection (docking) to enable transfer. Chapter 16 reviews advances and trade-offs inherent in autonomous ocean vehicle docking.

Autonomous ocean vehicles are tools for intervention and/or data acquisition. One useful class of intervention is based on manipulators. Manipulation is a mature technology on tethered, remotely operated vehicles. The use of manipulators on autonomous platforms is more challenging. Chapter 17 reviews recent progress in autonomous manipulation. Data acquisition is a function of sensor payload, which determines the

perception of the vehicle. Sensor technology is advancing rapidly, driven by miniaturization, fiber optics, and low-power electronics. Sensors span a broad domain of transduction techniques and signal processing, and can be categorized in various ways. Non-acoustic sensors, in particular, have progressed significantly in recent years, and Chap. 18 reviews this field of development.

There are many advantages of using multiple autonomous ocean vehicles to collaboratively accomplish a mission. Four chapters (19–22) examine various aspects of this approach. Chapter 19 reviews cooperative control of autonomous ocean vehicles for environmental monitoring. An array of vehicles can act as a mobile, reconfigurable, large aperture sensor (antenna). Using feedback control to coordinate relative vehicle motion, such an array can adapt efficiently to a changing environment. The methodology and advantages of such cooperative control are described in two unprecedented field experiments in the coastal ocean. Chapter 20 describes the concept of nested autonomy as an architecture for distributed ocean sensing. Hierarchical, distributed command and control enables deployment of an array of autonomous ocean vehicles over a wide area for a long period of time with little or no human supervision. Effective coverage, limited communication range, safe operation, and environmental uncertainty are all primary constraints. This approach provides a mechanism for balancing autonomy with the periodic need for operator intervention. Chapter 21 focuses on time-optimal path planning and adaptive sampling for groups of ocean vehicles in realistic ocean conditions. Path planning optimization and adaptive sampling both involve prediction of ocean variables, and thus ocean modeling is essential. The state-of-the-art using dynamical models in feedback control of autonomous ocean vehicles is documented for both idealized and realistic environments. The goal is to improve the predictive skill of the continuous (space and time) ocean fields being measured. Optimal path planning and adaptive sampling can also be used to improve the predictive skill of detecting and tracking discrete objects (targets) in the environment. In many sport, military, and biological endeavors, cooperative strategies have proven themselves to be advantageous over non-cooperative ones. For a multi-vehicle team operating in the continuous and transient ocean, cooperative behavior involves optimizing a high-dimensional parameter space, even more so in cases when the target has the ability to make intelligent choices to avoid being tracked. Chapter 22 examines recent progress to construct a theoretical framework and initial applications for maritime surveillance. A methodology is developed that can help bridge the gap between a top-down view starting with theoretical concepts and a bottom-up view based on the details of real experimentation and execution at sea.

Deploying autonomous vehicles for extended periods in the coastal ocean requires safe and reliable operation in potentially congested waterways. The existing rules of the road for marine vessel operation are evolving to accommodate autonomous vehicles. Simultaneously, designers of such vehicles are developing advanced autonomy behaviors that exhibit *human-like* performance. Chapter 23 provides a snapshot of this rapidly changing operational environment, with particular attention paid to the software shaping high-level autonomy and on the legal framework for acceptable integration into public water space. In contrast to coastal ocean operation, where collision is a principal risk, the risk in deploying autonomous vehicles for extended periods in the open ocean lies more in their potential disappearance. Indeed, testing of true autonomy has often been limited by risk-of-loss aversion. When evaluating risks, various stakeholders may weigh consequences differently. For example, the owner may focus on the risk of vehicle loss, whereas the user on risk of data loss. Other risks, such as hazards to navigation, affect all stakeholders. Chapter 24 describes a risk management process using several methods tailored specifically to deployment of autonomous ocean vehicles.

11. Highly Maneuverable Biorobotic Underwater Vehicles

Promode R. Bandyopadhyay

After receiving only a trickle of interest over the past 50 years or so, biolocomotion has more recently attracted broader attention. In this chapter, biology-inspired highly maneuverable underwater vehicles (of ≈ 1 m length scale) that have achieved some degree of maturity are reviewed. Primarily, the Naval Undersea Warfare Center (NUWC) works are described. The nonlinear theoretical foundation of animal-inspired hydrodynamics and control is summarized. Low-speed propulsion (including hovering and maneuvering), control, and integration with sensors are considered. High-lift flapping-fin propulsion technology, cycle-averaged and olivo-cerebellar temporal control, integration with interaural time-differencing bio-sonar, and basin demonstration of utility are discussed. While numerous variables affect the mechanisms and integration of actuators, controllers, and sensors, the principles of self-regulating nonlinear dynamics provide a common and simplifying framework for the development of biorobotic vehicles. The salient features of these vehicles and their control design laws are tabulated. These emergent, low-speed platforms complement existing higher speed naval capabilities and are not substitutes for any existing mature system; in this sense, their utility depends on the imaginative nature of future operation concepts – a process that is unpredictable. The greater potential of the biorobotics approach is limited by a serious lack of progress in so-called strong artificial muscle technology, as well as by a lack of understanding of nonlinear temporal control and sensing and how they should be integrated. Application to utility has been facilitated where understanding of the underlying science has been mathematically formulated, and such formulation needs to be more widely emulated.

11.1 **Biorobotics** 281
11.1.1 Flapping Fin Propulsion Technology 284
11.2 **Theoretical Foundation of Animal-Inspired Hydrodynamics and Control** 286
11.2.1 Hydrodynamics 286
11.2.2 Animal-Like Motion Control Laws and the Principles of Integrated Design 287
11.3 **Description of Biology-Inspired Vehicles of Emergent Maturity** 289
11.4 **Reliability, Low Power Consumption, and Disturbance Rejection of Bio-Inspired Propulsion** 296
11.5 **Demonstrated Maneuverings of NUWC Bio-Inspired Vehicles** 296
11.5.1 BAUV 296
11.5.2 SPLINE 297
11.5.3 RAZOR 297
11.6 **Discussion** 297
11.7 **Concluding Remarks** 298
11.8 **Nomenclature** 298
References 299

11.1 Biorobotics

The 1934 work of *Karman* and *Burgers* [11.1] laid the early theoretical foundation of biorobotics, followed in 1975 by *Lighthill* and others [11.2], who established a broader foundation. Karman and Burgers give the key insight into the hydrodynamic mechanism of biolocomotion [11.2, Fig. 8], which states that in a swimming fish, caudal fins need to flap just enough to produce vortex wakes to give rise to a thrust jet, and the Karman vortex street is present in reverse. However, energy is not wasted in excessive flapping in making the vortex wake and yet the jet cross-section is large, whereby *good efficiency* is achieved [11.2, p. 30].

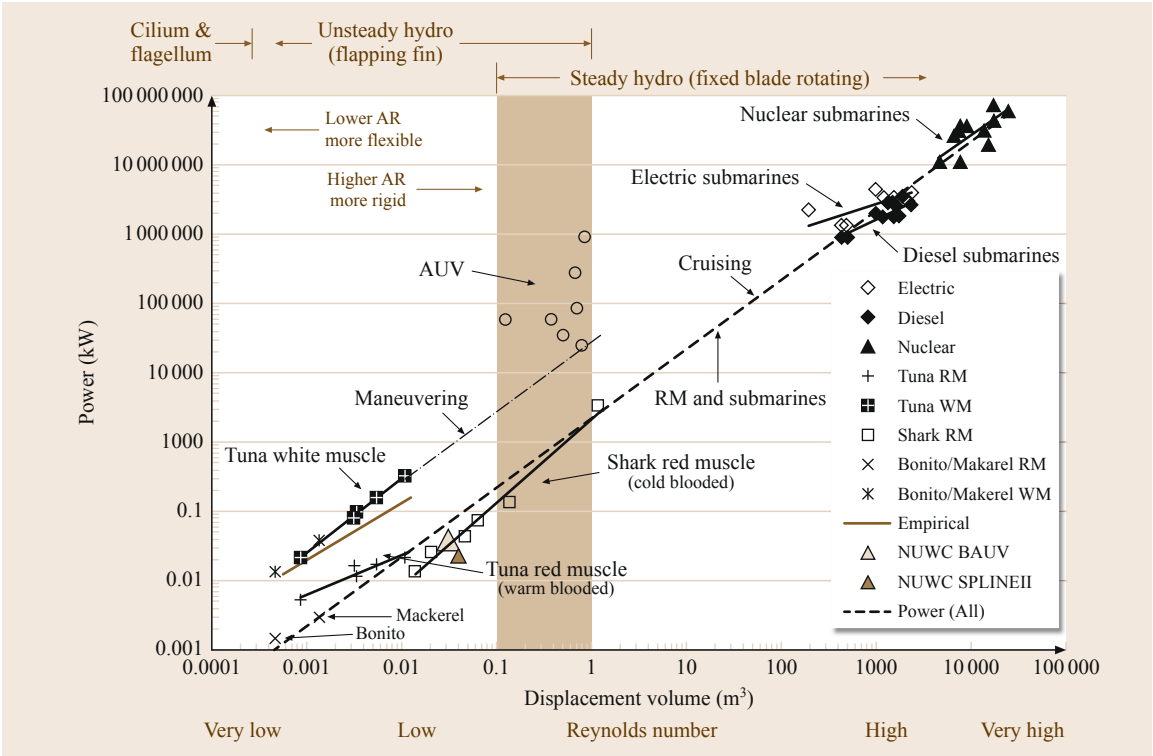


Fig. 11.1 Comparison of propulsion density of animal and manmade autonomous underwater swimmers; the latter shows both biorobotic vehicles (NUWC BAUV, and SPLINEII) and swimmers with conventional rotating propulsors. BAUV: biorobotic autonomous underwater vehicle; SPLINE: self-propelled line array; RM: red muscle (used for cruising); WM: white muscle (used for maneuvering, which follows a different trend (after [11.3])). For data source, see [11.3, 4]; the submarine data are from the open literature [11.5, 6] (reproduced after [11.4, Fig. 3] and after [11.3, Fig. 6])

Biolocomotion is now seen as more than hydrodynamics because animals are seen to be perfectly autonomous, with excellent maneuvering and sensing abilities – a perfect *unmanned underwater vehicle* so to speak. Reverse engineering of swimming animals is thus seen as instructive and the subject of biorobotics attempts to find utility in those animal properties. Here we review the more mature progress in this field.

Research on bio-inspired underwater vehicles seeks to expand the operational envelope by implementing *integrated* principles of hydrodynamics, control, materials, and sensing. The integrated nature of these principles and the differences between bio-inspiration and bio-mimicry are important aspects that we need to understand because success hinges on that understanding. The measure of success is readiness for demonstration on real missions (*transitionability* in naval context) to practical value, and realization of the potential depends on the maturity achieved over a period of 5–10 years, during which a substantial amount of investment will have been made. These aspects are considered below.

We are aware that in ocean engineering, bio-inspired vehicles are considered by some practically minded, influential people as just a curiosity. To rebut such notions, the focus in this review is on vehicles of emergent maturity, as described below. We do not consider the numerous biorobotic vehicles that have not made it out of the confines of the small tank or laboratory, including some of our own early works. Thus, this review leans more on our US Navy work over the last decade, and on the open literature in regard to others' work.

Figure 11.1 shows a meta-analysis of propulsive power (kW) versus displacement volume (m^3), extending over eight decades of these two parameters, has been carried out for engineering underwater vehicles and for animals [11.3–6]. Data from the open literature were utilized. For the animals, the propulsive power was separated for cruising and maneuvering; these functions are attributed to red and white muscles, respectively. The trend for cruising was identical between the engineering vehicles and the red muscles indicating that the drag levels are minimized. The ani-

mals shown use flapping caudal and/or pectoral fins for propulsion/maneuvering. Since the propulsive mechanism of these flapping lifting surfaces makes use of transitional vortex shedding, it must be that larger and larger biorobotic vehicles can also be built with flapping fin propulsion as long as the fin Reynolds number is within the critical range. From the graph in Fig. 11.1, this is a limit of displacement volume of the order of $1\,m^3$ (of the size of a large shark). (We have not been able to find animal data between 2 and $100\,m^3$, and a higher limit is not ruled out.) The tuna white muscle data shows that power required for maneuvering is 10 times greater than that required for cruising. Below, we review our biorobotic vehicles of displacement volume of the order of $0.05\,m^3$.

Let us begin with several important definitions. A *mature platform based on bio-inspiration or bio-mimicry* is one for which:

1. A user mission has been identified.
2. An instrumented version has been built and demonstrated both in the laboratory and in open/saltwaters.
3. Integration with mission sensors has been demonstrated.
4. Documentation or knowledge about improvements over conventional engineering solutions and mechanisms have been experimentally verified and theoretically backed.
5. The biology principle, mechanism, and design laws have been determined, are mathematically founded, and have been peer reviewed, and
6. Long-duration performance has been demonstrated.

Bio-inspiration is described as an undertaking in which the science has been distilled and implemented in a framework that is useful to a user where a bio-inspired platform does not necessarily resemble any animal. In contrast, *bio-mimicry* is an effort in which, first and foremost, biology is mimicked and whereby the resulting platform resembles the chosen animal. The latter, which tends to draw popular interest, may have value in camouflaging.

The subject of unmanned biorobotic underwater vehicles has been reviewed in the past [11.7, 8]. The underlying science and technology have been reviewed by several authors [11.2–4, 7–17]. From the larger biorobotic point of view ($\approx 1\,m$ scale), we separate the propulsive surfaces of animal swimmers in two categories as follows: smaller swimmers with highly flexible caudal and pectoral fins and with main bodies that are more on the flexible side (they have lower Reynolds number); generally larger swimmers whose fins are rather rigid (but not fully rigid) with bodies that are more on the rigid side (they have higher Reynolds number) [11.13]. An example of the former is sunfish (the kind of fish extensively studied in the laboratory [11.18]), and the examples of the latter are penguins, dolphins and whales (the fins are studied mostly biorobotically [11.3, 11, 14, 15, 19]). (Box fish with its rigid body and flexible fins straddles the two categories [11.20].)

Biorobotics received an early boost with the bio-mimicry of caudal fin swimming of tuna fish [11.21–23] in a similar flexible hull. This was followed by departure from bio-mimicry of form – by the application of flapping fin propulsion technology to rigid cylinders – the common engineering underwater hull forms of choice [11.19]. This chapter focuses mainly on flapping fin propulsion and control technology, and the fins are large and fairly stiff (or only slightly flexible).

In relation to swimming animals and from a practical point of view, the primary interest is in the hovering and maneuvering of flow-aligned cylinders or other platforms of about 1 m in length and 30 cm in diameter or depth. Hovering and other maneuverings such as yawing include station-keeping in sea states and in currents. Ideally, such platforms should have high efficiency, with minimal input of momentum into the water and very little acoustic radiation. It is also desirable to have novel requirements such as a built-in disturbance rejection capability (robustness) and rational integration of controllers, actuators, and sensors. Concepts that have a rigorous mathematical formulation are preferred, and a synthesis of bio-physical advances shows that this formulation needs to be nonlinear in character to achieve self-regulation and disturbance rejection properties. (Theoretical definition of self-regulation and why this is a nonlinear process is given in [11.15]; theory showing why the motion of animals is determined by olivo-cerebellar dynamics which is a nonlinear self-regulating process is given in [11.24–29]; and the theory of why sensing, control, and actuator hydrodynamics in swimming and flying animals are likely described by similar self-regulating equations, is given in [11.28].

Below, the flapping fin and the transitional vortices produced are treated as coupled nonlinear oscillators. The transitional (reverse Karman) vortex production process, which is the key to thrust, is shown schematically in Fig. 11.2 for a pectoral fin [11.15, 28]. A leading-edge vortex (LEV) on the fin (Fig. 11.2c) and a reverse Karman vortex train (Fig 11.2b) in the wake are produced when the Strouhal number (St) and pitch amplitude (and Reynolds number) are appropriate (Fig. 11.2f). Due to the motion of the stagnation point [11.15], symmetry is broken [11.29]. As a result, the lift force is enhanced and thrust is produced. This vortex becomes uniform along the span when an appropriate twist is applied (Fig. 11.2d,e). Figure 11.2f shows

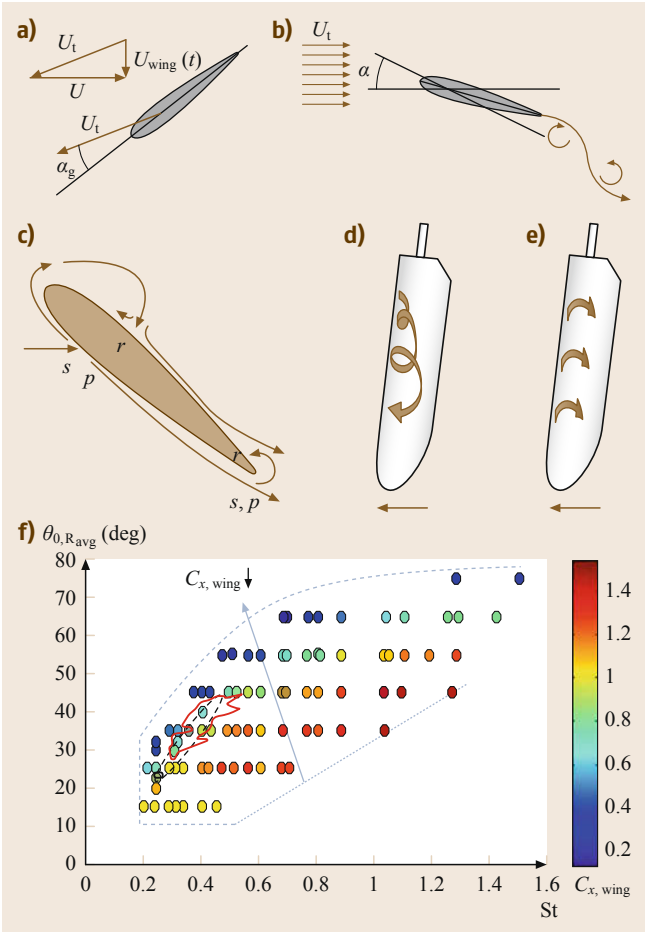


Fig. 11.2a–f Schematics of flow features in a penguin-like flapping pectoral fin: **(a, b)** fin–wake coupling; **(c)** flow bifurcations; and **(d, e)** leading-edge vortex (spirals on fins) in untwisted fins **(d)** and fins twisted spanwise **(e)**; the leading-edge vortex enhances lift force; **(a)** the geometric angle of attack of the fin α_g is determined from the forward fin velocity U and the wing velocity $U_{wing}(t)$, and U_t is total velocity; **(b)** the wake vortices, however, alter the motion of the stagnation points resulting in an effective angle of attack α ($\neq \alpha_g$), and the symmetry breaking by the motion of the stagnation point leads to thrust production; **(c)** streamline symbols are: s (stagnation point), r (reattachment), p (separation); **(d–e)** arrows: direction of fin motion. **(f)** Strouhal number (St) – pitch amplitude (θ_o) parameter space where cruising hydrodynamic efficiency in a single flapping and twisting fin is maximum (> 0.57; region marked by broken ellipse and red lines), or maximum forward thrust (deep red symbols) is produced; thrust is produced only in the region marked by the broken lines; the large arrow indicates direction where forward thrust ($C_{x,wing}$) drops (after [11.15, 29]) ◄

where a fin, hinged at one end, is operated for efficient cruising (area enclosed by red lines) or maximum thrust (red symbols). Efficiency and thrust reach maximum values at lower and higher values of St, respectively, while θ_o remains in the same range; the angle of attack is lower and higher, respectively. Note that the effects of the orthogonal fin roll (given by St) and pitch oscillations (θ_o) are coupled.

11.1.1 Flapping Fin Propulsion Technology

Flapping fins can impart a great deal of momentum to the water because the frequency of flapping can be increased without being constrained by the size of any internal reservoir. Rigid or nearly rigid fins can be scaled up better compared to highly flexible fins as in sunfish. Flapping fins when hinged at one end produce a LEV that remains pinned to the fin due to Coriolis forces [11.30]. As the measurements in [11.16, Fig. 9] show, due to delayed stall, higher lift forces are produced by flapping fins, which can operate at spatiotemporal angles of attack higher than the (first) stall angle. When the fin is flapped at the natural frequency of vortex shedding and the pitch amplitude is commensurate (between 20° and 40°) with the Strouhal number (St) in the range of 0.2–0.4, at a phase difference of about 90° between pitch and roll, efficiency reaches a maximum value (of about 0.60 in rigid fins of low aspect ratio) (Fig. 11.2f) [11.15]. Swim rules have been developed for optimal yawed swimming and yawing [11.3, 15–17, 19]; (the summary is presented later in Table 11.1). Further optimal swim rules have been developed by calibrating a cruising cylinder appended with six flapping fins in a tow tank [11.3]. The vehicle maturities considered here have been achieved by flapping the fins at these optimal conditions and by developing controllers to accomplish the flapping [11.14, 19].

In the vehicles considered here, there is no buoyancy control device because the multiple fins appended to the hull produce the required degrees of freedom with added simplicity. This is useful in rivers and estuaries where salinity can vary greatly. Forces and moments in all directions are controlled by phased oscillation of the flapping fins.

Table 11.1 Summary of flapping fin parameters for optimization of efficiency and thrust during cruising and yawing. Also included are characteristics of wall-shear stress (τ_w) patterns over the flapping fin for these flapping fin parameters: $\phi_0 = 30°$, $\psi = 90°$, $\theta_{t0} = 0°$

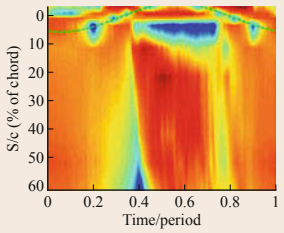
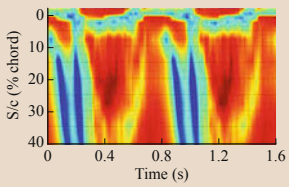
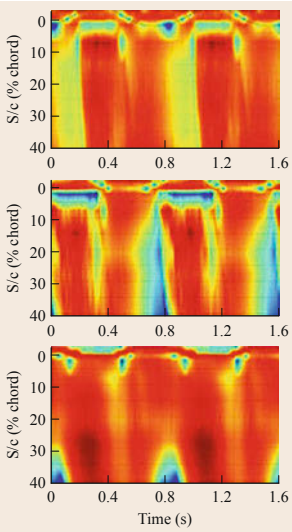
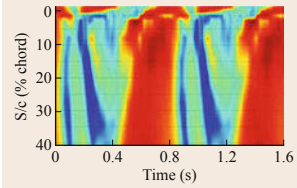
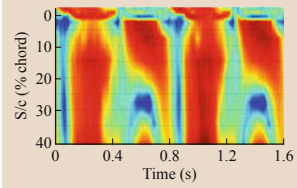
Cruising, hovering, or yawing	Fin kinematics: straight ($\theta_{Bias} = 0$), or yaw ($\theta_{Bias} \neq 0$) swim	Fin optimization parameters based on thrust or efficiency	Salient features of $\tau_w(t)$ on fin surface (blue: low, red: high wall-shear stress); both pressure and suction surfaces are shown; sinusoidal line: stagnation point trajectory; vertical axis: % chord, horizontal axis: time	
Cruising or hovering in straight path (that is, not yawing)				
Cruising: $U > 0$ $U_{wing}/U_t \to 0$ $Re_c = (30 - 130) \times 10^3$	$\theta_{Bias} = 0°$ $\phi_0 = 30°$ $\psi = 90°$ $\theta_{t0} = 0°$	$0.25 \leq St \leq 0.40$ $20° \leq \theta_0 \leq 45°$ $\eta > 0.57$ $C_{x,\eta} = fn(St, \theta_0)$ (higher θ_0 for higher St)	Spanwise bands of low τ_w near the leading edge, followed by high τ_w over the chord away from the leading edge when the stagnation point is advancing to the facing side of the fin during half the cycle of oscillation; during the remaining half of oscillation, as the stagnation point recedes to the other side, a thin band of high τ_w near the leading edge is followed by low τ_w over the remaining part of the chord	
Hovering: $U = 0$ $U_{wing}/U_t \to 1$	$\theta_{Bias} = 0°$, $\phi_0 = 30°$, $\psi = 90°$, $\theta_{t0} = 0°$,	$C_x = fn(f^2)$ $\theta_0 = 45°$	Bands of low τ_w along oscillating stagnation line; this is followed by one pair of a band of low τ_w along the chord per cycle, followed by high τ_w over $s_c/c > 0.05$	
Yawing while cruising or hovering				
Cruising: $U > 0$ $U_{wing}/U_t \to 0$ $Re_c = (30 - 130) \times 10^3$	$\lvert\theta_{Bias}\rvert < 20°$ $\phi_0 = 30°$ $\psi = 90°$ $\theta_{t0} = 0°$	$0.25 \leq St \leq 0.40$ $20° \leq \theta_0 \leq 45°$ $\eta > 0.57$ $C_{x,\eta} = fn(St, \theta_0)$ (higher θ_0 for higher St)	Oscillating leading-edge stagnation point for zero (top, 0°), positive (middle, 10°), and negative (bottom, −10°) pitch bias; bands of low τ_w for $0° < \theta_{Bias} < 20°$: spanwise near leading edge as for no yaw plus some chordwise patterns; for $0° > \theta_{Bias} > -20°$: mostly high τ_w and attached flow	

Table 11.1 (continued)

Cruising, hovering, or yawing	Fin kinematics: straight ($\theta_{Bias} = 0$), or yaw ($\theta_{Bias} \neq 0$) swim	Fin optimization parameters based on thrust or efficiency	Salient features of $\tau_w(t)$ on fin surface (blue: low, red: high wall-shear stress); both pressure and suction surfaces are shown; sinusoidal line: stagnation point trajectory; vertical axis: % chord, horizontal axis: time	
Yawing while cruising or hovering				
Hovering: $U = 0$ $U_{wing}/U_t \to 1$	$0° < \theta_{Bias} < 45°$ $\phi_0 = 30°$ $\psi = 90°$ $\theta_{t0} = 0°$	$C_x = fn(f^2)$ $\theta_0 = 45°$	With increasing positive pitch bias angles, regions of low τ_w increase; one pair of brief bursts of low τ_w extend over chord per cycle of fin oscillation; oscillating stagnation line is a band of low τ_w as when $U = 0$ and $\theta_{Bias} = 0°$	
Hovering: $U = 0$ $U_{wing}/U_t \to 1$	$0° < \theta_{Bias} < -45°$ $\phi_0 = 30°$ $\psi = 90°$ $\theta_{t0} = 0°$	$C_x = fn(f^2)$ $\theta_0 = 45°$	With decreasing pitch bias angles below 0°, the chordwise bands of low τ_w partly become compact and shift mostly down chord	

Note: When τ_w is very low, the local fin boundary layer approaches separation. When τ_w is very high, the boundary layer is attached and may be accelerating. The topology of the near-wall flow over the fin can be defined by critical points such as saddles and foci; in the flapping fins, these flow bifurcation points can be identified: spatiotemporal stagnation point, separation and reattachment points [11.15]

11.2 Theoretical Foundation of Animal-Inspired Hydrodynamics and Control

Flapping fins are nonlinear oscillators. In the case of a drag-producing cylinder in a stream, the lowest order solution of the Navier–Stokes equation of the wake instability reduces to a van der Pol equation [11.31, Appendix]. The latter equation applies also when the cylinder is freely suspended and is galloping (at the natural frequency) [11.32]. Obviously, the cylinder can be replaced by a low-drag foil and the van der Pol form of the Navier–Stokes equation would still describe the lowest order solution of the wake instability. It has been shown [11.15] that the van der Pol solution can be extended to a flapping fin (i. e., if the fin wake is perturbed by a flapping fin motion) oscillating at the natural frequency when the drag wake is replaced by a thrust wake after symmetry breaking [11.29]. In other words, in the transitional range of Reynolds number, vortex-propelled animals can be described by nonlinear oscillators as indeed was implied earlier by *Karman* and *Burgers* [11.1] and *Lighthill* [11.2] and *Bandyopadhyay* et al. [11.15].

11.2.1 Hydrodynamics

As depicted in Fig. 11.2a,b, in the swimming animal, the oscillating surface and the vortex street couple [11.28]. The wake circulation alters the geometric fin angle of attack given by the forward velocity and the fin oscillation velocity. This mechanism allows a freely suspended flexible fin oscillating at the natural frequency and its wake vortex train to adjust to each other (lock-in) in terms of phase and energy, iterating to a low-energy system (minimal losses).

For a single fin hinged at one end that is also twisting and flapping in the optimal manner (Fig. 11.2d,e) near the natural shedding frequency of the wake instability of the nonflapping fin, the instantaneous fluctuating force F'_x produced agrees well – to a leading order particularly at high Reynolds numbers (fin chord and cruise speed-based $Re_c \approx 70\,000$) – with the van der Pol formulation given below,

$$\ddot{F}'_x - \omega_s G(F'^2_{xo} - 4F'^2_x)\dot{F}'_x + \omega_s^2 F'_x = \omega_s T R_{avg}\dot{\xi}\,, \quad (11.1)$$

where

$$\xi(t) = \phi(t) + \theta(t) \,,$$
$$\phi = \phi_0 \sin(2\pi f t) \,,$$
$$\theta(t) = \theta_0 \sin(\omega t + \psi) + \theta_{\text{Bias}} \,,$$
$$\omega_s = 2\omega\pi \frac{StU}{A} \,,$$
$$A = 2\phi_o R_{\text{avg}} \,.$$

F'_x is the fluctuation in force, and F'_{xo} is the amplitude of F'_x; $G(=0.02)$ and $T(=0.5)$ are constants (determined by trial) that affect the shape and the blur of the oscillatory solution (the limit cycle), respectively. In other words, the flapping (rolling and pitching) fin is a nonlinear oscillator and has a limit cycle oscillation (LCO) [11.15]. (The roll, pitch, and twist oscillation of the hinged fin may be seen to act as the $I_{\text{exti}}(t)$ term (see olivo-cerebellar control Sect. 11.2.2 below) to the nonlinear oscillator arising out of the natural vortex shedding of the nonflapping fin.)

Flapping Fin Parameters for Cruising and Maneuvering

Table 11.1 summarizes the fin kinematics that a controller (in the BAUV, SPLINE, and Razor vehicles, Sect. 11.3) uses for optimization of thrust and efficiency during cruising, hovering, and yawing. The total forward velocity is much higher than the wing velocity (which is cross-stream) during cruising, and these velocities are nearly equal during hovering (the forward speed is zero during hovering). Biasing the fin pitch at half the required yawing angle is the key fin setting required for yawing. For optimization of efficiency, the controller makes use of the observed direct coupling between Strouhal number St, whose definition includes roll oscillation amplitude ϕ_0 and pitch amplitude θ_0; a higher pitch amplitude of about 40° is used when St is near 0.40, and a lower pitch amplitude of 20° is used when St is near 0.25 (Fig. 11.2f) [11.15]. Also, note that in a single flapping fin, thrust is a function of the square of the flapping fin frequency f for given pitch amplitude and forward speed [11.15, 19].

The flapping fin mechanism can be understood by observing that the forces produced are describable by van der Pol (11.3) [11.15] or Stuart Landau nonlinear oscillators [11.28]. Using the equation for the latter oscillator, it has been shown theoretically that a coupling exists between the fin and the wake [11.28]. Table 11.1 can be viewed as tuning of the oscillator under various performance requirements. The rightmost column in Table 11.1 summarizes the wall shear stress on the fin for the fin kinematics given in the other columns. Characteristic patterns can be observed:

1. The point of stagnation oscillates, shifting from one pressure side of the fin to another pressure side as the pressure and suction roles of the fin surfaces alternate.
2. Downstream of the stagnation line and on the suction side, a generally spanwise low-wall-shear region develops, indicating the presence of a LEV.
3. Chordwise low-wall-shear bands appear in the suction side in hovering and yawing fins, indicating a complex and less efficient surface flow.

11.2.2 Animal-Like Motion Control Laws and the Principles of Integrated Design

We describe the unmanned underwater vehicle (UUV) design to be integrated when the actuators (fins), controllers, and sensors have similar laws of dynamics. In the case of animal-inspired swimming, these will be equations of nonlinear oscillators.

We assume that the principles of integrated design are derived from the olivo-cerebellar motion control laws [11.24–27, 33], as evidenced by the dynamics of the inferior-olive (IO) neurons that are responsible for motion and balance in all mammals, from rats to human beings [11.24, 25]. We first consider the motion control laws and then how the controller, actuator, and sensors are integrated in a common framework of dynamic systems mathematics.

The motion control laws are self-regulating. In such control, a sensor is not needed for error estimation with respect to a reference because a conventional closed control loop is not present. In other words, these controllers reject disturbances, which make them robust. (See [11.3, 28] for discussion of self-regulation.) Owing to this robustness, the controllers have remained unchanged in many animals.

The olivo-cerebellar dynamics of an IO neuron ($i = 1, 2, \ldots$) can be represented by two coupled nonlinear oscillators (u_i, v_i) and (z_i, w_i) as given below [11.26, 27],

$$\begin{bmatrix} \dot{u}_i \\ \dot{v}_i \\ \dot{z}_i \\ \dot{w}_i \end{bmatrix} = \begin{bmatrix} k(\varepsilon_{\text{Na}})^{-1}(p_{iu}(u_i) - v_i) \\ k(u_i - z_i + I_{\text{Ca}} - I_{\text{Na}}) \\ p_{iz}(z_i) - w_i \\ \varepsilon_{\text{Ca}}(z_i - I_{\text{Ca}}) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ -\varepsilon_{\text{Ca}} \end{bmatrix} I_{\text{exti}}(t) \,, \tag{11.2}$$

where the variables (z_i, w_i) are associated with the subthreshold oscillations and low-threshold (Ca-dependent) spiking, and (u_i, v_i) describe the higher threshold (Na^+-dependent) spiking (this is similar to FitzHugh–Nagumo equation). The constant parameters ε_{Ca} and

ε_{Na} control the oscillation time scale; I_{Ca} and I_{Na} drive the depolarization levels; and k sets a relative time scale between the uv- and zw-subsystems. The nonlinear functions are

$$p_{iu}(u_i) = u_i(u_i - a)(1 - u_i) ,$$
$$p_{iz}(u_i) = z_i(z_i - a)(1 - z_i) .$$

The function $I_{exti}(t)$ is the extracellular stimulus, which can be used for the purpose of control, such as phase synchronization. Note that in a UUV, we do not need to be concerned with the ionic interpretations of the variables and constants; we are only concerned with the generic nonlinear dynamics [11.28].

The (u_i, v_i) oscillator can generate spikes (analogous to neural action potential), but this feature has not been used in the context of UUV maneuvering. The (z_i, w_i) oscillator, however, has been used successfully to synchronize the phase of a six-finned vehicle [11.33].

The salient features of this approach to control compared with conventional engineering controllers (i. e., proportional, integral, derivative (PID) controllers) are that no sensors are required, no reference is required, the controller has built-in disturbance rejection ability, and, in principle, the controller has no size or time scale limitation.

The lower oscillator in (11.2) can then be written as

$$\ddot{z}_i + F(z_i)\dot{z}_i + kz_i + \varepsilon I = 0 , \qquad (11.3)$$

where F is a cubic polynomial function and k is a constant if $I_{exti}(t) = 0$. Equation (11.3) resembles Lienard's oscillator (in contrast, the function F is a well-defined quadratic in the more familiar van der Pol oscillator [11.34, p. 13]). The oscillator exhibits a closed orbit Γ_i in the state space $(z_i - \dot{z}_i)$, that is, $(z_i - w_i)$, which is also known as LCO, with the constant parameters determining the form of Γ_i.

Below, we consider the notions of integrated design and nonlinear sensing because they are theoretically attractive.

Principles of Integrated Design

Swimming animals offer clue to homing, autonomy, and station keeping. The mechanisms are at a theoretical stage of understanding and we touch upon them notionally to spur future work. The principles of integration of the controller, actuator, and sensor are based on the hypothesis of *persistent synchrony with the environment* which after all should enhance mission effectiveness (pursuit of food, mating, and evasion in the case of animals). According to this hypothesis, not only the controller, but also the actuator and the sensor must follow similar dynamics as given essentially in (11.1)–(11.3). We have shown that the actuator does indeed follow laws similar to olivo-cerebellar dynamics [11.15]. It has been shown that it is beneficial for a UUV if olivo-cerebellar dynamics is a common principle of the three elements of the platform. In [11.28], it is shown that an integrated design accelerates homing if we build-in 10–15% of handedness.

Nonlinear Sensing as a Theoretical Requirement of Autonomy

Nonlinear acoustic sensing is another piece of the autonomy puzzle which however is the least developed of the three components, namely sensor, actuator, and controller. Most hydrophones are based on linear vibration of transducers which require lines of them for improved aperture. In [11.35], an alternative theoretical foundation is given based on nonlinear LCO of tiny three-dimensional transducer elements. (These are similar to cilium sensors – widely used by animals.) This transducer has a built-in disturbance rejection (of vibration, e.g.) property. Furthermore, by the use of the property of metachronism in a cluster of such nonlinearly oscillating transducers, acoustic sensing methodologies can be miniaturized and aperture can be increased.

The above rationale for the integration of actuators, controllers, and sensors, together with the need to develop nonlinear oscillatory versions of these three individual systems, gives a mathematically founded basis of autonomy.

11.3 Description of Biology-Inspired Vehicles of Emergent Maturity

Tables 11.2 and 11.3 summarize the salient features of the more mature bio-inspired and bio-mimicry vehicles, respectively. Brief descriptions of the bio-inspired vehicles and sample results are presented in Figs. 11.3 through 11.6. Table 11.4 summarizes the salient features of the engineering controllers (i. e., PID gain, adaptive, or cascading controllers) and olivo-cerebellar controllers that were crucial to demonstration of the low-speed maneuverability of the bio-inspired vehicles.

Figure 11.3a shows the first generation of the six-finned cylindrical BAUV; Fig. 11.3b shows the BAUV mated with a dolphin-inspired interaural time-differencing bio-sonar; and Fig. 11.3c presents representative results showing the tracks of the vehicle maneuvering around obstacles while pointing the bio-sonar at the target; these experiments were carried out in an acoustic test facility.

Figure 11.4a is a schematic of a cable with sensors (such as hydrophones and accelerometers) held taut between two flapping fin propulsion units. Figure 11.4b is a photograph of the components inside the self-propelled line array (SPLINE) hull. Figure 11.4c shows the cable being turned around a pole by one propulsion unit, while Fig. 11.4d gives representative measurements of turn radius, which can be held constant as desired.

In the meta analysis in Fig. 11.1, the trend line of the cruise data of the very large engineering vehicles and the red muscle data of the animals like sharks converge for displacement volumes from 0.01 to 1.0 m^3, although the force production mechanisms are different – steady-state lift properties of fixed blades in the former and pinned-LEV type modified dynamic stall of flapping fins in the latter. Based on this cue, a propulsor was

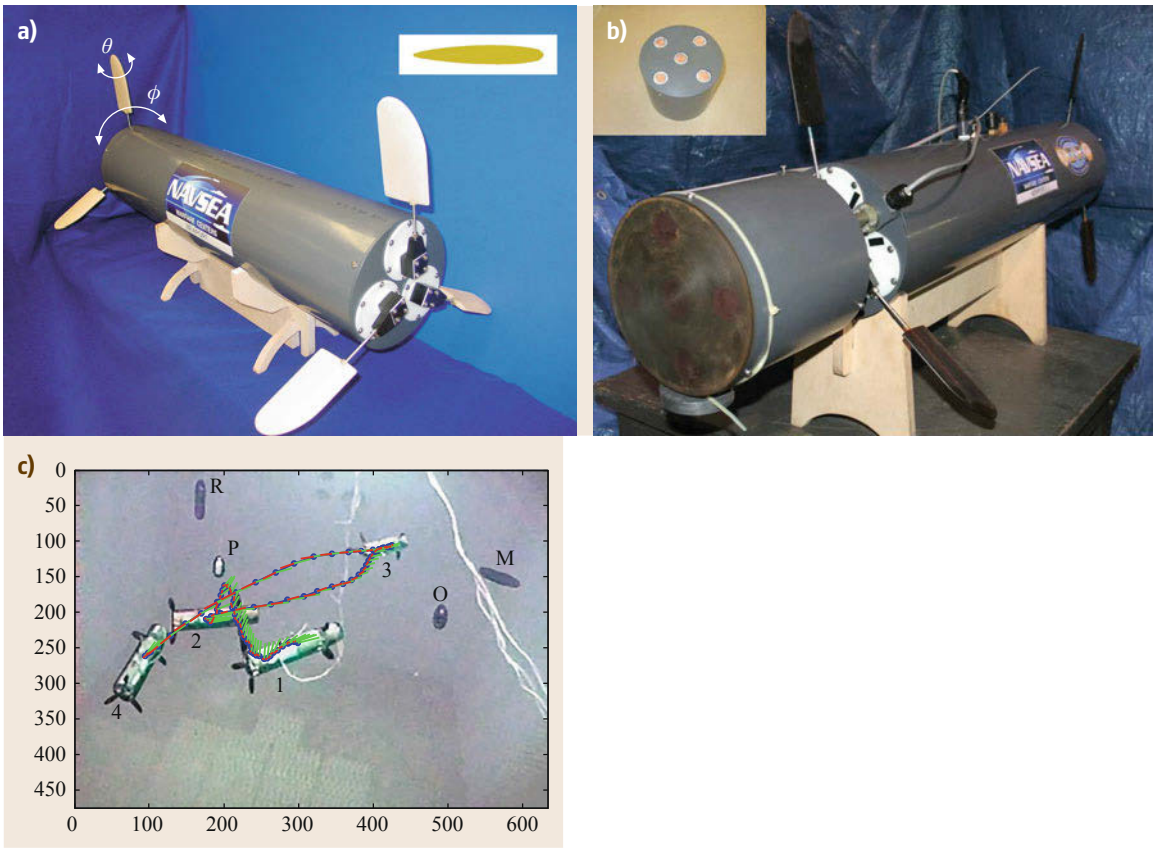


Fig. 11.3 **(a)** Photograph of BAUV, the first generation of the NUWC six-finned bio-inspired UUV. The inset shows the fin section (after [11.19]); fin pitch (θ) and roll (ϕ) oscillations are shown. **(b)** BAUV mated with approximated *bio-sonar* in front; inset shows the four (interaural) receivers and the central transmitter of the bio-sonar (after [11.36]). **(c)** Bio-sonar mounted BAUV navigating (1–4: increasing time) obstacles (M, O, P, R) in the NUWC Acoustic Test Facility (after [11.36]); water tank size: 10.6 m (depth) × 12.2 m × 18.3 m; axes: in pixels

Table 11.2 Summary of more mature bio-inspired underwater platforms

Platform name, institution, figure, and reference	Mission, purpose, comments	Typical dimensions	System and instrumentation	Biology principle, and modes of operation	Comparative performance statement with respect to conventional technology	Future work
BAUV, NUWC, Fig. 11.3 [11.19]	Exploratory development of flapping fin propulsion and olivo-cerebellar control technology	Cylindrical hull length 0.91 m, diameter 0.22 m; fin chord 7.5 cm, span 15 cm	Dolphin-inspired interaural time-differencing bio-sonar; depth sensor; onboard power source and data storage	Flapping fins: low-aspect ratio fins hinged at one end, pitches at 1/3rd chord from leading edge; LEV high lift (pinned-LEV type modified dynamic stall). Control: (a) olivo-cerebellar dynamics control – self-regulating dynamics (LCO); or (b) PID control using synthesis of cycle-averaged fin forces	Exquisite hovering and low-speed maneuverability; low cruise drag at low forward speeds; propulsion power/displacement is shark-like; more efficient than cross-tunnel thruster-based design	Mission-driven self-exploration of optimal maneuvering with olivo-cerebellar dynamics
SPLINE, NUWC, Fig. 11.4 [11.14]	Exploratory development of self-propelled line array concept of persistent surveillance	Cylindrical hull length 1.57 m, diameter 0.2 m; fin chord 9 cm, span 21 cm; cable length between propulsion units 7.3 m	Depth sensor, compass, inertial navigation unit, tilt sensor, line array tension load cell; onboard power source and data storage	Flapping fins: LEV high lift (pinned-LEV type modified dynamic stall); control: PID gain, adaptive, or cascading control using synthesis of cycle-averaged fin forces	Hovering and low-speed maneuverability; low cruise drag at low speeds; propulsion power/displacement is shark-like [11.3], Fig. 11.1; more efficient than cross-tunnel thruster-based design	Station-keeping in highly disturbed environment (high sea states)

Table 11.2 (continued)

Platform name, institution, figure, and reference	Mission, purpose, comments	Typical dimensions	System and instrumentation	Biology principle, and modes of operation	Comparative performance statement with respect to conventional technology	Future work
RAZOR, NUWC, Fig. 11.6 [11.8, 37, 38]	Multimission expeditionary autonomous UUV; harbor defense including swimmer interdiction, bottom surveys, hull inspection, and search, classify, and map missions with 360° sensor coverage of target; payload $> 0.0566\,m^3$; 100 m depth rating	Delta hull; modular design; overall length 1.7 m × depth 0.27 m × 0.76 m wide plus fin span; rear thruster; four rigid fins; two-man portable	Streaming of actuated imaging sonar and video camera; interdiction devices; radio frequency (RF) antenna, camera and global positioning system (GPS) on mast; depth sensor, doppler velocity log (DVL), inertial measurement unit (IMU), onboard power source, and data storage; onboard and tethered programming; field replaceable energy, endurance 6–20 hr with one-three batteries (1.5 kW/hr/tolerant lithium battery); WiFi 54 Mbs secure	Flapping fins: LEV high lift (pinned-LEV type modified dynamic stall); control: (a) hovering and low speed: PID control using synthesis of cycle-averaged fin forces; (b) higher speed: rotational speed control of conventional rotating propulsor; hull planform and fin layout bear similarity to turtles [11.39] but for the rear propulsor, the antenna, and the launch tubes	Exquisite hovering and low-speed maneuverability; more efficient than cross-tunnel thruster-based design; energy efficiency over conventional means of propulsion	Development of controller that transitions from flapping fins to rotating propulsion
SLOSHER, NUWC, Fig. 11.5 [11.40]	Platform mobility	Without shroud: NACA 0012-34 section fin chord 7.5 cm, 16 cm span, abstracted penguin wing planform; with shroud: 0.29 m internal diameter 7.5 cm chord shroud with 7.5 cm chord 8.9 cm span similar fin section	Power diagnostics for efficiency tracking	Fins can flap (roll and pitch) or spin as commanded; flapping fins: LEV high lift (pinned-LEV type modified dynamic stall); spinning fins: steady-state fin lift properties	Sloshing mode at hovering and low forward speeds complements high forward speeds with spinning mode	Simultaneous rolling, pitching, and twisting fins; vectored thrusting; integration with olivo-cerebellar dynamics controller

Table 11.3 Summary of more mature underwater platforms based on bio-mimicry

Platform name, institution, and reference	Mission, comments	Typical dimensions	System and instrumentation	Biology principle	Comparative performance statement with respect to conventional technology	Future work
Robotuna MIT [11.22, 23]	Purpose: Study of fish swimming in controlled manner; appears to support Gray's paradox that the drag of the swimming Robotuna (with oscillating tail) is less than that of the straight (that is, non-flapping, rigid) Robotuna	Seeks to mimic full scale fish: 70 km/hr; skipjack tuna (0.91 m × 35 lbs) – Blue fin tuna (4.27 m × 680 kg); Robotuna: six brushless DC servomotors, pulleys, and cable tendons drive eight discrete rigid vertebra; skin: *thin layer of flexible reticulated foam covered by a conformal Lycra sock*	Not available	Flapping caudal fin propulsion; flexible body shaped as Giant Dania; a wave of variable amplitude passes from nose to tail for propulsion	Directly demonstrated fish thrust production mechanism; tuna form has scalability	Burst of acceleration; fast start; understanding of hydrodynamic mechanism, reduction of power consumption due to the robotic drive; Robopike and GhostSwimmer are derivatives
Finnegan MIT [11.39]	Purpose: explore fin-based swimming in turtle-shaped vehicle configuration; 10 m depth operation	2 m × 0.6 m × 0.5 m; foils protrude 0.4 m from each side, 0.1 m average chord	Onboard power source and data storage; preprogrammed mission; inertial navigation; modified Rodrigues parameter linear control algorithm; open loop control; navigation sensor; pressure sensor; Doppler velocity logger	Mimics turtle; four flapping rigid fins for propulsion appended to a rigid turtle-like hull; linearized engineering control	Improvement in maneuverability by reduction in stability; banked and twisting turns (assisted by hull shape); exceeds turning performance of AUV by a factor of two; minimum turn radius = 0.7 times body length	Not available
GhostSwimmer Boston Engineering Corp. [11.21]	Reconnaissance; hull inspection; oceanographic mission	See Robotuna above	Compact electronics compared to that in Robotuna	Dorsal, pectoral, and caudal fins; flexible body shaped as blue fin tuna;	Small radius turning ability	Efficient actuation of flexible surface, speed enhancement

MIT – Massachusetts Institute of Technology

Table 11.4 Summary of features of controllers of bio-inspired underwater platforms

Platform name, institution, figure, and reference	Control problem scope	Control principle/laws	Summary of demonstration	Comparative statement with respective to conventional approach without bio-inspiration or biomimicry	Future work
BAUV, NUWC, Fig. 11.3 [11.19]	Exploration of maneuvering and hovering in a rigid cylindrical hull appended with six flapping fins (with internal actuator constraints)	Open loop control; cycle-averaged forces and moments are given by fin kinematic parameters based on experimental calibration; fin forces act at one point of fin at a distance of average radius (R_{avg}) from roll axis; net force and moment from synthesis of individual fin forces and moments, and fin placement on hull; PID	Coordination of multiple fins to achieve high maneuverabilities (listed in Sect. 11.5). Solution accommodates inherent constraints of actuator placement within hull.	Far more maneuverable and energy efficient by a factor of 2 compared to cross-tunnel thruster-based vehicle design; no separate buoyancy control is needed	Integration of olivo-cerebellar dynamics with nonlinear sensing and flapping fin propulsion
SPLINE, NUWC, Fig. 11.4 [11.14]	Tension and position control of a short-length cable in laboratory and tidal basin using flapping fin propulsors at each end	Open and closed loop control; approach similar to that in BAUV; PID	Turning of a short-length cable around a pole at a given depth while holding it taut with propulsion at one end; holding of commanded tension and position (horizontal or vertical) in cable with propulsion at one end; swimming of the short-length cable to commanded depth while holding it taut using propulsion at both ends	Minimal input of momentum into water; smoother transition between maneuverings; no separate buoyancy control is needed	Station-keeping in sea state while holding the cable taut
RAZOR, NUWC, Fig. 11.6 [11.8, 37, 38]	Synthesis of four flapping fins and a thruster	Open and closed loop control; approach similar to that in BAUV; control of a thruster for cruising; PID	Coordination of fin propulsion and thrusters; see list of surface and underwater maneuverings in Sect. 11.5.	Fin propulsion for hovering and low-speed maneuvering; thruster does higher speed cruising; no separate buoyancy control is needed	Controller for smooth transition from flapping fin propulsion to thrusters
BAUV, NUWC, Fig. 11.3 [11.33]	Transition of animal-like control to UUV	Limit cycle – olivo-cerebellar dynamics; phase synchronization using SPR property of inferior-olive neurons (SPR-self-referential phase reset [11.26])	Demonstrated faster swimming when phase difference between multiple fins is appropriate	Olivo-cerebellar dynamics has disturbance rejection property; also has phase synchronization property when external impulse is applied; no separate buoyancy control is needed	Station-keeping, gait change, integration with sensors and artificial muscles

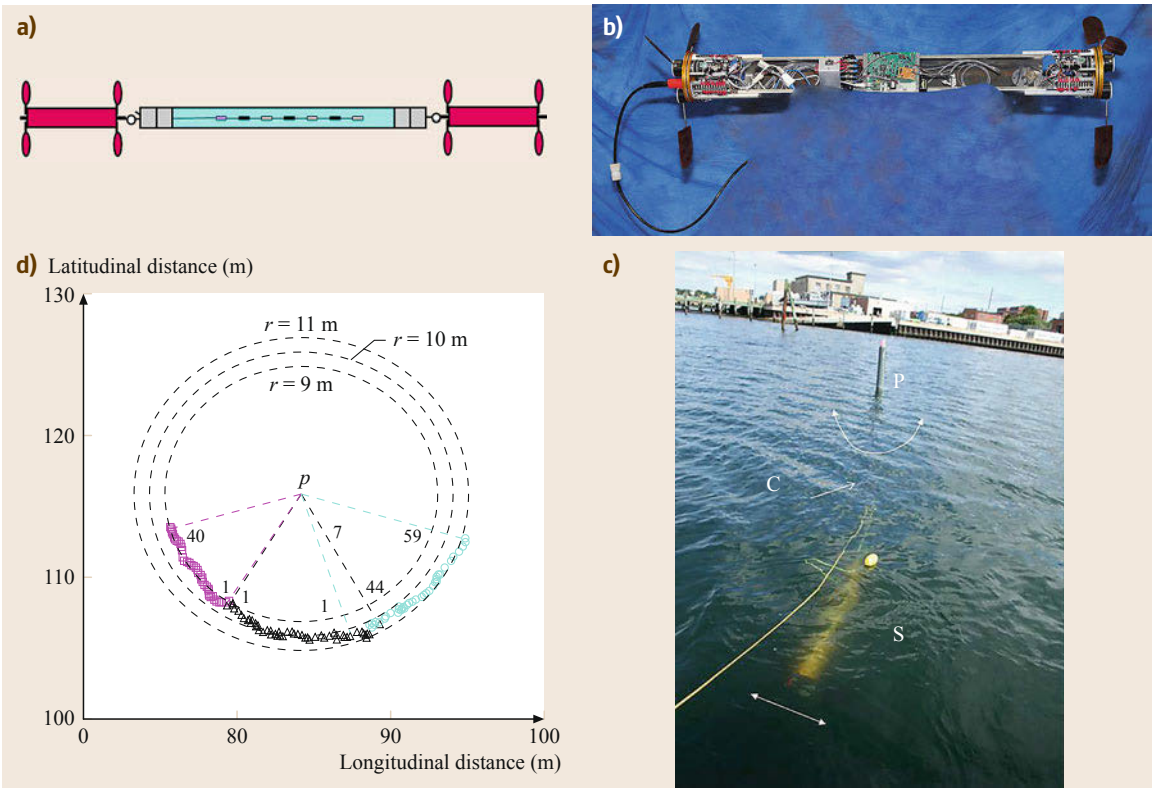


Fig. 11.4 **(a)** Schematic of line array held taut by a pair of SPLINE propulsors. **(b)** Photograph of the instrumented inside of the cylindrical hull of the SPLINE propulsor. **(c)** A SPLINE propulsor turning a short-length cable about a pole (after [11.14]). **(d)** Measurements of turn radius; P is the pole in **(c)**, 7 denotes the position of the 7-th hydrophone, the numerals indicate the direction of increase in time, r is radius, three different runs are shown. In **(c)**: P – pole, C – cable, S – SPLINE; ↔ indicate directions of turning

conceived where the blades could both flap and spin as needed. Figure 11.5a is a photograph of such a novel propulsor where the blades can spin as in conventional propulsors while holding the pitch angle fixed, or the blades can slosh (roll and pitch represented by ϕ and θ, respectively). Figure 11.5b shows the internal blade motor drives and controller housed in the hub. This version of the propulsor also shows a shroud normally used for reduction of acoustic radiation. In the configuration shown in Fig. 11.5a, the shroud is not present and the blades have a boundary layer fence on both sides near the tip. Figure 11.5c shows that thrust can indeed be produced in the spinning and sloshing modes – lower levels of thrust in the sloshing mode and higher levels in the spinning mode.

Figure 11.6a is an outer view of the RAZOR vehicle, and Fig. 11.6b is the RAZOR vehicle mated with the dolphin-interaural spacing bio-inspired sonar shown in Fig. 11.3b. This vehicle has undergone an extensive outdoor capability demonstration (Table 11.2 and Sect. 11.5).

Of the seven biorobotic platforms and devices in Tables 11.2 and 11.3, three disparate platforms – namely, the SPLINE, RAZOR, and GhostSwimmer – are more focused on missions. Judging from this small but varied sample, tentatively speaking, *uncommon low-speed mission in disturbed environments may be the forte of biorobotics* (these vehicles remain to be tested in challenging currents and sea states; Sect. 11.5 lists maneuverings that can be termed uncommon because regular UUVs propelled by rotating propulsors alone cannot undertake those motions efficiently and with minimal disturbance input to the water). Review of the controllers in Table 11.4 indicates that the self-referential phase reset (SPR) mechanism of olivo-cerebellar dynamics may have yet unexplored far-reaching potential in precision station-keeping and platform stabilization; this might facilitate the mat-

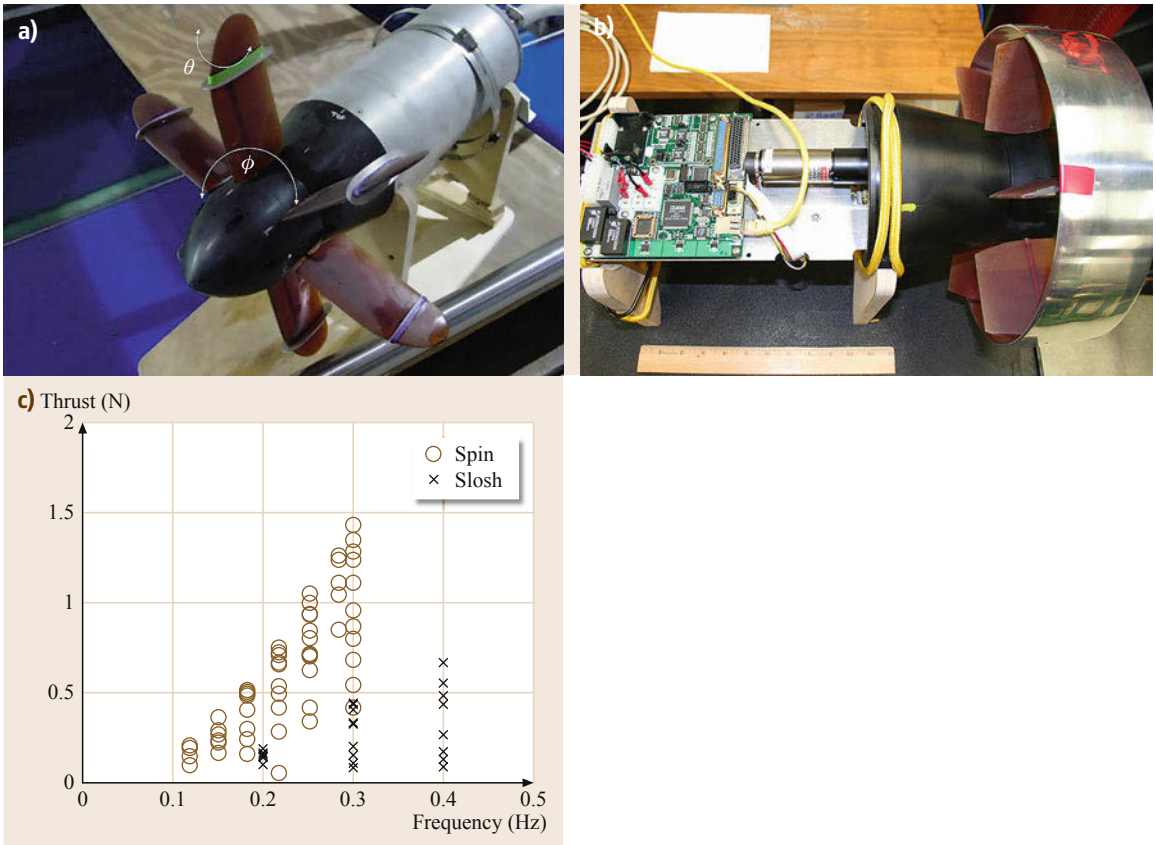


Fig. 11.5a–c Photographs of the slosher propulsor **(a)** and internal motor and controller **(b)**. Note the boundary layer fence in each blade in **(a)**. The shroud in **(b)** was not used in the experiments. **(c)** Thrust measurements when the fins are spinning, or sloshing (after [11.40]). The blade roll (ϕ) and pitch (θ) oscillations in the sloshing mode are shown in **(a)**

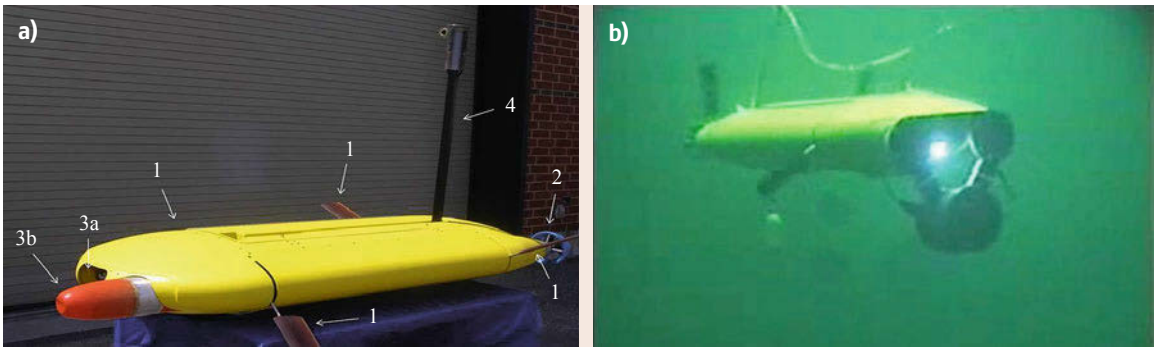


Fig. 11.6 **(a)** Photograph of the two-man-portable RAZOR. *1* – flapping fins, *2* – conventional rotating propulsor, *3a* – launch tube, *3b* – payload, and *4* – 0.76 m-long antenna). **(b)** RAZOR swimming with the approximated bio-sonar and front illumination (after [11.8, 37, 38])

ing of the platform with three-dimensional close-range sectional diagnostics of the target. Overall, the trend is what it should be for UUVs – that is, to provide a low-noise stable platform, in a disturbed environment, that can bring sensors in proximity to the target and make precise observations to assist bigger decision making – and, of course, do so in an autonomous manner.

11.4 Reliability, Low Power Consumption, and Disturbance Rejection of Bio-Inspired Propulsion

The six-finned SPLINE propulsor was operated in a water-filled tank held in position while pulling a fishing line. The line tension and energy consumption were recorded and are shown in Fig. 11.7. Nearly 450 h of continuous run is shown. The hull is capable of carrying the required batteries. The SPLINE propulsor has the ability to loiter at low speeds for several weeks.

The BAUV is a 40 W propulsor and the SPLINE is a 23 W propulsor. The extraordinarily low power consumption was further demonstrated by the trickle charge harvesting experiments described in [11.41, 42]. An electronic circuit was built to harvest very low levels of power available from benthic microbes in littoral basins. Power is stored in 58–450 F capacitors. Operation was automated, and once the capacitors were fully charged, the SPLINE propulsor fins were activated. The 58 F capacitor was able to operate the propulsor for 165 s in a tidal basin. While this is not a long duration, the amenability of bio-inspired propulsion to renewable energy sources opens up unconventional concepts of persistent autonomous operation in the littoral basin for discussion.

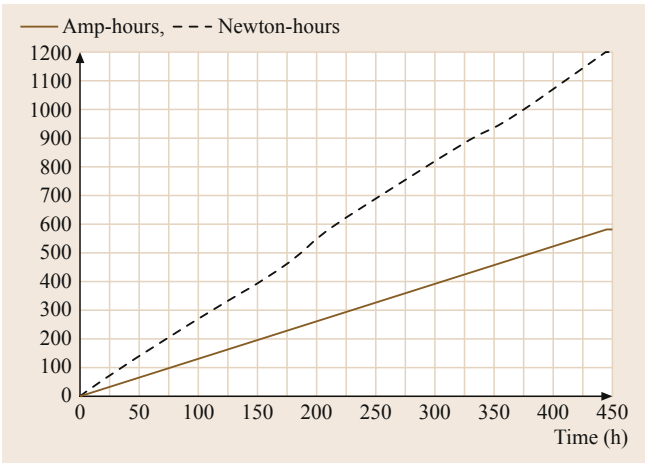


Fig. 11.7 Total energy consumption and cumulative force produced in the SPLINE propulsor, indicating long duration reliability

An experiment was carried out by driving the flapping fin roll and pitch oscillations using a van der Pol oscillator [11.33]. The fin was oscillated near the natural frequency of oscillation for maximum efficiency. The fin roll and pitch oscillation states exhibited a LCO. An artificial electrical disturbance (a square wave) was abruptly applied to the fin to dislodge it from the set LCO; this produced large levels of transient forces, causing the fin to visibly go off track momentarily (see video in [11.33]). Although no sensor was used and no error from the reference was estimated, the thrust force and fin motions autonomously returned to the states of the limit cycle after the disturbance was removed. This experiment demonstrated that a platform in which the actuator, controller, and sensors are integrated by the self-regulation principle (as discussed in Sect. 11.2) will be autonomously robust. *We therefore posit that autonomy in animals is a consequence of the self-regulation property of nonlinear oscillators and biorobotic platforms also need to have self-regulation property.*

11.5 Demonstrated Maneuverings of NUWC Bio-Inspired Vehicles

The NUWC bio-inspired vehicles have demonstrated a variety of precision maneuverings, as listed below. Video recordings of these maneuvering motions have been made and their precision has been determined. (Listed movement precision is in cm.) The unpublished video recordings of the maneuverings are available from the author.

11.5.1 BAUV

1. Near zero meter radius clockwise and counterclockwise horizontal plane turning [11.19]
2. Parallel parking in a narrow rectangular side chamber offset from a straight tank
3. Continuous pitch up and down motion keeping vehicle center at the same location [11.19]
4. Cruising broadside in a narrow tank with minor clearance between BAUV ends and tank walls
5. Cruising while remaining yawed
6. Cruising on water surface remaining partially submerged
7. Cruising, braking, and then making soft landing on tank floor
8. Making soft takeoff from tank floor and resumption of cruising
9. Braking after cruising
10. Swimming to commanded depth and holding that depth
11. Attaching and detaching of suction cups on walls [11.43]
12. Swimming among obstacles while pointing sensors at targets [11.36].

11.5.2 SPLINE

1. Tension and position control of a fishing line or cable [11.14]
2. Simultaneous tension and position control in a fishing line or cable [11.14]
3. Swimming of a 7.3 m-long cable to commanded depth, using two SPLINE propulsors, one at each end of the cable
4. Turning of a short-length cable while holding it taut and maintaining depth in a shallow water tidal basin [11.14].

Demonstrations (1–3) were carried out in large laboratory tanks; number of propulsors: one in one end only in (1–2, 4), and one at each end in (3).

11.5.3 RAZOR

1. Cruising and remotely operated vehicle (ROV)-like maneuvering, including hovering and station-keeping
2. Numerous demonstrations of underwater and surface maneuverings carried out in Narragansett Bay, RI, in more realistic mission environments than is possible in the laboratory.

11.6 Discussion

Exploratory transitioning to utility is a filtering process applied to basic research. The material in this chapter and the papers on the underlying science [11.9, 11, 12, 15] indicate the steepness of the transition pyramid, most being eliminated quickly. Anecdotally, the transition process of underwater vehicles appears to have attained greater maturity than that of flying vehicles, if non-biology-based micro air vehicles are excluded.

The section on integrated design principles shows how disturbance-rejecting nonlinear sensing, which is currently a large gap, should be built. In view of the infancy of nonlinear sensing, we are far off from the full potential of what biology-inspired integrated platforms can offer.

In our experience, the research that transitions to utility is that where understanding is developed to the extent that a mathematical formulation of the mechanism is possible. In our design and control, scaling laws and an understanding of the cycle-averaging process have been found to be useful. To develop reliable control laws, we have experimentally calibrated full-scale actuators in the operational range whenever possible. We have not used any computational fluid dynamics because, in practice, it is postdictive and not predictive.

The value of actuator materials such as strong artificial muscles has been emphasized in the past for actuation of shape deformation of animal-like propulsors [11.7, 44–48]. However, ionic polymer, or the much heralded carbon nanotechnology is yet to make an impact on platforms of 1 m scale.

The Naval utility of platform-length scales much smaller than 1 m has not been demonstrated. This may be due to boundary between the strengths of feasible actuation that is stiff versus highly flexible. System considerations indicate that some approaches such as jets (of squid) are inherently inefficient compared to flapping fins (of fish which impart more momentum to the water without the constraint of internal bladder size) and cannot be scaled up (the oxygen consumption rate of squid is five to seven times higher than that of fish) [11.49]. Furthermore, it does not seem to be useful to develop approaches that are vulnerable to saltwater corrosion and blockage due to biological growth, which always accompanies prolonged exposure to a natural oceanic environment. Caution is in order because, based on past experience of underwater research at NUWC, in some cases extrapolation of developments from clean water to oceanic environments can be fatal.

By many accounts, the NUWC flapping fin propulsion and control technology has matured. These technologies have been built in *affordable* modular form that is detachable/attachable to a platform. Future prospects include their evaluation for stabilizing (and recovering) platforms in distress and in high sea states, or for tautening towed cables. Future interest in biorobotics would be sustained if its mission value is continuously demonstrated to the users and if it is combined with conventional approaches. In basic research, an assessment needs to be made to identify the nagging obstacles to the development of strong muscle technology. In biorobotics basic research, integration of the SPR mechanism with flapping fins needs to be generalized. The results in Figs. 11.2 through 11.7 indicate that in the current stage of maturity, biorobotics is straddling between the laboratory and near-shore open waters.

11.7 Concluding Remarks

UUVs have wide-ranging naval utility. They are the main work horse for unmanned missions, and their long range, versatile payload capacity, and operational envelope offer advantage and value in many naval applications. However, our studies have documented that a large gap in turning ability exists between fish and tactical scale UUVs [11.50]. Advances in digital control have not been able to close this gap. As a result, the gap has been attributed to differences in the hydrodynamic mechanism of lift production – lift production in steady fins versus that in flapping fins – that is, attached fin flow versus pinned-LEV type modified dynamic stall. At NUWC, we have proceeded to implement the high-lift mechanism of flapping fins to naval utility in hovering and low-speed maneuvering, while others have proceeded to mimic swimming animals. These approaches to the development of biology-inspired UUVs have been reviewed by filtering those that we believe have matured the most.

The low-aspect-ratio, rigid flapping fin has been found to be the most successful actuator in the biology-inspired UUVs. We have also been able to mate the flapping fins to rigid cylinders – the most widely used underwater hull form, and also to noncylindrical sections.

Maneuverability and controllability have been demonstrated in both bio-inspired and bio-mimicry UUVs. Mission concepts, duration, payload capabilities, and sensor integration have been demonstrated, and the former type of UUV appears to be somewhat more advanced than the latter. A six-finned cylindrical UUV has demonstrated long-duration reliability, performing incessant swimming for nearly 3 weeks with low power consumption (Fig. 11.7).

While conventional PID controllers have been successfully developed for use with flapping fins, which produce oscillatory forces, control of the instantaneous unsteady forces due to flapping fins using olivocerebellar dynamics should make the UUVs more animal-like [11.24–26, 33, 51, 52]. This would give a more logical foundation to the integration of nonlinear sensors and would also make the UUV controller robust and most responsive to unsteady environmental disturbances. Autonomy is theoretically described as a consequence of the self-regulating property of nonlinear oscillators.

To date, significant progress has been lacking in the development of strong artificial muscles [11.44, 47, 48, 53, 54] and nonlinear sensing [11.35] for underwater application. When these developments move forward, their integration with flapping fins controlled by olivo-cerebellar dynamics would lead to radically new naval precision and other capabilities.

11.8 Nomenclature

A	Amplitude of fin oscillation
$C_{x,\mathrm{wing}}$	Coefficient of axial thrust in a single flapping fin [11.15]
F	Cubic polynomial function
F'_x, F'_{xo}	Instantaneous fluctuating force and force amplitude, respectively
f, ω, ω_s	Frequency of fin oscillation
G, T	Constants
I_{Ca}, I_{Na}	Factors driving depolarization level of inferior-olive neuron
$I_{\mathrm{exti}}(t)$	External stimulus
k	Factor that sets a relative time scale between the uv- and zw-subsystems
p_{iu}, p_{iz}	Nonlinear functions of u_i and z_i, respectively
R_{avg}	Average fin radius
$\mathrm{Re_c}$	Fin chord (c) Reynolds number Uc/υ
St	Strouhal number of a flapping fin, $\mathrm{St} = \frac{2f\phi_0 R_{\mathrm{avg}}}{U}$
U, U_{wing}, U_t	Cruise speed, wing speed, and total forward speed
u_i, v_i, z_i, w_i	States of oscillator i
α	Effective fin angle of attack considering fin–wake coupling
α_g	Geometric fin angle of attack (excluding fin–wake coupling)
Γ_i	Limit cycle orbit
$\varepsilon_{\mathrm{Ca}}$, $\varepsilon_{\mathrm{Na}}$	Factors controlling oscillator time scale of oscillation of inferior-olive neuron
$\theta(t)$, θ_o, θ_{Bias}	Pitch at time t, pitch amplitude, and pitch bias angle, respectively
$\xi(t)$	Combined roll and pitch angle
$\tau_\mathrm{w}(t)$	Temporal local wall-shear stress on the flapping fin surface
$\phi(t)$, ϕ_o	Roll angle at time t and roll amplitude, respectively
ψ	Phase difference between roll and pitch oscillations

References

11.1 T. von Karman, J.M. Burgers: *General Aerodynamic Theory: Perfect Fluids*, Aerodynamic Theory, Vol. II (Springer, Leipzig 1934), ed. by W. F. Durand

11.2 J. Lighthill: *Mathematical Biofluiddynamics*, CBMS-NSF Regional Conference Series in Applied Mathmatics (IAM, Philadelphia 1975)

11.3 P.R. Bandyopadhyay, H.A. Leinhos: Propulsion efficiency of bodies appended with multiple flapping fins: When more is less, Phys. Fluids **25**(4), 041902 (2013)

11.4 P.R. Bandyopadhyay, M. Boller: Convergence in underwater swimming between nature and engineering, Proc. 14th Unmanned Untethered Submers. Technol. (2005)

11.5 S. Saunders (Ed.): *Jane's Fighting Ships 2004–2005* (Janes Information Group, Coulsdon 2004)

11.6 T. Hooton (Ed.): *Jane's Naval Weapon Systems* (Janes Information Group, Coulsdon 2005)

11.7 P.R. Bandyopadhyay: Trends in biorobotic autonomous undersea vehicles, IEEE-JOE **30**(1), 109–139 (2005)

11.8 T.M. McKenna: Commentary: Developing bioinspired autonomous systems, Marine Tech. Soc. J. **45**(4), 19–23 (2011)

11.9 P.R. Bandyopadhyay: Review: Swimming and flying in nature – the route toward applications: The Freeman scholar lecture, ASME J. Fluids Eng. **129**, 31801–31829 (2009)

11.10 P.R. Bandyopadhyay: Biology-inspired science and technology for autonomous underwater vehicles, IEEE-JOE **29**(3), 542–806 (2004)

11.11 M.S. Triantafyllou, A.H. Techet, F.S. Hover: Review of experimental work on biomimetic foils, IEEE-JOE **29**, 585–594 (2004)

11.12 M.S. Triantafyllou, F.S. Hover, A.H. Techet, D.K.P. Yue: Review of hydrodynamic scaling laws in aquatic locomotion and fishlike swimming, Appl. Mech. Rev. **58**, 226–237 (2005)

11.13 M. Sfakiotakis, D.N. Lame, J.V.C. Davis: 1999 Review of fish swimming modes for aquatic locomotion, IEEE-JOE **24**(2), 237–252 (1999)

11.14 P.R. Bandyopadhyay, H.A. Leinhos, J.D. Hrubes, N. Toplosky, J.C. Hansen: Turning of a short-length cable using flapping fin propulsion, IEEE-JOE **36**(4), 571–585 (2011)

11.15 P.R. Bandyopadhyay, D.N. Beal, J.D. Hrubes, A. Mangalam: Relationship of roll and pitch oscillations in a fin of low aspect ratio and rounded leading edge flapping at transitional to high Reynolds numbers, J. Fluid Mech. **702**, 298–331 (2012)

11.16 P.R. Bandyopadhyay, D.N. Beal, A. Menozzi: Biorobotic insights into how animals swim, J. Exp. Biol. **211**, 206–214 (2008)

11.17 D.N. Beal, P.R. Bandyopadhyay: Comparison of steady-state and unsteady hydrodynamic mechanisms, Proc. 14th Unmanned Unethered Submers. Technol. (2005)

11.18 G.V. Lauder, E.G. Drucker: Morphology and experimental hydrodynamics of fish fin control surfaces, IEEE-JOE **29**(3), 556–571 (2004)

11.19 A. Menozzi, H.A. Leinhos, D.N. Beal, P.R. Bandyopadhyay: Open-loop control of a multi-fin biorobotic underwater vehicle, IEEE-JOE **33**(2), 59–68 (2008)

11.20 J. Walker: Kinematics and performance of maneuvering control surfaces in teleost fishes, IEEE-JOE **29**(3), 572–584 (2004)

11.21 M. Rufo, M. Smithers: Ghostswimmer AUV: Applying biomimetics to underwater robotics for achievement of tactical relevance, Marine Tech Soc. J. **45**(4), 24–30 (2011)

11.22 J. Piotrowski: A Fin-tuned Design, http://www.economist.com/node/12494717?story_id=12494717 (2008), The Economist

11.23 M. S. Triantafyllou: D. Isla robotuna project to model real fish, The Tech Online Edition, **115**(115), (1995), http://tech.mit.edu/V115/N49/robotuna.49n.html, also https://en.wikipedia.org/RoboTuna

11.24 R.R. Llinas, Y. Yarom: Electrophysiology of mammalian inferior-olive neurons in vitro, different types of voltage dependences in conductances, J. Physiol. **315**, 549–567 (1981)

11.25 R.R. Llinas: *I of the Vortex* (MIT Press, Cambridge 2001)

11.26 V.B. Kazantsev, V.I. Nekorkin, V.I. Makarenko, R. Llinas: Olivo-cerebellar cluster-based universal control system, Proc. Nat. Acad. Sci. **100**(22), 13064–13068 (2003)

11.27 R.R. Llinas, E. Leznik, V.I. Makarenko: The olivo-cerebellar circuit as a universal control system, IEEE-JOE **29**(3), 631–639 (2004)

11.28 P.R. Bandyopadhyay, H.A. Leinhos, A.M. Hellum: Handedness helps homing in swimming and flying animals, Sci. Rep. **3**, 1128 (2013)

11.29 N. Vandenberghe, N.J. Zhang, S. Childress: Symmetry breaking leads to forward flapping flight, J. Fluid Mech. **506**, 147–155 (2004)

11.30 D. Lentink, M.H. Dickinson: Rotational accelerations stabilize leading edge vortices on revolving fly wings, J. Exp. Biol. **212**, 2705–2719 (2009)

11.31 P. Albarède, P.A. Monkewitz: A model for the formation of oblique shedding and Chevron patterns in cylinder wakes, Phys. Fluids A **4**, 74–756 (1992)

11.32 R.A. Skop, S. Balasubramanian: A new twist on an old model for vortex-induced vibrations, J. Fluids Struct. **11**, 39–412 (1997)

11.33 P.R. Bandyopadhyay, S.N. Singh, D.P. Thivierge, A.M. Annaswamy, H.A. Leinhos, A.R. Fredette, D.N. Beal: Synchronization of animal-inspired multiple fins in an underwater vehicle using olivo-cerebellar dynamics, IEEE-JOE **33**(4), 563–578 (2008)

11.34 H.K. Khalil: *Nonlinear Systems* (Prentice Hall, Upper-Saddle River 1996)

11.35 P.R. Bandyopadhyay, J.C. Hansen: Breakup and then makeup: A predictive model of how cilia self-regulate their hardness for posture control, Sci. Rep. **3**, 1956 (2013)

11.36 S. Forsythe, H.A. Leinhos, P.R. Bandyopadhyay: Dolphin-inspired maneuvering and pinging for

short distance echolocation, J. Acous. Soc. Am. Express Lett. **124**, EL255–EL261 (2008)
11.37 R. Berube: H. Leinhos, personal communication (2011)
11.38 Autonomous Undersea Vehicles Application Center: AUV System Spec Sheet RAZOR Configuration, http://auvac.org/configurations/view/218 (2015)
11.39 S.C. Licht: Biomimetic Oscillating Foil Propulsion to Enhance Underwater Vehicle Agility and Maneuverability, Ph.D. Thesis (MIT, Cambridge 2008)
11.40 P.R. Bandyopadhyay: A novel large slosh-or-spin low speed underwater propulsor bridges the unsteady and steady propulsion mechanisms of nature and engineering, IEEE-JOE (2016), doi:10.1109/JOE.2015.2497879
11.41 D.P. Thivierge, P.R. Bandyopadhyay: Power conditioner for microbial fuel cells, US Patent 8 232 761 (2012)
11.42 P.R. Bandyopadhyay, D.P. Thivierge, F.M. McNeilly, A. Fredette: An electronic circuit for trickle charge harvesting and its evaluation in littoral basin, IEEE-JOE **38**(1), 32–42 (2013)
11.43 P.R. Bandyopadhyay, J.D. Hrubes, H.A. Leinhos: Biorobotic adhesion in water using suction cups, Inst. Phys. J. Bioinspiration Biomim. **3**, 016003 (2008)
11.44 P.R. Bandyopadhyay: Emerging approaches to flow control in hydrodynamics, Proc. 38th IEEE CDC'99 (1999) pp. 2845–2850
11.45 P.R. Bandyopadhyay, W.P. Krol, D.P. Thivierge, W.H. Nedderman, M.A. Mojarrad: *Biomimetic Propulsor for Active Noise Control: Experiments*, NUWC Tech. Rep. 11351 (Naval Undersea Warfare Center, Newport 2002)
11.46 W.P. Krol, A. Annaswamy, P.R. Bandyopadhyay: A Biomimetic Propulsor for Active Noise Control, NUWC Techn. Rep. 11350 (Naval Undersea Warfare Center, Newport 2002)
11.47 P.G.A. Madden, J.D.W. Madden, P.A. Anquetil, N.A. Vandesteeg, I. Hunter: The relation of conducting polymer actuator material properties to performance, IEEE-JOE **33**(3), 696–705 (2004)
11.48 J.D.W. Madden, N.A. Vandesteeg, P.A. Anquetil, P.G.A. Madden, A. Takshi, R.Z. Pytel, S.R. Lafontaine, P.A. Wieringa, I. Hunter: Artificial muscle technology: Physical principles and naval properties, IEEE-JOE **33**(3), 706–728 (2004)
11.49 I.K. Bartol, R. Mann, M.R. Patterson: Aerobic respiratory costs of swimming in the negatively buoyant brief squid *Lolliguncula brevis*, J. Exp. Biol. **204**, 3639–3653 (2001)
11.50 P.R. Bandyopadhyay: Maneuvering hydrodynamics of fish and small underwater vehicles, J. Integ. Comp. Biol. **42**(1), 102–117 (2002)
11.51 A. Menozzi, P.R. Bandyopadhyay, S. Warren: Interfacing olivo-cerebellar neuron hardware to fin actuators of a biorobotic autonomous underwater vehicle, Proc. IASTED Conf. Intell. Syst. Control (2005), Paper No. 497-104
11.52 P.R. Bandyopadhyay, A. Menozzi, D.P. Thivierge, D.N. Beal, A. Annaswamy: Auto-catalytic oscillators for locomotion of underwater vehicles, US Patent 7 869 910 (2011)
11.53 J.W. Paquette, K. Kim: Ionomeric electroactive polymer artificial muscle for naval applications, IEEE-JOE **33**(3), 729–737 (2004)
11.54 K.J. Kim, X. Tan, H.R. Choi, D. Pugal: *Biomimetic Robotic Artificial Muscles* (World Scientific, Hackensack 2013)

12. Autonomous Underwater Gliders

Scott A. Jenkins, Gerald D'Spain

This chapter discusses the characteristics, design considerations, and performance of autonomous underwater (UW) gliders. These buoyancy-propelled, winged vehicles can be categorized as: (1) *profiling gliders* that traverse in bobbing trajectories to collect vertical profiles of ocean properties and (2) *cross-country gliders* designed for point-to-point horizontal transport efficiency. Horizontal transport efficiency is quantified by net transport economy and specific energy consumption. The latter metric for a glider is equal to its inverse lift-to-drag ratio (also called *finesse*) and is equivalent to the glide slope in steady-state, nonturning glides. Increases in efficiency can be obtained by:

1. Increasing the loaded mass (with larger buoyancy engines) and increasing the overall size of the glider, which increases the glider's speed and maintain sufficiently high Reynold's numbers to avoid the drag crisis.
2. Reducing the ratio of the total vehicle wetted area to wing area, via use of flying wing or blended wing body shapes, and
3. Increasing the wing aspect ratio, within structural strength and stiffness limitations.

Gliders have an intrinsic advantage in transport efficiency over conventional prop-driven autonomous underwater vehicles (AUVs) due to the simpler vortex dynamics of a wing versus a propeller. As a result, gliders can fly cooperatively with other winged vehicles or employ multielement wings to further improve transport efficiency. Although a glider must change depth to move forward, these depth changes not only allow the collection of vertical profiles of ocean properties, but also enable the extraction of energy from the ocean's vertical temperature gradients (*thermal glider*).

12.1 **Concept** 301
12.2 **Hydrodynamics of Wings Versus Propellers** 305
12.3 **Underwater Glider Attributes and Limitations** 306
12.3.1 Depth Unlimited Roaming 307
12.3.2 Depth Limited Roaming 307
12.3.3 2-D Station Keeping 307
12.3.4 Payload/Cargo Delivery 307
12.3.5 Level Flight Hybrids 307
12.4 **Optimal Size and Shape for Horizontal Transport Efficiency** 308
12.4.1 Net Transport Economy 308
12.4.2 Size Factors 310
12.4.3 Shape Factors 311
12.4.4 Glide Polar 313
12.5 **Thermal Glider** 318
12.6 **Discussion and Conclusions** 319
References 320

12.1 Concept

The underwater (UW) glider is a buoyancy-propelled, winged vehicle, analogous to a glider in air. The mechanical power of locomotion needed to overcome the drag on the vehicle as it moves through a fluid medium is supplied by gravity in the form of net buoyancy (positive or negative). Horizontal motion using the vertical force of gravity is made possible by the action of lift produced by a wing that acts perpendicular to the trajectory of the vehicle. Therefore, horizontal translation only occurs when the flight path is inclined at a glide angle (Fig. 12.1) that deviates from the horizontal plane in the direction of the vertical net force of gravity (upward for a positive net buoyancy and downward for negative net buoyancy). Inclination of the flight path along some glide angle allows the net hydrodynamic force of lift and drag to balance the net buoyancy in steady-state flight. The inclined flight path that produces this force balance also gives rise to a net vertical motion.

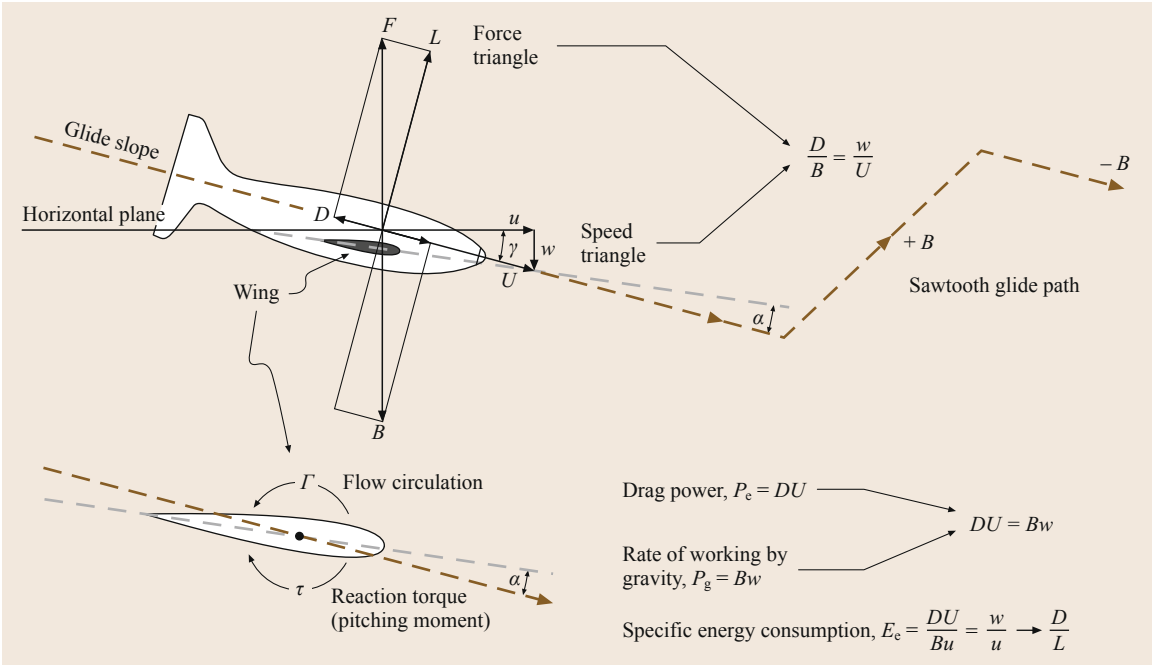


Fig. 12.1 Force balance and energetics for sawtooth glide path (after [12.1]). B = net buoyancy, L = lift, D = drag, F = resultant of lift and drag, u = horizontal speed, w = vertical speed, U = glide speed = magnitude of resultant of horizontal and vertical velocity, Γ = flow circulation, τ = pitching moment that is a reaction torque to the flow circulation, γ = glide angle from the horizontal, α = angle of attack, γ–α = pitch angle

Although analogous, a few differences do exist between air and underwater gliders. First, whereas an airborne glider only executes descending glides, and therefore must create lift only in the upward direction, a UW glider flies ascending as well as descending glides. In order to change the buoyancy between descending and ascending glides, a UW glider must be equipped with a buoyancy engine (Fig. 12.2) that effectively changes the displaced volume of the glider (equivalent to a change in average density for constant mass). This requirement to change the direction of lift – from upward on descending glides to downward on ascending glides – places constraints on the wing design (fixed camber usually is not designed into the wings of UW gliders) and additional demands on vehicle flight control. It also results in some interesting effects; for example, to turn to starboard, an ascending glider must bank to port, opposite the direction of a descending glider. These changes in buoyancy for an underwater glider occur about the point of neutral buoyancy, where the dry bulk weight is balanced by the weight of the water displaced by the vehicle. Therefore, the weight of an UW glider's loaded mass (its net buoyancy) can be remarkably different than its dry bulk weight, whereas these two weights are nearly the same for platforms in air. Another difference in flying in air versus underwater is the stability of the fluid medium. The troposphere, the lowermost 10 km of the earth's atmosphere, is a convective boundary layer with warmer, less dense air underlying colder, denser air. The resulting convective overturn causes turbulence which can make controlled flight very challenging. In contrast, the ocean is stably stratified for the most part, so that once an UW glider is traveling on its desired heading, very little additional flight control is required.

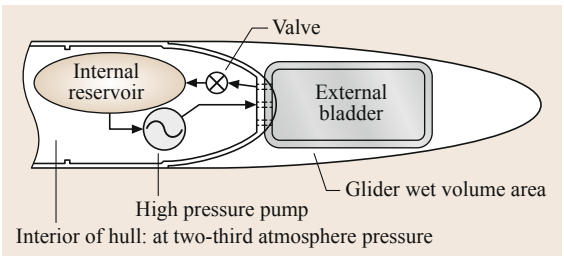


Fig. 12.2 Closed-loop oil-based buoyancy engine (after [12.2])

Underwater gliders originated in the early 1960s with the *Concept Whisper*, a prototype 2-man swimmer delivery vehicle (SDV) built by General Dynamics Corporation [12.5]. A prototype of Concept Whisper was built and tested in shallow dives in San Diego Bay in 1964. Concept Whisper was a classified project

and for several decades, nothing of its concepts and testing was known to the outside world. In the 1970s, analysis of the energetics of UW gliders and an evaluation of various applications was performed at the Naval Electronics Laboratory. At the end of the 1980s, *Stommel* independently proposed a fleet of autonomous gliders referred to as *Slocums* to profile the ocean's water properties [12.6]. Shortly thereafter, underwater gliders started being developed into useful sensor platforms as a result of sustained funding from the Office of Naval Research. These underwater gliders were developed and tested primarily for the purpose of profiling ocean water properties, commensurate with the role first envisioned by *Stommel* [12.6], and so are referred to herein as *profiling gliders* (Fig. 12.3).

The profiling gliders include *Spray*, *Seaglider*, and *Slocum* [12.2–4]. All three designs are based on a conventional winged body of revolution. Emphasis was placed on optimizing the performance of the profiling gliders to travel up and down through the water column at steep glide angles to obtain vertical profiles of water mass properties, rather than optimizing for cross country performance as with conventional gliders in air. The profiling gliders also were designed to accommodate the limited deck space available on small oceanographic vessels, and be two-person portable. As a result, all are similar in size and weight; order 2 m in length, 1 m in wingspan, 50 L in total vehicle volume, and operating at a net buoyancy of about 1–3 N. These three types of UW gliders have logged hundreds of at-sea days during a given mission, collecting ocean profiles. When running *down-wind* with western boundary currents such as the Kuroshio or Gulf Stream Currents, they have covered distances of 1000 km and more.

In the latter half of the 2000s, another type of UW glider, referred to as *cross-country gliders*, was developed for long horizontal range, long duration missions of passive ocean monitoring. These gliders have

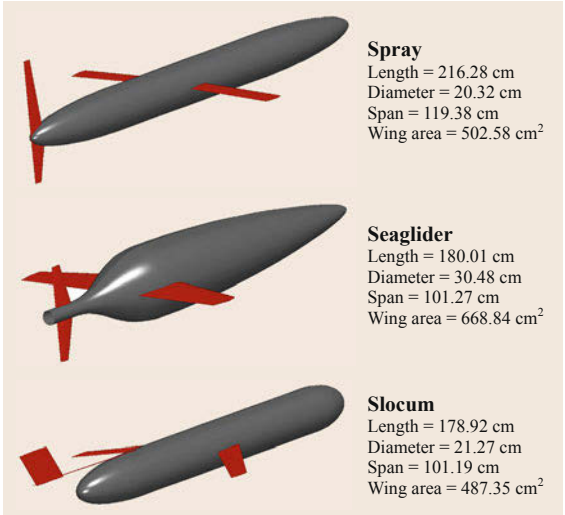


Fig. 12.3 *Profiling* underwater gliders, post *Concept Whisper*. Developed for vertical profiling of ocean water mass properties (after [12.2–4])

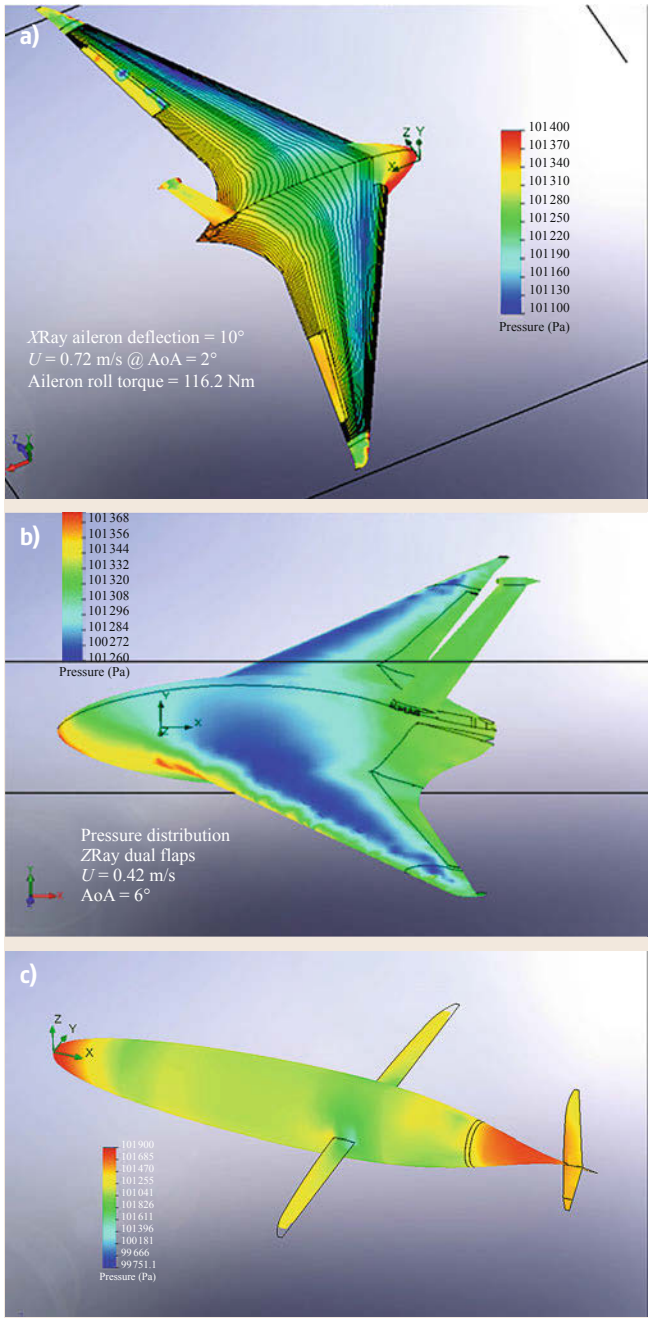


Fig. 12.4a–c Cross-country underwater gliders: Liberdade/XRay **(a)**, Liberdade/ZRay **(b)**, winged body of revolution **(c)**

larger payload and cargo-carrying capacity, and emphasize point-to-point, *cross-country*, transport efficiency. Various designs of these new cross-country gliders have been built and are currently undergoing sea trials, including two examples of flying wings, the Liberdade/XRay and Liberdade/ZRay gliders (Fig. 12.4). The XRay and ZRay designs use a seawater-based buoyancy engine analogous to the ballast system on submarines, while other cross-country glider designs use larger versions of the type of oil-based buoyancy engine diagrammed in Fig. 12.2.

As discussed earlier, underwater gliding is a buoyancy driven form of locomotion in which the power needed to overcome the drag (D) on the vehicle as it moves at a speed U through water is supplied by gravity in the form of positive or negative net buoyancy ($+/-B$). Horizontal translation using the vertical force of gravity is made possible by the lift (L) produced by a wing that acts perpendicular to the trajectory of the vehicle. Inclination of the glide slope from the horizontal gives rise to a horizontal component of lift that provides the force of forward propulsion. In steady-state flight, this force is balanced by the horizontal component of the drag, which yields the relationship that the glide slope is equal to the inverse of the lift-to-drag ratio. The lift-to-drag ratio (L/D) of a wing is also referred to as its *finesse* [12.7]. In the vertical direction, inclination of the glide slope from the horizontal also allows the net hydrodynamic force of lift and drag (F) to balance the net buoyancy in steady-state flight, but implies that a net vertical motion will result, (Fig. 12.1). This vertical motion is referred to as the sink rate (w). During each of the descending or ascending slopes of the sawtooth glide path, the power needed to overcome drag ($P_e = DU$) is equal to the rate of work by gravity acting down (or up) ($P_g = Bw$). Thus

$$P_g = P_e = DU = Bw\,. \tag{12.1}$$

Because the force triangle and the speed triangle in Fig. 12.1 are proportional, the power expenditure per horizontal distance traveled scales in direct proportion to the glide slope ($w/u \approx D/L$), or inversely with the lift-to-drag ratio (L/D). Therefore, by imparting the underwater glider with low drag and high lift properties, its energy consumed to produce locomotion in the horizontal direction can be minimized. In other words, almost all the energy in forward propulsion is consumed at the bottom of the dive cycle where the glider's displaced volume must be increased against ambient pressure. By decreasing the glide slope, i. e., increasing L/D, the number of times this change in displaced volume must occur is minimized for a given horizontal distance traveled.

The *buoyancy engine* generates a variable displaced volume increment, or net buoyancy volume $\pm V_b$; such that the total displaced volume of the glider is $V_d = V_s \pm V_b$, where V_s represents the volume of the glider's rigid body. By varying the net buoyancy volume, the buoyancy engine causes the net buoyancy force to alternate between positive and negative states, $B = \pm\rho g V_b$. Typically, the underwater glider is trimmed to be approximately neutrally buoyant in seawater when $V_b = 0$; so that the average density of the glider approaches the ambient seawater density, $\bar{\rho}_s \to \rho$, and the net buoyancy reduces to $B = \rho g(V_d - V_s) = 0$. In addition to the volume of the glider's rigid body and buoyancy engine displacement volume, the total volume of the vehicle, V_0, also includes void-space water V_{void}, that enters into the freely flooding internal spaces. Thus, the total volume of the vehicle is represented by: $V_0 = V_s \pm V_b + V_{void}$. Because the total vehicle volume is fixed, the void water is a function of the net buoyancy volume, $V_{void} = f(V_b)$, and the hull must accommodate ventilation of the void water with the outside water. Under neutral buoyancy, the total vehicle mass becomes $M_0 = \rho(V_s \pm V_b + V_{void})$. A variety of buoyancy engine technologies have been employed, including closed-loop liquid-based engines [12.2–4] for which buoyancy engine capacity is in the neighborhood of $0.1\% \le V_b/V_0 \le 4\%$; open-loop seawater-based buoyancy engines, as with XRay and ZRay having capacities typically ranging from $1\% \le V_b/V_0 \le 8\%$; open-loop compressed gas-based systems [12.8] having $5\% \le V_b/V_0 \le 20\%$; as well as the open-loop gas-based buoyancy engines that consume gas-generating compounds, such as those used in Concept Whisper [12.5] for which $11\% \le V_b/V_0 \le 28\%$. The profiling gliders use a closed-loop oil-based buoyancy engine as diagrammed in Fig. 12.2, while the cross-country gliders typically use open-loop seawater-based buoyancy engines. The closed-loop liquid-based engines have the disadvantage that the weight of the working fluid (oil) is always onboard, limiting the net buoyancy capacity to about half that of an open circuit buoyancy engine that uses seawater (e.g., as in a submarine buoyancy system). Also, oils can undergo phase changes from liquid to vapor under large decreases in pressure, which can occur in a closed-loop system when the oil is transferred from the external bladder to the internal reservoir to initiate ascent from deep depths. This phase change phenomenon limits the ability to evacuate all of the oil in the external bladder. Regardless, both close- and open-loop buoyancy engines typically are driven by small electric pumps that produce only low levels of intermittent self-noise, ideally suited for passive underwater monitoring applications or those requiring stealthy behavior.

12.2 Hydrodynamics of Wings Versus Propellers

Buoyancy-propelled winged vehicles are intrinsically more efficient than propeller-driven vehicles because the power needed to overcome drag (P_e) is minimized by the simplicity of the vortex system of the wing (Fig. 12.5). From the first law of thermodynamics, P_e is given by the rate of dissipation of the flow kinetic energy as a consequence of the generation of vorticity $\boldsymbol{\Omega} = \nabla \times \boldsymbol{u}$. On a winged vehicle, the preponderance of vorticity is generated during the action of making lift. From [12.11], P_e is given by

$$P_e = \iiint \mu(\boldsymbol{\Omega} \bullet \boldsymbol{\Omega})\,\mathrm{d}x\mathrm{d}y\mathrm{d}z\,, \tag{12.2}$$

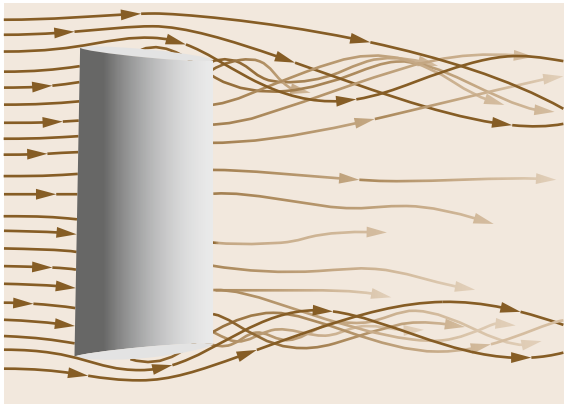


Fig. 12.5 Flow visualization shows the trailing line vortices shed from the wing tips of a constant chord wing as viewed from the top looking downward (after [12.9]). Flow is from *left* to *right*

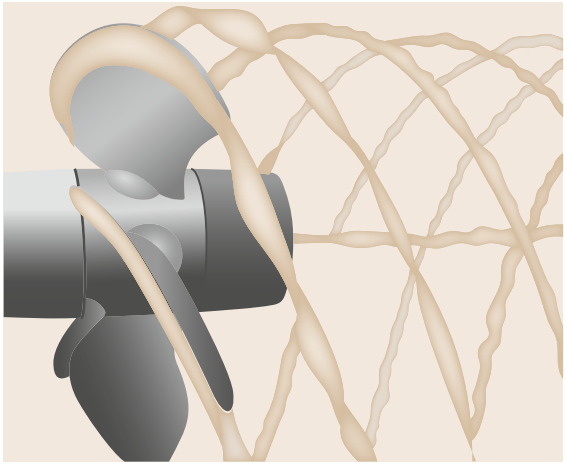


Fig. 12.6 Flow visualization of the trailing helical vortices shed from the tips of a propeller (after [12.9])

where (x, y, z) are the Cartesian coordinates fixed on the wing, with x being the horizontal coordinate in the chord-wise direction, y is the horizontal coordinate in the span-wise direction, and z is the vertical coordinate positive in the upward direction. The quantity μ is the dynamic viscosity, and the flow circulation, Γ, is related to the vorticity induced by a wing with vector airfoil sectional area $\bar{A}$ according to $\Gamma = \int \boldsymbol{\Omega} \bullet \mathrm{d}\boldsymbol{A}$. The vorticity released into the mean flow by a wing is at lowest order one-dimensional (1-D), in the form of a pair of trailing line-vortices shed from the wing tips (Fig. 12.5). In this case, the expression for the power needed to overcome drag reduces to

$$P_e = \mu \iiint \left[\left(\frac{\partial w}{\partial y} - \frac{\partial v}{\partial z}\right)^2\right] \mathrm{d}x\mathrm{d}y\mathrm{d}z \quad \text{(wing)}\,, \tag{12.3}$$

where (u, v, w) are the components of fluid velocity in the (x, y, z) directions, respectively.

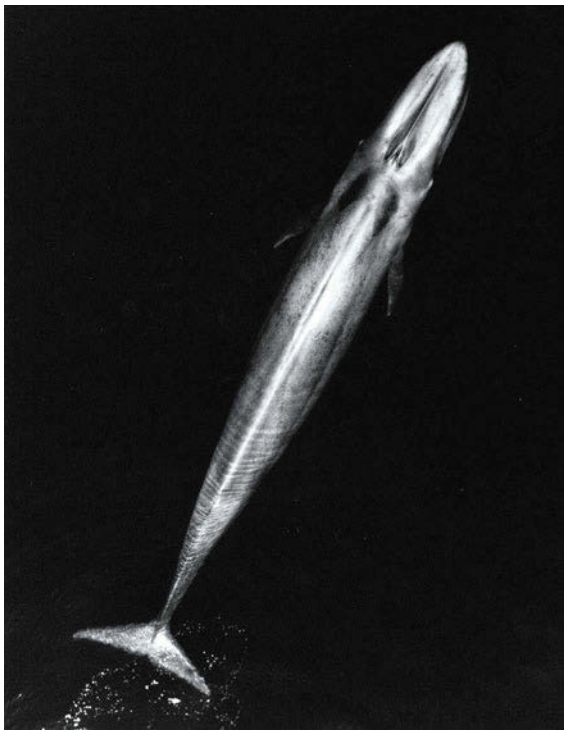


Fig. 12.7 The blue whale evolved winged tails (flukes), not propellers, for propulsion for their annual migrations over thousands of miles between feeding grounds and birthing grounds, as well as for endurance during high-speed pursuits from predators such as killer whales (after [12.10])

In contrast, a propeller is a rotating wing which releases the vorticity into the mean flow. It produces a more complex helical system of trailing vortices imparting swirl to the mean flow (Fig. 12.6). The trailing vortex system from a propeller is fully three dimensional; as a result, the power needed to overcome drag expression contains two additional terms

$$P_e = \mu \iiint \left[\left(\frac{\partial w}{\partial y} - \frac{\partial v}{\partial z} \right)^2 + \left(\frac{\partial u}{\partial z} - \frac{\partial w}{\partial x} \right)^2 + \left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \right)^2 \right] \mathrm{d}x\mathrm{d}y\mathrm{d}z \quad \text{(propeller)} . \tag{12.4}$$

From the additional vorticity components in the vortex trail of a propeller comes higher dissipation rates that require greater expenditures of onboard power to maintain a given vehicle speed through the water, U ($= |\boldsymbol{u}|$). The advantage in transport efficiency of a wing over a propeller in water is illustrated nature through the evolution of the propulsion systems for marine mammals, for example the blue whale (Fig. 12.7). Whales evolved winged tails (flukes) not propellers as the basis of their propulsion systems. Natural selection of winged tails did not occur for lack of a joint capable of 360° rotation, as the shoulder joint in primates clearly proves biomechanically viable.

12.3 Underwater Glider Attributes and Limitations

The most compelling attribute of an underwater glider from the point of view of other types of subsurface vehicles is its high endurance/long on-station time capability. This attribute arises because an UW glider can readily operate over its full flight envelope, from top speed to neutrally buoyant (determined by the net vehicle buoyancy created by its buoyancy engine). Therefore, it can conserve onboard propulsion energy by moving as slowly with respect to the surrounding medium as possible while still accomplishing its objectives. In addition, an UW glider has very low levels of self-noise (acoustically, electrically, and magnetically), because self-noise is created primarily by the buoyancy pump that is activated only episodically for short intervals. Once the underwater glider changes its net buoyancy, it can glide silently with no machinery or hydrodynamic noise until reaching the next reversal in the sawtooth glide path. (Again, the ocean's stable stratification minimizes disturbance from turbulence and so minimizes the need for actuating flight controls.) The intermittency of a glider's self-noise is in sharp contrast to the continuous self-noise emissions from propulsion of prop-driven vehicles. Finally, in contrast to prop-driven platforms, winged structures such as UW gliders can fly cooperatively to improve flight efficiency in horizontal transport. This point is illustrated later in this chapter.

The primary weakness of an UW glider compared with other types of underwater vehicles arises from its inability to maintain level flight in the water column – it must change depth to propel itself forward. This weakness actually is a strength when an objective is to collect vertical profiles of ocean properties.

As with all underwater vehicles, an UW glider is vulnerable to damage – from surface ship collisions, harsh weather, and entanglement – while on the ocean surface. For missions that require covertness, the probability of platform detection is also much greater at the surface; all of a glider's stealthy attributes are compromised during surfacing periods. However, wideband, inexpensive (energy-wise, size and weight-wise, and cost-wise), two-way communications can be achieved only after surfacing. In addition, most environmental sources of energy in the ocean – solar, wind, and wave – are available only at the sea surface. Once an underwater vehicle descends below the ocean surface, it must carry onboard all of the energy supplies required to accomplish its objectives (one exception is discussed in Sect. 12.5).

The characteristic of being *autonomous* for any underwater vehicle also imparts certain benefits and limitations. The primary benefit of autonomy is providing the platform with the ability to accomplish useful objectives (collection of ocean measurements, cargo transport, military-relevant missions, etc.) without direct human input. This ability almost always results in large cost savings, and enables transits in areas too dangerous or inaccessible for humans. The major limitation, on the other hand, is associated with this lack of direct human input. Once it leaves the ocean surface, an autonomous underwater platform must have onboard all of the artificial intelligence required to accomplish its goals. Improving the level of onboard intelligence is the primary challenge to future applications in underwater robotics.

Given these attributes and limitations, UW gliders (and autonomous underwater gliders in particular) are capable of performing a variety of functions and missions. These functions can be categorized as follows [12.1].

12.3.1 Depth Unlimited Roaming

In depth unlimited roaming, the glider is not confined to a particular depth regime while transiting cross-country in a deep ocean environment. Often such roaming will require trans-basin round-trip excursions. The Seaglider and Spray are among this functional class, being designed primarily for the role of gathering ocean vertical profiles of water mass properties. Design adaptations for such depth unlimited roaming require some or all of the following capabilities: long range, cruise speeds adequate to penetrate large-scale ocean circulation but otherwise kept to a minimum to conserve onboard propulsion energy, small-to-moderate payload volume, minimized hotel loads, and deep dive capability with neutral hull compressibility.

12.3.2 Depth Limited Roaming

This function requires cross-country capability within a limited depth regime. This constraint arises when operations are required in the shallow water regimes of the littoral zone, or when the vehicle must operate in a prescribed sector of the water column such as a sound channel or avoid penetration across the thermocline. The Slocum has design adaptations for this role such as its piston pump that allows it to rapidly reverse dive direction in close proximity to the free surface or seabed in confined shallow water areas. Useful design adaptations for this role include: flat glide slope (high L/D) which simultaneously maximizes range for a minimum number of dive cycles, high sprint speed capability to penetrate strong coastal currents, small to moderate payload volume, minimized hotel loads, rapid high-resolution pitch and roll response to avoid broaching or grounding in confined depth regimes, adequate flight control authority for suppressing ocean-surface-wave induced flight oscillations, and avoidance measures for fishing trawlers.

12.3.3 2-D Station Keeping

Station keeping requires the ability to maintain position at a prescribed point in the ocean. Other than the case of free drifting in stagnation flow or grounding on the seabed, underwater gliders must execute depth excursions in order to maintain station at a fixed latitude and longitude. Hence, the station keeping ability is two-dimensional (2-D) and is referred to as 2-D station keeping. All three profiling gliders have demonstrated 2-D station keeping, some within a watch circle of several meters for several weeks at a time. Some station keeping roles may require the glider to profile the water column at a fixed latitude and longitude, which might require maximum sprint and dive speed. Other station keeping roles may involve the glider maintaining station in a certain depth regime, which would benefit from the glider having minimum sink rate properties. In all cases, the glider needs adequate cruise speed to penetrate and hold station against ocean currents. In addition, the glider may be required to hold station on the seabed, in which case it must generate sufficient negative net buoyancy and/or have provisions in its hull and wing shape for anchoring against currents. Other vehicle qualities for station keeping would be long on-station capability, small to moderate payload volume, minimized hotel loads, rapid high-resolution pitch and roll response to execute grounding maneuvers or maneuvers near the seabed, adequate flight control authority when operating near the sea surface, and avoidance measures for fishing trawlers.

12.3.4 Payload/Cargo Delivery

The stealthy, high endurance (long distance combined with long duration) capabilities demonstrated by profiling gliders suggest that delivery of payloads and cargo would be logical function when rapid delivery time is not required. Among the most important characteristics for this breed of underwater glider would be the ability to move large payloads/cargo over significant horizontal distances while minimizing onboard energy consumption (i. e., using minimal numbers of dive cycles). Winged structures such can fly cooperatively to improve flight efficiency. Therefore, an UW glider can be equipped with cargo-carrying exterior compartments so that the total is *more* energy efficient in point-to-point transport than the glider alone, as discussed later in this chapter. This function may require both deep water and depth-limited operational capability (high L/D) with adequate cruise speeds to penetrate both large scale ocean circulation and coastal currents. In addition, large payload volume is required along with minimal hotel loads.

12.3.5 Level Flight Hybrids

This concept is posed to compensate for the major limitation of underwater gliders: the inability to maintain fixed depth in forward transit. This *motor–glider* is a glider with an auxiliary motor-driven screw. Prop-driven AUVs typically are ballasted for safety reasons to be slightly positively buoyant and use either vectored thrust or dive planes to achieve level flight. The motor–glider is a more efficient means for accomplishing the same force balance, using the lift from a high aspect ratio wing to provide the compensating trim

forces against buoyancy to maintain level flight. The concept is envisioned for circumstances where operations require short-term burst speed in very narrow depth regimes, or when level flight is required. Level flight capability may be required for proper operation of certain sensors like seafloor mapping sensors. High burst speed over a limited depth extent may be needed for certain avoidance maneuvers or for penetration in strong shallow water currents. This hybrid propulsion system approach provides the basis for a long range, high endurance prop-driven AUV (autonomous underwater vehicle). Design adaptations for this role require many of the same characteristics as depth limited roaming, with the addition of low drag at high cruise speed to maximize prop-powered flight performance.

12.4 Optimal Size and Shape for Horizontal Transport Efficiency

Transport efficiency varies inversely with the cost of locomotion. It is commonly measured in transportation science by *specific energy consumption* or *net transport economy*, which is equal to the energy consumed per horizontal distance traveled for each unit of vehicle weight [12.7, 12]. Specific energy consumption differs from net transport economy in that the former is based only upon the energy expended in propulsion, whereas net transport economy includes all forms of vehicle energy expenditure including the hotel loads (energy consumed by control systems, communications, payload power requirements, onboard data processing, launch and recovery systems, etc.). Similarly, the vehicle weight can be defined in various ways, for example in terms of total dry bulk weight of the vehicle (including payload), $F_g = \bar{\rho}_s g V_s$. In the next part of this section, it is defined as the weight (net buoyancy force) $Mg = B$, due to the loaded mass M applied to a wing (Fig. 12.1). An alternative weight normalization for specific energy consumption considers only that of the payload/cargo, to eliminate any potential bias toward larger underwater vehicles.

12.4.1 Net Transport Economy

Analytic surveys of natural and man-made fliers provide a starting point for identification of present operating regimes of the profiling gliders and rules of scaling to other performance regimes from those gliders. Design surveys in aeronautical design and bio-mechanics has produced a number of useful works for this purpose [12.7, 12–15].

Analytic flight survey literature reveals that a leading order variable controlling regimes of scale is the loaded mass. In an underwater glider, the loaded mass is supplied by the buoyancy engine. The loaded mass (net buoyancy/g) can be used as the normalization factor in the *net transport economy* (NTE)

$$\mathrm{NTE} = \frac{P}{(Bu)} = \frac{P}{(Mgu)} , \tag{12.5}$$

where P is the total time-averaged power consumed by the flier. The net transport economy is dimensionless and smaller values indicate more efficient transport. The dimensionless NTE is commonly compared to the loaded mass in kilograms [12.13]. Figure 12.8 plots NTE over 12 orders of magnitude variation in loaded mass, covering regimes of scale from insects to jet transports. The NTE values in Fig. 12.8 are based on total power consumption P including both the flight power spent overcoming drag ($P_e = DU$) as well as internal power consumption ($P - P_e$), which includes basal metabolic rates in the case of natural fliers and all subsystem energy consumption in the case of man-made fliers (hotel loads). In [12.13] and [12.15], estimates are made of the basal metabolic rates of birds based on the body mass, resulting in an empirical formulation of NTE

$$\mathrm{NTE} = 0.898(Mg)^{-0.227} \quad \text{(Tucker Fliers)} . \tag{12.6}$$

References [12.12] and [12.14] also examined a limited set of data on birds and derived a similar empirical relation

$$\mathrm{NTE} = 0.914(Mg)^{-0.293} \quad \text{(Schmidt–Nielsen Fliers)} . \tag{12.7}$$

These two empirical relations are indicated by the solid sloping lines in Fig. 12.8. The most apparent scale dependent feature of these empirical relations and the NTE data is that the energy consumed per meter traveled decreases as the loaded mass is increased – bigger fliers (greater loaded mass) are more efficient fliers. The profiling gliders presently operating in the range of 100–300 g of loaded mass are overlaid on Fig. 12.8 as a purple triangle. It appears that the profiling gliders are consuming relatively higher levels of energy per horizontal distance traveled than their bird/bat counterparts operating at equivalent loaded mass. Natural fliers are used as a standard for ultimate efficiency because natural selection tends to eliminates all but the most efficient mutations [12.13]. However, several factors

contribute to this apparent disparity in efficiency. First, the profiling gliders were not designed for efficiency in *horizontal* transport, but rather to move primarily up and down to collect vertical profiles of water column properties. For this case, then, a more useful form for (12.5) is to normalize by U, the speed along the glide slope, rather than the horizontal speed u. This change results in the net transport economy values from (12.5) being multiplied by $u/U = \cos(\gamma)$, where γ is the glide angle from the horizontal (Fig. 12.1). Since the profiling gliders are flown at glide angles of 20° to 30°, this factor results in only a 10% reduction in NTE. Second, rather than normalize the energy consumed per distance traveled by the weight of the loaded mass, NTE can be defined by normalizing by the dry bulk weight. Whereas this change in normalization usually has little effect on NTE for fliers in air (dry bulk weight usually is equal to the weight of the loaded mass), it causes a significant decrease in NTE for underwater platforms since the weight of the loaded mass can be remarkably different than the dry bulk weight due to the positive buoyancy of displaced water. Profiling gliders may be intrinsically less efficient relative to birds because of the extra energy that is consumed when gliding through ocean stratification, particularly when crossing the thermocline. Density changes in the ocean water mass cause corresponding changes in net buoyancy B, resulting in additional rate of working by gravity as defined by the right hand side of (12.1). A comparison of NTE for profiling gliders with (purple triangle) and without (yellow square) ocean stratification is provided in Fig. 12.8. Hull compressibility resulting from depth changes in the ocean cause additional changes in net buoyancy and in the work rates by gravity that ultimately factor against the total power consumption of the UW glider. No counterpart to this increment of energy consumption exists for birds flying in air. To minimize energy consumption due to hull compressibility, the optimal design solution is to match the hull compressibility with seawater compressibility. This match has been done on the deep-diving profiling gliders Seaglider and Spray that operate at very low values of loaded mass (net buoyancy), and where even small changes in net buoyancy can be critical to overall net transport economy.

For comparison, the NTE for the 6.2 m wingspan ZRay flying wing glider (presented later in Fig. 12.12), based on energy consumption measurements made at sea, is plotted in Fig. 12.12 as a red circle. Its NTE value of 0.65 is also somewhat above the Schmidt–Nielson Fliers curve, but approaches it if the ZRay payload is removed (NTE drops to less than 0.45).

To further reveal relative differences in efficiency of the flying shapes, *specific energy consumption* (E_e) can be used. It is a transport economy formulation based

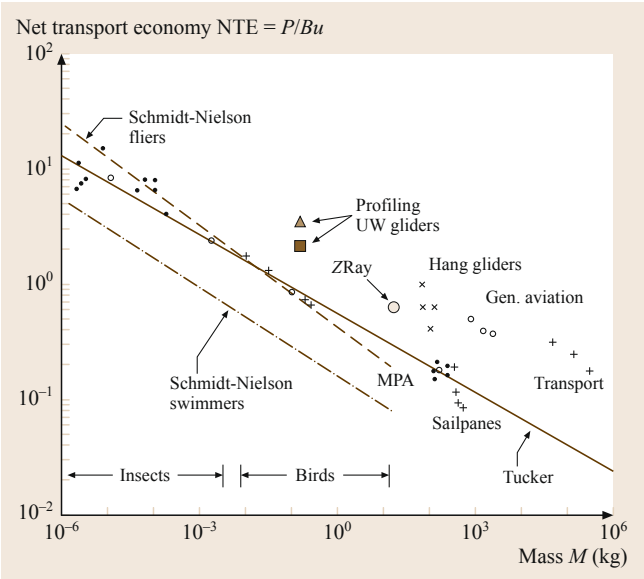


Fig. 12.8 Net Transport Economy (NTE) for natural and man-made fliers (after [12.12]) versus the profiling underwater gliders with (*purple triangle*) and without (*yellow square*) ocean stratification and the ZRay flying wing glider from at-sea measurements (*red circle*). NTE is based on total energy (propulsion plus hotel load and payload) consumption

only on the rate of expenditure of flight energy required to overcome drag ($P_e = DV = Bw$) ([12.7] and (12.1)). Energy is consumed by the buoyancy engine to generate a variable displaced volume increment $\pm V_b$, for forward propulsion. Only the horizontal component of the glide speed, u, results in horizontal distance traveled point-to-point. Consequently, the specific energy consumption (net transport economy) for horizontal transport of an underwater glider is

$$E_e = \frac{P_e}{\rho g V_b u} = \tan\gamma = \left(\frac{L}{D}\right)^{-1} . \qquad (12.8)$$

The glide slope, $\tan\gamma$, is equal to the reciprocal of the lift-to-drag ratio $(L/D)^{-1}$, and provides a physical metric of horizontal point-to-point transport efficiency. In other words, specific energy consumption is minimized by achieving the flattest possible glide slopes. A flat glide slope allows an underwater glider to travel the greatest distance point-to-point for a given number of buoyancy engine cycles. A flat glide slope also enables shallow water operations and other kinds of depth limited applications. Using the conventional quadratic formulation in the β-plane for lift L, and drag D, forces normalized to the wing area A_0 [12.16]

$$L = \frac{1}{2}\rho C_L A_0 U^2 \quad \text{and} \quad D = \frac{1}{2}\rho C_D A_0 U^2 . \qquad (12.9)$$

Here, C_L and C_D are the quadratic lift and drag coefficients, respectively. The drag coefficient is made up of two terms, a profile drag term C_{D0}, that depends on Reynolds number R_e, and an induced drag term C_{Di}, that increases with increasing lift coefficient and decreases with wing aspect ratio, N_R [12.17]

$$C_D = C_{D0} + C_{Di} = K_0 R_e^{-\xi} N_A + K_i \frac{C_L^2}{\pi N_R} . \quad (12.10)$$

In (12.10), K_0 is the profile shape factor, N_A is the ratio of the total wetted surface area to the wing area $N_A = A_t/A_0$, and K_i is the wing plan form factor. The aspect ratio is defined as $N_R = S/\bar{c} = S^2/A_0$ where S is the wingspan and $\bar{c} = A_0/S$ is the mean aerodynamic chord of the glider's wing. The power-law dependence of profile drag is $\xi = 1/2$ for completely laminar, unseparated boundary layer flow over the vehicle, and $\xi = 1/5$ for fully turbulent un-separated boundary layers. In (12.10), the Reynolds number is the size-dependent scale factor $R_e \equiv U\bar{c}/\nu$, where ν is the kinematic viscosity of the fluid. Since the glider is considered a streamlined body, the Reynolds number dependence of the profile drag term in (12.10) is an approximation of friction acting on the total wetted area, A_t. Specific energy consumption, E_c, can be explicitly examined for efficient gliders that achieve small glide angle during coordinated wings-level flight ($v = v' = 0$; $\beta = 0$). With the small glide angle assumption, the lift-to-drag ratio can be written after [12.18] as

$$\frac{L}{D} = \frac{C_L}{C_D} = \frac{u^2}{K_3 u^4 + K_4} = \frac{1}{E_e} , \quad (12.11)$$

where

$$K_3 = \frac{A_0}{2gV_b} K_0 R_e^{-\xi} N_A ,$$

$$K_4 = \frac{2gV_b}{A_0} \frac{K_i}{\pi N_R} .$$

Taking $\partial(L/D)/\partial u = 0$, the speed at which the lift-to-drag ratio is maximized becomes $U = (K_4/K_3)^{1/4}$ giving

$$\left(\frac{L}{D}\right)_{max} = \frac{1}{2}(K_3 K_4)^{-1/2} = \frac{1}{(E_e)_{min}} . \quad (12.12)$$

It can be readily shown that the small glide angle approximation ($L \approx B = \rho g V_b$) also reduces the Reynolds number to

$$R_e \equiv \frac{U\bar{c}}{\nu} \cong \frac{1}{\nu}\left(\frac{2gV_b}{C_L N_R}\right)^{1/2} .$$

Using this approximation to eliminate the Reynolds number in (12.10) and (12.11) gives the minimum specific energy consumption for nonturning, steady-state flight in (12.12) as [12.18]

$$(E_e)_{min} = 2\left[\frac{\nu^\xi K_i K_0 C_L^{\xi/2} N_A}{\pi (2gV_b)^{\xi/2} N_R^{1-\xi/2}}\right]^{1/2} = \left(\frac{L}{D}\right)_{max}^{-1} . \quad (12.13)$$

12.4.2 Size Factors

The minimum specific energy consumption from (12.13) decreases as $V_b^{-\xi/4}$, i. e., it decreases with increasing net buoyancy. Because net buoyancy is some fraction n_b of the total vehicle volume V_0, where $V_b = n_b V_0$, bigger buoyancy-driven vehicles generally are more transport efficient. Actually, surveys of natural and man-made fliers by [12.12] and [12.14] demonstrate that specific energy consumption monotonically decreases across 12 orders of magnitude of size increase. This size advantage is accentuated in an underwater glider because the buoyancy volume factor, n_b, increases with increasing vehicle volume approximately as $n_b \approx 1.2 \times 10^{-5} V_0^{7/6}$, due to economies of scale in packing efficiency [12.7]. Larger n_b permits higher glide speeds (speed increases as the square root of the increase in net buoyancy) and higher wing section Reynolds numbers, which results in higher wing section lift-to-drag ratios (Fig. 12.9).

Equation (12.13) also indicates that specific energy consumption decreases with increasing aspect ratio of the wing, as $N_R^{-(1-\xi/2)/2}$, favoring long tapered wingspans with relatively small wing chords. The aspect ratio of the profiling gliders varies from a maximum of $N_R = 9.75$ for Spray to a minimum of 4.4 for Seaglider. Specific energy consumption decreases as the lift to drag ratio (L/D) increases, but L/D suffers a precipitous decline if the wing section chord is made too small in an effort to achieve a high aspect ratio. Figure 12.9 shows that this L/D crisis occurs when the wing section Reynolds number drops into the mid 10^4 regime. This phenomenon arises because of laminar separation on the suction side of the wing section, which destroys a large percentage of the lift. Profiling gliders with their present 10–20 cm wing chords and 30 cm/s cruise speeds are operating within the laminar separation regime where their wings will not be able to realize a higher L/D and lower E_e by simply flying at higher angles of attack (Fig. 12.9). This laminar separation phenomenon suggests a need to go to bigger wing chords to get above the mid 10^4 Reynolds number regime in order to improve flight efficiency. A primary design philosophy of the larger cross-country gliders (Fig. 12.4) is to achieve

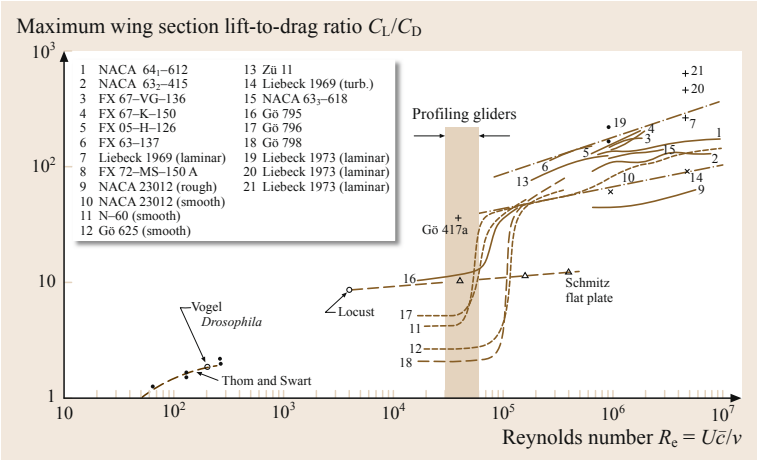


Fig. 12.9 Variation in maximum airfoil section lift-to-drag ratio with Reynolds number (after [12.12]). Numbered curves according to airfoil sections in the inset table

sufficiently large wing chords to avoid the L/D crisis shown in Fig. 12.9. Since Fig. 12.10 indicates that the wing area of profiling gliders is properly sized, larger wing section chords would reduce the wing aspect ratio ($N_R = A_0/\bar{c}^2$).

Increases in wing aspect ratio to reduce specific energy consumption are not only constrained by Reynolds number effects, but by material strength properties as well. As wing aspect ratio increases for a given wing area, the mean wing chord and so the Reynolds number of the wing section decreases, causing degradation of the maximum lift-to-drag ratio of the wing section according to wind tunnel measurements [12.19–21]. If the aspect ratio is increased by increasing the span, the weight of the wing, F_W, will increase as $F_W \approx S^{5/3}$. As the wing weight increases, the thickness-to-chord ratio of the wing section, $\bar{t}/\bar{c}$, must be increased as $\bar{t}/\bar{c} \approx S^{1/6} \approx V_0^{1/9}$ to provide adequate span-wise bending strength and torsional stiffness [12.22]. If $\bar{t}/\bar{c}$ is made excessively large to satisfy strength requirements of a high aspect ratio wing, then the maximum lift-to-drag ratio of the wing section will further degrade [12.19, 20] and [12.21]. For natural fliers, wing dimensions scale with total volume as $S \approx V_0^{1/3}$ and wing area as $A_0 \approx 0.165 V_0^{2/3}$ according to the *square-cubed law* originally proposed by Cayley, and critiqued later in [12.12]. Based on dimensional analysis, the basic square-cubed law specifies

- Wing semispan: $b \sim M^{1/3}$,
- Wing area: $A_0 \sim 0.165 M^{2/3}$, (12.14)

where M again is the loaded mass.

Figure 12.10 suggests that nature has found a compromise between transport efficiency and structural limitations at an aspect ratio of about $N_R = S^2/A_0 = 6.1$. Inspection of Fig. 12.10a indicates the wing semispan of the profiling gliders may be a bit excessive for the loaded mass at which they operate, but that the wing area of the profiling gliders in Fig. 12.10b compares closely with nature, fitting almost exactly the square-cube law formulation in (12.14), refined by Tucker (contained and discussed in [12.12]) from measurements of birds. Hence, the lack of comparable efficiency of the shapes of the profiling gliders appears not to be due to insufficient wing area, but rather because the loaded mass M is too small or the wetted surface area A_t is too big, or both. From analysis of observational measurements of birds in wind tunnels and in natural environments, Tucker has extracted an empirical relation for specific energy consumption of natural fliers [12.12]

$$E_e = 0.109(Mg)^{-0.185} . \quad (12.15)$$

Note that the exponent in this expression does not differ significantly from that in the empirical fit to NTE in (12.6); they only differ in the size of the leading exponent. From this empirical formulation, it appears that profiling gliders are less efficient fliers in horizontal transport (larger E_e when normalizing by loaded mass) than Tucker's equation would predict. Hence, it appears that horizontal transport efficiency can be further improved in profiling UW gliders (they were designed to profile vertically, not optimize horizontal transport efficiency). This improvement in E_e was a primary guiding factor in developing the larger underwater gliders shown in Fig. 12.4.

12.4.3 Shape Factors

Of all the geometric properties of the glider, the wetted surface to wing area ratio $N_A = A_t/A_0$, has the strongest influence on the specific energy consumption, increasing as $N_A^{1/2}$ (12.13). This result suggests that design focus on reducing N_A will achieve the greatest improvements

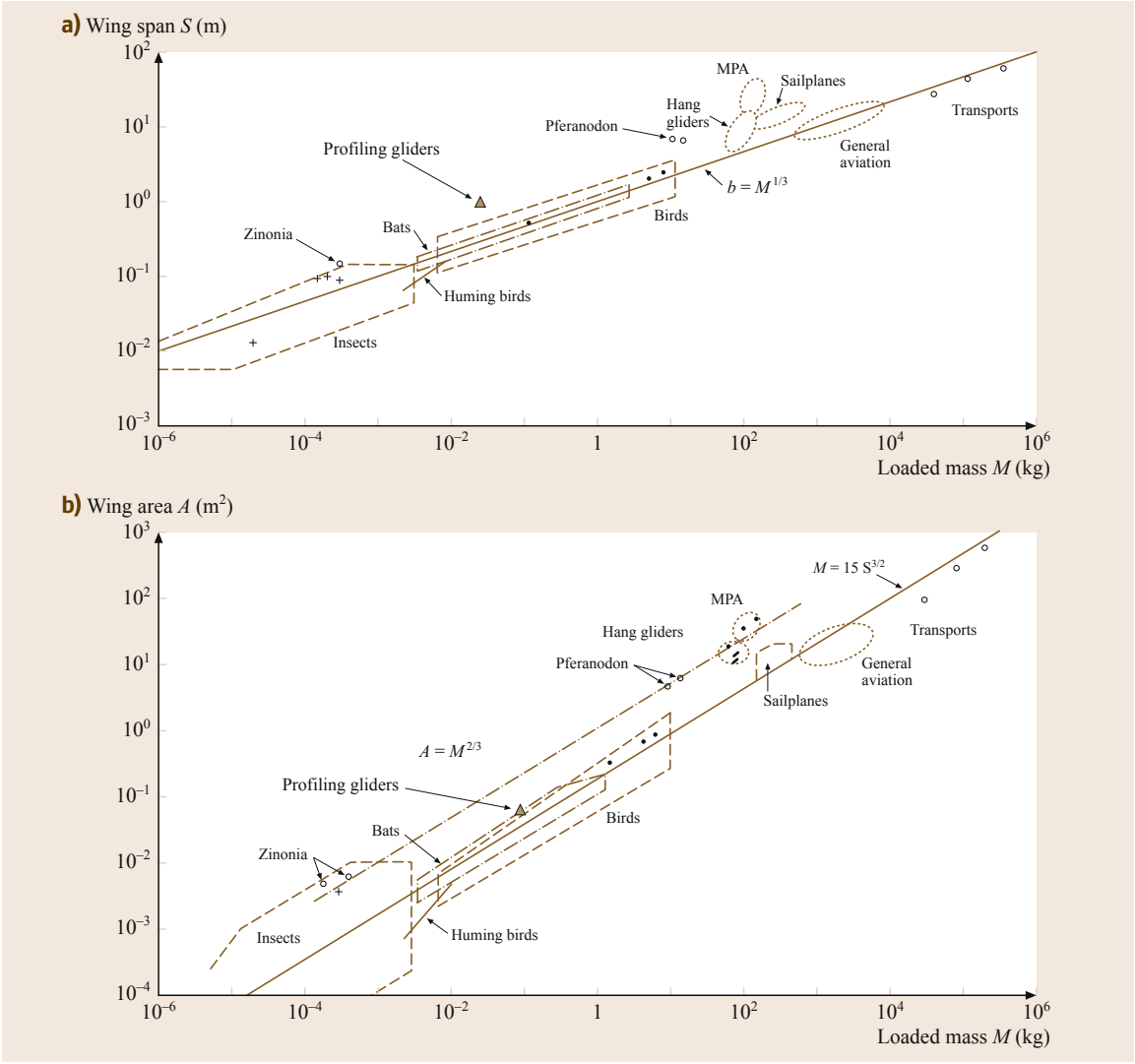


Fig. 12.10a,b Characteristic wing dimensions as a function of loaded mass. **(a)** Scaling of wingspan as a function of loaded mass; **(b)** wing area as a function of loaded mass for natural and man-made fliers (after [12.12, 18])

in horizontal transport economy. As a benchmark in underwater gliders, *Seaglider* has an $N_A = 21.1$ [12.3]. However, the smallest N_A values are associated with *flying wing* and *blended wing/body* geometries, such as utilized by birds, for which typically $N_A \approx 2.2$ to 2.4. The other benefit derived from concentrating the vehicle volume in the wing itself is a large wing area that reduces the magnitudes of C_L and the associated induced drag (the largest component of the drag at minimum E_e). Equation (12.13) indicates that specific energy consumption grows as $C_L^{\xi/4}$. However, increasing wing area indefinitely to achieve a low C_L becomes mutually exclusive with high aspect ratio, N_R. In (12.13), the factor $C_L^{\xi/2}/N_R^{1-\xi/2}$ indicates that a large N_R exerts a greater reduction in E_e than does a proportionally smaller C_L, subject to the structural limits mentioned above.

Another issue with concentrating vehicle volume in the wing is the effect on the profile drag shape factor K_0. Shape efficiency comparisons based on K_0 between a 2-D flying wing and a three-dimensional (3-D) winged body of revolution should be based on the same vehicle volume, V_0. Considering both shapes are streamlined bodies, K_0 scales in proportion to the wetted surface area relative to volume as $A_t/V_0^{2/3}$. Taking an ellipse as the canonical cross section for both 2-D and 3-D shapes, and following the assumption taken in [12.23] and [12.24] that the skin friction per unit area is the same for each shape, the shape factor ratio at con-

stant volume becomes

$$\frac{K_0(\text{2-D})}{K_0(\text{3-D})} = \frac{\pi}{4}\frac{A_t(\text{2-D})}{A_t(\text{3-D})} \approx \pi\sqrt{\frac{2(1-e^2)}{2-e^2}}\,, \quad (12.16)$$

where $e(0 \le e \le 1)$ is the eccentricity of the ellipse taken here as a proxy for the length-to-thickness ratio (*fineness*) of the body. Equation (12.16) is consistent with empirical data on drag of streamlined shapes at high Reynolds numbers (turbulent flow regime) [12.22]. These empirical data relatively show higher profile drag for 2-D shapes at small fineness ratios, as is the case in (12.10) when $e \to 0$. The data for 3-D shapes develop relatively higher profile drag at large fineness ratios, as occurs with (12.16) when $e \to 0.99$. Both the empirical data and (12.16) indicate a 2-D flying wing may suffer as much as a 3-fold increase in K_0 relative to a 3-D body of revolution. However, that increase is more than offset in (12.13) by a 9-fold reduction in N_A when a flying wing is compared to a winged body of revolution. For comparison, Fig. 12.11 shows a computational fluid dynamical (CFD) simulation of the Liberdade/XRay flying wing glider with vehicle volume of $V_0 = 1000\,\text{L}$ versus a Seaglider scaled up to a comparable size of $V_0 = 1000\,\text{L}$. The winged body of revolution in the 1000 L class of glider has about a 30% advantage in maximum cross country speed over the flying wing, but the flying wing is about 43% more efficient in horizontal transport.

Comparative analyses of transport economy in large subsonic transport aircraft concur with this conclusion, generally finding a 20–25% advantage in lift-to-drag ratio for the flying wing over conventional winged bodies of revolution, with additional benefits in gross takeoff weight, operating weight per passenger, and fuel consumption per passenger [12.25]. Other factors not immediately apparent in (12.13) that favor the 2-D flying wing geometries for underwater gliders are: higher Reynolds number on the wing section (due to larger wing chord), leading to higher maximum lift-to-drag ratios (thereby avoiding the 10^4 Reynolds number regime in Fig. 12.9) [12.12]; and increased structural depth of the center section allowing increases in wingspan (and aspect ratio) with fewer weight penalties compared to winged bodies of revolution [12.25, 26].

Together, (12.13) through (12.16) indicate four distinct adjustments that can be made to vehicle characteristics to get energetically more efficient in horizontal transport than the profiling gliders:

1. Increase the buoyancy engine volume to the maximum extent possible for the given internal volume
2. Make the underwater gliders bigger
3. Reduce the total vehicle wetted area A_t relative to the wing area, and

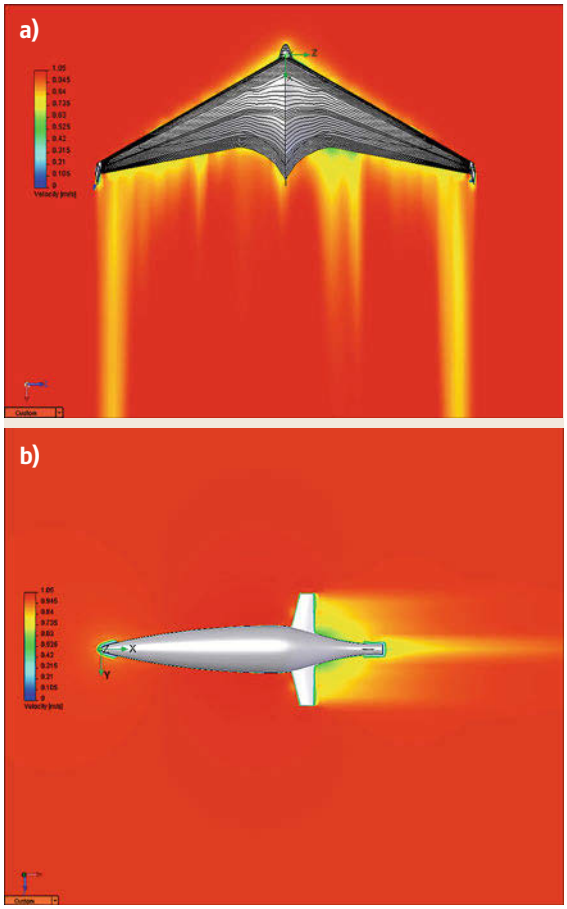


Fig. 12.11a,b Computational fluid dynamical (CFD) simulation of velocity field (using after [12.27]) in the horizontal plane for **(a)** Liberdade/XRay at $(L/D)_{max} = 19$, and **(b)** profiling-type Seaglider scaled up to $V_0 = 1000\,\text{L}$ at $(L/D)_{max} = 11$. The surface roughness is 10 μm in both simulations

4. Increase the wing aspect ratios to the maximum extent possible without reducing wing chord to such a degree that it operates in the mid 10^4 Reynolds number regime.

The wetted area could be reduced by copying birds and designing underwater gliders with flying wing or blended wing body shapes (Fig. 12.12). However, the form factors of such shapes have lower packing efficiency for the glider subsystems that typically fit more readily into cylindrical or spherical shapes.

12.4.4 Glide Polar

Glider flight efficiency is not only just a function of vehicle characteristics, but also a function of how the

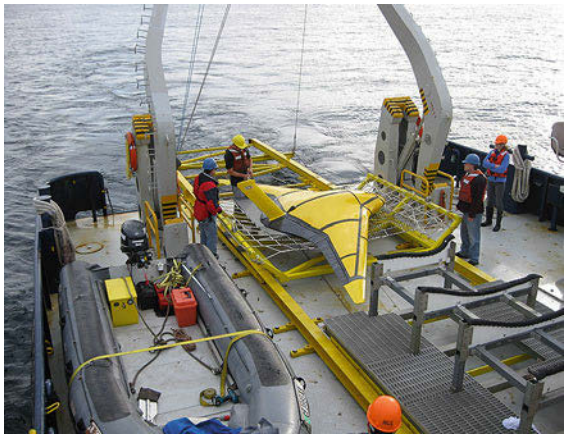


Fig. 12.12 Liberdade/ZRay blended wing body glider aboard R/V Sproul, January, 2011

glider is flown. The expression for specific energy consumption in (12.13) shows that flight energy consumption is minimized by maximizing the lift-to-drag ratio, L/D. From the proportionality between the force and speed triangles in Fig. 12.1, a relation exists between the maximum L/D achievable for a given glider and the angle of the glide path. Since the profiling gliders are flown at glide path angles between 20° and 30° in order to profile the ocean temperature and salinity fields, they will not achieve a specific energy consumption any better than $E_e \approx 0.3$ to 0.5 no matter how optimal their physical characteristics are made. In most ocean environments, the horizontal scales of variability are sufficiently large that vertical profiles of the water column can be collected at significantly shallower glide slopes. So the gliders in the functional class of *depth unlimited roaming* would benefit from improved horizontal flight efficiency, assuming their existing desirable characteristics (e.g., two-person portability) can be retained. However, in other functional classes, maximum flight efficiency in the horizontal is much more important. In particular, *depth limited roaming* requires a glider to travel long distances in a restricted depth interval, *payload/cargo delivery* vehicles must transport large payload and cargo over long distances point-to-point, and *level flight hybrids* must fly at a very flat glide path angle to minimize the expenditure of energy on auxiliary propulsion. All these types of applications for UW require minimum NTE and E_e, and hence maximum L/D. The specific energy consumption for the ZRay flying wing glider demonstrated at sea is 0.05 and the NTE less than 0.65.

While the L/D varies with glide path angle, a corresponding change occurs in the proportions of the speed triangle in Fig. 12.1. This proportional change in the speed triangle with changing glide slope angle yields a continuous relationship between the horizontal and vertical components (w versus u) of the glide velocity, U. This relation, known as the *glide polar*,is readily derivable by balancing the forces in the vertical and horizontal for steady-state flight. Based on the quadratic formulation of lift and drag in (12.9) and (12.10), it is given by

$$w = \pm \left(\frac{\rho g V_b \sqrt{1 + \tan^2 \beta_z}}{(1/2)\rho A_0 (1 + H^2(\alpha)) \sqrt{C_L^2(\alpha) + C_D^2(\alpha)}} \right)^{1/2} , \tag{12.17}$$

where the polynomial H is a function of angle of attack α

$$H(\alpha) \equiv \frac{L}{D} = \frac{C_L(\alpha)}{C_{D0}(\alpha) + C_{Di}(\alpha)} , \tag{12.18}$$

and β_z is the bank angle from [12.28] given as,

$$\beta_z = \arcsin[\cos\varphi \sin\beta \cos\eta + (\cos\alpha \sin\varphi - \sin\alpha \cos\varphi \cos\beta) \sin\eta] . \tag{12.19}$$

In (12.19), φ is the pitch angle, β is the roll angle, and η is the yaw angle. In coordinated flight, $\eta = 0$, the bank angle approaches the angle of roll at small pitch angles, i. e., $\beta_z \to \beta$ as $\phi \to 0$. Positive values of w in (12.17) correspond to ascending glides and $w < 0$ corresponds to descending glides. The corresponding expression for u is the same as (12.17) except that (a) only positive values of the horizontal speed are considered (the glider does not fly backward) and (b) $H^2(\alpha)$ changes to $H^{-2}(\alpha)$. The glide polars of the XRay flying wing glider (Fig. 12.11a) are plotted as solid lines in Fig. 12.13 for noncircling, wings-level flight at four different wing loadings within a potential range of net buoyancy volume, $n_b = V_b/V_0 = 3.8$ to 27.6%. The four polars in Fig. 12.13 (red, black, green, and blue curves) are representative of various buoyancy engine technologies, where the red curve ($V_b = 38.36\,L$, $n_b = 3.8\%$) is representative of the upper end of the closed-loop liquid-based engine technology [12.2, 3], the black curve ($V_b = 50.0\,L$, $n_b = 5.0\%$) represents the upper range open-loop liquid-based engine technology [12.1, 4, 29], the green curve ($V_b = 122\,L$, $n_b = 12.2\%$) is a proxy for the open-loop compressed gas-based systems [12.8], and the blue curve ($V_b = 275\,L$, $n_b = 27.6\%$) approximates the upper end of the open-loop gas-based buoyancy engines that consume gas generating compounds [12.5]. The magnitude of the glide velocity, U, and the velocity components (u, v, w) increase as the square root of the buoyancy engine volume increase and the associated wing loading (ratio

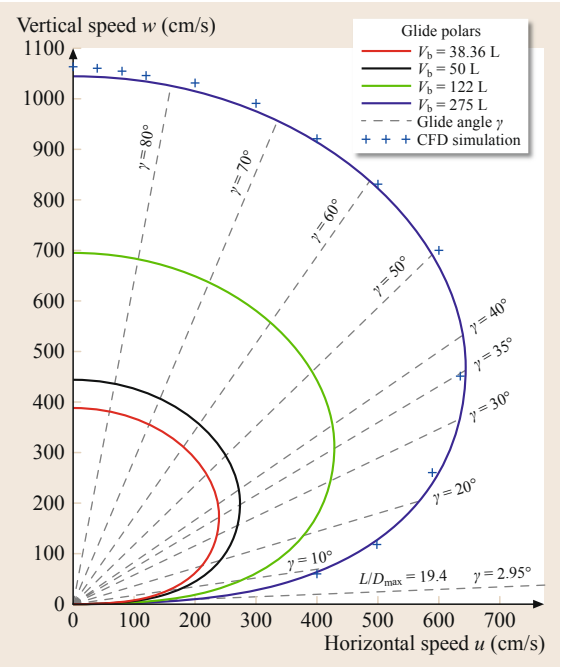


Fig. 12.13 Liberdade/XRay glide polars in the z-plane for wings level ($\beta = 0°$) attitude over a range of net buoyancy between 38.36 and 275 L. Glide angles γ shown as *dashed gray lines* are invariant with net buoyancy change. *Blue crosses* are derived from CFD simulations at $V_b = 275$ L using (after [12.27]). See also Fig. 12.11a

of net buoyancy to wing area) [12.28, 30]. The maximum lift-to-drag ratio over ground (minimum specific energy consumption) is given by the point of tangency on the glide polar to a straight line drawn from the origin [12.28, 30]. The dashed gray tangent line in Fig. 12.13 indicates XRay has a predicted $(L/D)_{max} = 19.4$, which corresponds to a best glide angle $\gamma = 2.95°$ and a specific energy consumption of $E_e = 0.051$. With 10 μm surface roughness, the CFD simulations like those in Fig. 12.11a gave approximately the same results, $(L/D)_{max} = 20.0$. At-sea results for ZRay without the use of trailing edge flaps to change camber confirm these maximum lift-to-drag ratio results.

For comparison, the published polar of the Seaglider [12.3] gives a $(L/D)_{max} = 7.0$, or $E_e = 0.143$. If the Seaglider with net buoyancy of $V_b = 0.33$ L was scaled up from its present volume of $V_0 = 66$ L to an equivalent volume of XRay of $V_0 = 1000$ L and given a net buoyancy $V_b = 38.36$ L, (Fig. 12.11b), then by (12.17), the maximum lift-to-drag ratio would increase by a factor of 1.81. This increase is based on the assumption of all laminar boundary layers ($\xi = 1/2$) in (12.17) and would give the scaled-up Seaglider a $(L/D)_{max} = 12.7$. For all turbulent boundary layers ($\xi = 1/5$), (12.11) gives a $(L/D)_{max} = 8.9$ for the scaled-up Seaglider. A CFD simulation (using [12.27]) for the scaled-up Seaglider (Fig. 12.11b) finds that about 60.5% of the boundary layer is laminar for a surface roughness of 10 μm, giving a $(L/D)_{max} = 11.1$.

When the XRay glider is flown at steeper glide angles (shown by the other dashed gray radial lines in Fig. 12.13 above the tangent line of the best glide angle $\gamma = 2.95°$), the cross-country speed, u, increases, reaching a maximum at a glide angle $\gamma = 35°$. This maximum arises because C_D is nearly invariant with C_L, when C_L becomes small at the small angles of attack during steep glides. In this case, an equation for u analogous to (12.17) reduces to u being dependent on $\sin\gamma \cos^2\gamma$, where $\gamma = \arctan(H^{-1})$. This expression has a maximum at $\gamma = 35°$ [12.1]. Within its range of possible buoyancy variation, the maximum country speeds of XRay can theoretically vary over a wide range, from $u_{max} = 241$ cm/s (4.7 kts) at a net buoyancy of $V_b = 38.36$ L (red curve), reaching $u_{max} = 644$ cm/s (12.5 kts) at a vehicle net buoyancy of $V_b = 275$ L (blue curve). XRay's glide speeds along the $\gamma = 35°$ glide slope, referred to as $\hat{U}$, are found in Fig. 12.13 to be $\hat{U} = 296$ cm/s (5.8 kts) at a net buoyancy $V_b = 38.36$ L, reaching $\hat{U} = 793$ cm/s (15.4 kts) at the maximum vehicle net buoyancy of $V_b = 275$ L. These numerical results, confirmed by at-sea tests, indicate that the XRay flying wing glider fitted with moderately sized buoyancy engines (the as-built XRay glider has $V_b = 30$ L) is capable of traveling point-to-point at horizontal speeds comparable to those of commercially available, mid-size prop-driven autonomous underwater vehicles (AUVs).

The glide polars for Seaglider [12.3] give $u_{max} = 46$ cm/s (0.89 kts) at a net buoyancy $V_b = 0.33$ L, ($n_b = 0.5\%$). For the scaled-up Seaglider at $V_0 = 1000$ L and $V_b = 38.36$ L, the wing loading is increased by a factor of 47, thereby increasing cross-country speed to $u_{max} = 315$ cm/s (6.1 kts) based on application of the wing loading relationship expressed in (12.17) to the published polar. As stated earlier, the winged body of revolution in the 1000 L class of glider has about a 30% advantage in maximum cross-country speed over the flying wing, although the flying wing is about 43% more efficient in horizontal transport.

Maximum along-coarse speed in still water is always obtained at a 35° glide angle regardless of vehicle shape or other hydrodynamic properties [12.1]. Figure 12.14 compares the potential cross-country speeds as a function of loaded mass (mass equivalent net buoyancy) of the four classes of underwater gliders when optimally sized for their intended missions. These include the winged body of revolution carrying single (small) payloads (*red*), the winged body of revolution carrying variable (large) payloads (*blue*), the flying

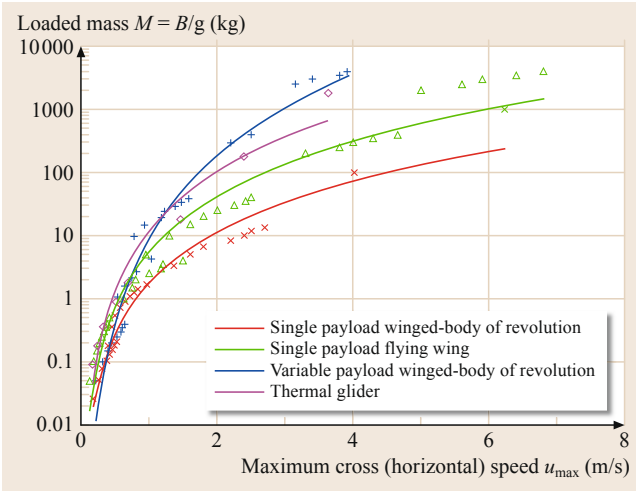


Fig. 12.14 Maximum cross-country (horizontal) speed for an underwater glider due to variation in the loaded mass for a winged body of revolution carrying single (small) payloads (*red*), a winged body of revolution carrying variable (large) payloads (*blue*), the flying wing carrying single (small) payloads (*green*), and a thermal glider in a winged-body-of-revolution configuration with single payload (*purple*)

wing carrying single (small) payloads (*green*), and the thermal glider in a winged-body-of-revolution configuration with single payload (*purple*). Winged bodies of revolution with maximum buoyancy engine capacity are the optimal combination for maximum speed (Fig. 12.15). For a given n_b, flying wings of equivalent vehicle volume are slower than winged bodies of revolution, but have superior transport economy, requiring fewer dive cycles (and less near surface exposure time) for a given distance traveled.

To exploit the superior horizontal transport economy of the flying wing for large payloads and cargo, vehicle packing must deal with the planar form factor of the flying wing and distribute the vehicle weight such that the critical vehicle balance factors that provide stability are met both before and after payload/cargo delivery. Those critical balance factors are (1) adequate vertical separation between the centers of buoyancy and mass, and (2) adequate horizontal separation between the centers of pressure and mass. A large payload/cargo concept for the flying wing glider is based on the use of multiple wing systems, and is referred to as the *Coanda triplane*, Fig. 12.16. The concept is based on adding two auxiliary wings to the basic ZRay flying wing glider hull form. To assure maximum vertical separation between the centers of buoyancy and mass, the lower wing carries the negatively buoyant payload/cargo while the upper wing carries buoyancy-compensating foam. This arrangement allows the glider to deliver the payload/cargo

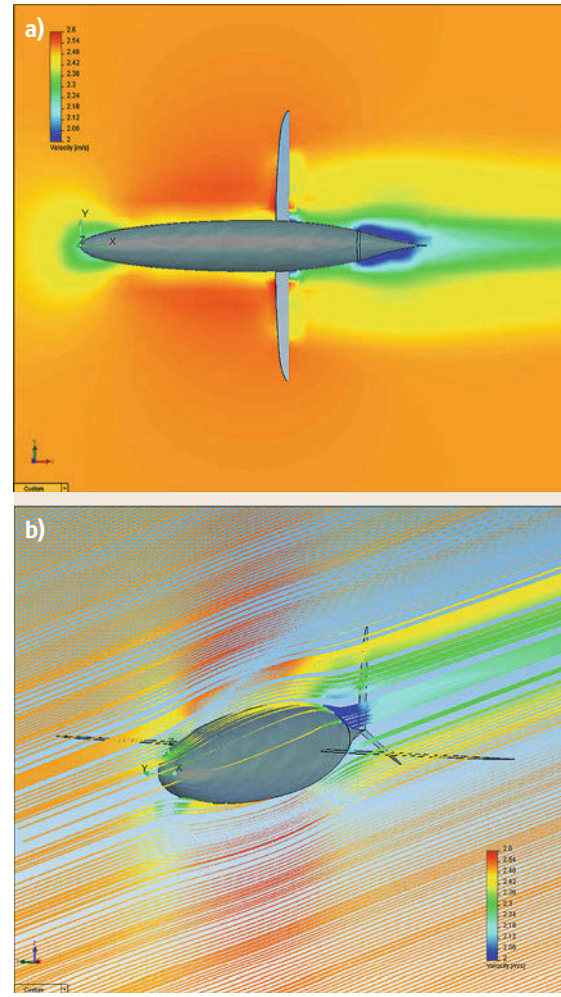


Fig. 12.15a,b CFD simulations of *The Bus* concept, a large winged body of revolution for carrying large payloads, used for creating the blue curve in Fig. 12.14. **(a)** Velocity contour plot in the horizontal plane; **(b)** streamlines in the vertical plane of *The Bus*

to the seafloor without having to perform unusual flight behaviors; upon arrival at the deployment site, the upper and lower wings are released, transforming the vehicle into a conventional ZRay glider.

The upper and lower external wings have high thickness-to-chord cambered airfoil sections. Computational fluid dynamical simulations (Fig. 12.17) show that for a particular-sized separation between these cambered auxiliary wings and the symmetric ZRay main-body wing section, remarkably high lift coefficients (large aggregate flow circulation) can be obtained at low angles of attack during both descent and ascent. This phenomenon is attributable to the *Coanda effect*, whereby the high velocity flow through the two gaps

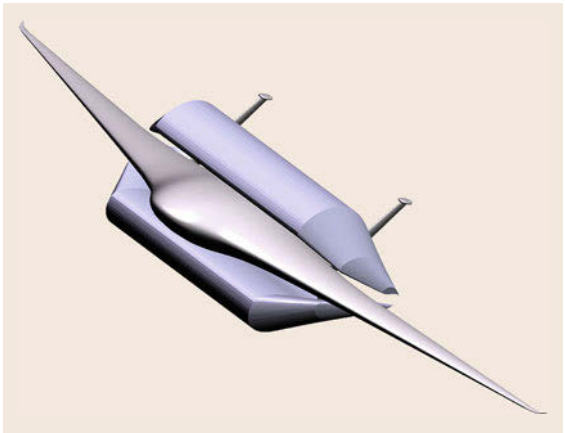


Fig. 12.16 The *Coanda triplane* is a large flying wing glider based on the ZRay glider (Figs. 12.4b and 12.12) for carrying large payloads and cargo to be delivered on the seafloor. The lower wing carries the negatively buoyant payload/cargo while the upper wing carries the foam for buoyancy compensation ◄

between the three wing sections remains attached in the diverging flow over the aft surfaces. This phenomenon results in a high degree of flow circulation in a closed-ended wake behind the aggregate three-wing system. The circulation is clearly visible by the vertical asymmetry of the velocity contour plots in Fig. 12.17 and by the progressive vector arrows. The lift coefficient that results from this circulation is on the order of $C_L = 2.8$, and translates into a near doubling of the L/D over that of ZRay alone. The glide polar for the Coanda triplane concept, computed from CFD simulations, is plotted as the dashed black curve in Fig. 12.18. A maximum L/D of 37.7 is achieved, equivalent to a ZRay glider using 10° of camber-changing, trailing-edge flaps. However, the triplane is about 10 cm/s to 20 cm/s slower than ZRay, and achieves its best L/D at a glide speed of only

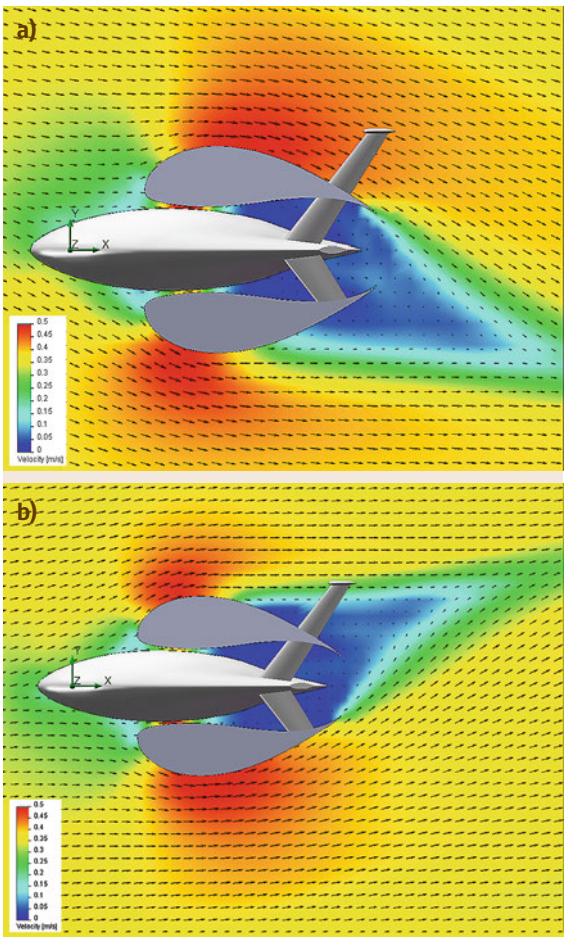


Fig. 12.17a,b A CFD study of the wing circulation for the *Coanda triplane*. **(a)** Shows the velocity contours during descending dives at negative angles of attack while **(b)** shows the velocity contours during ascending dives at positive angles of attack

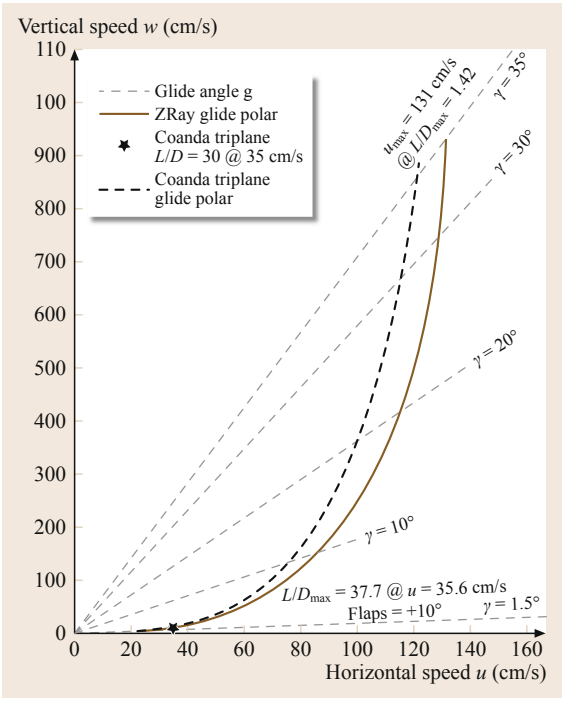


Fig. 12.18 Low-speed portion of the glide polar of the Coanda triplane concept (*dashed black curve*) and the ZRay glide polar (*red curve*). The glide polars are for wings level ($\beta = 0°$) attitude for a net buoyancy of 13.15 L. Glide angles γ are shown as *dashed gray lines*

35.6 cm/s. The low speed characteristics at maximum L/D are largely a consequence of the triplane's relatively high wetted surface area.

In summary, the ZRay glider becomes *more* energy efficient in horizontal transport of large cargo and payload than when flying by itself, although it does so at slower speeds. More generally, vehicles with wings can reduce their propulsion energy consumption by flying cooperatively with other wings, in contrast to prop-driven vehicles. Cooperative flight is enabled by the simplicity of the wing's trailing edge vortex system – Sect. 12.2.

12.5 Thermal Glider

This vehicle is applicable for ultra-long range, depth-unlimited roaming and for high endurance station keeping [12.4] (Table 12.1). The concept gives the glider the ability of renewing its onboard energy stores by harvesting environmental energy from the heat reservoir of the ocean, specifically from the temperature differences of the cold deep water and the warmer surface water (available in 80% of the world's oceans). Ranges of 30 000 to 40 000 km, circumnavigating the world, then become conceivable. Harvesting this thermal energy depends on the volume change associated with the state (phase) change of a material with a melting/freezing point in the range of ocean temperatures. Heat is absorbed from the warm surface water, causing a change of state from solid to fluid (melting), and released to the cooler, deeper water during the vehicle's transit through the thermocline, resulting in a state change back to the solid state (freezing). The heat exchange volume occurs inside tubes that run the vehicle's length and provide a large surface area for rapid heat flow. Almost all materials have a positive thermal expansion coefficient so that melting causes an increase in volume and freezing results in a decrease in volume (water below 4 °C is an exception). Because these volume changes are opposite to those required of a glider's buoyancy engine, they cannot be used directly for forward propulsion. Rather, the thermal buoyancy engine must include unique design features to account for the positive thermal expansion coefficient (as in Fig. 12.19) or the energy from expansion/contraction must be stored onboard over a half dive cycle to be useable for propulsion. The four stages of the thermodynamic cycle are shown Fig. 12.19. Environmental energy is harvested by heat flowing into and out of the working fluid in chamber 1, which contracts on freezing and expands on melting. The resulting work is transmitted around the system by the transfer fluid, typically mineral oil. Chamber 2 is an energy storage accumulator, with the transfer fluid pressurized by nitrogen at a pressure greater than the maximum external ocean pressure. In Fig. 12.19a, the vehicle is in stable thermal equilibrium in the warm surface water, N_2 is compressed, the external bladder is inflated, and working fluid is expanded. Descent begins by opening the three-way valve (Fig. 12.19b), venting the external bladder to the internal bladder. Maintaining the hull interior slightly below atmospheric pressure creates the pressure differential for this flow. As the vehicle reaches cold water, heat flows out of the working fluid, which freezes and contracts, and draws in mineral oil from the internal reservoir. The beginning of ascent (Fig. 12.19c) results from opening the three-way valve, the pressurized oil in the accumulator moves to the external bladder and the vehicle changes from negative buoyancy to positive buoyancy. During ascent (Fig. 12.19d), the vehicle ascends to warm waters, heat flows into the working fluid, which melts and expands, and oil flows to recharge the accumulator.

Because the thermal glider is nearly identical in shape and dimensions to the winged body of revolution for single payloads, the glide polar data from Fig. 12.14 can be used for hydrodynamic input to an analysis of

Table 12.1 Analysis results for the net transport economy of various sizes of a thermal glider designed for the ultra-long range, depth-unlimited roaming application (after [12.1])

Scale [L]	Vol. [L]	Net buoyancy [L]	u	w [m/s]	Range [km]	NTE	Re vol$^{(1/3)}$	Cd vol$^{(2/3)}$
0.5	62	0.09	0.17	0.11	35 000	0.61	3.2×10^{4}	0.062
1	64	0.18	0.24	0.16	35 000	0.22	4.6×10^{4}	0.062
2	67	0.36	0.33	0.22	35 000	0.08	6.6×10^{4}	0.062
5	76	0.9	0.51	0.33	35 000	0.02	1.1×10^{5}	0.06
10	91	1.8	0.69	0.45	35 000	7.5×10^{-3}	1.7×10^{5}	0.059
100	358	18	1.46	0.95	35 000	3.6×10^{-4}	8.8×10^{5}	0.049
1000	3036	180	2.39	1.55	35 000	2.2×10^{-5}	5.9×10^{6}	0.036
10 000	29 809	1.80×10^{3}	3.63	2.36	35 000	1.4×10^{-6}	4.2×10^{7}	0.032

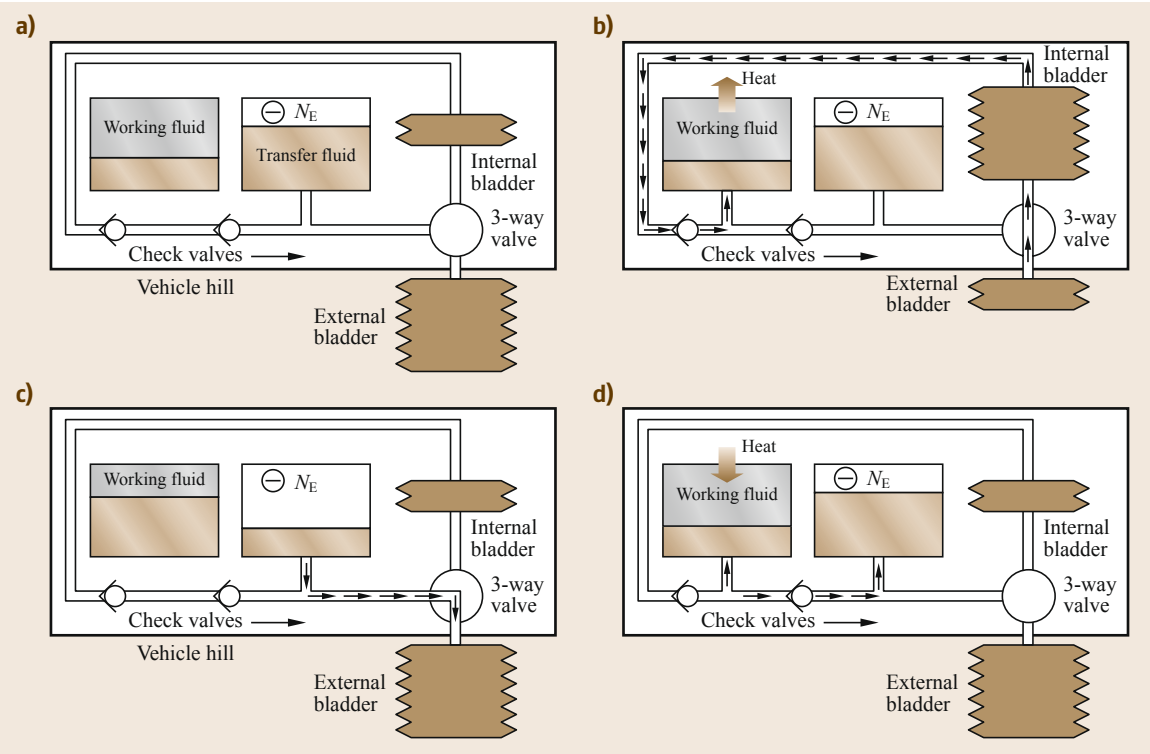


Fig. 12.19a–d Thermodynamic cycle of the thermal glider heat pump: **(a)** at surface in thermal equilibrium, **(b)** descending, evacuating external bladder and pumping into hull, pressurizing hydraulic accumulator **(c)**, at depth, working fluid frozen and beginning ascent by releasing pressure from hydraulic accumulator **(d)** fully developed ascent using thermal expansion of melting working fluid (after [12.4])

energetics and transport economy. To make the transport economy analysis problem tractable, the following assumptions are made regarding a long range, depth unlimited roaming type of application:

- Range = 35 000 km
- Profile depth = 1300 m
- Dive angle = 33°.

The hull weight scales with volume. For thermal engines, the oil required and compensator scale with the drive force required. The drive force (provided by the net buoyancy, B) is a function of the glide angle and is the on-axis component of the buoyancy desired. The payload weight is fixed in the analysis. Energy consumption is based solely on hotel load, i. e., all propulsion energy is assumed to be harvested from the ocean temperature gradients. As a result, as the mission time decreases with increasing velocity, the onboard energy needed decreases. The volume is calculated for the scaled buoyancy, scaled thermal engines, and scaled oil in bladders and compensator. The velocity is then calculated based on the estimated volume and the C_D. Energy is recalculated based on the new velocity and the volume is readjusted. The results of taking these steps are in [12.1].

The NTE figures above for some of the larger thermal glider sizes demonstrate that the thermal glider is capable of transport economies unmatched by any existing man-made flier.

12.6 Discussion and Conclusions

Gliders occupy a unique niche in the universe of autonomous underwater vehicles. They have an intrinsic advantage in transport efficiency over conventional prop-driven AUVs due to the simpler vortex dynamics of a wing compared to a propeller. (Propulsion systems for long-distance persistence found in nature, e.g., birds and marine mammals, are based on wings, not propellers.) Due to a wing's simple vortex dy-

namics, gliders with wings can fly cooperatively with other winged vehicles and structures to further improve transport efficiency, in contrast to prop-driven vehicles. In addition, gliders are capable of readily operating throughout their full speed regime, from top sprint speed to creeping flight to free-drifting at neutral buoyancy in which no propulsion energy is consumed. Consequently, gliders offer a persistent, high-endurance solution for many ocean sampling and surveillance missions. Buoyancy-driven gliders also are silent (acoustically and electromagnetically) throughout most of the dive cycle, including when operating at top sprint speed. They create low levels of self-noise for a few percent of the total dive time due primarily to buoyancy engine operation. However, gliders are generally slower than prop-driven AUVs, typically operating in the speed regime below 3 kts. (Much of a glider's on-station persistence is attributable to these low operating speeds.) In addition, a glider must change depth in order to move forward and so is incapable of level flight. On the other hand, these depth changes allow vertical profiles of ocean properties to be collected and for energy to be extracted from the temperature gradients in the ocean.

In addition to being slower and an inability to conduct level flight, gliders are less maneuverable and more balance sensitive than prop-driven AUVs. However, possibly the greatest disadvantage is the additional vehicle interior volume consumed by, and the additional complexity of, a buoyancy engine compared to a prop-driven system.

The calculus of gliding has been explored for ways of increasing the horizontal point-to-point transport efficiency and speed of underwater gliders. An analytic solution for the minimum specific energy consumption (maximum L/D) indicates that increasing the net buoyancy volume of the vehicle V_b increases the speed as $\sqrt{V_b}$ while specific energy consumption E_e declines as $E_e \approx V_b^{-\xi/4}$, where $1/5 \leq \xi \leq 1/2$. Since buoyancy engine capacity is some fraction n_b of vehicle volume, V_0, this finding immediately argues for larger gliders for improved horizontal transport efficiency. The challenge therefore reduces to finding the most efficient geometries for large volumes. Of all the geometric properties of the glider, the wetted surface to wing area ratio, $N_A = A_t/A_0$, has the strongest influence on specific energy consumption, with E_e decreasing as $E_e \approx N_A^{1/2}$ with decreasing N_A. The flying wing glider offers a 9-fold reduction in the wetted surface to wing area ratio over existing profiling gliders, which offsets the 3-fold (at most) increase in profile drag resulting from the thick wing section of a flying wing. The other benefit derived from concentrating the vehicle volume in the wing is that a large wing area reduces the magnitudes of lift coefficient C_L required to support a given V_b, thereby reducing the associated induced drag, the largest component of drag at maximum L/D. However, increasing wing area indefinitely becomes mutually exclusive with high aspect ratio, $N_R = S^2/A_0$, where S is the wingspan. Specific energy consumption was shown to decline with decreases in the ratio of induced drag factors at a rate given by $E_e \approx C_L^{\xi/4}/N_R^{(1/2-\xi/4)}$.

Other factors that favor flying wing geometries for many classes of underwater gliders are higher Reynolds number on the wing section due to larger wing chord, leading to higher maximum lift-to-drag ratios and avoidance of rapid L/D degradation occurring in the 10^4 Reynolds number regime, and increased structural depth of the center section allowing increases in wingspan (and aspect ratio) with fewer weight penalties compared to winged bodies of revolution [12.31–33]. Analytic and numerical comparisons, supported by at-sea results, for large gliders ($V_0 \approx 1000$ L) of comparable net buoyancy demonstrate that a flying wing glider is about 43% more efficient in horizontal transport, but that a winged body of revolution glider has about a 30% advantage in maximum cross-country speed. All gliders designed for long-duration, long-distance flights certainly would benefit significantly from using a buoyancy engine that can harvest energy from the temperature gradients in the ocean.

References

12.1 S. A. Jenkins, D. E. Humphreys, J. Sherman, J. Osse, C. Jones, N. Leonard, J. Graver, R. Bachmayer: *Underwater Glider System Study, Scripps Institution of Oceanography*, Tech. Rep. 57 (University of California, San Diego, La Jolla 2003) online available at http://repositories.cdlib.org/sio/techreport/53/

12.2 J. Sherman, R.E. Davis, W.B. Owens, J. Valdes: The autonomous underwater glider spray, IEEE J. Ocean. Eng. **26**(4), 437–446 (2001)

12.3 C.C. Eriksen, T.J. Osse, R.D. Light, T. Wen, T.W. Lehman, P.L. Sabin, J.W. Ballard, A.M. Chiodi: Seaglider: A long-range autonomous underwater vehicle for oceanographic research, IEEE J. Ocean. Eng. **26**(4), 424–436 (2001)

12.4 D.C. Webb, P.J. Simonetti, C.P. Jones: SLOCUM: An underwater glider propelled by environmental energy, IEEE J. Ocean. Eng. **26**(4), 447–452 (2001)

12.5 R. H. Oversmith, R. E. Leadon: *Concept Whisper: A buoyancy-propelled, multiple cycle under-*

sea glide vehicle, Tech. Rep. GD/C-62-206A (General Dynamics/Convair, San Diego 1962)
12.6 H. Stommel: The SLOCUM mission, Oceanography **2**(1), 22–25 (1989)
12.7 H. Tennekes: *The Simple Science of Flight* (MIT Press, Cambridge 1997)
12.8 C. P. Rains: *Analytical Investigation of the Performance Characteristics of Cyclic Glide Undersea Vehicles*, Tech. Rep. GD/C419-69-006 (General Dynamics Corp., San Diego 1968)
12.9 M. Van Dyke: *An Album of Fluid Motion* (Parabolic, Stanford 1982)
12.10 T. Bjornstad, NOAA Fisheries, Protected Resource Division, Southwest Fisheries Science Center, https://commons.wikimedia.org/wiki/File:Blue_Whale_001_body_bw.jpg, (2007)
12.11 H. Glauert: *Elements of Aerofoil and Airscrew Theory*, 2nd edn. (Cambridge Univ. Press, Cambridge 1948)
12.12 J.H. McMasters: An analytic survey of low speed flying devices: Natural and man-made, Tech. Soar. **3**(4), 17–42 (1974)
12.13 C.J. Pennycuick: Gliding flight of the Fulmar Petrel, J. Exp. Biol. **37**, 330–338 (1960)
12.14 J.H. McMasters, R.M. Cummings: Airplane design and biomechanics of flight – A more completely multi-disciplinary perspective, AIAA 42nd Aerosp. Sci. Meet. (2004), AIAA2004-532
12.15 C.H. Greenewalt: Dimensional relationships for flying animals, Smithson. Misc. Collect. **144**, 1–46 (1962)
12.16 B.W. McCormick: *Aerodynamics, Aeronautics and Flight Mechanics* (Wiley, New York 1979)
12.17 L. Prandtl, O.G. Tietjens: *Applied Hydro- and Aeromechanics* (McGraw-Hill, New York 1934)
12.18 S.A. Jenkins, J. Wasyl: Optimization of glides for constant wind fields and course headings, J. Aircr. **27**(7), 632–638 (1990)
12.19 R. Eppler: Laminarprofile fur Segelflugzeuge, Z. Flugwiss. **3**, 346–353 (1955)
12.20 D. Althaus: *Stuttgarter Profilkatalog I* (Vieweg, Braunschweig 1981)
12.21 S.R. Hoerner: *Fluid Dynamic Drag* (Hoerner Fluid Dynamics, Bakersfield 1965)
12.22 F.A. Cleveland: Size effects in conventional aircraft design, J. Aircr. **7**(6), 483–511 (1970)
12.23 N.K. Delany, N.E. Sorensen: *Low-Speed Drag of Cylinders of Various Shapes*, Tech. Note, Vol. 3038 (NACA, Washington 1953)
12.24 H. Schlichting: *Boundary Layer Theory* (McGraw-Hill, New York 1960)
12.25 A.L. Bolsunovsky, N.P. Buzoverya, B.I. Gurevich, V.E. Denisov, A.I. Dunaevsky, L.M. Shkadov, O.V. Sonin, A.J. Udzhuhu, J.P. Zhurihin: Flying wing – Problems and decisions, Aircr. Des. **4**, 193–219 (2001)
12.26 V.E. Denisov, A.L. Bolsunovsky, N.P. Buzoverya, B.I. Gurevich: Recent investigations of the very large passenger blended-wing-body aircraft, Proc. Int. Conf. Acoust. Speech, Vol. 98 (1998), paper 98–4.10.3
12.27 COSMOS (FlowWorks, Los Angeles 2001)
12.28 R.F. Stengel: *Flight Dynamics* (Princeton Univ. Press, Princeton 2004)
12.29 J.G. Graver, R. Bachmayer, N.E. Leonard: *ONR Underwater Glider Systems Study Glider Design Notes, Part 1*, Tech. Rep. (Princeton Univ., Princeton 2003)
12.30 H. Reichmann, P. Lert: *Cross-Country Soaring* (Thomson, Santa Monica 1978)
12.31 R. Horten: Toward the theory of flying wings, http://www.nurflugel.com/Nurflugel/Horten_Nurflugels/horten_nurflugels.html (2008) translated by Y. Leshinski
12.32 J. K. Northrop: The development of all-wing aircraft, Roy. Aeronaut. Soc. **51**, 481–510 (1947) 35th Wilbur Wright Memorial Lecture
12.33 J.H. McMasters, I.M. Kroo: Advanced configurations for very large transport airplanes, Aircr. Des. **1**(4), 217–242 (1998)

13. Autonomous Sea Surface Vehicles

Stefano Brizzolara, Robert A. Brizzolara

Capitalizing on the experience and technology developments gained with underwater autonomous vehicles, the current research frontier in the field of autonomous marine vehicles has moved from under water to the sea surface, i. e., autonomous sea surface vehicles. Current and future perspectives of these types of autonomous vehicles are given in Sect. 13.1, with particular attention paid to US navy current interests. These were initiated a decade ago and still actively drive developments in this sector. The ability to design craft specialized for particular tasks to reach the best performance at sea, going beyond size and operational/safety limitations currently imposed by manned ships, offers unique opportunities to the naval architect. In this respect, the basic naval architecture principles that drive the selection of a type of hull with respect to operational requirements are given in the Sect. 13.3. The selection of the type of hull is a preliminary essential activity for the successful design or acquisition of an autonomous surface craft. Finally, as a practical example, Sect. 13.4 summarizes the main results of extensive research done at the MIT-iShip lab to develop a new class of autonomous unmanned surface vehicles, based on a highly specialized and optimized design of an unconventional SWATH (small waterplane area twin hull) hull, able to achieve superior performance and operational capabilities in real sea state conditions.

13.1 **Platforms** 324
13.2 **Autonomous Maneuvering and Navigation** 326
13.3 **Naval Architecture of AUSV Design** 327
13.3.1 Froude Number and Hull Typologies 327
13.4 **Optimized Class of Autonomous Unmanned Surface Vehicles** 330
13.5 **Conclusions** 337
References 339

Unmanned surface vehicles (USVs) provide significant performance advantages relative to their counterparts in air and undersea domains. USVs have the ability to carry large, heavy payloads (or fuel) and maintain them on-station for long periods of time, for a platform of a given size and cost. Displacement and planing hulls typically used for USVs can be optimized for high payload fraction. Air-breathing propulsion means that the platform can be designed to have a long range and endurance or high speed. USVs can be designed to be quite stealthy or they can be designed for high speed. By virtue of being located on the sea surface, USVs are able to communicate with air and undersea platforms. Finally, the USV platform can and should be of low complexity and low cost.

USV development is still in its early stages, despite the fact that the use of unmanned boats in military operations is far from new. Some of the first uses of unmanned boats took place in 1946 for the collection of water samples following *operation crossroads* nuclear weapon tests [13.1] and subsequently in Vietnam for mine sweeping operations. Despite these early activities, concerted USV development did not begin until the late 1990s. This is typical of the development timelines for many complex technologies, which can be several decades long. For example, in the case of naval air power, it was almost four decades between the Wright brothers' first flight in 1903 and the first decisive uses of naval air power in World War II, with key milestones in between such as the first airplane takeoff from a warship in 1910 and the first aircraft carrier (early 1920s) [13.2, 3]. The first decisive military uses of radar occurred in World War II, for example in the Battle of Britain in 1940, following the seminal work on electromagnetic waves by Maxwell, Hertz, and others that occurred in the late 1800s [13.4]. Although very different to naval aviation and radar, the full capability of USVs will be enabled by solutions to technical issues that have a substantial degree of difficulty (for example, sensors, guidance and navigation, artificial intelligence, computer processing, and reduced size and weight of payloads of operational interest); these issues will still require time to be solved.

USVs have greatly advanced in capability over the past 10 years and can be of substantial use in missions today. As these developments continue, USVs are expected to soon have capabilities formerly seen only in much larger, manned vessels and aircraft. USVs are now being sold commercially, under development, used operationally, and are considered for operational use by a number of nations. The utility of USVs can be thought of as dependent on the complexity of mission they are called on to do, the complexity of environment, and the degree of involvement of human operators. Today, USVs are on the lower end of the mission and environmental complexity scales; therefore, a substantial degree of human operator involvement is required. In the future, as technology matures, operations with greater mission and environmental complexity will become possible and less human involvement will be required. For example, currently USVs generally operate in fair weather and low sea states, and in areas of little or no other maritime traffic. Capability is degraded in higher sea states, in part due to limitations of sensors and algorithms. In the future, USVs will be called on to operate in degraded weather conditions and higher sea states. Missions that require patrol and running tracks are fairly straightforward and can be undertaken today. Missions that require the USV to adjust on-the-fly to changing situations are still in the future. Today, the human operator is commonly responsible for hazard avoidance, controlling the USV remotely based on video streamed from the USV. As the autonomy matures, more functionality will be assumed by the autonomous control system, freeing the USV from the tether of a communications system and reducing the workload of the human operator.

USVs are a system composed of the platform, the navigation/maneuvering control system, and a payload. Additional system components may include, for example, subsystems for launch, recovery, and refueling. The following sections will address USV platforms and control systems; payloads are outside the scope of this chapter.

13.1 Platforms

The US Navy USV Master Plan has classified USVs into four size ranges: Fleet Class (approximately 11 m in length), Harbor Class (approximately 7 m in length), X-Class (small USVs deployable from an 11 m RHIB (rigid hull inflatable boat)) and Snorkler (semi-submersible platforms) [13.5]. All of these size classes are designed to operate from, and in close proximity to, conventional ships, or close to shore-based pier facilities. The USV mission is typically in support of conventional ships, and their size is limited by host ship launch and recovery constraints. In its anti-submarine warfare unmanned surface vehicle (ACTUV) (Fig. 13.1) program, the Defense Advanced Research Projects Agency (DARPA) is investigating the technical viability of a much larger unmanned naval vessel that will operate independently [13.6].

USVs platform design is constrained by the same *iron triangle* as manned boats and ships: the tradeoff between speed, range, and endurance. For example, an increase in the top speed of a particular craft design necessitates a decrease in payload capacity or endurance. In other words, every USV platform design, such as for manned craft, has been optimized for a particular set of performance parameters. USV platforms are developed by designing a purpose-built platform or by converting a manned platform to be unmanned (by adding a control system and payload). Purpose-built platforms have the advantage of being optimized for the intended mission and environment, while USVs that are based on converted manned platforms provide cost savings by obviating the need for new design activity and

Fig. 13.1 DARPA ACTUV (anti-submarine warfare continuous trail unmanned surface vehicle) (courtesy of DARPA)

Fig. 13.2 ONR unmanned sea surface vehicle – high tow force (USSV-HTF) (courtesy of ONR)

Fig. 13.3 ONR unmanned sea surface vehicle – high speed (USSV-HS) (courtesy of ONR)

Fig. 13.4 Remote multi-mission vehicle (RMMV) (courtesy of PEO LCS)

have readily available spare parts. Many developmental USVs and 7 m classes have used an existing manned platform, the rigid hull inflatable boat (RHIB). RHIBs are designed for high speeds in benign sea states; therefore, the RHIB may be an appropriate platform for missions that fit these parameters. However, many envisioned missions do not fit these parameters – they do not require such high speeds, but they do require operation in a seaway. For these missions, the RHIB is not the optimal platform and may, in fact, not be capable of performing the mission at all. To *optimize* the performance of the USV, for the intended mission and environment, the hull form must be optimized, which necessitates a purpose-built design. An example of a purpose-built design is the Office of Naval Research's (ONR) unmanned sea surface vehicle with high tow force (USSV-HTF) (Fig. 13.2), which was optimized for tow capacity and payload fraction, within particular structural constraints [13.7].

Currently, most USVs are monohulls, either planing or semi-planing. For example, the USV-HTF is a semi-planing craft and the RHIB has a planing hull. There are numerous alternative hull forms that might be considered. Each has advantages and disadvantages relative to planing hulls in terms of payload capacity, seakeeping, speed, launch and recovery considerations, and others. Depending on the particular mission and environment, an alternate hull form may be an effective choice. There have been a few examples of USV designs that explore the use of alternate hull forms. For example, ONR has designed and built a hydrofoil USV [13.7]. The unmanned sea surface vehicle–high speed (USSV-HS) (Fig. 13.3) is optimized to maintain its top speed in higher sea states than would be possible with a similarly-sized planing hull. As a semisubmersible, the remote multi-mission vehicle (RMMV) (Fig. 13.4) operates below the sea surface, which provides more stability [13.8]. Another hull form that provides improved stability and is being investigated for use in USVs [13.9, 10] is the small waterplane area twin hull (SWATH). Later in this chapter we will present an example of a purpose-built USV design based on a SWATH hull form.

In addition to the investigation of alternate hull forms, the pursuit of reduced lightship weight is a key area of research and development for USVs. Reduced lightship weight would result in the ability to carry additional payload or fuel, which would provide direct benefits to mission effectiveness. The use of advanced hydrodynamic computational tools to determine hull pressures, particularly during slamming events, as well as the computational prediction of extreme environmental structural loading may lead to optimized structural design and reduced structural weight while maintaining structural integrity [13.11, 12]. Computational fluid dynamical (CFD) methods for hydrodynamic performance prediction will be discussed later in this chapter.

13.2 Autonomous Maneuvering and Navigation

Most USVs today require substantial, remote human assistance to maneuver and navigate in anything but a very highly controlled environment. Consequently, a reliable, high-bandwidth radio link to transmit imagery from the USV to the human operator for hazard avoidance is required; they are not capable of autonomous operation. On the other hand, an autonomous USV (AUSV) is capable of perceiving its internal and external environment, capable of performing a task or function in a complex environment with minimal human intervention, and capable of responding effectively to a dynamic situation. A non-autonomous USV includes control approaches such as remote-control, waypoint navigation, and scripted operations, and has no organic perception – it is not capable of sensing or responding to its environment. In contrast, an AUSV is capable of hazard avoidance in compliance with the International Regulations for Preventing Collisions at Sea 1972 (COLREGs) without human assistance, can arbitrate between multiple, competing objectives, and can do planning/replanning based on changing mission objectives, situations, and resource levels.

The state-of-the-art for AUSVs is on-water demonstrations in good visibility with avoidance of fixed hazards for own ship speeds of > 20 kn, avoidance of moving hazards with COLREGs compliance for own boat speeds of > 20 kn, implementation of COLREGs for head-on, overtake, and crossing situations, and tactically relevant behaviors, including go-to-waypoint, avoid hazard, obey COLREGs, trail, intercept, and patrol. Real-time adaptive response to vehicle faults and dynamic re-planning have also been demonstrated on-water. An example of arbitration between multiple, competing objectives is the autonomous maneuvering and navigation of a USV in achieving a particular waypoint while avoiding hazards and complying with COLREGs.

The sea surface environment presents significant challenges for autonomous maneuvering and navigation. Issues include intermittent obscuration of sensors in rough seas by sea spray and wave slap, limitations of radar close to the USV, the effect of platform motions on sensor performance, the effect of sea surface topography on sensor line-of-sight, and limitations imposed by atmospheric weather conditions (fog, rain, snow, marine layer) on sensor performance. There is work underway to develop advanced autonomy for USVs and address the above technical challenges [13.13–15]. This is enabled by continuing advancements in technologies such as sensors (EO (electro-optic), IR (infrared), radar, LIDAR (light detection and ranging)), sensor processing, sensor fusion, perception capable of dynamically adjusting to environmental conditions, determination of intent (of other vessels), decision-making using uncertain and intermittent perception information, and look-ahead planning. A highly dynamic and unpredictable environment combined with mission complexity motivates a hybrid autonomy architecture [13.14–16] that includes a reactive decision-making component and a deliberative decision-making component. The reactive component provides a fast answer and is used for quick response to close-aboard hazards not detected by the perception system until the last moment and is necessary for safe operations in high risk environments. Deliberative decision-making uses more information and provides an answer that is valid over a longer period of time (unless the situation changes!), but it takes longer for the system to provide that answer. Deliberative planning is an enabler for long-duration operations. A hybrid autonomous architecture has been demonstrated on-water [13.14–16].

Widespread adoption of autonomous maneuvering and navigation will require that significant non-technical challenges be overcome to properly address safety/liability concerns. Yet to be defined are standard metrics of autonomy in terms of core competencies with associated proficiency tests. Approaches consisting purely of on-water testing to a high level of confidence in all conditions are prohibitively expensive; therefore, modeling and simulation will likely play a key role. Cultural issues will be more difficult for AUSVs than for unmanned systems. Non-autonomous unmanned systems have had to overcome the issue of taking the human out of the vehicle and locating him/her remotely from the platform, but still having a high degree of involvement with the operation of the platform. AUSVs will confront the issue of the human being much less involved in the operation, with the involvement being at a supervisory level only. Acceptance of this mode of operation will occur through a gradual approach in which AUSVs are initially used in an autonomous mode, but with a high degree of human supervision to establish trust in the system's performance. As confidence grows, the degree of human oversight will be decreased, allowing AUSVs to reach their full potential.

13.3 Naval Architecture of AUSV Design

This section will first define the general criteria used to select a hull form as a function of operational speed and size and then it will continue to describe in some detail the hydrodynamic design and performance of a new class of vehicles whose optimized design is to become a reference for the next generation of autonomous unmanned surface vehicles for ocean and coastal exploration.

13.3.1 Froude Number and Hull Typologies

Essential for selection of the hull form is the calculation of the relative speed of advance with respect to a representative dimension (the length) and weight (displacement volume) of the hull.

In non-dimensional form these two parameters are defined as the length Froude number Fr_L and the volumetric Froude number Fr_∇

$$\mathrm{Fr}_L = \frac{U}{\sqrt{gL}}\,, \quad \mathrm{Fr}_\nabla = \frac{U}{\sqrt{g\nabla^{1/3}}}\,, \tag{13.1}$$

where g is the acceleration of gravity, L is the submerged length of the hull, U, is the speed of the hull and ∇ is the hull displacement volume. The length Froude number is often used to qualitatively discuss the speed regime of a boat. From a purely physical point of view, the Froude number is the ratio between inertial forces due to the ship advancing at a certain speed U and the gravitational force, which are responsible for the free surface waves and the wave resistance. In fact, using the dispersion relation valid for first-order plane progressive (airy) waves, it can be demonstrated that the length-based Froude number is related to the ratio between the fundamental wavelength generated by the hull and its own length.

If we approximate the length of a transverse wave produced by a pressure region at the bow to that of regular wave traveling with hull speed, we obtain the following relation between the Froude number and the relative wave length

$$2\pi\lambda = gT^2 \rightarrow 2\pi\lambda = g\frac{\lambda^2}{U^2} \rightarrow \frac{\lambda}{L} = 2\pi\mathrm{Fr}_L^2\,, \tag{13.2}$$

where in addition to the variables appearing in (13.1), T is the period and λ is the length of the fundamental transverse progressive wave.

Figure 13.5 compares the fundamental wavelengths relative to the hull profile at four different Froude numbers representative of four different sailing regimes and related hull types. In naval architecture, the length Froude number is often used as an indication to select the most suitable hull typology for the particular application, as it is graphically presented with the shaded areas in Fig. 13.6. This is a preliminary distinction and must correctly consider the volumetric Froude number, especially to distinguish the regime of planing hulls from that of semi-displacement hull forms. This will be explained in more detail later on in this section. Slow speed heavier displacement hull forms operate in the lower Froude number range (Fig. 13.5a, upper panel), at which the fundamental wave length is much lower than the hull length and the resistance is dominated by viscous effects (friction and form drag). The shape of these types of hulls is optimized in order to reduce to a minimum the frictional and form drag components. Wavelength and wave resistance both increase with the Froude number. Typically, at $\mathrm{Fr} = 0.25$ the wave resistance and viscous resistance share the same relative importance in the total resistance. For $\mathrm{Fr} = 0.4$, the relative speed is representative of the limit speed for a fast displacement hull. Most fast naval units (destroyers, frigates) are at these Froude numbers. At this speed, the hull is running on its own wave length ($\lambda/L = 1$). Bulbous bows are very beneficial at these Froude numbers, since their generated waves can interfere directly with the aft wave train (Fig. 13.5b, lower panel). At $\mathrm{Fr} > 0.4$, the fundamental wavelength is larger than the hull length and the hull begins to severely squat and trim down at the stern. The ship literally climbs up the back of its own bow wave. At these speeds the flow along the convex geometry of the hull bottom develops large suction pressures that further increase the squat and trim of the vessel. The hull resistance begins to increase dramatically and becomes a practical barrier to further increases in speed. The wave pattern generated by a displacement hull is, of course, more complex as it derives from the continuously variable distribution of dynamic pressure along the hull (which depends on its shape), but to a first-order approximation, the final transverse wave pattern can be thought as a superposition of a series of transverse wave trains, each one generated at regions of the hull close to the free surface with high pressure gradients.

A sketch of two transverse and divergent wave systems created at the bow and stern of a displacement hull form (Fig. 13.7) is illustrative of the wave interference effect. A solution to the complex free surface wave pattern was first attempted by Lord Kelvin and solved for simple analytical hull shapes by *Michell* [13.17] and later by *Wigley* [13.18]. These studies led the way to the development of thin ship theory, which is the simplest theoretical or numerical method to predict ship wave patterns, based on the assumption that the ship

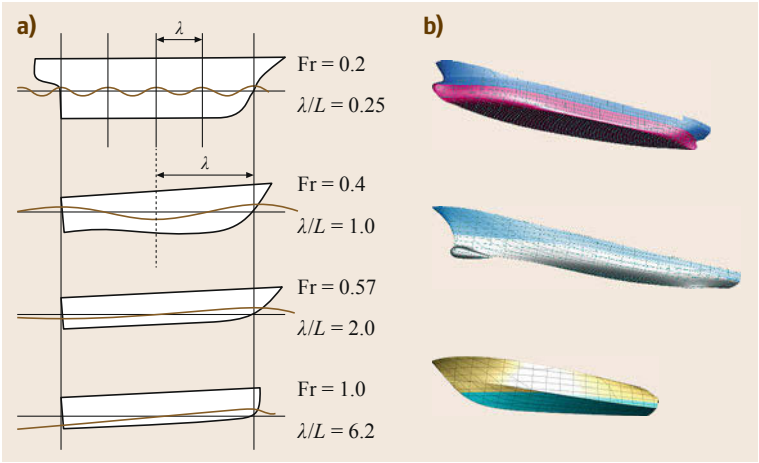


Fig. 13.5 (a) Fr = 0.2 corresponds to slow displacement hulls (e.g., large cargo ships ($L = 150-300$ m, 12–15 kn), but also to small (10 m long) boats at low speeds (3 kn). The weight of the vessel is mostly supported by buoyancy forces. For Fr $\lesssim 0.25$, the dominant resistance component is due to viscous effects, mainly friction and form drag, and the hull shape is normally adapted to this. ((b), top hull) An emblematic hull shape designed to operate at these speeds is a 160 m chemical tanker. For Fr = 0.4 ((a), second panel), the relative speed is representative of the limit speed for fast displacement ships. Most of fast naval vessels ((b), middle hull) operate in this Froude number range. At this relative speed, the hull is running on its own wave length ($\lambda/L = 1$). At Fr > 0.4 (e.g., (a), third panel), the fundamental wavelength is larger than the hull length. This Froude number corresponds to the so called *hull-speed*. Only semi-displacement and planing hulls ((b), lower hull), also characterized by lighter displacements, can overcome the speed barrier Fr = 0.4–0.5 and have a shape adapted to develop hydrodynamic lift force on their bottom surface. At Fr = 1, the hull starts to be supported primarily by hydrodynamic (instead of hydrostatic) forces, the transverse wave trains almost disappear, and the free surface flow around the hull significantly change its nature

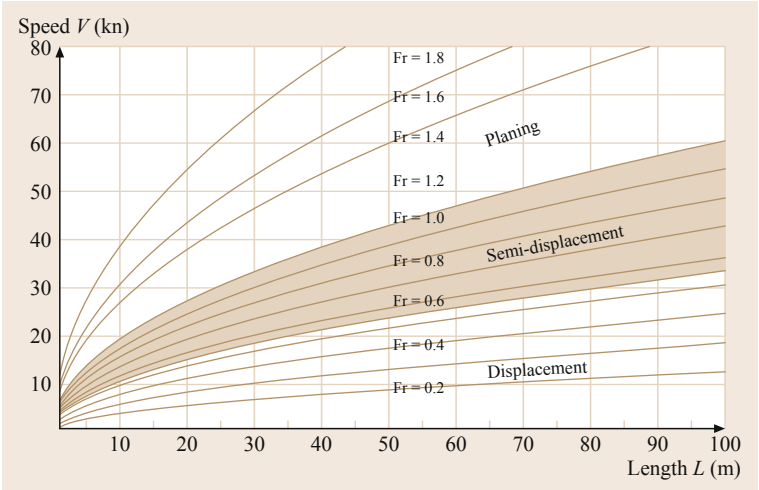


Fig. 13.6 Iso-Froude number curves in the ship speed–length plane and related hull typology

is a long slender body so the perturbation sources can be collapsed on the symmetry plane of the ship. The most recent applications of this theory with surprisingly good correlations were presented by *Tuck* [13.19], *Noblesse* [13.20], and *Doctors* and *Day* [13.21], who studied an appropriate extension of the theory for high-speed displacement hulls.

The approximations and inherent limitations of thin ship theory have been overcome by boundary element methods (BEM) or *panel methods*, which consider the real three-dimensional (3-D) shape of the hull as a boundary of the potential flow domain with linear [13.22, 23] or non-linear [13.24–27] free surface boundary conditions for the solution of the free

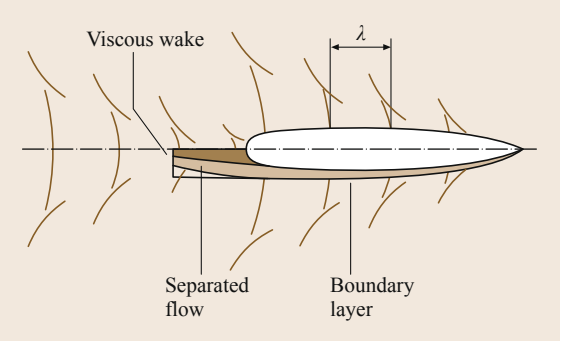


Fig. 13.7 Idealization of the of flow regions types around a ship advancing in calm water

wave patterns. Higher-order versions of these methods that use curved panels defined by B-spline surfaces [13.28] have been developed, and they have shown superior convergence and stability characteristics. Panel methods have been extensively tested and validated both in the case of slow and fast displacement hull forms [13.29] and for semi-displacement mono and multi-hulls, where the effect of dynamic trim and sinkage is essential [13.30–32].

Spectral analysis of wave pattern formation applied to experimental measurements [13.33] and to numerical simulations [13.30–32] has yielded insight into the wave generation characteristics of displacement hull forms, such as the relative importance of divergent and transverse wave trains at different speeds. In terms of energy, the relative importance of the divergent waves, negligible at low speeds with most of the energy concentrated in the transverse waves, progressively increases with the Froude number to become dominant for high-speed semi-displacement hull forms (Fr > 0.55). This ratio is directly dependent on the shape of the hull.

The solution of the viscous free surface flow around the hull is today possible with numerical methods that are able to solve the Reynolds averaged Navier–Stokes equations (RANS). The sharp free surface separating air with water was initially solved with surface tracking algorithms and level set methods [13.35] that are based on the assumption of a smooth and continuous surface on which dynamic and kinematic wave boundary conditions are satisfied, while more recent surface capturing algorithms, based on the volume of fluid technique [13.36] result in a simpler and faster numerical implementation and can inherently deal with wave breaking or spray flows. In recent validation studies, these methods have given good results both in cases of displacement hull forms [13.37] and for planing hulls [13.38–41]. Fully viscous free surface flow solvers avoid empirical estimation of the friction and form drag

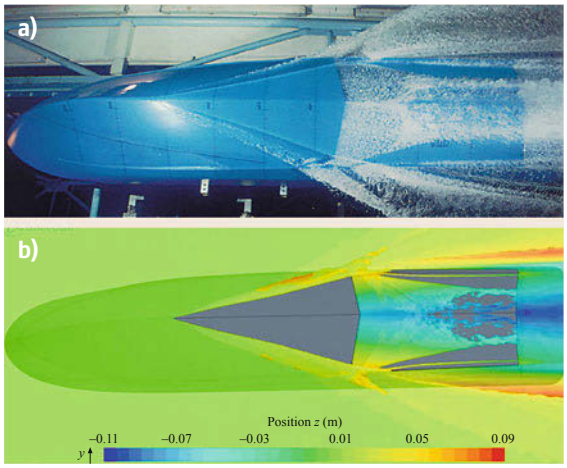


Fig. 13.8a,b Two-phase (air/water) flow field around a stepped planing hull with a partially ventilated afterbody. **(a)** Experiments, **(b)** numerical prediction with RANS solver (with permission from [13.34])

and include the possibility of simulating the non-linear interaction between the boundary layer flow and the flow induced by free surface waves, and they are general enough to be applied also to planing hulls (Fig. 13.8) for which the formulation of traditional boundary element methods, initially developed for displacement hulls, need to be radically changed to reproduce the different physics of the planing flow regime [13.42, 43].

Today it is possible to optimize the shape of the ship hull by means of CFD simulations (simulation based design and optimization). Some examples include fast displacement yachts [13.44, 45] and fast displacement navy hulls [13.46–48], but they are limited to local hull form optimization. More recently, global hull form optimization has been achieved on fast displacement monohulls with realistic design constraints [13.49] and on innovative types of fast catamarans with SWATH [13.50] or semi-SWATH shapes [13.51].

The effective power (defined as the product of the total resistance and the ship speed) normalized to displacement can be plotted as a function of the Froude number for different hull typologies (Fig. 13.9). The sharp increase in resistance for a displacement hull at $Fr_L > 0.4 \sim 0.5$ is sometimes called the resistance wall, beyond which speed is limited regardless of propulsion power. A substantial displacement reduction (Fnvol > 1.5) and specific semi-displacement hull shapes permit the resistance wall for a displacement hull (Fig. 13.9). However, at $Fr_L = 0.9 \sim 1.0$, the resistance of the semi-displacement hull increases very rapidly, delineating a practical upper limit for the operational speed of such hulls. At $Fr_L > 1$ and $Fr_\nabla > 3.0$, provided the hull

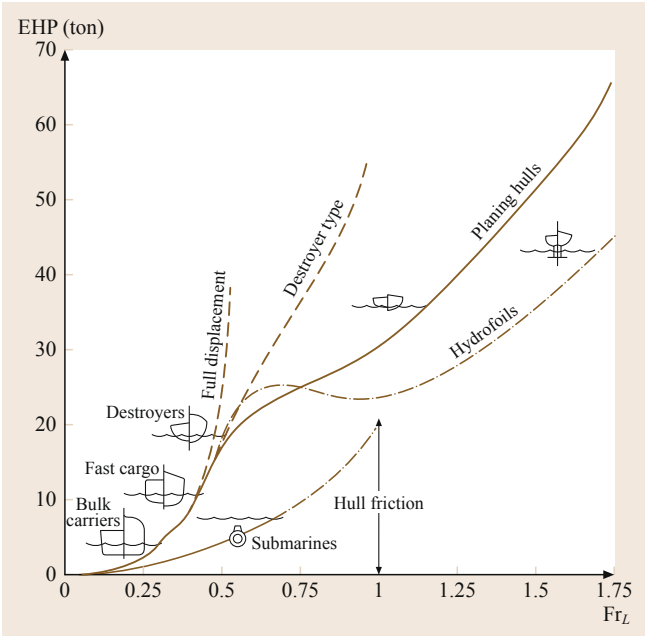


Fig. 13.9 Indicative effective horsepower (EHP) over ton of displacement as a function of the relative speed for different typologies of marine vehicles

shape is adequately changed (bottom hull in Fig. 13.5b), the planing regime occurs in which 80–90% of the weight is supported by the dynamic pressure of the high speed flow field around it. Although intended to operate at high Froude numbers, the planing hull must pass through the entire range of regimes prior to attaining its top speed. As confirmed by *Savitsky* [13.52], at $Fr_L < 1.0$, the planing hull has a somewhat larger resistance than displacement or semi-displacement hulls. This is attributed primarily to the sharp chines and submerged wide and sharp transom that promote flow separation and hence increase the form and wave-making drag component of the hull. The planing hull easily passes through the speed *barriers* for a displacement ship ($Fr_L = 0.4–0.5$) and the limit for a semi-displacement hull ($Fr_L = 1.0$). Within the three main hull categories described so far, naval architects have designed different families that can be distinguished by the shape and the typology of the hull. For example, in the last three decades the semi-displacement category, previously including only slender round bilge monohulls, has seen one of the largest unconventional design proliferations, with new hull shapes (deep-V, vertical or reversed stem, etc.) and typologies (multihulls such as catamarans and trimarans), which have caused a substantial performance increase at sea of this category of vessels compared to the previous state-of-the-art. An example of an unconventional hull form design, specifically for the application to a new generation unmanned surface crafts, is given in the next section.

13.4 Optimized Class of Autonomous Unmanned Surface Vehicles

While there has been a long history of vessels designed as manned platforms and a much shorter history of such vessels converted into unmanned vessels, vessels designed ab initio to perform optimally as AUSVs are an embryonic class. In addition to vessels designed for stand-alone missions, an AUSV is an essential, yet not often currently deployed, component of a persistent four-dimensional (4-D) ocean monitoring and surveillance system (Fig. 13.10).

The main characteristics of vessels in this sub-class are influenced by some key requirements:

- The ability to cover the assigned range (100 nmi) at design speed (12 kn) with onboard fuel and power generation, and conversion devices.
- Continuous operability up to sea state 3, corresponding to a significant wave height between 0.5 and 1.25 m and a wind speed up to 30 kn.
- The ability to transport, launch, and recharge subsurface AUVs (autonomous underwater vehicles) of lengths up to 2.2 m and maximum weights of about 100 kg.
- Sufficient static transverse stability to withstand 50 kn wind heeling moment without exposing the underwater hull or the propeller. Positive dynamic stability up to the heeling angle corresponding to the exposure of the underwater hull.
- Sufficient longitudinal metacentric height (greater than the transversal one) to reduce the dynamic trim

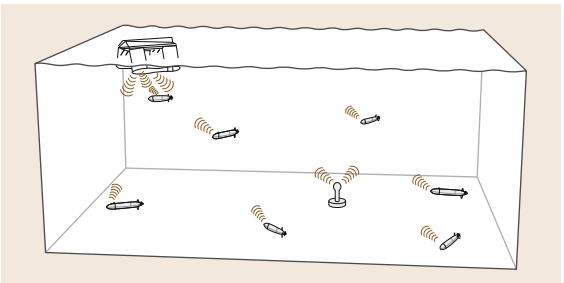


Fig. 13.10 The small AUSV-SWATH as a key element of an integrated, cooperative monitoring and surveillance system

by the bow, typical of high speed SWATHs, and withstand realistic shift of loads.
- The ability to serve as a launch, recovery, and recharging platform for unmanned aerial vehicles.
- Transportability inside a standard container (ISO LWH $12.2 \times 2.44 \times 2.59$).

The speed versus length relation of vessels belonging to this sub-class can be plotted along a constant Froude number (Fig. 13.11). We will consider one such a vessel here in detail as an example of the many design trade-offs involved. The principal geometric characteristics of this small unconventional ASV-SWATH hull are given in Table 13.1 (NURC), together with those of the dual mode hybrid hydrofoil SWATH (ONRG) and of another fast single-strut unconventional SWATH (NH04) optimized in [13.50]. To best satisfy the requirements, an unconventional SWATH hull form was selected, and a prototype was designed (Figs. 13.12–13.14). The advantages of a SWATH include superior seakeeping ability in high sea states with respect to other conventional displacement hull topologies (including catamarans or trimarans). The main drawback of SWATHs, however, has always been the higher powering requirement with respect to equivalent monohulls or catamarans, especially at the highest Froude numbers. This problem has been solved by adopting a particular unconventional shape of the underwater hull, described later in detail.

The new vessel configuration consists of a main central body with the keel line suspended more than 1 m above sea level by vertical struts that connect to the underwater hulls [13.55, 56]. The junctions between the top of the struts and the sides of the main body are fastened with removable and adjustable rods for easy disassembly and regulation of the strut inclination (canting angle). Two diesel generators, each 22 kW, provide the required propulsion power of 39 kW at the design speed as well as power for the estimated hotel load. Two fuel tanks, placed inside the middle of the underwater hulls, provide a range of more than 100 nmi (at full speed) as required for medium and short range missions.

Although the SWATH configuration offers superior seakeeping, the hydrostatic stability index must be evaluated carefully because of the high sensitivity to weight distribution. Particular attention has been paid to match the longitudinal position of the center of buoyancy and the center of gravity (LCB–LCG), with some trim control capability built in (ballast water tanks). Because of the high sensitivity to load changes, two mirrored ballast tanks have been fitted inside the underwater hulls, in the same position as the fuel tanks, to compensate

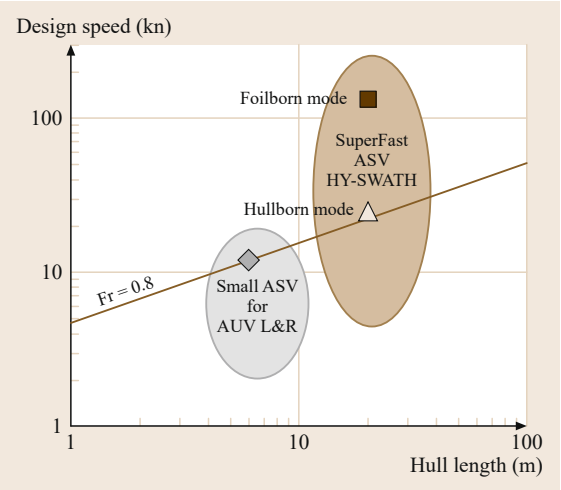


Fig. 13.11 Size and speed of two members of the AUSV class. The larger vessel (20 m long, *triangle*) has the same unconventional SWATH hull form in hull-borne mode, but it is designed to be convertible into a ultra-high speed hydrofoil to reach 120 kn (after [13.53])

Table 13.1 Main geometrical characteristics of unconventional SWATH designs with optimized underwater hull shapes. NURC and ONRG belong to the same family (twin canted struts) presented in Fig. 13.11. NH04 has a single vertical strut

	L/D	CP	$L/V_1^{1/3}$	$L/V_{tot}^{1/3}$	Fr	Fr_{V1}	Fr_{Vtot}
NH04	13.3	0.65	7.04	4.89	0.50	1.32	1.10
NURC	10.0	0.55	6.15	3.74	0.80	2.00	1.56
ONRG	16.7	0.42	9.44	5.80	0.92	2.82	2.21

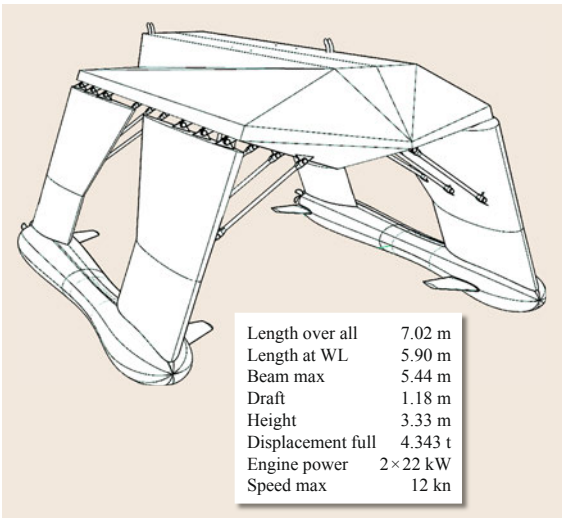


Fig. 13.12 SWATH-AUSV, external view and main characteristics (after [13.54])

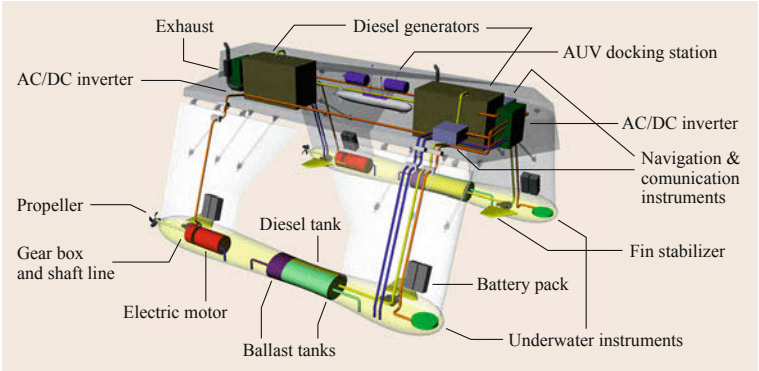


Fig. 13.13 SWATH-AUSV (after [13.54]). General arrangement of main machineries and systems

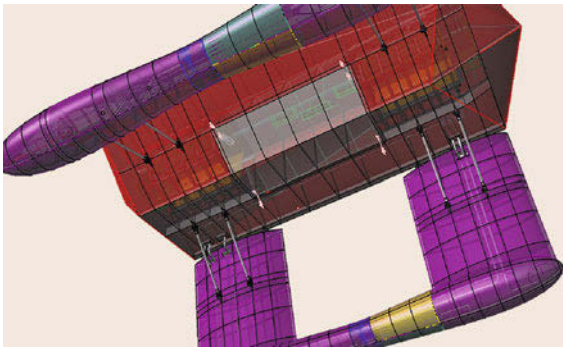


Fig. 13.14 SWATH-AUSV underside view showing AUV launch and recovery bay

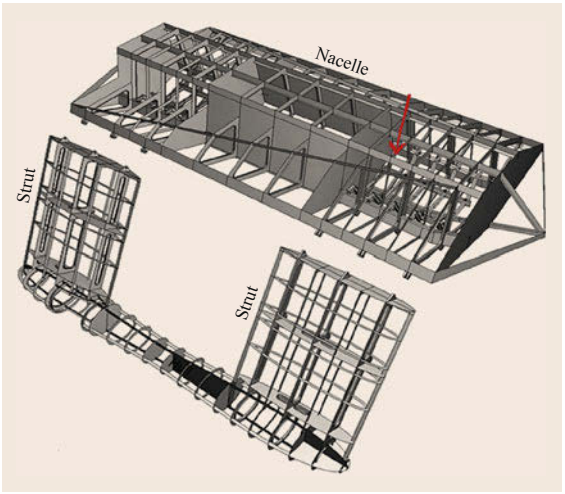


Fig. 13.15 Layout of main girders and structural elements

for the loss of weight and trim moment due to fuel consumption. The sensitivity is a consequence of the slender geometry of the struts in the horizontal plane, which do not offer much buoyancy reserve to counterbalance weight changes. To decrease sensitivity, the struts have been tapered just above the waterplane, increasing in width and length, to gain volume when subjected to a pitching moment. A similar approach has been adopted for forces and moments in the transverse plane.

Hydrostatic calculations indicate that the initial metacentric heights of the vessel in the transverse and longitudinal planes are in close proximity, in contrast to a conventional monohull, where the longitudinal metacentric height is in the order of the vessel length, while the transversal metacentric height is in the order of its draft. This fact implies no preferable direction of heeling for the vessel.

To manage the dynamic trim and sinkage of the vessel at high speed, as well as the dynamic stability of the vessel in waves, four active stabilizer fins have been installed, which will guarantee the optimal attitude when the vehicle is advancing and good damping while it is stationary. These two pairs of lifting surfaces are placed on the inner side and far ends of the hulls to obtain the maximum trimming moment.

The aft end of each hull houses an electric motor, gear box, and shaft, whereas the fore end is reserved for sensor payloads. The central body has been divided into three separate compartments by two transversal watertight bulkheads. Inside the fore and aft parts there are diesel generators, frequency converters, and control computers; in the mid compartment a docking space for AUVs has been constructed (Fig. 13.14). The weight of AUV recovery has been compensated by two additional seawater tanks with adequate reserve buoyancy. The framing consists of such a light metal shell (aluminum alloy 5083-H321, 2.7 kg/dm^3) with transverse main girders (Fig. 13.15). The diesel engines are mounted in a rubber mounted, highly damped cradle.

Several transverse, quasi-static (large angle) stability tests have been executed to verify the response of the vessel to the combined effect of wind and waves in terms of heeling righting moments versus heeling angle (Fig. 13.16). Worst case scenarios (tanks half full) were assumed. Three different strut design shapes were

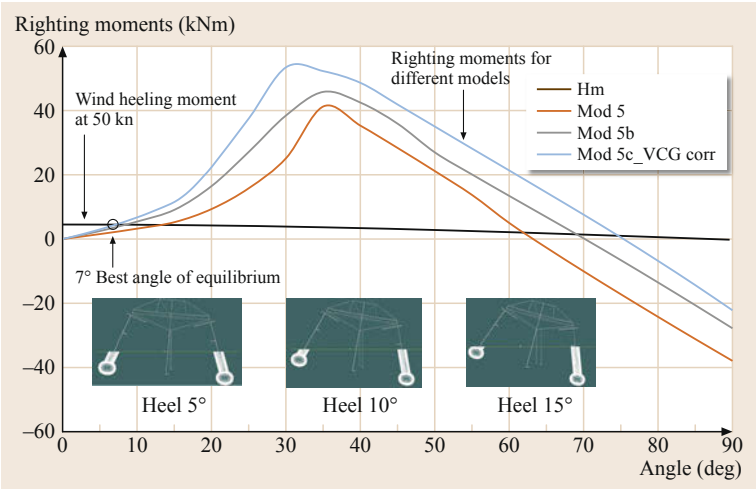


Fig. 13.16 Righting moment curves calculated for three different strut layouts and verification of the extreme wind stiffness heeling moment (after [13.55])

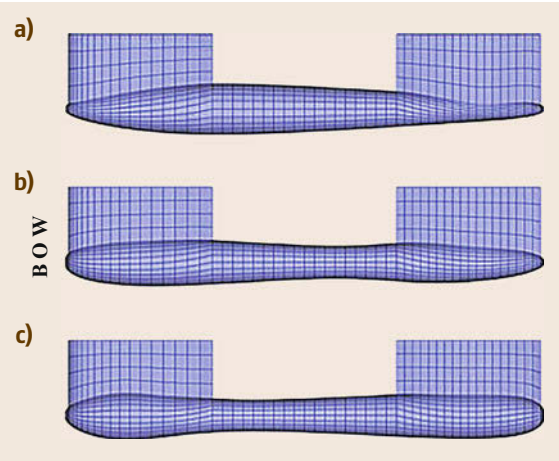


Fig. 13.17 Three different underwater hull shapes of the AUSV-SWATH considered during optimization: **(a)** conventional shape (no intermediate contraction), **(b)** unconventional reference design, **(c)** optimized unconventional shape

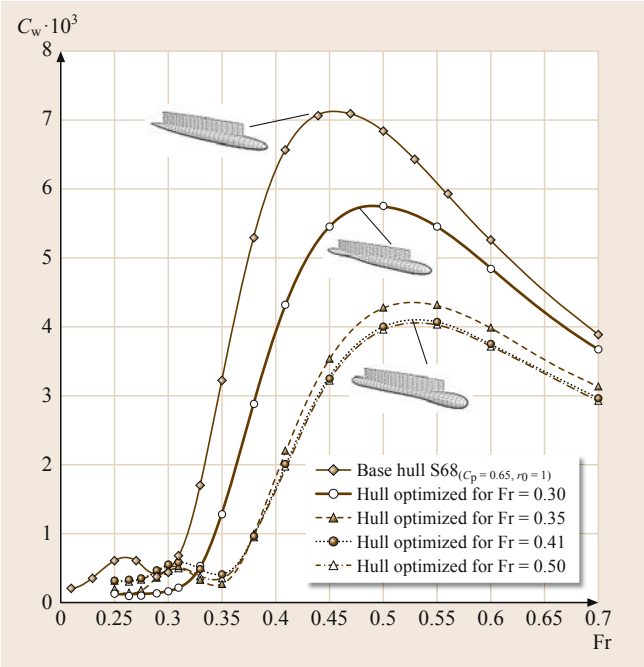


Fig. 13.18 Wave resistance reduction of unconventional SWATHs optimized according to *Brizzolara* [13.50]

considered. The lowest curve (Fig. 13.16) is for straight vertical struts, and the highest one for the final strut configuration with flared sides above waterline and longitudinal inclined stem profile.

For the second generation autonomous vehicles, an unconventional shape for the underwater hulls is utilized with two expanded main sections and an intermediate contraction (Fig. 13.17). Minimization of the resistance at a given design speed is primarily dependent on the position of the intermediate contracted section, as was first demonstrated by *Brizzolara* [13.50] for a single strut SWATH vessel. The optimum longitudinal position of the contracted section changes as a function of the Froude number. In general, for a single vertical strut SWATH, at lower speeds, the contracted section moves aft. This trend is maintained also for the twin canted strut SWATH design of the autonomous surface craft [13.57].

The advantages of shaping the lower hulls to reduce wave resistance (which in displacement mode is the main resistance component) can be as large as 50% as in Fig. 13.18, in the case of the single strut SWATH (NH04 in Table 13.1) and close to 35% as in Fig. 13.19 for the twin canted struts SWATH (NURC in Table 13.1),

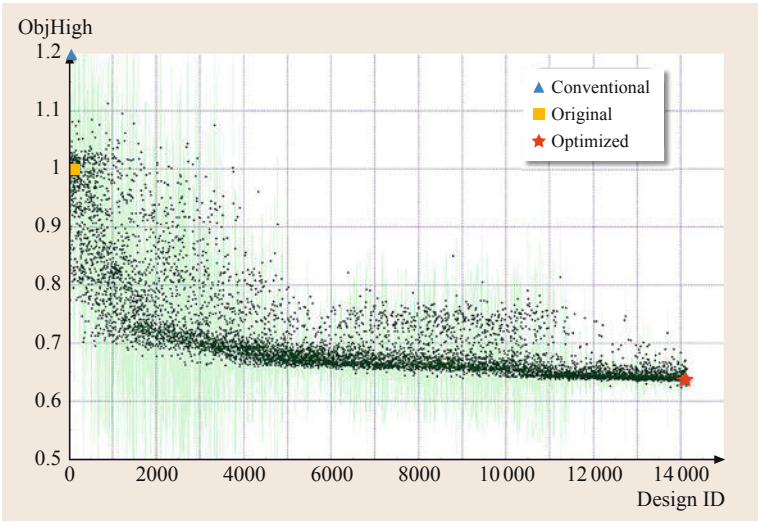


Fig. 13.19 Relative reduction of the wave resistance (function ObjHigh) during the optimization procedure for different hull shapes. The points corresponding to the original (first unconventional hull design), conventional (series-58 torpedo shaped hulls) and optimized hull forms are highlighted with different symbols. Each *black point* corresponds to a different hull geometry generated and evaluated by the optimization. The total considered cases are in excess of 14 000

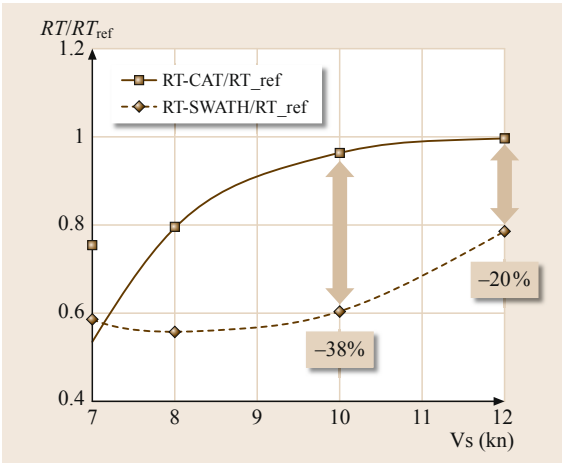


Fig. 13.20 Comparison of the total resistance predicted for the AUSV-SWATH and an equivalent catamaran. Values on the *y*-axis are relative to the total resistance of the catamaran predicted at 12 kn

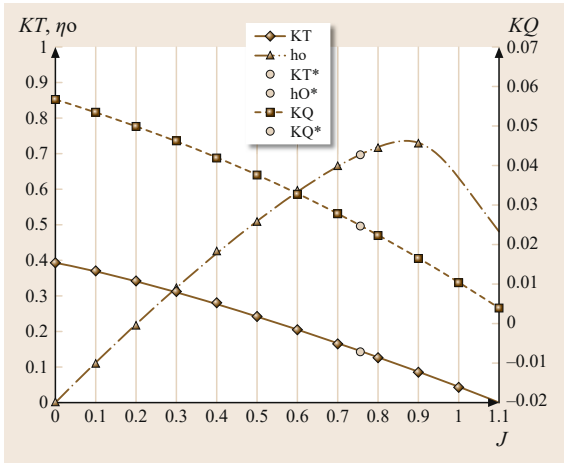


Fig. 13.21 Open water characteristics of the optimum propeller

similarly to its sister design variant in hull-borne mode (ONRG in Table 13.1). A fully automatic, computer driven, optimization procedure based on a new fully 3-D parametric model of the hull surface [13.57] was used to find the unconventional shape. This procedure is based on a viscous-inviscid flow solver for the prediction of the wave and viscous resistance of a SWATH vessel as validated in *Brizzolara* [13.50].

After the optimization of the underwater hull forms at the design cruise speed, the resistance–speed curve of the vessel was computed in a complete range of speeds by high fidelity CFD simulations with a fully turbulent, viscous, free surface flow model, validated against experimental tests in a towing tank on different occasions [13.58], also in the case of unconventional SWATH vessels [13.59]. At hump speed, a characteristic relative maximum of the resistance coefficient curve versus speed appears, which is typical of high speed SWATHs and corresponds to a critical bow-down dynamic attitude, which must be corrected by the action of active fin stabilizers. The two couples of fins, arranged at the bow and stern sections of the interior of the underwater hulls, are controlled to dampen roll and pitch motions in waves, as well as to regulate the attitude of the vessel at speed. Steep and high waves are generated at a hump speed corresponding to 5–6 kn for the AUSV presented in Fig. 13.12 ($Fr_L = 0.33–0.4$). This critical speed range, similar to the pre-planing regime of hard chine fast crafts, should be avoided during transfer. As a consequence of the hull form optimization for

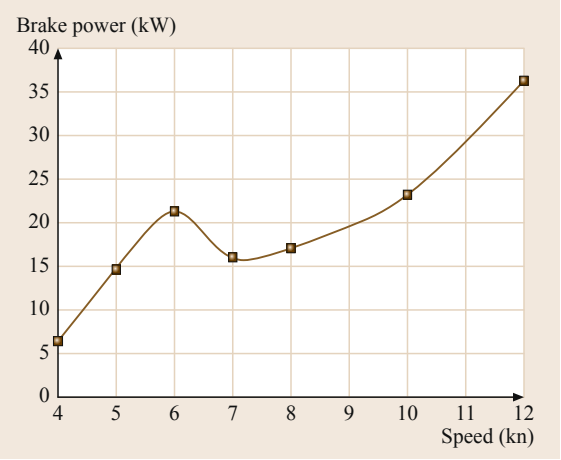


Fig. 13.22 Estimated total brake power with optimum propellers and twin shaft

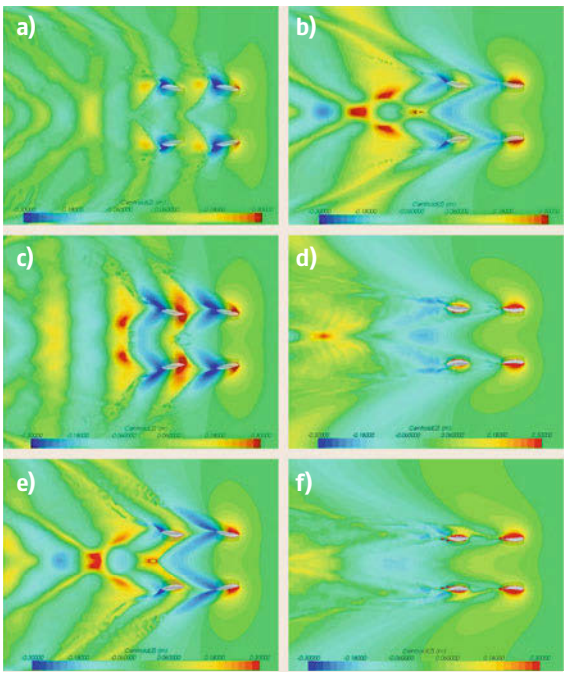


Fig. 13.23a–f Free near-field wave patterns calculated with the RANS solver for the optimized hull form at different speeds: **(a)** 5 kn, **(b)** 8 kn, **(c)** 6 kn, **(d)** 10 kn, **(e)** 7 kn, **(f)** 12 kn

higher speeds, the resistance drops considerably at 7 kn and then rises with a milder slope than equivalent conventional SWATH or catamarans (Fig. 13.20).

Figure 13.20 also presents the results of a comparison study with the resistance predicted for an equivalent conventional catamaran. The catamaran hull was obtained by transformation in affinity from the Marintek catamaran [13.60, p. 279], which is representative of a good high speed modern catamaran hull form. The original Marintek hull was non-uniformly scaled with two different scale factors, one for length and the second one for beam and draft, to reach an overall length of 7 m and the same displacement of the SWATH. Hull spacing of the catamaran was selected in order to reach the same deck area. The final catamaran hull is presented in Fig. 13.24, overlapped in 3-D onto the SWATH hull. The shape optimization procedure adopted to design the AUSV-SWATH hull allows for 38% drag reduction at 10 kn (the actual speed considered for the optimization) and 20% drag reduction at top speed (12 kn in this case). These are considerable values and imply a significant increase of endurance with respect to the current state-of-the-art.

On the basis of the effective power estimated with the RANS solver, selection of the optimum propeller of the Wageningen B-series was performed (Fig. 13.21). With the known load coefficient K_T/J^2, calculated at design speed, the optimum propeller has a diameter of 700 mm, a maximum rotational speed of 490 rpm at 12 kn, and an open water efficiency of about 70%, which is quite good for this size of craft. This high level of efficiency can be reached due to the relatively low rotational speed of the propeller, which is possible due to the reduction gear box fit in between the electric motor and the propeller.

The optimum propeller selection was based on a blade expanded area ratio of 0.45, calculated to satisfy standard cavitation criteria. An empirically derived thrust deduction factor of 0.95 and a wake fraction of 0.7 have been also assumed. The final brake power estimated on the basis of the predicted optimum propeller efficiency (Fig. 13.22) was used to select the characteristics of the two electric motors and the power of the two diesel generators, rated 22 kWe each (Fig. 13.13).

To examine seakeeping capabilities, unsteady RANS calculations [13.61, 62] with the SWATH free to heave and pitch but restrained in surge were systematically repeated for different incoming regular wavelengths, with values ranging from the SWATH hull length L, up to four times the SWATH length (Fig. 13.25, for example). All waves had the same height, nominally 3.5% of the SWATH hull length L, to be representative of extreme, non-linear conditions.

To compare the performance of the optimized AUSV-SWATH against an equivalent vessel, the seakeeping CFD simulations were repeated for a catamaran with a hard chine, deep-V hull form, previously also used for the resistance comparison and represented in Fig. 13.24 as overlapped on the SWATH.

Figure 13.25 shows a typical snapshot of the unsteady wave pattern predicted by the viscous flow solver

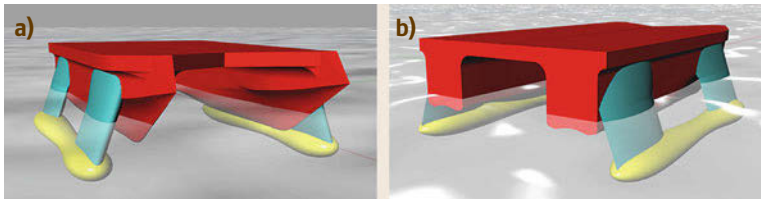


Fig. 13.24a,b Comparison by overlap of the AUSV-SWATH hull with the equivalent catamaran considered for resistance and seakeeping comparative studies. **(a)** Perspective view from bow quartering and **(b)** from stern quartering

around the AUSV-SWATH hull (superstructures are not represented) subject to an incoming wave with a length (λ) 2.5 times that of the SWATH hull length. Although the pitch and heave motions of the AUSV-SWATH, as presented later, are smaller than those of the catamaran, the radiated and diffracted waves are very well visible in the picture, consistently with the larger motion damping of the SWATH with respect to the catamaran. This is the result of the twin canted struts, which have a distinct capacity of dissipating the kinetic energy of the vessel moving in waves, by radiated waves and shed vorticity around the hull.

A hybrid strip theory method, based on two-dimensional (2-D) viscous non-linear free surface unsteady RANS calculation of radiated forces, demonstrated that the viscous effects responsible for vortex shedding are absolutely non-negligible for a correct prediction of the vertical plane motion of SWATH hull forms [13.62]. This study also confirmed that the peaks of heave and pitch response amplitude operators (RAOs) move to very high relative wavelengths (about $\lambda/L = 4.5$) in the case of this special SWATH design. At such low oscillation frequencies, the particular configuration of canted struts effectively limits the peak of the resonant response, which would be singular if evaluated with potential flow methods.

A comparison of the maximum motion amplitude during the simulations reveals that the SWATH heave and pitch motions are noticeably smaller than those of the equivalent catamaran over a wide range of incoming wavelengths (Figs. 13.26 and 13.27). As an example of the time history of the simulated pitch motion of the two crafts in the same incoming regular waves with $\lambda = 1.5L$ are given in Fig. 13.28. The heave RAO is calculated as the amplitude of the first harmonic of heave motion signal (elaborated with Fourier analysis) divided by the incoming wave amplitude. The pitch RAO is calculated as the ratio of the pitch first harmonic amplitude to the product of the wave number $k = 2\pi/\lambda$

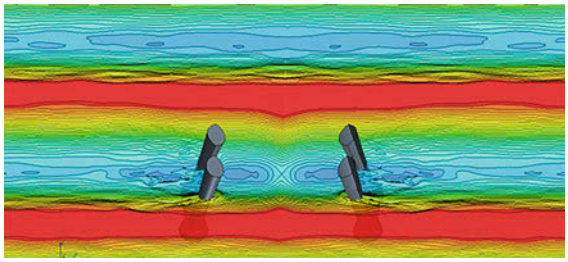


Fig. 13.25 Snapshot of the diffracted incoming regular waves around the SWATH predicted by the URANSE method ($\lambda/L = 2.5$)

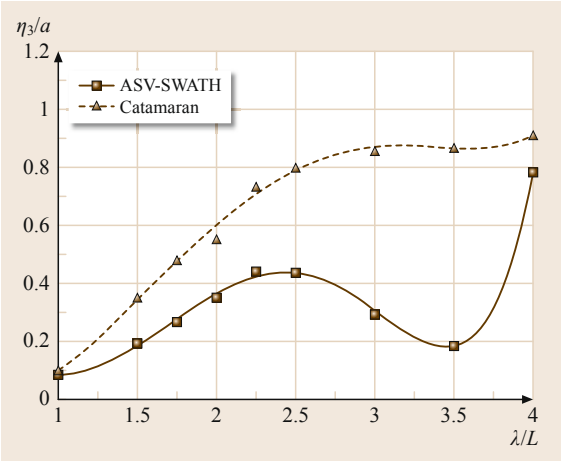


Fig. 13.26 Comparison of the heave RAO calculated for the two equivalent vessels

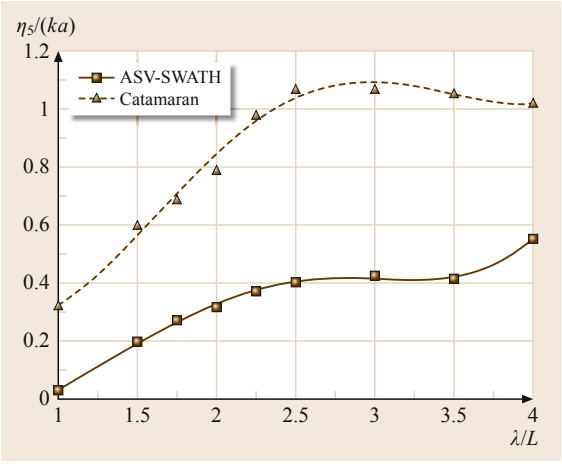


Fig. 13.27 Comparison of the pitch RAO calculated for the two equivalent vessels

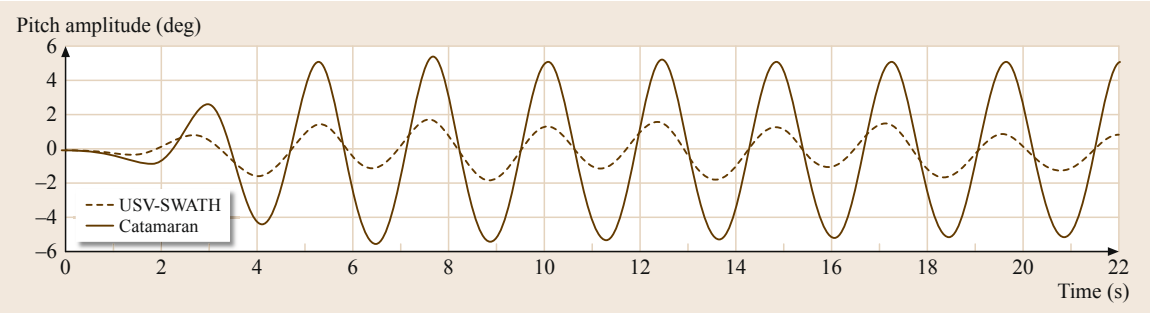


Fig. 13.28 Pitch angle time histories for the catamaran and the SWATH subject to the same regular progressing waves coming from the bow ($\lambda/L = 1.5$)

and the wave amplitude a. In the range investigated, a first relative maximum of the SWATH heave response happens for wavelengths between 2 and 2.5 times the hull length, but it still measures half of the catamaran heave amplitude. At the same time, the maximum SWATH pitch motion is less than 1/3 of that of the catamaran over the entire wave frequency range investigated. No distinct peak is noted in the pitch RAO for the SWATH; again this is a desirable effect consequence of the twin canted strut design of the unconventional SWATH, which smooths out and shifts the resonance peak towards very high relative wavelengths.

The calculated RAOs which correctly accounts for 3-D and viscous effects can be used to predict the motions of the SWATH design in irregular waves [13.62], according to linear seakeeping theory. This procedure can be conveniently repeated at different ship scales to evaluate the size of the unconventional autonomous vessel that best adapts to the particular sea state in the mission area.

13.5 Conclusions

With the advent of autonomous maritime systems, the opportunity exists to design a new class of surface vessels from the keel up to perform optimally as an integral component of such systems. Greater autonomy will enable the design of AUSV platforms for which no human will ever set foot on board. This will afford the opportunity to revisit approaches to platform design and naval architecture, and could result in radical changes to surface platforms [13.6]. Ships' hull and superstructure geometry, internal structure, and volume might be optimized for payload or fuel fraction rather than human habitability. Design limits and margins for structural adequacy, reserve buoyancy, stability, dynamics, and platform orientation might be different for AUSVs than for manned platforms, and result in additional mission capability. Higher risk mission profiles, untenable for human-inhabited ships, could be considered. New platform designs that result in increased platform performance may in the future lead to capabilities and missions not anticipated today.

As with all systems engineering, trade-offs abound, and mission requirements are necessary to constrain a design. Here, as an example, the requirements of a persistent, 4-D ocean monitoring and surveillance system have been defined, and an unconventional SWATH hull form has been designed to satisfy those requirements (Fig. 13.28). The advantages of a SWATH include superior seakeeping capability in high sea states with respect to other conventional displacement hull topologies, as well as better resistance and powering characteristics relative to conventional shapes. Such advantages are at a premium due to the functioning of this vessel as a critical interface between subsurface autonomous vehicles and unmanned aerial vehicles (thus the stability requirements). In addition to operating as a control node for such a network of autonomous platforms, the SWATH AUSV also functions as a long haul freighter for transport of such a system to its theater of operation (thus the efficiency requirements). In considering this requirement, transport efficiency comes into play (Fig. 13.29).

To compare the energy efficiency of different means of transport and assess the limit of a new technology, often a metric is used that was first introduced by *Gabrielli* and *von Kármán* in the middle of the last century [13.63]. It was defined as the specific power, relative to the vehicle displacement or the payload weight, and it was plotted versus speed in a logarithmic diagram. Such a diagram is useful to identify the technological efficiency limits of different types of vehicles (terrestrial, aerial, etc.) and hence to assess their transport efficiency at a glance. Different formulations

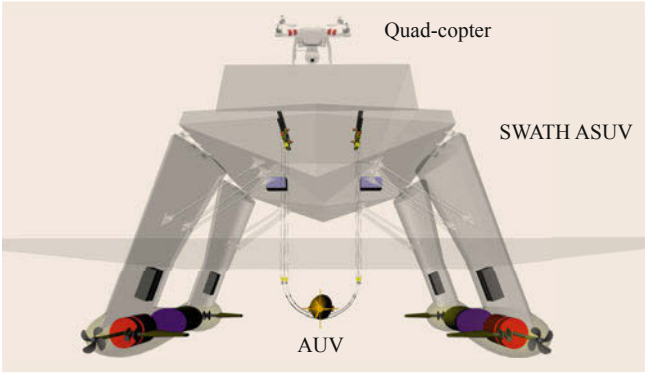


Fig. 13.29 Unconventional SWATH autonomous surface vessel as support platform of the 4-D autonomous ocean monitoring and surveillance system

of the diagram have been created since then and applied to different fast ship databases, for instance, *Papanikolaou* [13.64]. Here we refer to an updated version of the diagram, reported by *Almeter* [13.65], which is based on the most recent data for different typologies of marine crafts.

The transportation efficiency factor (TEF) in its non-dimensional form is defined as

$$\mathrm{TEF} = \Delta \frac{V}{P_\mathrm{B}} , \tag{13.3}$$

where Δ = vehicle weight, V = vehicle speed, and P_B = total brake power used to drive (propel and eventually sustain) the vessel at the given speed. All parameters are expressed in consistent units.

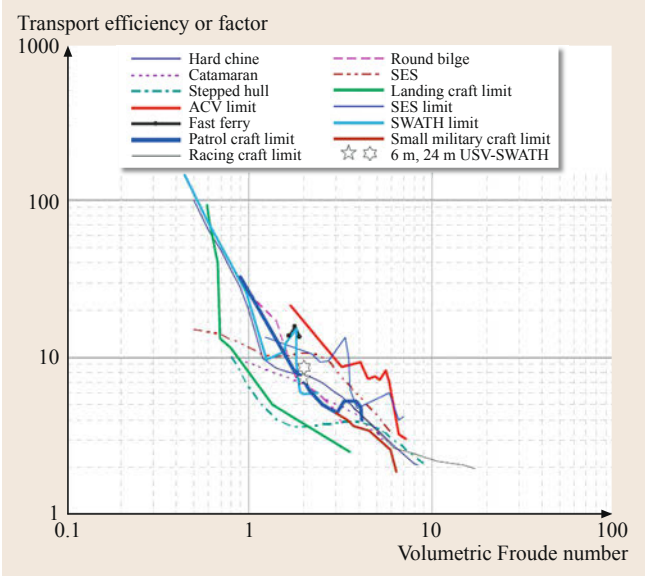


Fig. 13.30 Transport efficiency for different types of high speed crafts (after [13.65])

The transport factor is plotted versus the volumetric Froude number in the logarithmic chart of Fig. 13.30 for different technologies of high speed marine crafts. It should be kept in mind that the database of these crafts is composed of usually medium to large vessels and that the scale of the vessel has an impact (the smaller the worse) on the efficiency factor. Two different sizes of AUSV-SWATH designs – a 6 m long version briefly presented in the previous section and a larger scaled version 24 m long – are reported and compared in the figure with the other vehicle technologies. As is evident, at their design Froude number $\mathrm{Fr}_\nabla = 2.0$, the presented vessels have a higher efficiency than the conventional curve of high speed SWATH vessels, even though the considered sizes are surely smaller than the equivalent ones of the SWATH vehicles. In general, the efficiency index reduces when a ship is scaled to a smaller size according to Froude similitude; the Reynolds number, in fact, increases and with it friction resistance and form drag (relative to displacement). This is confirmed by the two different positions that the points representing the two AUSV-SWATH vehicle sizes assume on the graph of Fig. 13.29. The larger SWATH shows a higher efficiency getting closer to more sophisticated technologies such as surface effect ships. Generalizing, we can conclude that the transport efficiency attained by the optimized SWATH design is more than competitive with conventional monohull patrol crafts, catamaran, and fast planing hulls. This gap is expected to increase even further when added resistance in waves is considered. Experimental and numerical investigations are being carried out in this respect at MIT iShip lab.

However, the transport factor, although significant to assess the merit of the hull form optimization, is not the main performance goal for the autonomous surface craft, which has to be sought in the ability to perform launching and recovery missions in higher sea states than equivalent vessels. In this respect, the drastically lower RAO of pitch and heave motions of the unconventional SWATH design with respect to equivalent catamarans for a large range of incoming wave lengths, as previously presented, surely brings the particular design presented to a preferential place.

The AUSV, in turn, must interface with global scale transport infrastructure for cost-effective forward deployment. From here, the selection of the presented small size of particular unconventional SWATH design is oriented to compact (and standard) packaging requirements (ISO container size). The cost efficiency of the AUSV mainly derives from their easy and quick availability through the global transport infrastructure

network, which is nowadays based on standardized container modules. A highly portable, economically viable, persistent 4-D network based on the presented AUSV will be a major enabler for new and more extensive testing, leading to greater reliability and higher transition-to-operations potentials in many different marine offshore and coastal research fields. This revolution is not far away in time from us as it is envisaged at MIT.

References

13.1 Operation Crossroads, http://en.wikipedia.org/wiki/Operation_Crossroads (2016)
13.2 USS Langley (CV-1), http://en.wikipedia.org/wiki/USS_Langley_(CV-1) (2016)
13.3 Naval Aviation, http://en.wikipedia.org/wiki/Naval_aviation (2016)
13.4 L. Brown: *A Radar History of World War II* (Taylor Francis, New York 1999)
13.5 The Navy Unmanned Surface Vehicle (AUSV) Master Plan (PEO-LMW, July 2007), http://www.navy.mil/navydata/technology/AUSVmppr.pdf (2007)
13.6 R.R. Burgess: Follow that sub, Sea Power Mag. **56**, 28–29 (2013)
13.7 R. Brizzolara, W. Sokol, S. Littlefield, J. Corrado: Unmanned sea surface vehicle technology development, Proc. ASNE Ship Ship Syst. Technol. Symp. (2006)
13.8 Remote Multi-mission Vehicle (RMMV), http://www.navy.mil/navydata/fact_display.asp?cid=2100&tid=453&ct=2 (2016)
13.9 SAM 3 Minesweeping AUSV, http://www.kockums.se/en/products-services/naval-surface-ships/mine-countermeasures/sam-3/ (2016)
13.10 S. Brizzolara, M. Bovio, A. Federici, G. Vernengo: Hydrodynamic design of a family of hybrid SWATH unmanned surface vehicles, Proc. 11th Int. Conf. Fast Sea Transp. (2011) pp. 821–828
13.11 D.-H. Kim, A.H. Engle, A.W. Troesch: Estimates of long-term combined wave bending and whipping for two alternative hull forms, Trans. SNAME **119**, 137–166 (2011)
13.12 M. Jiang, V. Lien, D. Lesar, A. Engle, R. Lewis: A validation of various codes using hydrodynamic wedge impact data, Proc. ASME 31th Int. Conf. Offshore Mech. Arct. Eng. (2012) pp. 743–752
13.13 L. Elkins, D. Sellers, W.R. Monach: The autonomous maritime navigation (AMN) project: Field tests, autonomous and cooperative behaviors, data fusion, sensors, and vehicles, J. Field Robotics **27**(6), 790–818 (2010)
13.14 T. Huntsberger, H. Aghazarian, A. Howard, D. Trotz: Stereo vision–based navigation for autonomous surface vessels, J. Field Robotics **28**(1), 3–18 (2011)
13.15 M.T. Wolf, C. Assad, Y. Kuwata, A. Howard, H. Aghazarian, D. Zhu, T. Lu, A. Trebi-Ollennu, T. Huntsberger: 360-degree visual detection and target tracking on an autonomous surface vehicle, J. Field Robotics **27**(6), 819–833 (2010)
13.16 R. Arkin: *Behavior-Based Autonomy* (MIT, Cambridge 1998)
13.17 J.H. Michell: On the wave–resistance of a ship, Philos. Mag. **45**(5), 106–123 (1989)
13.18 W.C.S. Wigley: The theory of the bulbous bow and its practical application, Trans. North East Coast Inst. Eng. Shipbuild. **LII**, 65–88 (1936)
13.19 E.O. Tuck: The wave resistance formula of J.H. Michell and its significance to recent research in ship hydrodynamics, J. Aust. Math. Soc. B **30**, 365–377 (1989)
13.20 F. Noblesse: *Analytical Representation of Ship Waves 23rd Weinblum Memorial Lecture* (Naval Surface Warfare Center, Carderock 2000), Report NSWCCD-TR2000/011
13.21 L.J. Doctors, A.H. Day: Resistance prediction for transom-stern vessels, Proc. 4th Int. Conf. Fast Sea Transp. (FAST) (1997) pp. 743–750
13.22 C.W. Dawson: A practical computer method for solving ship-wave problems, 2nd Int. Conf. Numer. Ship Hydrodyn. (1977) pp. 30–38
13.23 D. Bruzzone: Numerical evaluation of the steady free surface waves, Proc. CFD Workshop (1994) pp. 126–134
13.24 D.E. Nakos: Ship Wave Patterns and Motions by a Three Dimensional Rankine Panel Method, Ph.D. Thesis (MIT, Cambridge 1990)
13.25 F. Xia: Numerical Calculation of Ship Flows, with Special Emphasis on the Free Surface Potential Flow, Ph.D. Thesis (Chalmers Univ. Technology, Goteborg 1986)
13.26 S.Y. Ni: Higher Order Panel Methods for Potential Flows with Linear and Nonlinear Free Surface Boundary Conditions, Ph.D. Thesis (Chalmers Univ. Technology, Goteborg 1987)
13.27 H.C. Raven: A Solution Method for the Nonlinear Ship Wave Resistance Problem, Ph.D. Thesis (Delft Univ. Technology, Delft 1986)
13.28 D. Kring: Time Domain Ship Motions by a Three-Dimensional Rankine Panel Method, Ph.D. Thesis (MIT, Cambridge 1995)
13.29 D.E. Nakos, P.D. Sclavounos: Kelvin wakes and wave resistance of cruiser and transom-stern ships, J. Ship Res. **38**(1), 9–29 (1994)
13.30 S. Brizzolara, D. Bruzzone, P. Cassella, A. Scamardella, I. Zotti: Wave resistance and wave patterns for high speed crafts. Validation of numerical results by model tests, Naval Hydrod. Conf. Proc. ONR98 (1998) pp. 69–83
13.31 S. Brizzolara, D. Bruzzone: Numerical wave resistance and dynamic trim of high speed crafts, NAV 2000 Int. Conf. Ship Shipp. Res. (2000) pp. 4.2.1–4.2.13
13.32 S. Brizzolara, D. Bruzzone: Near and distant waves of fast ships in limited and unlimited bottom

depths, 7th Int. Conf. Fast Sea Transp. (FAST) (2003) pp. H1–H13
13.33 K.W.H. Eggers, S.D. Sharma, L.W. Ward: An assessment of some experimental methods for determining the wavemaking characteristics of a ship form, Trans. SNAME **75**, 112–157 (1967)
13.34 S. Brizzolara, A. Federici: CFD modeling of planning hulls with partially ventilated bottom, William Froude Conf. Adv. Theor. Appl. Hydrodyn. Past Future (2010)
13.35 F. Stern, R.V. Wilson, H. Coleman, E. Patterson: Comprehensive approach to verification and validation of CFD simulations Part 1: Methodology and procedures, ASME J. Fluids Eng. **123**(4), 793–802 (2001)
13.36 S. Muzaferija, M. Peric: Computation of free surface flows using interface-tracking and interface-capturing methods. In: *Nonlinear Water Wave Interaction*, ed. by O. Mahrenholtz, M. Markiewicz (WIT, Southampton 1999)
13.37 L. Larsson, F. Stern, V. Bertram: Benchmarking of computational fluid dynamics for ship flows: The Gothenburg 2000 Workshop, J. Ship Res. **47**(1), 63–81 (2003)
13.38 L. Larsson, F. Stern, M. Visonneau: Numerical ship hydrodynamics – results of the Gothenburg 2010 workshop, MARINE 2011. IV International Conference on Computational Methods in Marine Engineering (Springer, Dordrecht 2013)
13.39 M. Caponnetto: Practical CFD simulations for planing hulls, Proc. 1st Euro Conf. High Perform. Mar. Veh. (2001) pp. 128–138
13.40 S. Brizzolara, D. Villa: CFD simulations of planing hulls, 7th Int. Conf. High-Perform. Mar. Veh. (HIPER) (2010)
13.41 S. Brizzolara, A. Federici: Designing of V-Stepped planing hulls: CFD in support of traditional semi-empirical methods, Design Constr. Super Mega Yachts, Genua (2013)
13.42 C. Lai, A.W. Troesch: Modelling issues related to the hydrodynamics of three dimensional steady planing, J. Ship Res. **39**(1), 1–24 (1995)
13.43 S. Brizzolara, G. Vernengo: A three-dimensional vortex method for the hydrodynamic solution of planing cambered dihedral surfaces, Eng. Analysis Bound. Elem. **63**, 15–29 (2016)
13.44 P. van Oossanen, J. Heimann, J. Henrichs, K. Hochkirch: Motor yacht hull form design for the displacement to semi-displacement speed range, 10th Int. Conf. Fast Sea Transp. (FAST) (2009) pp. 629–643
13.45 S. Brizzolara, S. Gaggero, D. Grassi, D. Villa: CFD modeling for powering and propulsion of motor yachts. Recent developments and applications of the marine CFD group, Des. Constr. Oper. Super Mega Yachts (2009) pp. 21–34
13.46 E. Campana, D. Peri, Y. Tahara, F. Stern: Shape optimization in ship hydrodynamics using computational fluid dynamics, Comput. Methods Appl. Mech. Eng. **196**, 634–651 (2006)
13.47 I. Biliotti, S. Brizzolara, M. Galliussi, A. Manfredini, D. Ruscelli, G. Vernengo, M. Viviani: Automatic parametric hull form optimization of fast naval vessels, 7th Int. Conf. High-Perform. Mar. Veh. (HIPER) (2010) pp. 294–301
13.48 I. Biliotti, S. Brizzolara, M. Viviani, G. Vernengo, D. Ruscelli, M. Galliussi, M. Manfredini: Automatic parametric hull form optimization of fast naval vessels, 11th Int. Conf. Fast Sea Transp. (FAST) (2011)
13.49 K. Nestoras: *A Tool to Create Hydrodynamically Optimized Hull-Forms with Geometrical Constraints from Internal Arrangements*, M. S. (MIT, Cambridge 2013)
13.50 S. Brizzolara: Parametric optimization of SWAT-hull forms by a viscous-inviscid free surface method driven by a differential evolution algorithm, 25th Symp. Naval Hydrodyn. (2004) pp. 47–64
13.51 G. Vernengo, S. Brizzolara, D. Bruzzone: Resistance and seakeeping optimization of a fast multi-hull passenger ferry, Int. J. Offshore Polar Eng. **25**(1), 26–34 (2015)
13.52 D. Savitsky: On the subject of high-speed monohulls, Greek Sect. Soc. Naval Archit. Mar. Eng. (2003)
13.53 S. Brizzolara: Watercraft Device, Italian Patent # GE2011A000012 (2011), US Patent 8820260 B2 (2014)
13.54 S. Brizzolara Watercraft Device, Italian Patent # GE2011A000011 (2011), US Patent 8763546 B2 (2014)
13.55 S. Brizzolara, T. Curtin, M. Bovio, G. Vernengo: Concept design and hydrodynamic optimization of an innovative SWATH AUSV by CFD methods, Ocean Dyn. **62**(2), 227–237 (2011)
13.56 S. Brizzolara, T. Curtin, M. Bovio, G. Vernengo: *Concept Design and Hydrodynamic Optimization of an Innovative AUSV by CFD methods* (NURC NATO Int. Rep., La Spezia 2011)
13.57 S. Brizzolara, G. Vernengo: Automatic optimization computational method for unconventional S.W.A.T.H. ships resistance, Int. J. Math. Models Methods Appl. Sci. **5**, 882–889 (2011)
13.58 S. Brizzolara, D. Villa: Multiphase URANS simulations of surface combatant using Star-CCM+, Proc. Workshop CFD Ship Hydrodyn. (2010)
13.59 S. Brizzolara, D. Villa: Numerical and experimental hydrodynamic performance of slice hull forms, 16th Int. Conf. Ships Shipping Res. (2009)
13.60 O.M. Faltinsen: *Hydrodynamics of High Speed Mar. Vehicles* (Cambridge Univ. Press, Cambridge 2005)
13.61 S. Brizzolara, C. Chryssostomidis: Design of an unconventional ASV for underwater vehicles recovery: Simulation of the motions for operations in rough seas, ASNE Int. Conf. Launch Recover. (2012)
13.62 S. Brizzolara, L. Bonfiglio, J. Seixas de Medeiros: Influence of viscous effects on numerical prediction of motions of SWATH vessels in waves, Int. J. Ocean Syst. Eng. **3**(3), 219–236 (2013)
13.63 G. Gabrielli, Th. von Kármán: What price speed? Specific power required for propulsion of vehicles, Mech. Eng. **72**(10), 775–781 (1950)
13.64 A. Papanikolaou: Review of advanced marine vehicles concepts, Proc. 7th Int. High Speed Mar. Veh. Conf. (HSMV) (2005)
13.65 J. Almeter: Avoiding common errors in high speed craft powering predictions, Int. Conf. Hi-Perform. Mar. Veh. (HIPER) (2008)

14. Autonomous Underwater Vehicle Navigation

John J. Leonard, Alexander Bahr

This chapter surveys the problem of navigation for autonomous underwater vehicles (AUVs). Navigation is critical for the safety and effectiveness of AUV missions. The unavailability of global positioning system (GPS) underwater makes AUV navigation a challenging research problem. Recent years have seen considerable improvements in performance and reduction in the cost and size of the various sensor devices available for ocean vehicle navigation. In concert with these developments, advances in algorithms such as simultaneous localization and mapping, and cooperative navigation have enabled dramatic improvements in the navigation capabilities of AUVs. These improvements in AUV navigation have contributed to the successful deployment of AUVs for a wide variety of applications over the past decade.

14.1 **Sensors** 343
14.1.1 Depth 343
14.1.2 Compass 343
14.1.3 Gyroscopes 343
14.1.4 Attitude Heading Reference Systems 344
14.1.5 Inertial Navigation Systems 344
14.1.6 GPS 344
14.1.7 Doppler Velocity Log (DVL) 344
14.1.8 Acoustic Ranging Methods 345
14.2 **Algorithms** 346
14.2.1 Dead-Reckoning and Inertial Navigation 346
14.2.2 Acoustic Navigation 347
14.2.3 Geophysical Map-Based Navigation 349
14.2.4 Simultaneous Localization and Mapping 350
14.2.5 Cooperative Navigation of Multiple Vehicles 351
14.3 **Summary** 352
14.3.1 Glider with Very Low Power Sensor Suite 352
14.3.2 Low-Cost AUV Sensor Suite 352
14.3.3 Standard AUV Sensor Suite 353
14.3.4 High-End AUV 353
14.3.5 Special-Task AUV Using Visual SLAM 353
14.4 **Conclusion** 353
References 354

This chapter surveys the problem of navigation for autonomous underwater vehicles (AUVs). We review the major types of sensors available for underwater navigation, and then describe some of the key techniques employed, including long baseline (LBL) navigation and simultaneous localization and mapping (SLAM) algorithms. Subsequently, we describe several example navigation systems utilized in recent AUV systems. Figure 14.1 shows some examples of the wide range of AUVs that are in use today. The selection of appropriate navigation sensors and algorithms, based on cost, power, size, and mission constraints, is a key element of AUV system design [14.1].

Navigation is an important requirement for any type of mobile robot, but this is especially true for AUVs. Good navigation information is essential not only for the safe operation and recovery of the AUV, but also for the data gathered by an AUV to be of value. For many types of AUV missions, such as seabed mapping and mine countermeasures, the quality of the data acquired by the vehicle depends critically on the accuracy of the vehicle's navigation system. Cost can be a major factor in AUV navigation system design. This is especially true for applications that involve the coordinated operation of multiple vehicles.

The absence of global positioning system (GPS) measurements underwater makes AUV navigation a difficult challenge. Without an external reference in the form of acoustic beacons at known positions, the vehicle has to rely on proprioceptive information obtained through a compass, a Doppler velocity logger (DVL) and/or an inertial navigation system (INS) to perform dead reckoning (DR). Independent of the quality of the sensors used, the error in the position estimate based on

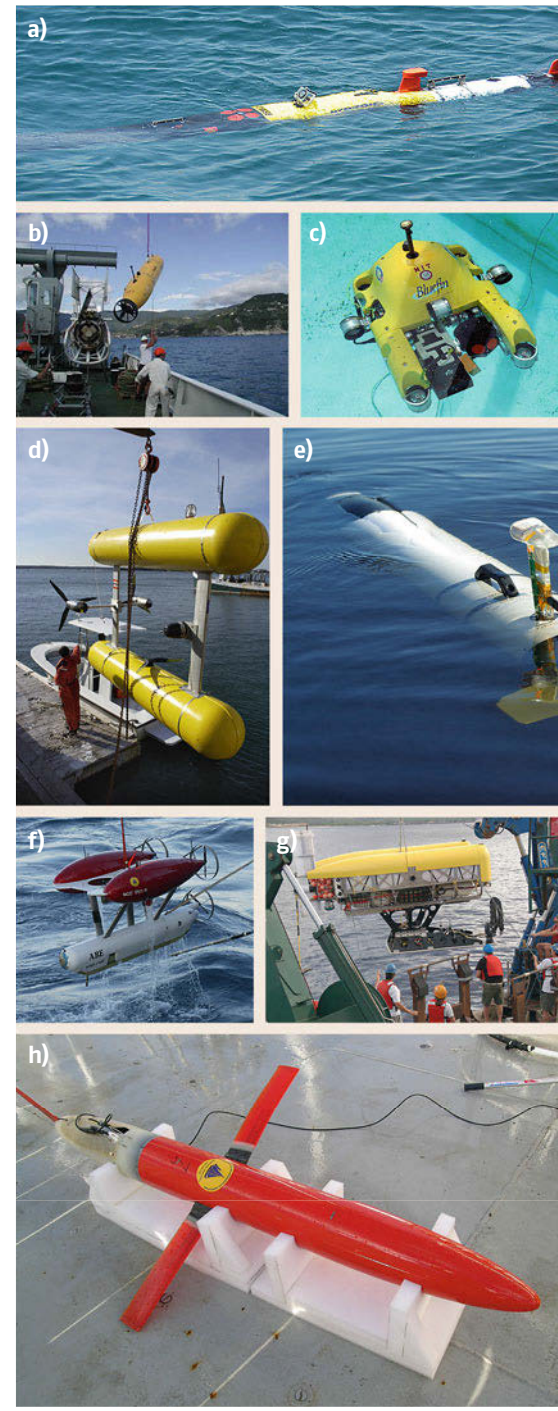


Fig. 14.1a–h The AUV systems that employ the diverse range of navigation sensors and algorithms described in this chapter include the REMUS [14.2], Bluefin 21 [14.3], HAUV [14.4], Seabed [14.5], Iver2 [14.6], ABE [14.7], and Nereus [14.8], AUVs and the Spray glider [14.9]. Photos **(a–c, e)** by MIT; **(f,h)** by WHOI; **(d)** Tom Kleindinst, WHOI; **(g)** Christopher Griner, WHOI ►

DR information grows without bound. Typical navigation errors are 0.5–2% of distance traveled for vehicles traveling within a few hundred meters of the sea floor such that their DVL has a lock on the bottom. Errors as low as 0.1% can be obtained with large and expensive INS systems, but for vehicles relying only on a compass and a speed estimate, the rate of error can be as high as 10%. By surfacing the AUV can obtain a position update through its GPS, but this is impossible or undesirable for many applications. The use of static beacons in the form of an LBL array limits the operation area to a few km^2 and requires a substantial deployment effort before operations, especially in deep water.

As underwater vehicles become more reliable and affordable, the simultaneous use of several AUVs has become viable, and it is anticipated that multivehicle deployments will become standard in the upcoming years. This will not only make entirely new types of missions which rely on cooperation possible, but will also allow each individual member of the group to benefit from the navigation information obtained from other members. For optimal cooperative localization, a few dedicated Navigation Aid-AUVs (NAs), which maintain an accurate estimate of their positions through sophisticated DVL and INS sensors, can enable a much larger group of vehicles with less sophisticated sensor suites to maintain an accurate position.

Navigation (along with communications, power, and autonomy) has been one of the fundamental challenges in the development of AUV technology. Underwater vehicle navigation is a challenging problem for several reasons. Due to the absorption of electromagnetic radiation in the ocean, GPS is only available at the surface. Unfortunately, there is no *silver bullet* solution for the AUV navigation problem. There are five primary technologies that an AUV designer has to draw upon in selecting an AUV navigation system:

1. Proprioceptive sensing
2. GPS
3. Acoustic transponder navigation
4. Map-based navigation
5. Cooperative navigation of multiple vehicles.

Proprioceptive navigation refers to using measurements of the vehicle's self-motion to deduce the vehicle's position. There are two major categories, based on price: (a) INS combined with DVL, and (b) magnetic compass/attitude heading reference systems. Integrated INS/DVL systems typically cost up to $100 000 and have been integrated on many large-scale, high-cost AUV systems. In the past two decades, the cost and

size of proprioceptive navigation sensors have reduced dramatically, making them available to smaller and lower-cost AUV systems. The performance claims of better than 0.1% of distance traveled have been made in the literature [14.10].

Regardless of sensor cost and quality, the problem with exclusive reliance on proprioceptive sensing is that the position error increases without bound as the distance traveled by the vehicle increases. The rate of increase will be a function of ocean currents, the vehicle speed, and the quality of dead reckoning sensors. If a vehicle can surface, then GPS can be used for a position fix. Indeed, many AUVs have demonstrated this capability. However, frequent surfacing is impractical for deep-water missions and is undesirable for other types of missions.

In acoustic navigation, transponders serve as *beacons* to constrain INS/DR error growth without the need for resurfacing. Two types of systems have been primarily employed [14.11–13]: LBL and ultra-short baseline (USBL). Both systems employ external transducers or transducer arrays as aids to navigation. Acoustic navigation is a well-established and widely used technique. However, in littoral waters, a number of sources of error occur. These include: multipath, drop-outs, fading, and reverberation. Further development is required to attain better outlier rejection and improve robustness for autonomous operations [14.14]. A longer term goal is to fully integrate navigation, communications, and tomography [14.15].

For many applications, deployment of acoustic beacons is undesirable, and map-based navigation presents an alternative. If an accurate a priori map of the environment is available, one approach to globally-referenced position estimation is to use measurements of geophysical parameters, such as bathymetry, magnetic field, or gravitational anomalies [14.16, 17]. In practice, an up-to-date, high-quality map may be unavailable in the operating area of interest. This motivates research into the problem of SLAM, in which the goal is for the AUV to build a map of its environment and to use that map to navigate in real time.

The structure of this chapter is as follows. Section 14.1 describes available sensors for measuring depth, heading, velocity, acceleration, acoustic range and bearing, and GPS. Section 14.2 describes the main navigation algorithms including DR and inertial navigation, LBL and USBL navigation, geophysical and map-based navigation, and cooperative localization. Section 14.3 summarizes the performance of several typical types of AUV systems in use today, and provides an outlook for future developments in this area.

14.1 Sensors

14.1.1 Depth

All submersible vehicles are outfitted with a pressure sensor [14.18] that allows an accurate determination of absolute depth using the known properties of sea water [14.19]. As a result, all other underwater navigation systems are only used to resolve the two-dimensional (2-D) position, (i. e., longitude and latitude) and underwater-vehicle-related localization problems are typically stated in 2-D. Quartz crystal pressure sensors can typically attain accuracies of 0.01% or better, but achieving full accuracy requires careful calibration and thermal compensation [14.1].

14.1.2 Compass

Like the pressure sensor, a compass is part of the basic navigation sensor suite of most underwater vehicles, as it is an inexpensive and low-power device. It provides the three-dimensional (3-D) vector of the local magnetic field. Before computing the heading from the magnetic field vector, it is necessary to carefully calibrate the compass each time the vehicle's area of operation changes, as the difference between the orientation of the 3-D magnetic field vector and the direction of true north (called *variation*) depends on the geographic location. In addition to the spatially slow variation, there are highly localized *magnetic anomalies*. The compass output is also affected by its position in the vehicle as electrical currents create magnetic fields which cannot be discerned from the earth's magnetic field. On-line compass calibration algorithms [14.3, 20] can substantially improve performance.

14.1.3 Gyroscopes

Gyroscopic systems measure changes in vehicle orientation by exploiting physical laws that have predictable effects under rotation [14.21]. Available sensors include mechanical gyrocompasses, ring laser gyroscope (RLG), fiber optic gyroscope (FOG), dynamically tuned gyroscope (DTG) and micro-electromechanical (MEMS) devices [14.22]. Optical gyroscopes have become popular as an accurate angular rate sensor, and operate on the SAGNAC principle [14.23]. However, sensible DR navigation requires high gyroscope performance increasing the associated costs significantly.

The power consumption and size of common optical gyroscopes have limited their use to larger and more expensive AUVs [14.1].

14.1.4 Attitude Heading Reference Systems

An attitude-heading reference system (AHRS) unit typically consists of a 3-axis gyroscope as well as a 3-axis linear acceleration sensor and a heading sensor (magnetic or gyrocompass). Combining the measurements from these sensors, the AHRS maintains a common algorithmic orientation estimate between the sensor cluster and the global world navigation frame. The orientation is maintained through integration of the gyroscope measured rotation rates. Long term pitch and roll stability of the orientation estimate is achieved through selective use of earth's gravity vector, which is measured by a triad of accelerometers. The long-term heading accuracy is maintained from either magnetic compass sensor readings or direct measurement of the earth rotation through the gyroscopes.

14.1.5 Inertial Navigation Systems

The sensors of an INS are the same as those of the AHRS described earlier. In addition to the AHRS, measured accelerations are rotated to the common navigation frame, as discussed above, and then double integrated to compute a position estimate for the vehicle. The initial position of the INS is obtained from an absolute position sensor (such as GPS or LBL). The process of inertial navigation is discussed in more detail in Sect. 14.2.1.

14.1.6 GPS

Almost all underwater vehicles today are equipped with a GPS receiver as it can be used to get a position fix before the start of the mission or during intermittent surfacings. GPS measurements obtained at different points during a mission can be used to constrain the error growth of position estimates derived from inertial and Doppler sensors [14.3]. Farrell provides a comprehensive summary of the mathematical techniques for integrating GPS measurements with high-rate inertial and acoustic sensors [14.24].

Fig. 14.2 (a) Doppler velocity log (DVL); (b) REMUS AUV with DVL

14.1.7 Doppler Velocity Log (DVL)

A DVL (Fig. 14.2) is a device which typically has four transceiver units that emit acoustic pulses. When a DVL is used for navigation purposes, it is usually mounted on a vehicle such that the transceivers are facing downward. If the DVL is close enough to the bottom, the transceiver will receive the reflected pulses (*bottom lock*) and as the transceivers are mounted at an angle with respect to the sea-floor plane, the received pulses will be subject to a Doppler shift if the vehicle is moving. Combining the measured Doppler shifts from all four sensors with the built-in roll, pitch, and heading sensors the DVL can then compute the vehicle's 3-D-speed vector $\boldsymbol{v}_v = (\dot{x}, \dot{y}, \dot{z})$ in a world-referenced frame.

The maximum distance between the DVL unit and the sea floor depends on the operating frequency of the transceivers. A low-frequency (150 kHz) DVL can obtain bottom-lock for ranges up to 500 m, while a high-frequency DVL (1200 kHz) can obtain bottom-lock up to 30 m.

The ranges indicated above can only be obtained under ideal conditions. A soft sea floor or vegetation can absorb most of the energy of the incoming pulse and thereby significantly decrease the maximum range. Another option is to mount the DVL in an upward-looking configuration such that the acoustic pulses are reflected at the water/air interface (*surface-lock*). Then, the vehicle measures its speed relative to the water surface, but this strategy may introduce errors in the case of significant surface currents. Figure 14.2 shows a REMUS 100-AUV with a double-DVL configuration. If bottom-lock cannot be obtained with the downward-looking DVL the vehicle tries to determine its speed using the upward looking unit. Recent developments greatly increased the accuracy of DVL-systems and errors as low as 0.2% (1200 kHz) or 1% (150 kHz) of distance traveled can be obtained.

In deeper waters, where the distance from the seabed is beyond the range of a DVL, an alternative is to use a correlation speed log sonar [14.25]. Whereas a DVL employs four angled beams, a correlation sonar utilizes a single widebeam transducer, pointing straight downward toward the seabed. By cross-correlating the received waveforms on two closely separated transducers, the displacement of the crosscorrelation peak provides a measure of the vehicle velocity. Correlation sonar navigation has been demonstrated by *Griffiths* and *Bradley* for performing long-distance under-ice excursions in deep water in polar regions with the Autosub AUV, achieving ranges greater than 1000 m [14.26].

14.1.8 Acoustic Ranging Methods

The most commonly used way to obtain absolute position information underwater is through the use of beacons. These beacons are at known locations and the AUV obtains the range and/or bearing to several of these and then calculates its position through trilateration or triangulation. Based on the location of the transceivers, we can identify three different baseline systems.

Electromagnetic energy cannot propagate appreciable distances in the ocean except at very low frequencies. Acoustic energy, however, propagates well in the ocean, and hence acoustic transponders can be used as *beacons* to guide the motion of an AUV without the need for resurfacing. Two types of system have been primarily employed [14.11–13]: LBL and USBL. Both systems employ external transducers or transducer arrays as aids to navigation.

Standard LBL

A typical LBL-configuration is shown in Fig. 14.3a. Two or more beacons are deployed around the perimeter of the area in which the AUV will operate. These beacons are anchored and float on the surface or, particularly in deeper water, a few meters above the sea floor. Each unit listens to acoustic query pings on a common receive channel. After receiving a query ping from an AUV, each unit waits for a unique turn around time (TAT) t^{TAT} and then sends out a reply ping on its individual transmit channel. The AUV then receives the reply pings. The transmit channel as well as the TAT are different for each unit. A unique TAT ensures that two beacons will not interfere by transmitting at the same time and by using different transmit frequencies the beacons provide a way for the AUV to identify from which unit a reply ping was sent. The time difference Δt_i between sending out the query ping and receiving a reply can then be used to determine the one-way travel-time (OWTT) t_i^{owtt}.

$$t_i^{\mathrm{owtt}} = \frac{\Delta t_i - t_i^{\mathrm{TAT}}}{2}$$

The distance d_i between a beacon i and the AUV is then given by

$$d_i = \frac{c}{t_i^{\mathrm{owtt}}} ,$$

where the speed of sound c is either a pre-programmed value or measured on-board. Using range measurements to several beacons and the beacon positions stored in the vehicle before deployment, the AUV can now trilaterate its position.

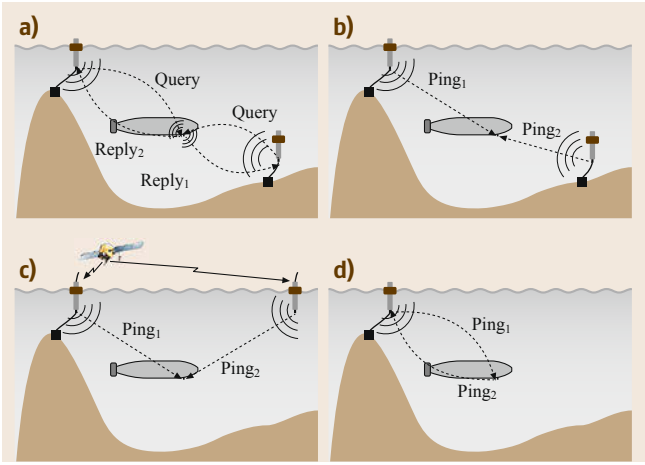


Fig. 14.3a–d Beacon-based underwater localization techniques. **(a)** Conventional (two-way) Long Baseline (LBL) navigation; **(b)** time-synchronized (one-way) LBL; **(c)** GPS-buoy navigation [14.27]; **(d)** ultra-short baseline (USBL) navigation

The maximum possible distance between the AUV and a beacon as well as the localization accuracy depends on the the frequency band used for query and reply pings. Long-range LBL-systems using the 12 kHz band work over distances as long as 10 km [14.28] and can provide an absolute position with an error between 1 and 10 m. Short-range LBL systems using frequencies up to 300 kHz band can achieve subcentimeter precision, but the maximum range is limited to 100 m [14.28]. The indicated errors assume that large outliers have been filtered out. These outliers, which can be seen in Fig. 14.4, are due to multipath and other acoustic propagation effects.

LBL Variants

Standard LBL systems such as the one described earlier are not well suited for large groups of AUVs because only one vehicle at a time can query the beacon network and get a position update. Thus, the position update interval increases with the number of vehicles. Newer LBL systems, like the one developed by ACSA [14.27, 30] and shown in Fig. 14.3b, have synchronized clocks in the beacons and the AUV transceiver units. The beacons broadcast a ping containing a unique identifier at fixed time intervals. When the AUV receives this ping, the beacon's known broadcast schedule and the synchronized clock's time ensure that the vehicle knows when a ping was sent and can directly compute the OWTT. The synchronized clocks thereby eliminate the need for query pings and allow all vehicles within the range of the beacons to get a range to the broadcasting beacon. As a result, the ping interval is independent of the number of vehicles relying on the beacon network.

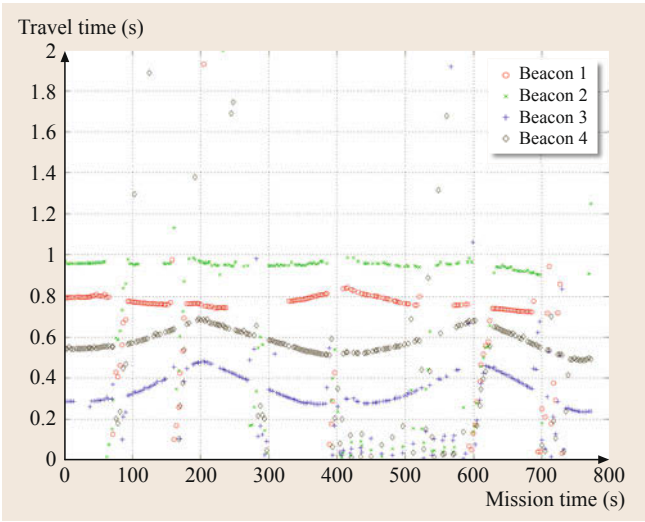


Fig. 14.4 Time-of-flight measurements obtained from four LBL beacons at the GOATS 2002 experiment (after [14.29]). The plot shows significant outliers for all beacons, particularly between 400 s and 600 s. This data illustrates the importance of techniques for outlier rejection in autonomous operations (after [14.7, 14])

Another improvement over conventional LBL is the system depicted in Fig. 14.3c. Building on the setup in Fig. 14.3b, beacons now transmit their GPS position along with the unique identifier. As with the system described previously, the vehicles do not need to query the beacons. With the position of the beacons embedded in the ping the beacons can float freely and it is not necessary to store their coordinates in the AUV before deployment.

USBL

Another variant of beacon-based navigation systems is USBL (Fig. 14.3d). Here the beacon is of the same kind as in a standard LBL system, but the transceiver on the AUV contains several receiving elements which are very close to each other. After querying the beacons, the reply ping is captured by all receiving elements. The phase difference between the signals coming from the different receiving elements allows the AUV to compute a bearing to the beacon. Combined with the beacon position stored in the AUV and the distance d obtained from the OWTT, the vehicle can compute its absolute position using only a reply from a single beacon.

Modern beacon-based systems, such as the ones shown in Fig. 14.3, significantly decrease the predeployment effort when compared to early beacon-based systems such as the standard LBL. However, all beacon-based systems confine the operating area of the vehicles to a polygon of beacons or, as in the case of USBL, to the coverage radius of a single beacon. Thus, beacon-based navigation is only feasible for operating areas of $\mathcal{O}$ (10 km^2) in size.

14.2 Algorithms

We now review some of the basic algorithms employed in AUV navigation. These are divided into:

1. Dead-reckoning and inertial navigation
2. Acoustic navigation
3. Map-based navigation
4. SLAM.

14.2.1 Dead-Reckoning and Inertial Navigation

The most obvious and the longest established navigation technique is to integrate the vehicle velocity in time to obtain new position estimates [14.31, 32]. This process is called DR. For low-cost vehicles, the measurement of the velocity components of the vehicle is usually accomplished with a compass and a water speed sensor. The principal problem is that the presence of an ocean current will add a velocity component to the vehicle which is not detected by the speed sensor. In the vicinity of the shore, ocean currents can exceed 2 km. Consequently, DR for power-limited AUVs, operating at small speeds (3–6 km), integrating water-relative speed measurements can generate extremely poor position estimates.

In inertial navigation, rotation rate measurements from gyroscopes are integrated to estimate the vehicle attitude, and measurements from accelerometers are integrated twice in time to compute the change in the vehicle position from a known initial location [14.33]. Inertial navigation is a widely studied field with a fascinating history [14.22]; *Titterton* provides a comprehensive description of inertial sensors and algorithms [14.34].

Position drift rates for current high-quality commercial grade INS units are of the order of several kilometers per hour. Initialization of the INS system for marine systems can be difficult. Cost and power consumption have historically made INS systems unattractive for small AUVs; however, this may change as systems get smaller and cheaper in the future.

As the linear and angular acceleration sensors are subject to noise, the position derived from these sensors in the absence of GPS or LBL is subject to a cumulative error and the obtained position will drift with respect to

the true position. The drift (error) e between the vehicle's true position x_{true} and the position obtained with DR x_{DR} are expressed as *drift over time* or *drift over distance traveled*

$$e = \frac{\|x_{\text{true}} - x_{\text{DR}}\|_2}{\Delta t} \quad \text{or} \quad e = \frac{\|x_{\text{true}} - x_{\text{DR}}\|_2}{\Delta x} .$$

Typically the heading and rate sensors of an INS are less noisy than those of a comparably inexpensive AHRS, which reduces the effect of accumulated drift. An INS that fits into the hull of a typical AUV shows typical drift rates of 1 km/h [14.35]. The exact performance of the most precise INS available are those developed for nuclear submarines; the drift rates for these sensors are not published but are expected to be $\mathcal{O}$ (0.01 km/h) [14.22].

For operations near the seabed, DVL sensors can be used to measure the vehicle's velocity relative to the ground. The integration of this information in the navigation Kalman filter can greatly improve the performance. For example, the DARPA Autonomous Minehunting and Mapping UUV developed in the early 1990s achieved a navigation performance of 0.01% of distance traveled using an integrated INS/DVL system [14.10]. The MARPOS system developed by Mariden A/S and the Technical University of Denmark for the MARIDAN series of AUVs represents a recent state-of-the-art commercially available AUV system, achieving 0.02% accuracy for site surveys and 0.10% accuracy for straight-line transits [14.36].

The problem with exclusive reliance on DR or inertial navigation is that the position error increases without bound as the distance traveled by the vehicle increases. The rate of increase will be a function of ocean currents, the vehicle speed, and the quality of DR sensors. Radio and satellite navigation systems can provide an accurate position update provided the vehicle can travel at or near the surface periodically for a position fix. The maximum vehicle travel time between surfacing for a position update will be governed by DR/inertial navigation accuracy. Poor quality DR will dictate an unacceptably high frequency of surfacing. Also, vehicles operating close to the coast are in appreciable danger of collision with surface vessels if they need to frequently approach the surface for position fixes. For deep water applications, the time and energy needed by a small AUV for transiting to the surface from near the bottom are very unfavorable. Finally, surfacing is impossible in ice-covered oceans.

14.2.2 Acoustic Navigation

When acoustic transponder measurements are available, a variety of algorithms are possible. Most techniques are based on recursive least-squares estimation, often with an extended Kalman filter. As described above, in an LBL navigation system, an array of transponders is deployed and surveyed into position. The array is usually calibrated through use of an additional acoustic transponder that is hung from a surface ship and interrogates the array from various locations.

The vehicle sends out an acoustic signal which is then returned by each beacon as it is received. Position is determined by measuring the travel time between the vehicle and each beacon, measuring or assuming the local sound speed profile, and knowing the geometry of the beacon array. With this information, the relative distances between the vehicle and each array node can be calculated. The two primary techniques are (1) to compute position fixes by locating the intersection point of spheres of appropriate radii from the beacons in the array, and (2) to integrate the raw time-of-flight (TOF) measurements into an appropriate Kalman filter.

The basic solution for computing a fix from spherical ranges to three transponders is as follows [14.14]. The local earth frame origin is at the surface. The x-axis points north, the y-axis points east, and the z-axis points down. The vehicle position in this frame is (x, y, z) and the coordinates of beacon i are (x_i, y_i, z_i). The measured round-trip travel times between the vehicle and the beacons are t_i and the associated distances are d_i. The distances are computed from the travel times by the approximate relation: $d_i = c(t_i - \tau_i)/2$, where c is the average speed of sound (Note: spatial variations in the speed of sound can be significant, especially variations with depth [14.11]).

The analytical solution based on measurements from three beacons consists of solving the following nonlinear system of equations for the vehicle coordinates

$$(x-x_i)^2 + (y-y_i)^2 + (z-z_i)^2 = d_i^2, \quad i = 1, 2, 3 . \tag{14.1}$$

In order to make the computations easier, we first compute the solution in an intermediate frame obtained by shifting the local earth frame at the location of beacon 1. The beacon coordinates in this frame are (x_i', y_i', z_i') and the vehicle position is (x', y', z'). The intermediate solution is then transformed back to the local earth frame. Advantage is also taken of the accurate knowledge of the vehicle depth, leading to a linear over-constrained problem (three equations known and two unknowns). The vehicle position is then

$$x = \frac{b_1 c_2 - b_2 c_1}{a_1 b_2 - a_2 b_1} + x_1 , y \quad = \frac{a_2 c_1 - c_2 a_1}{a_1 b_2 - a_2 b_1} + y_1 , \tag{14.2}$$

with

$$
\begin{aligned}
a_1 &= -2x_2' \,, \\
b_1 &= -2y_2' \,, \\
c_1 &= x_2'^2 + y_2'^2 + z_2'^2 - 2z'z_2' - d_2^2 + d_1^2 \,, \\
a_2 &= -2x_3' \,, \\
b_2 &= -2y_3' \,, \\
c_2 &= x_3'^2 + y_3'^2 + z_3'^2 - 2z'z_3' - d_3^2 + d_1^2 \,.
\end{aligned}
$$

When measurements from four beacons are available, least-squares minimization can be performed [14.37].

In more difficult acoustic environments, such as in shallow water or in the Arctic, it becomes difficult to distinguish between the direct arrival and multipath interference, and the rejection of outliers becomes a key issue, as shown in Fig. 14.4. In a fix computation scheme, one can rule out spurious fixes, whereas in a Kalman filter-based system, one can gate the raw TOF values [14.14].

A variant of this system is *hyperbolic navigation*, in which the vehicle does not actively ping but instead listens to an array of beacons whose geometry is known [14.38]. Each beacon pings in a specific sequence relative to the others at its specified frequency. By knowing which beacon pings when and the geometry of the array, the vehicle can reconstruct where it must be in space in order to hear the ping sequence as recorded. This system has the advantage of saving the vehicle the power expenditure of active pinging, and is especially useful for multiple AUV operations. With an active (spherical) LBL system, multiple AUV operations require careful sequencing of the pings between vehicles [14.39].

Most LBL systems work at a frequency of $\approx 10\,\mathrm{kHz}$ and provide position accuracy to within a few meters with a maximum range of the order of a few kilometers. Alternative systems operating at 300 kHz have been created that can provide positioning repeatability down to 1 cm resolution in a triangular operating area that is 100 m to a side [14.40].

In USBL navigation, the vehicle has a multielement receiver array that enables it to measure the angle as well as the range to an acoustic beacon. This system is a variant of a popular system for tracking an underwater vehicle from a surface ship [14.41]. By measuring the arrival time (phase) difference of a single sonar ping between two or more hydrophones, the bearing from the vehicle to the beacon can be determined. If the beacon responds to vehicle interrogation, then the time delay (and hence distance, as with an LBL array) can be calculated. Knowing the distance and direction to the beacon allows for local navigation [14.42]. Knowing the latitude/longitude of the beacon allows for geodetic navigation. This type of system is especially effective for homing and docking operations, which are important for Arctic [14.43] and autonomous ocean sampling network [14.44] deployments.

Errors in both LBL and USBL arrays come from many sources. The key sources of error can be broken down into two primary categories: errors in the assumed array geometry and errors in the assumed sound speed profile. Positioning error comes from inadequately or improperly surveying the relative and/or geodetic positions of the array beacons. In the event that only local navigation is desired, then only relative beacon positions are relevant. If the navigation is to be geodetic-referenced, then the beacons must be located globally as well. Sophisticated software packages are available for accomplishing this. Self-calibrating beacons simplify the task by reducing the surveying task to only one beacon with the others determining their own positions relative to the first. However, this raises the possibility of relative position errors due to errors in the assumed local sound speed.

A significant difficulty in acoustic navigation can be caused by an error in the assumed sound speed profile. An inaccurate sound speed profile will appear as a distance bias in the calculations. Reflection or multipath errors will result in incorrect TOF values and hence erroneous position fixes. Typically, LBL works well in deep water and with array separations of a few kilometers. Over longer distances in shallower water, more complex propagation effects come into play and increase the frequency of bad position fixes. If the topography is sufficiently severe, beacons may be occluded by rocks or other seabed formations. Even if the sound speed profile is known at the start of an AUV mission, the acoustic propagation environment can change during the mission [14.45].

Acoustic tomography refers to the goal of using travel time information between one or more vehicles and vertical hydrophone arrays to estimate the sound speed profile and other information at various places in the intervening water column [14.46]. These techniques are being investigated from scales of a few kilometers [14.47] to the global scale [14.48]. To perform moving source acoustic tomography with an AUV, one needs to know the location of the vehicle to a high precision. In addition, one needs to be able to identify the different propagation paths of the different arrivals received by each hydrophone [14.46]. This same information (effectively, knowing the channel impulse response of the environment) is vital for providing effective acoustic communications [14.49]. Hence, the problems of acoustic tomography, communication, and navigation are closely intertwined. An interesting idea

for future research may be to employ techniques from acoustic time reversal [14.50] to achieve higher accuracy in range estimation while concurrently estimating ocean acoustic propagation conditions.

14.2.3 Geophysical Map-Based Navigation

For some applications of AUVs, the use of acoustic beacons is undesirable or impractical. If an accurate a priori map of the environment is available, one approach to globally-referenced position estimation is to use measurements of geophysical parameters, such as bathymetry, magnetic field, or gravitational anomaly [14.16, 17, 51, 52]. These approaches are based on matching sensor data with an a priori environment map, under the assumption that there is sufficient spatial variation in the parameter(s) being measured to permit accurate localization.

All forms of map-based navigation are motivated by the desire to operate at an arbitrary location without the additional expense or problems associated with the installation of artificial beacons. In principle, the process appears straightforward: gather information about the surrounding terrain and match that information to an on-board map or database of terrain information. When the vehicle has a match to the database, then it knows its location on the map. This is analogous to the method which humans use to navigate; we find our way to our destination by locating and identifying landmarks which are familiar to us – either from past experience or via a map which has been constructed for our benefit.

In practice this form of navigation is not so simple. The vehicle is attempting to navigate by matching a set of sensed data with an a priori map or dataset of stored data. Two key problems are the cost and difficulty of generating the a priori maps and the computational complexity of searching for a peak in the n-dimensional correlation surface, where n is the number of dimensions in the map or sensor data set. Typically, map making expense is governed by both the type of data being collected and the desired resolution of that data. Determining the map resolution has a direct effect on the size and level of detail of the search needed to locate the vehicle in space. Since the vehicle could be in any of a large number of possible orientations relative to the original dataset, the search must be performed over all possible locations and orientations. This is a potentially large search space, necessitating some simplifications and/or simplifying assumptions in order to make the search more tractable. Typical simplifications are: restricting the types of map data stored (what sensor values, how many different sensors), lowering map resolution, *patchy* maps (maps of key areas only), restricting vehicle orientations (to reduce the correlation problem), and using inertial navigation or dead reckoning systems to limit the valid search area.

Evidence exists that geomagnetic navigation is employed by birds, fish, and other animals for migration and general navigation [14.53]. The magnetic flux density of the earth varies according to latitude, the presence of man-made and natural anomalies, and even one's depth in the ocean, increasing from 6 to 30 nT/km depth, depending on location [14.54]. Additionally, there are small but predictable variations in the earth's magnetic flux from day to night, and large arbitrary changes during magnetic storms; magnetic maps can be rendered useless for the duration of such storms. Useful magnetic maps, generated by satellites or surface ships, can be employed by underwater vehicles by accounting for the daily field variations and by calculating the effective magnetic field at depth using a Laplace field equation, setting the boundary conditions at the ocean surface [14.55].

Research into the nature of the earth's gravitational field has demonstrated that it is far from uniform and indeed possesses a varied topography [14.56, 57]. These variations are due to a variety of factors, especially the effects of local topography [14.58] and density inhomogeneities [14.59]. Variations in the earth's gravitational field on the ocean's surface relative to a regular ellipsoidal model have been measured to be on the order of 30–50 mgal [14.60]. Gravity maps were originally gathered on behalf of the US Navy for the purposes of INS calibration [14.61]. To an INS, the effects of a change in the local gravitational field are indistinguishable from accelerations of the vehicle itself. *Gerber* [14.62] proposed the use of a gravity gradiometer as an aid to INS. *Jircitano* et al. extended this idea to the AUV community, performing navigation simulations using a model of the Bell Aerospace Textron Gravity Gradiometer System [14.60] with good preliminary results. The drawbacks to such a system are the size, expense, and complexity of a gradiometer. In addition, the gradiometer must be mounted on an inertially stabilized and vibrationally isolated platform, making its use difficult on small, low-cost scientific AUVs.

Geophysical navigation algorithms have origins in techniques of navigating at sea using depth soundings that have been in use for centuries [14.31]. *Kamgar-Parsi* has developed techniques for performing geophysical navigation that are based on fitting contour lines to sensor data and matching these curves to an a priori map using matching techniques from computer vision [14.63, 64]. *Lucido* et al. have also investigated the segmentation and registration of bathymetric profiles [14.65]. *Tuohy* et al. have investigated geophysical navigation using maps of multiple geophysical parameters based on contour intersection methods [14.17]. The

reliability of any of these approaches will depend on the accuracy of the a priori map.

AUV navigation based on bathymetric data has been successfully achieved by *Bergem* [14.66]. In this system, depths are measured at different angles using a multibeam sonar. This gives an accurate profile of the sea floor, and the absolute position is determined by matching this profile against an a priori known detailed bathymetric map of the actual area. This idea is motivated by the successful employment of this technique to missile guidance systems.

14.2.4 Simultaneous Localization and Mapping

In practice, an up-to-date, high-quality map may be unavailable in the operating area of interest. This motivates research into the use of SLAM to enable an AUV to build a map of its environment while concurrently using that map to navigate in real time [14.67]. SLAM is a probabilistic estimation problem in which noisy sensor measurements are combined into a probabilistic representation of the state of the sensor and the observed surroundings. This model needs to be updated and extended sequentially by integrating new sensor measurements as they become available.

The variables to be estimated and the sensor measurements can be described in a graph [14.68] that captures their relation. We use a factor graph [14.69] as a general graphical model representation of the SLAM problem. Formally, a factor graph is a bipartite graph $G = (\mathcal{F}, \Theta, \mathcal{E})$ with two node types: *factor nodes* $f_i \in \mathcal{F}$ and *variable nodes* $\theta_j \in \Theta$. Edges $e_{ij} \in \mathcal{E}$ are always between factor nodes and variable nodes. A factor graph G defines the factorization of a function $f(\Theta)$ as

$$f(\Theta) = \prod_i f_i(\Theta_i) \,, \tag{14.3}$$

where Θ_i is the set of variables θ_j adjacent to the factor f_i, and independent relationships are encoded by the edges e_{ij}: each factor f_i is a function of the variables in Θ_i. Our goal is to find the variable assignment Θ^* that maximizes (14.3)

$$\Theta^* = \arg\max_{\Theta} f(\Theta) \,. \tag{14.4}$$

The general factor graph formulation of the full SLAM problem is shown in Fig. 14.5, where the landmark measurements m, loop closing constraints c and odometry measurements u are the examples of factors. Note that the factor graph formulation supports general probability distributions or cost functions of any number of variables, allowing, for example, the inclusion of calibration parameters.

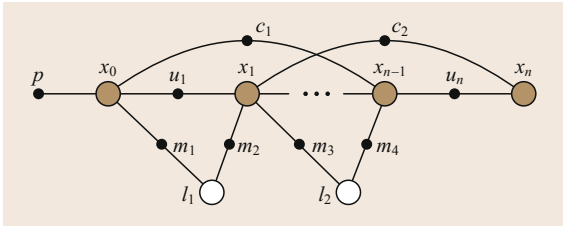


Fig. 14.5 Factor graph for the pose graph formulation of the SLAM problem

It is standard in the SLAM literature [14.70, 71] to assume Gaussian measurement noise models

$$f_i(\Theta_i) = \mathcal{N}(h_i(\Theta_i); z_i, \Sigma_i) \propto \exp\left(-\frac{1}{2}\,\|h_i(\Theta_i) - z_i\|^2_{\Sigma_i}\right) \tag{14.5}$$

because the factored objective function to maximize (14.4) reduces to a nonlinear least-squares problem

$$\arg\min_{\Theta}\,(-\log f(\Theta)) = \arg\min_{\Theta} \frac{1}{2}\sum_i \|h_i(\Theta_i) - z_i\|^2_{\Sigma_i} \tag{14.6}$$

for which efficient solutions are available. Here, $h_i(\Theta_i)$ is a measurement function and z_i a measurement, and $\|e\|^2_{\Sigma} \triangleq e^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}e$ is the squared Mahalanobis distance with the covariance matrix $\boldsymbol{\Sigma}$.

Recently, least-squares solutions to the full SLAM problem dominate, but many other solutions have been proposed in the past, mostly aimed at making computational complexity manageable. For example, Kalman filter solutions to SLAM achieve greater computational efficiency by keeping only the most recent robot pose (through marginalization of previous poses). However, this approach has been shown to be inconsistent [14.72] when applied to the inherently nonlinear SLAM problem, i. e., the estimate is biased and the error covariance is smaller than the actual one. State-of-the-art approaches achieve better performance by estimating the entire robot trajectory, which is known as full SLAM. The first solution to the full SLAM problem has been presented in [14.68] and a range of iterative least-squares solvers have been applied including relaxation [14.73], gradient descent [14.74], conjugate gradient [14.75], multilevel relaxation [14.76], and loopy belief propagation [14.77]. Exploiting the sparsity of the information form [14.78], Gauss–Newton type solvers using sparse matrix factorization have since become very popular [14.37, 79–81] because of their faster convergence rate. Full SLAM incurs a computational burden that grows in time with the duration

of the mission. Strategies to mitigate this are discussed in *Johannsson* et al. [14.82].

SLAM techniques have been used successfully on a wide range of AUVs, using sonar and image measurements. A key requirement for SLAM is to be able to extract and match features from measurements obtained from different vantage points, to extract constraints that can be used to constrain error growth. Figures 14.6 and 14.7 show an illustration of the constraint matching process for features extracted from consecutive camera images in a seabed mapping task [14.83]. Visual SLAM has been successfully applied by a variety of researchers, using either sonar [14.67, 84–86] or camera data [14.87–89].

14.2.5 Cooperative Navigation of Multiple Vehicles

An exciting development over the past fifteen years has been the integration of ranging and communication capabilities in modern undersea acoustic modems, such as the WHOI micro-modem [14.90]. Given recent advances in AUV communications technology [14.64], the technology exists now for two or more AUVS to establish communication links with each other. This new capability allows the development of fundamentally new approaches to navigation based on multiple cooperating vehicles. Improved position estimation via collaboration enables the development of cooperative behaviors, permitting teams of AUVs to perform adaptive and more efficient survey missions. Ultimately, one can envision a system in which one AUV can direct other AUVs to revisit targets of interest, using complementary sensors. Two or more vehicles can be used to establish a mobile transponder network to provide positioning support for each other and for a fleet of other vehicles [14.91].

If more than one AUV is available to carry out a task and these vehicles are at least occasionally within communication range, both vehicles can cooperate in order to improve their navigation accuracy. Generally speaking, cooperative navigation relies on the fact that one vehicle may have a more accurate position estimate than another one. It then broadcasts it's own position estimate, possibly along with additional information such as the uncertainty associated with that estimate. Other vehicles, which are within the communication range of the broadcasting vehicle and receive this information, are able to obtain a relative position estimate to the broadcasting vehicle (range and/or bearing) and can incorporate this information in order to improve their own position estimate.

Systems

One key advantage of cooperative navigation is that most AUVs are already fitted with the necessary gear. In order to exchange information, a standard acoustic modem, found on most of today's AUVs, can be used. The only other information required – a range and/or bearing to the broadcasting vehicle can be obtained by slightly modifying existing modem hardware. By adding several transducers to a modem, it can determine the incident angle of an incoming transmission through the small differences in the arrival time. In addition, all vehicles can carry a very accurate clock

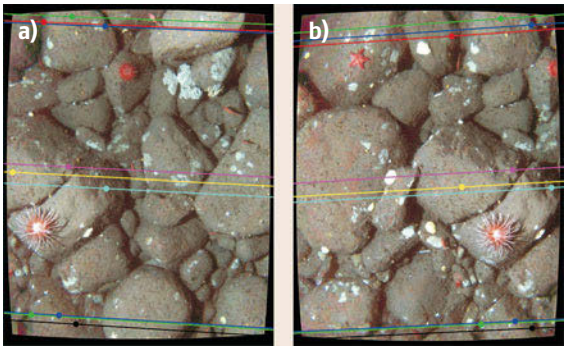


Fig. 14.6 Two consecutive pictures after being processed by the visually aided navigation (after [14.83]). From the several hundred features identified in each picture, only nine, marked by the *colored dots* have correspondences in both pictures and fit within the epipolar constraints (courtesy of R. Eustice)

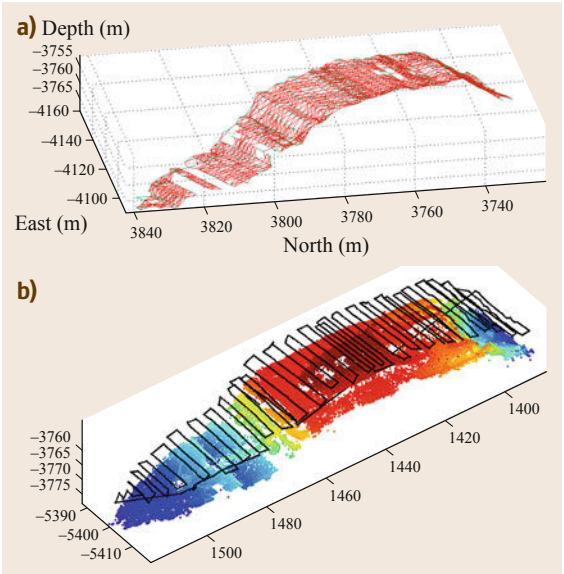


Fig. 14.7 Network of contraints **(a)** and estimated vehicle trajectory **(b)** for visual mapping of the RMS Titanic (after [14.87], courtesy of R. Eustice)

which is synchronized to a global clock via GPS on the surface [14.92–94].

Accurate time keeping has been central to precision navigation at sea throughout history [14.32]. Recent implementations of synchronous-clock one-way-travel-time acoustic navigation by *Eustice* et al. [14.92] and *Bahr* et al. [14.95] have been enabled by using a low-power temperature compensated crystal oscillator (TCXO), from SeaScan Inc., which has a typical drift of 20 ns/s, corresponding to a drift in position of approximately 2.6 m/day [14.96]. A common timebase and embedding a time-stamp indicating the broadcast time of each outgoing transmission allow a receiving vehicle to determine the time which the signal took to reach the receiver. With an accurate estimate of the speed of sound in water, the vehicle can then determine the distance to the broadcasting vehicle.

Strategies

Two different kinds of strategies can be employed for cooperative navigation. The first assumes a strict hierarchy where a group of AUVs acts as dedicated navigation aids (NAs). Their sole purpose is to provide navigation information to the mission AUVs (MAs). The NAs are thus equipped with very sophisticated sensors such that the drift of their position estimate is minimized. They may also operate very close to the surface such that they can easily surface to update their position using a GPS fix. In this scenario, only NAs broadcast navigation information and serve as mobile LBL-beacons to all MAs in communication range. As the sole purpose of the NAs is to maximize the MAs' navigation accuracy, their path can be adapted to best accomplish that goal [14.97].

The second strategy does not impose any particular hierarchy. Each vehicle occasionally broadcasts its position estimate and incorporates other overheard broadcasts. This approach does not require an extra set of dedicated vehicles, but due to the stochastic nature of received updates it is very difficult to predict, let alone guarantee, the performance of the navigation improvement of this cooperation strategy. In addition, the fact that each vehicle may transmit and receive navigation information may lead to cyclical updates. This occurs when a vehicle A incorporates information from a vehicle B and at a later point in time vehicle B then incorporates information from A. The dependencies arising between the position estimates may lead to overconfidence in the position estimate of each vehicle and must be carefully mitigated [14.98].

Algorithms

By treating the information received from a cooperating AUV as the noisy observation of landmark with an uncertain position, we can draw on a vast amount of algorithms and techniques developed by the SLAM community. Their work provides various frameworks to represent a state estimate (in this case a position) together with the uncertainty of this estimate. It also provides several methods to incorporate landmark observations, in this case received broadcasts from cooperating vehicles. Several of these SLAM techniques have been adapted for cooperative navigation [14.95, 99, 100].

14.3 Summary

Figure 14.8 shows typical navigation system performance for current AUVs based on five typical vehicle configurations. The sensor suite employed for an AUV depends on the navigation accuracy required for the mission as well as the available power, space, and the cost constraints.

14.3.1 Glider with Very Low Power Sensor Suite

Autonomous ocean gliders must operate for extended periods of time without being able to recharge their batteries [14.9, 101]. As a result, power consumption is the limiting factor for the selection of navigation sensors, and the navigation suite of a glider usually consists of a GPS, an AHRS, and a pressure gauge. While submerged the glider uses the AHRS combined with a vehicle model to estimate its heading and forward velocity and dead-reckon its position. The high noise and the unobservable variables in the vehicle model lead to a very high drift of 30% or even more if strong currents are present. On the surface, the vehicle resets its position estimate using GPS. The navigation accuracy achieved is typically more than adequate for providing data into ocean circulation models [14.102].

14.3.2 Low-Cost AUV Sensor Suite

A low-cost AUV such as the *IVER* uses a flow meter to obtain a measurement of their forward speed u_x^V. This information combined with an AHRS leads to a significant improvement of the navigation accuracy when compared to that of a glider.

14.3.3 Standard AUV Sensor Suite

The standard AUV adds a DVL to the list of sensors. When the DVL is able to obtain bottom lock a very accurate vehicle-referenced velocity vector $\boldsymbol{u}^V$ is available and the navigation accuracy improves by an order of magnitude. Drift rates as low as 1% of the distance traveled can be obtained with a well calibrated magnetic compass. Standard AUVs operating in a confined area are often outfitted with an LBL system. When operating within the polygon established by the position of the LBL beacons, the position drift will remain bounded.

14.3.4 High-End AUV

The dominant source of error in the standard AUV sensor suite described above is introduced during the transformation of vehicle-referenced velocities to world-referenced velocities as a result of errors in the heading measurements. Replacing the simple magnetic compass with an FOG improves the navigation by two orders of magnitude (0.1% of the distance traveled). When the DVL is not able to obtain bottom lock all of the vehicles described so far can only rely on the vehicle's linear acceleration sensors to obtain velocities. Due to the large noise introduced by these sensors the navigation accuracy decreases dramatically.

14.3.5 Special-Task AUV Using Visual SLAM

The special-task AUV has the same sensors as the standard AUV. Additionally, it uses a bottom-looking camera to take a series of pictures of the sea-floor. When revisiting a point it has taken a picture of before, it is able to recognize that fact and the navigation algorithm is able to reset the drift. As a result, the drift is bounded. An example, of this technique has been illustrated in Fig. 14.6. This method however requires the AUV to revisit points and stay close enough to the sea floor (< 10 m) to acquire the images.

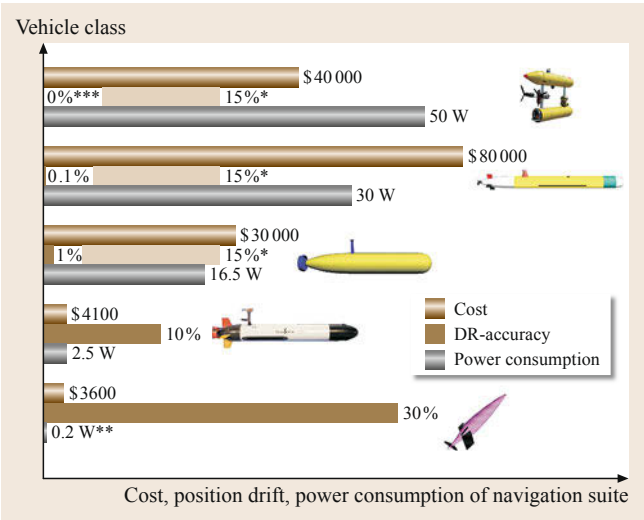


Fig. 14.8 Navigation accuracy, power consumption and price of various AUV sensor suites (*bottom to top*): (1) Glider with compass and attitude sensor; (2) low-cost AUV with compass, attitude sensor and flow meter; (3) medium-range AUV with INS, DVL and LBL; (4) high-end AUV with FOG-based INS, DVL and LBL; (5) special-task AUV with INS, DVL and SLAM. * Drift in mid water-column when DVL cannot obtain bottom or surface lock. ** Assuming a 10% duty cycle during which the navigation sensors are powered. *** Assuming that the vehicle was close enough to the sea floor throughout the entire mission to take pictures and revisit places.

14.4 Conclusion

In summary, the design of an AUV navigation system will depend on the mission requirements and cost constraints. For missions that are performed in shallow water, and for which frequent GPS surfacing is acceptable, a combination of GPS and dead-reckoning/inertial navigation will be acceptable in many situations. Cooperative navigation will become even more important in the future, as improved temperature compensated oscillators, and possibly even miniature atomic clocks, are anticipated to become widely available at low cost. This can greatly improve available options for hyperbolic and one-way time-synchronized AUV localization algorithms.

In deep water, the current state of the art is represented by the Nereus vehicle, whose sensor suite contains a Paroscientific pressure sensor, a Teledyne-RDInstruments 300 kHz Doppler sonar, an IXSEA Phins IMU, a WHOI LBL transceiver, a WHOI Micro-Modem, and a Microstrain gyro-stabilized attitude and magnetic-heading sensor [14.8]. Nereus navigates using the NavEst navigation software package developed by Woods Hole Oceanographic Institution and Johns Hopkins University, which integrates the DVLNav [14.103] package with the ABE LBL navigation suite [14.7]. This high-performance system embodies the current state of the art in deep ocean AUV navigation.

For further information on navigation sensor device performance, the reader is directed to a detailed survey by *Kinsey* et al. [14.1]. A tutorial introduction to SLAM is provided by the authors of [14.104, 105]. *Fallon* et al. gives a detailed presentation for several recent SLAM applications, including mine neutralization, real-time cooperative navigation, and ship hull inspection [14.106].

For future research in AUV navigation, a key challenge will be to develop techniques for Arctic and Antarctic operations. Polar regions pose major difficulties for AUV navigation including: high magnetic declination, inability to surface for GPS, infeasibility of deploying transponders, potentially moving ice canopy, and limited recovery options. Recent work in AUV navigation for polar regions includes the work of *Kunz* et al., who have deployed variants of the Seabed AUV in both the Arctic and Antarctic [14.107, 108].

Another ambitious goal for future research is to develop robust navigation techniques, combined with autonomous docking [14.44] and advanced communication networks [14.109], to enable long-term AUV deployments for persistent ocean sampling missions [14.110] using underice or deep ocean observatories [14.111, 112].

References

14.1 J.C. Kinsey, R.M. Eustice, L.L. Whitcomb: A survey of underwater vehicle navigation: Recent advances and new challenges, IFAC Conf. Manoeuvering Control Mar. Craft (2006)

14.2 C. von Alt, B. Allen, T. Austin, R. Stokey: Remote environmental monitoring units, Proc. AUV '94 (1994)

14.3 J.W. Rish, S. Willcox, R. Grieve, I. Montieth, J. Vaganay: Operational testing of the battlespace preparation AUV in the shallow water regime, Proc. IEEE Oceans, Vol. 1 (2001) pp. 123–129

14.4 F.S. Hover, R.M. Eustice, A. Kim, B.J. Englot, H. Johannsson, M. Kaess, J.J. Leonard: Advanced perception, navigation and planning for autonomous in-water ship hull inspection, Int. J. Robotics Res. **31**(12), 1445–1464 (2012)

14.5 H. Singh, A. Can, R. Eustice, S. Lerner, N. McPhee, C. Roman: Seabed AUV offers new platform for high-resolution imaging, Eos Trans. Am. Geophys. Union **85**(31), 289 (2004)

14.6 B. Anderson, J. Crowell: Workhorse AUV – A cost-sensible new autonomous underwater vehicle for surveys/soundings, search and rescue, and research, Proc. MTS/IEEE OCEANS (2005) pp. 1–6

14.7 D. Yoerger, M. Jakuba, A. Bradley, B. Bingham: Techniques for deep sea near bottom survey using an autonomous underwater vehicle, Int. J. Robotics Res. **26**(1), 416–429 (2007)

14.8 L.L. Whitcomb, M.V. Jakuba, J.C. Kinsey, S.C. Martin, S.E. Webster, J.C. Howland, C.L. Taylor, D. Gomez-Ibanez, D.R. Yoerger: Navigation and control of the Nereus hybrid underwater vehicle for global ocean science to 10,903 m depth: Preliminary results, Proc. IEEE Int. Conf. Robotics Autom. (ICRA) (2010) pp. 594–600

14.9 J. Sherman, R.E. Davis, W.B. Owens, J. Valdes: The autonomous underwater glider, IEEE J. Ocean. Eng. **26**(4), 437–446 (2001)

14.10 J.G. Paglia, W.F. Wyman: DARPA's autonomous minehunting and mapping technologies (AMMT) program: An overview, Proc. IEEE Oceans, Vol. 2 (1996) pp. 794–799

14.11 D.B. Heckman, R.C. Abbott: An acoustic navigation technique, Proc. IEEE OCEANS '73 (1973) pp. 591–595

14.12 M. Hunt, W. Marquet, D. Moller, K. Peal, W. Smith, R. Spindel: *An acoustic navigation system*, Tech. Rep. WHOI-74-6 (Woods Hole Oceanographic Institution, Falmouth 1974)

14.13 P.H. Milne: *Underwater Acoustic Positioning Systems* (Gulf Publishing, Houston 1983)

14.14 J. Vaganay, J.G. Bellingham, J.J. Leonard: Outlier rejection for autonomous acoustic navigation, Proc. IEEE Int. Conf. Robotics Autom. (1996) pp. 2174–2181

14.15 M. Deffenbaugh, H. Schmidt, J. Bellingham: Acoustic positioning in a fading multipath environment, Proc. IEEE Oceans (1996) pp. 596–600

14.16 E. Geyer, P. Creamer, J. D'Appolito, R. Gains: Characteristics and capabilities of navigation systems for unmanned untethered submersibles, Proc. Int. Symp. Unmanned Untethered Submers. Technol. (1987) pp. 320–347

14.17 S.T. Tuohy, J.J. Leonard, J.G. Bellingham, N.M. Patrikalakis, C. Chryssostomidis: Map based navigation for autonomous underwater vehicles, Int. J. Offshore Polar Eng. **6**(1), 9–18 (1996)

14.18 J.M. Paros: Digital pressure tranducer, US Patent 4 455 874 (1984)

14.19 N.P. Fofonoff, R.C. Millard: Algorithms for computation of fundamental properties of seawater, UNESCO Tech. Paper Mar. Sci. **44**, 1–53 (1983)

14.20 A.J. Healey, E.P. An, D.B. Marco: Online compensation of heading sensor bias for low cost AUVs, Proc. IEEE AUV 98 (1998) pp. 35–42

14.21 G. Dudek, M. Jenkin: Inertial sensors, GPS, and odometry. In: *Springer Handbook of Robotics*, ed. by B. Siciliano, O. Khatib (Springer, Berlin, Heidelberg 2008)

14.22 D. Mackenzie: *Inventing Accuracy* (MIT, Cambridge 1990)

14.23 R.A. Bergh, H.C. Lefevre, H.J. Shaw: All-single-mode fiber-optic gyroscope, Optics Lett. **6**(4), 198–200 (1981)

14.24 J. Farrell: *Aided Navigation: GPS with High Rate Sensors* (McGraw-Hill, New York 2008)

14.25 P.N. Denbigh: Ship velocity determination by doppler and correlation techniques, Commun. Radar Signal Process. IEE Proc. F **131**(3), 315–326 (1984)

14.26 G. Griffiths, S.E. Bradley: A correlation speed log for deep waters, Sea Technol. **39**(3), 29–35 (1998)

14.27 A. Alcocer, P. Oliveira, A. Pascoal: Study and implementation of an EKF GIB-based underwater positioning system, Control Eng. Pract. **15**(6), 689–701 (2007)

14.28 L.L. Whitcomb, D.R. Yoerger, H. Singh: Combined Doppler/LBL based navigation of underwater vehicles, Proc. Int. Symp. Unmanned Untethered Submers. Technol. (UUST) (1999)

14.29 R. Rikoski, J. Leonard, P. Newman, H. Schmidt: Trajectory sonar perception in the ligurian sea. In: *Experimental Robotics IX*, ed. by M.H. Ang Jr., O. Khatib (Springer, Berlin, Heidelberg 2006) pp. 557–570

14.30 H. Thomas: Commercial offer: GPS intelligent buoy system, ACSA (2001) http://www.underwater-gps.com

14.31 E.S. Maloney (Ed.): *Dutton's Navigation and Piloting* (Naval Institute, Annapolis 1985)

14.32 D. Sobel: *Longitude: The True Story of a Lone Genius Who Solved the Greatest Scientific Problem of His Time*, Vol. 3 (HarperCollins, London 2007)

14.33 M. Kuristsky, M. Goldstein: In: *Inertial navigation. In Autonomous Robot Vehicles*, ed. by I. Cox, G. Wilfong (Springer, Berlin, Heidelberg 1990)

14.34 D. Titterton: *Strapdown Inertial Navigation Technology*, Vol. 17 (IET, London 2004)

14.35 IXSEA: Specifications of PHINS inertial navigation system. http://www.ixsea.com/en/products/002.001.001.004/phins-6000.html

14.36 M.B. Larsen: High performance doppler inertial navigation – Experimental results, Proc. MTS/IEEE OCEANS, Vol. 2 (2000) pp. 1449–1456

14.37 M. Kaess, A. Ranganathan, F. Dellaert: iSAM: Incremental smoothing and mapping, IEEE Trans. Robotics **24**(6), 1365–1378 (2008)

14.38 J.G. Bellingham, T.R. Consi, U. Tedrow, D. Di Massa: Hyperbolic acoustic navigation for underwater vehicles: Implementation and demonstration, Proc. AUV '92 (1992) pp. 304–309

14.39 D.K. Atwood, J.J. Leonard, J.G. Bellingham, B.A. Moran: An acoustic navigation system for multiple vehicles, Proc. Int. Symp. Unmanned Untethered Submers. Technol. (1995) pp. 202–208

14.40 Imetrix Inc.: http://www.imetrix.com, 1998

14.41 ORE Inc.: http://www.orehouston.com, 1998

14.42 B.H. Tracey: Design and Testing of an Acoustic Ultra-Short Baseline Navigation System, M.S. Thesis (MIT, Cambridge 1992)

14.43 J.G. Bellingham, J.J. Leonard, J. Vaganay, C. Goudey, D. Atwood, T. Consi, J. Bales, H. Schmidt, C. Chryssostomidis: AUV operations in the arctic, Proc. Sea Ice Mechanics Arct. Model. Workshop (1995)

14.44 H. Singh, J. Catipovic, R. Eastwood, L. Freitag, H. Henriksen, F.F. Hover, D. Yoerger, J.G. Bellingham, B.A. Moran: An integrated approach to multiple AUV communications, navigation and docking, Proc. MTS/IEEE Oceans '96, Vol. 1 (1996) pp. 59–64

14.45 M. Deffenbaugh: A Matched Field Processing Approach to Long Range Acoustic Navigation, M.S. Thesis (MIT, Cambridge 1994)

14.46 W. Munk, P. Worcester, C. Wunsch: *Ocean Acoustic Tomography* (Cambridge University Press, Cambridge 1995)

14.47 H. Schmidt, J. Bellingham, M. Johnson, D. Herold, D. Farmer, R. Pawlowcisz: Real-time frontal mapping with AUVs in a coastal environment, Proc. IEEE OCEANS, Vol. 3 (1996) pp. 1094–1098

14.48 C. Wunsch: *The Ocean Circulation Inverse Problem* (Cambridge University Press, Cambridge 1996)

14.49 J. Catipovic: Performance limitations in underwater acoustic telemetry, IEEE J. Ocean Eng. **15**(3), 205–216 (1990)

14.50 W.A. Kuperman, W.S. Hodgkiss, H.C. Song, T.A.C. Ferla, D.R. Jackson: Phase conjugation in the ocean: Experimental demonstration of an acoustic timereversal mirror, J. Acoust. Soc. Am. **103**, 25 (1998)

14.51 M.B. May: Gravity navigation, Record 1978 Position Locat. Navig. Symp. (1978) pp. 212–218

14.52 C. Tyren: Magnetic anomalies as a reference for ground-speed and map-matching navigation, J. Navig. **35**(2), 242–254 (1982)

14.53 T.H. Waterman: *Animal Navigation* (Scientific American Library, New York 1989)

14.54 J. Myers, C. Holm, R. MacAllister (Eds.): *Handbook of Ocean and Underwater Engineering* (McGraw-Hill, New York 1969)

14.55 S.T. Tuohy: Geophysical Map Representation, Abstraction, and Interrogation for Underwater Vehicle Navigation, Ph.D. Thesis (MIT, Cambridge 1993)

14.56 W.D. Kahn: *Accuracy of mapping the Earth's gravity field fine structure with a space borne gravity gradiometer mission*, Tech. Rep. N.84-30473 (NASA Goddard Geodynamic Branch, Greenbelt 1984)

14.57 R.B. Whitmarsh, L.M. Pinheiro, P.R. Miles, M. Recq, J.C. Sibuet: Thin crust at the western Iberia ocean-continent transition and ophiolites, Tectonics **12**, 1230–1239 (1993)

14.58 R. Forsberg: Topographic effects in airborne gravity gradiometry, Proc. 15th Gravity Gradiometer Conf. (1987)

14.59 D. Turcotte, G. Schubert: *Geodynamics: Applications of Continuum Physics to Geological problems* (Wiley, New York 1982)

14.60 A. Jircitano, J. While, D. Dosch: Gravity based navigation of AUVs, Proc. AUV '90 (1990) pp. 177–180

14.61 D.T. Sandwell: Geophysical applications of satellite altimetry, Rev. Geophys. Suppl. **29**, 132–137 (1991)

14.62 M.A. Gerber: Gravity gradiometer: Something new in inertial navigation, Astronaut. Aeronaut. **16**, 18–26 (1978)

14.63 B. Kamgar-Parsi: Registration algorithms for making accurate geophysical maps, Proc. IEEE Oceans '97, Vol. 2 (1997) pp. 974–980

14.64 B. Kamgar-Parsi, L. Rosenblum, F. Pipitone, L. Davis, J. Jones: Toward an automated system for a correctly registered bathymetric chart, IEEE J. Ocean Eng. **14**(4), 314–325 (1989)

14.65 L. Lucido, B. Popescu, J. Opderbecke, V. Rigaud: Segmentation of bathymetric profiles and terrain matching for underwater vehicle navigation, Proc. 2nd Annu. World Autom. Conf. (1996)

14.66 O. Bergem: A multibeam sonar based positioning system for an AUV, 8th Int. Symp. Unmanned Untethered Submers. Technol. (AUSI) (1993) pp. 291–299

14.67 J.J. Leonard, R. Carpenter, H.J.S. Feder: Stochastic mapping using forward look sonar, Robotica **19**, 467–480 (2001)

14.68 F. Lu, E. Milios: Globally consistent range scan alignment for environmental mapping, Auton. Robots **4**, 333–349 (1997)

14.69 F.R. Kschischang, B.J. Frey, H.-A. Loeliger: Factor graphs and the sum-product algorithm, IEEE Trans. Inform. Theory **47**(2), 498–519 (2001)

14.70 G. Dissanayake, P. Newman, S. Clark, H.F. Durrant-Whyte, M. Csorba: A solution to the simultaneous localisation and map building (SLAM) problem, IEEE Trans. Robotics Autom. **17**(3), 229–241 (2001)

14.71 R. Smith, M. Self, P. Cheeseman: A stochastic map for uncertain spatial relationships, Proc. Int. Symp. Robotics Res. (ISRR) (1988) pp. 467–474

14.72 S.J. Julier, J.K. Uhlmann: A counter example to the theory of simultaneous localization and map building, Proc. IEEE Int. Conf. Robotics Autom. (ICRA), Vol. 4 (2001) pp. 4238–4243

14.73 S. Thrun, W. Burgard, D. Fox: *Probabilistic Robotics* (MIT, Cambridge 2005)

14.74 J. Folkesson, H. Christensen: Closing the loop with graphical SLAM, IEEE Trans. Robotics **23**(4), 731–741 (2007)

14.75 K. Konolige: Large-scale map-making, Proc. AAAI Conf. Artif. Intell. (2004) pp. 457–463

14.76 U. Frese, P. Larsson, T. Duckett: A multilevel relaxation algorithm for simultaneous localisation and mapping, IEEE Trans. Robotics **21**(2), 196–207 (2005)

14.77 A. Ranganathan, M. Kaess, F. Dellaert: Loopy SAM, Int. Joint Conf. Artif. Intell. (2007) pp. 2191–2196

14.78 F. Dellaert, M. Kaess: Square Root SAM: Simultaneous localization and mapping via square root information smoothing, Int. J. Robotics Res. **25**(12), 1181–1203 (2006)

14.79 M. Kaess, H. Johannsson, R. Roberts, V. Ila, J.J. Leonard, F. Dellaert: iSAM2: Incremental smoothing and mapping using the Bayes tree, Int. J. Robotics Res. **31**, 217–236 (2012)

14.80 K. Konolige, G. Grisetti, R. Kümmerle, W. Burgard, B. Limketkai, R. Vincent: Efficient sparse pose adjustment for 2D mapping, Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS) (2010) pp. 22–29

14.81 R. Kümmerle, G. Grisetti, H. Strasdat, K. Konolige, W. Burgard: G^2O: A general framework for graph optimization, Proc. IEEE Int. Conf. Robotics Autom. (ICRA) (2011) pp. 3607–3613

14.82 H. Johannsson, M. Kaess, M.F. Fallon, J.J. Leonard: Temporally scalable visual SLAM using a reduced pose graph, IEEE Int. Conf. Robotics Autom. (ICRA) (2013) pp. 54–61

14.83 R. Eustice: Large-Area Visually Augmented Navigation for Autonomous Underwater Vehicles, Ph.D. Thesis (MIT, Cambridge 2005)

14.84 N. Fairfield, A.G. Kantor, D. Wettergreen: Real-time SLAM with octree evidence grids for exploration in underwater tunnels, J. Field Robotics **24**(1), 3–22 (2007)

14.85 J. Folkesson, J. Leonard: Autonomy through SLAM for an underwater robot. In: *Robotics Research*, Springer Tracts in Advanced Robotics, Vol. 70, ed. by C. Pradalier, R. Siegwart, G. Hirzinger (Springer, Berlin, Heidelberg 2011) pp. 55–70

14.86 C. Roman, H. Singh: Improved vehicle based multibeam bathymetry using submaps and SLAM, Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS) (2005) pp. 3662–3669

14.87 R.M. Eustice, H. Singh, J.J. Leonard: Exactly sparse delayed-state filters for viewbased SLAM, IEEE Trans. Robotics **22**(6), 1100–1114 (2006)

14.88 N.R. Gracias, S. Van Der Zwaan, A. Bernardino, J. Santos-Victor: Mosaic-based navigation for autonomous underwater vehicles, IEEE J. Ocean. Eng. **28**(4), 609–624 (2003)

14.89 I. Mahon, S.B. Williams, O. Pizarro, M. Johnson-Roberson: Efficient view-based SLAM using visual loop closures, IEEE Trans. Robotics **24**(5), 1002–1014 (2008)

14.90 L. Freitag, M. Grund, S. Singh, J. Partan, P. Koski, K. Ball: The WHOI micromodem: An acoustic communications and navigation system for multiple platforms, Proc. MTS/IEEE OCEANS Conf. Exhib., Vol. 1 (2005) pp. 1086–1092

14.91 J. Vaganay, J.J. Leonard, J.A. Curcio, J.S. Willcox: Experimental validation of the moving long base line navigation concept, Proc. IEEE/OES AUV (2004) pp. 59–65

14.92 R.M. Eustice, L.L. Whitcomb, H. Singh, M. Grund: Experimental results in synchronous-clock one-way-travel-time acoustic navigation for autonomous underwater vehicles, Proc. IEEE Int. Conf. Robotics Autom. (ICRA) (2007) pp. 4257–4264

14.93 M.F. Fallon, G. Papadopoulos, J.J. Leonard, N.M. Patrikalakis: Cooperative AUV navigation using a single maneuvering surface craft, Int. J. Robotics Res. **29**(12), 1461–1474 (2010)

14.94 S.E. Webster, R.M. Eustice, H. Singh, L.L. Whitcomb: Preliminary deep water results in single-beacon one-way-travel-time acoustic navigation for underwater vehicles, Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS) (2009) pp. 2053–2060

14.95 A. Bahr, J.J. Leonard, M.F. Fallon: Cooperative localization for autonomous underwater vehicles, Int. J. Robotics Res. **28**(6), 714–728 (2009)

14.96 S.E. Webster, L.L. Whitcomb, R.M. Eustice: Advances in decentralized single-beacon acoustic

navigation for underwater vehicles: Theory and simulation, Proc. IEEE/OES AUV (2010) pp. 1–8
14.97 A. Bahr, J.J. Leonard, A. Martinoli: Dynamic positioning of beacon vehicles for cooperative underwater navigation, IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS) (2012) pp. 3760–3767
14.98 A. Bahr, M. Walter, J.J. Leonard: Consistent cooperative localization, Proc. IEEE Int. Conf. Robotics Autom. (ICRA) (2009) pp. 3415–3422
14.99 R. Eustice, L. Whitcomb, H. Singh, M. Grund: Recent advances in synchronousclock one-way-travel-time acoustic navigation, Proc. IEEE OCEANS (2006) pp. 1–6
14.100 S.E. Webster, R.M. Eustice, H. Singh, L.L. Whitcomb: Advances in single-beacon one-way-travel-time acoustic navigation for underwater vehicles, Int. J. Robotics Res. **31**(8), 935–949 (2012)
14.101 D.C. Webb, P.J. Simonetti, C.P. Jones: Slocum: An underwater glider propelled by environmental energy, IEEE J. Ocean. Eng. **26**(4), 447–452 (2001)
14.102 S.R. Ramp, R.E. Davis, N.E. Leonard, I. Shulman, Y. Chao, A.R. Robinson, J. Marsden, P.F.J. Lermusiaux, D.M. Fratantoni, J.D. Paduan, F.P. Chavez: Preparing to predict: The second autonomous ocean sampling network (AOSN-II) experiment in the monterey bay, Deep Sea Res, II **56**(3), 68–86 (2009)
14.103 J.C. Kinsey, L.L. Whitcomb: Preliminary field experience with the DVLNAV integrated navigation system for oceanographic submersibles, Control Eng. Pract. **12**(12), 1541–1549 (2004)
14.104 H.F. Durrant-Whyte, T. Bailey: Simultaneous localisation and mapping (SLAM): Part I, IEEE Robotics Autom. Mag. **13**(2), 99–110 (2006)
14.105 T. Bailey, H.F. Durrant-Whyte: Simultaneous localisation and mapping (SLAM): Part II, Robotics Autom. Mag. **13**(3), 108–117 (2006)
14.106 M.F. Fallon, H. Johannsson, M. Kaess, J. Folkesson, H. McClelland, B.J. Englot, F.S. Hover, J.J. Leonard: Simultaneous localization and mapping in marine environments. In: *Marine Robot Autonomy*, ed. by M. Seto (Springer, New York 2013)
14.107 C. Kunz, C. Murphy, R. Camilli, H. Singh, J. Bailey, R. Eustice, M. Jakuba, K. Nakamura, C. Roman, T. Sato, R. Sohn: Deep sea underwater robotic exploration in the icecovered arctic ocean with AUVs, Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS) (2008) pp. 3654–3660
14.108 C.G. Kunz: Autonomous Underwater Vehicle Navigation and Mapping in Dynamic, Unstructured Environments, Ph.D. Thesis (Massachusetts Institute of Technology and Woods Hole Oceanographic Institution, Falmouth 2012)
14.109 D.B. Kilfoyle, A.B. Baggeroer: The current state-of-the-art in underwater acoustic telemetry, IEEE J. Ocean Eng. **25**(1), 4–27 (2000)
14.110 T. Curtin, J.G. Bellingham, J. Catipovic, D. Webb: Autonomous ocean sampling networks, Oceanography **6**(3), 86–94 (1993)
14.111 J. Delaney, G.R. Heath, A. Chave, H. Kirkham, B. Howe, W. Wilcock, P. Beauchamp, A. Maffei: Neptune: Real-time, long-term ocean and earth studies at the scale of a tectonic plate, Proc. IEEE OCEANS MTS/IEEE Conf. Exhib., Vol. 3 (2001) pp. 1366–1373
14.112 R. Krishfield, J. Toole, A. Proshutinsky, M.-L. Timmermans: Automated ice-tethered profilers for seawater observations under pack ice in all seasons, J. Atmos. Oceanic Technol. **25**(11), 2091–2105 (2008)

15. Acoustic Communication

Milica Stojanovic, Pierre-Philippe J. Beaujean

In this chapter, we discuss acoustic communication methods which are used to provide wireless connection between remote nodes operating in an underwater environment. We begin with an introductory overview of the history of acoustic communication, and an outline of current and emerging applications. We then provide a summary of communication channel characteristics, with an eye towards acoustic propagation mechanisms and the ways in which they differ from radio propagation. The main focus of our treatment is on two major aspects of communication system design: the physical link and the networking functions. On the physical link level, we discuss noncoherent and coherent modulation/detection methods, paying attention to both single-carrier modulation and multicarrier broadband techniques. Specifically, we discuss signal processing methods for synchronization, equalization, and multichannel (transmit and receive) combining. On the networking level, we discuss protocols for channel sharing using both deterministic division of resources (frequency, time, and code-division multiple access) and random access, and we also overview recent results on routing for peer-to-peer acoustic networks. We conclude with an outline of topics for future research.

15.1 **A Brief History** 360
15.2 **Current and Emerging Modem Applications** 360
15.3 **Existing Technology** 361
15.3.1 System Requirements 361
15.3.2 Commercially Available Modems 362
15.3.3 Field Tests 363
15.4 **Propagation Channel** 364
15.4.1 Attenuation and Noise 364
15.4.2 Multipath Propagation 365
15.4.3 Time Variability: Motion-Induced Doppler Distortion 369
15.4.4 Time Variability: Random Effects (Fading) 370
15.4.5 System Constraints 372
15.5 **Point-to-Point Links: Signal Processing** 374
15.5.1 Noncoherent Modulation/Detection 374
15.5.2 Coherent Modulation/Detection 375
15.5.3 Data Link Reliability 378
15.5.4 Turbo Equalization 378
15.5.5 Adapting to the Environment 379
15.5.6 Networks 379
15.5.7 Channel Sharing 381
15.5.8 Routing and Cross-Layer Integration 382
15.6 **Future Trends** 383
References 383

Wireless transmission of signals underwater over distances in excess of a 100 m relies almost exclusively on acoustic waves. Radio waves do not propagate well underwater, except at low frequencies and over extremely short distances (a few meters at 10 kHz) [15.1]. Optical signals, which are best used in the blue–green region (around 500 nm), also suffer from attenuation and do not propagate beyond about 100 m, although they do offer high bandwidths (on the order of MHz) [15.2]. Hence, sound is used for wireless transmission over anything but very short distances. Sound propagates as a pressure wave, and it can thus easily travel over kilometers, or even hundreds of kilometers, but to cover a longer distance, a lower frequency has to be used. In general, acoustic communications are confined to bandwidths that are low compared to those used for terrestrial radio communications. Acoustic modems that are in use today typically operate in bandwidths on the order of a few kHz, at a comparably low center frequency (5 centered at 10 kHz) [15.3]. While such frequencies will cover distances on the order of a kilometer, acoustic frequencies in the 100 kHz region can be used for shorter distances, while frequencies below a kHz are used for longer distances. Underwater acoustic communication over basin scales (several thousand kilometers) can be established in a single hop as well; however, the attendant bandwidth will be only on the

order of 10 Hz [15.4]. Horizontal transmission is notoriously more difficult due to the multipath propagation, while vertical channels exhibit less distortion [15.5]. Frequency-dependent attenuation, multipath propagation, and low speed of sound (about 1500 m/s), which results in a severe Doppler effect, make the underwater acoustic channel one of the most challenging communication media.

15.1 A Brief History

Among the first operational underwater acoustic systems was the submarine communication system developed in the United States around the end of the Second World War. This system used analog modulation in the 8–11 kHz band (single-sideband analog modulation AM) [15.6]. Technology has since advanced, pushing digital modulation/detection techniques into the forefront of modern acoustic communications. In the early 1980s, the advent of digital signal processing sparked a renowned interest in underwater acoustic communications, leading a group of scientists at the Massachusetts Institute of Technology and the Woods Hole Oceanographic Institution (WHOI) to propose a system based on frequency shift keying (FSK) [15.7]. The system became known as DATS (digital acoustic telemetry system) and provided a basis for the first generation of commercial digital acoustic modems [15.8]. Today, coded FSK is used in several acoustic modems, including the WHOI *micro-modem* and the Teledyne (formerly Benthos) *telesonar type B* modem [15.9]. While FSK relies on simple energy detection (noncoherent detection), and thus offers robustness to channel impairments, its bandwidth utilization is not efficient. Motivated by this fact, research in the 1990s focused on investigating phase shift keying (PSK) and quadrature amplitude modulation (QAM) for underwater acoustic channels. These modulation methods offer more bits/sec per Hz of occupied bandwidth, but require a receiver that can track the channel and compensate for the time-varying multipath and phase distortion (coherent detection). Work carried out at Northeastern University and WHOI resulted in a channel equalization/synchronization method [15.10], which forms the basis of a second generation of *high-speed* acoustic modems. Through the last decade, these modems have been used in operations involving both stationary platforms and autonomous underwater vehicles (AUVs), over vertical and horizontal links at bit rates of about 5 kbps. [15.11] gives an impressive account of a 2009 deployment near the 11 km deep Mariana Trench. A detailed summary of existing technology is given in Sect. 15.3.

Bit rates in excess of those available with operational modems have been demonstrated as well, but these results are in the domain of experimental research. At the time of this writing, research is active on improved, and ever more sophisticated channel estimation and equalization methods for single-carrier broadband systems, as well as on multicarrier modulation/detection techniques which hold a promise of reducing the implementation complexity of high-speed acoustic modems. The success of various communication techniques largely depends on our understanding of the acoustic communication channel, i. e., our ability to identify a proper model for signal distortion. We thus begin our treatment of acoustic communications links by outlining the channel characteristics in Sect. 15.4, and then move on to discuss the basic as well as the emerging concepts of signal processing in Sect. 15.5.

As technology advances, companies around the world engage more easily in modem development, and the legacy of US manufacturers such as Teledyne-Benthos, WHOI, and Link-Quest are joined by new ones such as the French Thales and the German Evo-Logics. As a result, standardization efforts are becoming necessary to ensure inter-operability between acoustic modems of different manufacturers [15.12]. While it is ultimately the need for a certain technology that will dictate the usefulness of its presence on the market, we must keep in mind that application-driven technology development is not the only way forward. More often than not, technology-driven applications arise – wireless radio industry being the prime example. In other words, as the acoustic modems' capabilities grow, applications that previously were not thought possible may start to emerge.

15.2 Current and Emerging Modem Applications

Modem applications range from ocean observation to the oil industry and aquaculture, and include gathering of sensor data from remote instruments for pollution control, climate recording and prediction of natural dis-

turbances, as well as the detection of objects on the ocean floor, and transmission of images from remote sites. Implicitly, wireless signal transmission is also crucial for monitoring and control of remote instruments, and for communicating with robots and AUVs, which can serve on isolated missions, or as parts of a mobile network. Applications involving AUVs range from search and survey missions to supervisory control and collection of mapping data, be they images, video, or side-scan sonar.

Autonomous systems are envisioned both for stand-alone applications and as wireless extensions to cabled systems. For example, a stand-alone 15 kbps acoustic link, intended to provide data transfer from bottom-mounted, battery-powered instruments such as seismometers to a surface buoy is described in [15.5]. A different example is that of ocean observatories, which are being built on decommissioned submarine cables to deploy an extensive fiber optic network of sensors (cameras, wave sensors, seismometers) covering miles of ocean floor [15.13]. These cables can also carry communication access points, very much as cellular base stations are connected to the telephone network, allowing users to move and communicate from places where cables cannot reach. Another example are cabled submersibles, also known as remotely operated vehicles (ROVs). These vehicles, which may weigh more than 10 metric tons, are connected to a mother-ship by a cable that can extend over several kilometers and deliver high power to the remote end, along with high-speed communication signals. A popular example of an ROV/AUV tandem is the WHOI's Alvin/Jason pair of vehicles that were deployed in 1985 to discover the Titanic. Such vehicles were also instrumental in the discovery of hydro-thermal vents, sources of extremely hot water on the bottom of deep ocean, which revealed forms of life different from any others known previously. The first vents were found in the late 1970s, and new ones are still being discovered. The importance of such discoveries is comparable only to space missions, and so is the technology that supports them. Today, vehicle technology, sensor technology, and acoustic communication technology are mature enough to support the visionary idea of an autonomous oceanographic sampling network [15.14]. Recent years have seen a proliferation of research on underwater sensor networks, which we discuss in Sect. 15.5.6.

15.3 Existing Technology

The field of underwater acoustic communication has become fairly mature. Many commercial units are available, along with a significant number of prototypes developed in research laboratories. The large number of configurations available complicates the selection process of a specific underwater acoustic modem. The purpose of this section is to offer some guidance in selecting an underwater acoustic modem. The authors do not recommend any specific product, rather they highlight what is important in the selection process. Following this, a brief overview of typical performance and field tests is presented.

15.3.1 System Requirements

Underwater acoustic modem users should always keep in mind that the performance of these systems changes dramatically depending on the application. For example, a system designed to operate at tens of kilometers will typically have a much lower throughput than a system designed to relay data at a few hundred meters. Also, some underwater acoustic modems are designed to operate in a complex network, while others are designed for point-to-point communication between two units.

The following list is by no mean exhaustive, but covers the main factors that should be used in selecting an underwater acoustic modem:

- *Application*: The user should always keep the application in mind when selecting an acoustic modem. Voice communication constitutes a category of its own, with well-proven technology specifically designed to relay intelligible voice with minimal delay and simultaneous two-way communication capability. Acoustic modems designed for the transmission of command-and-control messages operate more slowly, but very reliably. Image streaming can be achieved using the latest generation of high-bit rate (one-way) acoustic modems. Voice communication modems and command-and-control modems technology is much more mature than image streaming modems.
- *Cost*: Typically between $5,000 and $50,000 (US) per unit, depending on the complexity and performance.
- *Size*: Underwater acoustic modems are usually cylindrical in shape, with a pressure vessel containing the electronics and one or multiple transducers mounted on one end. The size varies from 0.05 m

diameter by 0.1 m in length, to 0.1 m in diameter by 0.5 m in length, depending on the frequency of operation (from a few to hundreds of kHz), maximum range and operating depth.

- *Power*: The power consumption depends on the range and modulation: traditionally, the total power consumption is of the order of 0.1–1 W in receive mode (depending on the receiver complexity), and 10–100 W in continuous transmission mode.
- *Data rate*: (1) underwater acoustic modems designed to operate very reliably in challenging environments, on moving platforms and over distances of several kilometers typically operate under 1000 bps; (2) in less challenging environments (e.g., vertical communications in deep waters), acoustic modems can operate at data rates well over 1000 bps.
- *Horizontal versus vertical transmission*: It is critical to understand that commercial underwater acoustic modems come in different versions depending on the type of operation. In particular, acoustic modems designed for vertical transmission in deep water use directional transducers and are capable of high data rates. A traditional mistake is to purchase this type of acoustic modems for horizontal transmission in shallow water, especially on moving platforms; in this case, acoustic modems with omnidirectional and a lower data rates should be selected.
- *Shallow water versus deep water communications (horizontal or slanted)*: In deep waters, fluctuations in sound speed (caused by changes in temperature, hydrostatic pressure, and chemical content) change the direction of propagation of sound (a phenomenon known as sound refraction). This often results in shadow zones: the sound produced by a modem will not reach the other modem, even if both modems are within operating range of one another. Echoes reflected off the sea surface and sea bottom may also be an issue at low frequencies, along with the presence of thermoclines or thermohalines. In shallow waters, the main source of limitation fades, which results from the many echoes reflected off the sea surface and sea bottom. As a result, underwater acoustic modems designed to operate in shallow waters are less likely to be in a shadow zone, but the received signal is often badly distorted and requires specific signal processing techniques (equalization, spread spectrum signaling, power, error coding) to operate reliably.
- *Range*: for the most part, the maximum operational range of an underwater acoustic modem depends on the operating frequency band of the acoustic modem, power, signaling (modulation, error coding, equalizing), environment, and transducer directivity. At low frequencies (a few kHz), commercial acoustic modems achieve reliable acoustic communications beyond 10 km horizontally, albeit at a low data rate (less than 400 bps). As frequency and bandwidth increases, some acoustic modems can transmit data at 10,000 bps or more, but the practical range is limited to 1000 m or less.
- *Operations in dynamic environments*: (1) surface activity causes Doppler spread, which limits the data rate and reliability; (2) platform dynamics cause Doppler shift, which limits the acoustic modem performance when excessive. Moving platforms usually require the use of omnidirectional transducers to avoid loss of acoustic communication.
- *Protocol for point-to-point quality of service* (*QoS*): reliability, delay, retransmission.
- *Networking capability*: Most acoustic modems are designed for point-to-point operation using a point-to-point protocol (PPP). Some series are also designed to support multiple access (MA), where two or more units can be operated within the same environment. In this case, the acoustic modems are usually operated sequentially and can be taken in and out of the water with minimal disruption to the other units. The most sophisticated acoustic modem now support routing capability between nodes, although this feature is rarely available commercially. In this case, one or more acoustic modems can operate together to relay messages over longer distances.
- *Compatibility requirements*: Interoperability between underwater acoustic modems produced by various manufacturers is still work in progress. Protocols and standards exist, but every manufacturer uses its own proprietary modem signaling and protocol. As a result, mixing and matching acoustic modems from different manufacturers is not recommended.

15.3.2 Commercially Available Modems

Figure 15.1 provides an overview of the nominal data rate versus range resulting from a large literature survey. In general, the high rate or high range results are for deep channels while the cluster of low range, low rate are for shallow channels. Modems developed by the research community are represented with diamonds, while stars denote commercially available systems. The range-times-rate bound represents an estimate of the existing performance envelope. While there are exceptions, most reviewed systems are bounded by this performance limit. Table 15.1 and Fig. 15.2 provide a few examples of state-of-the-art acoustic modems. The information is obtained from published technical specifications. Every acoustic communication system

Table 15.1 A few acoustic modems

Product name	Max bit rate [bps]	Range [m]	Frequency band [kHz]
Teledyne Benthos ATM-916-MF1 [15.15]	15360	6000	16–21
WHOI Micromodem [15.16]	5400	3000	22.5–27.5
Linkquest UWM 1000 [15.17]	7000	1200	27–45
Evologics S2C R 48/78 [15.18]	31200	2000	48–78
Sercel MATS 3G 34 kHz [15.19]	24600	5000	30–39
L3 Oceania GPM-300 [15.20]	1000	45000	not specified
Tritech Micron Data Modem [15.21]	40	500	20–28
FAU Hermes [15.22]	87768	180	262–375

presented here is of excellent quality and reputation and uses field-tested technologies. This short section cannot provide a complete description of the acoustic communications field. Many other acoustic communication systems have been developed over the past three decades that could not be included due to space limitations. Because underwater acoustic modems are very specialized pieces of equipment produced in limited numbers, standardizing hardware and protocols is a work in progress. Although most developers and manufacturers use proprietary hardware and protocols, a compact control language [15.23] has been developed to make optimal use of the limited acoustic modem data throughput installed on unmanned underwater vehicles. In addition, three frequency bands, common between a large number of underwater acoustic modems, have been identified. Finally, some efforts have been made in developing some level of inter-operability between underwater acoustic modems produced by different manufacturers. Although these efforts are encouraging, combining devices originating from different companies in the same network is usually not desirable.

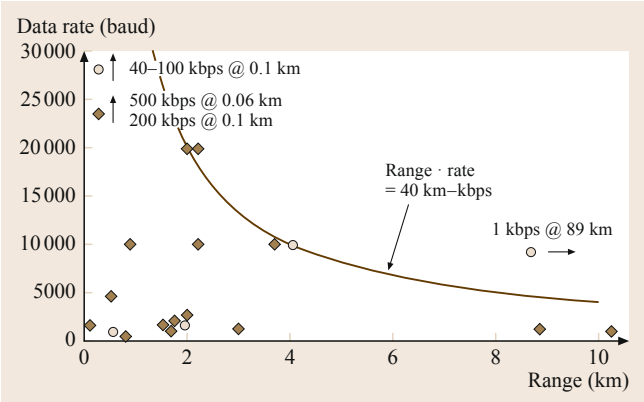


Fig. 15.1 Published experimental performance of underwater acoustic telemetry systems is summarized in this plot of rate [kbps] versus range [km] (after [15.24]). The channels vary from deep and vertical to shallow and horizontal

15.3.3 Field Tests

Field testing is a key step in evaluating the performance of an underwater acoustic communication system. Because of the extensive cost of such tests (especially in deep waters), numerous numerical models have been developed to predict the system performance and reduce the required in-water time for test and evaluation.

Underwater acoustic wave propagation is a complex combination of energy dissipation due to viscosity and chemical reactions, fluid–structure interaction, surface wave activity, biological activity, boat traffic, and source and receiver characteristics [15.25, 26]. The acoustic channel characteristics change with time, spa-

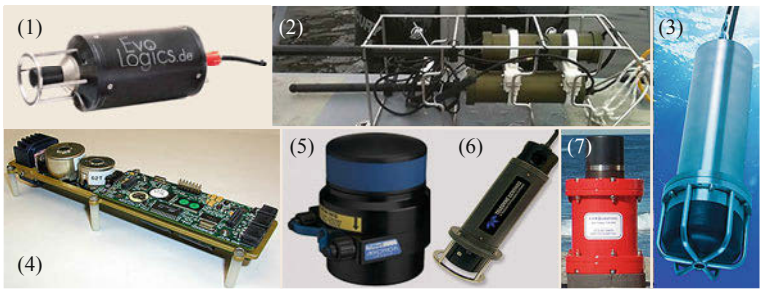


Fig. 15.2 A few examples of state-of-the-art underwater acoustic modems: (1) Evologics S2C R 48/78 (after [15.18]); (2) FAU Hermes (after [15.22]); (3) Sercel MATS 3G 34 kHz (after [15.19]); (4) WHOI Micromodem (after [15.16]); (5) Tritech Micron Data Modem (after [15.21]); (6) Teledyne Benthos ATM-916-MF1 (after [15.15]); (7) Linkquest UWM 1000 (after [15.17])

tial location, and frequency. The accurate representation of such a complex and dynamic system at any time, location, and frequency is simply not possible. As a result, even the most sophisticated models often display significant difference with measurements. Significant efforts are made to adapt the statistical models developed for radio communications, but much work remains in making these models sufficiently accurate.

Various scientific and governmental organizations (spearheaded by the United States Office of Naval Research), have developed series of comparative tests for existing technology and signal waveforms (Signal-ex, Modem-ex, KAM'11 [15.27]) and underwater acoustic networks (e.g., SeaWeb [15.28]). One key benefit of such tests is the careful monitoring of environmental characteristics during the test.

15.4 Propagation Channel

15.4.1 Attenuation and Noise

A distinguishing property of acoustic channels is the fact that the path loss depends on the signal frequency. This dependence is a consequence of absorption, i. e., transfer of acoustic energy into heat. In addition to the absorption loss, signal experiences a spreading loss which increases with distance. The overall path loss is given by [15.29]

$$A(l,f) = (l/l_{\text{ref}})^k \mathrm{a}(f)^{l-l_{\text{ref}}} , \tag{15.1}$$

where f is the signal frequency and l is the transmission distance, taken in reference to some l_{ref}. The path loss exponent k models the spreading loss, and its usual values are between 1 and 2, for cylindrical and spherical spreading, respectively. The absorption coefficient can be expressed empirically, using Thorp's formula, which gives $\mathrm{a}(f)$ in dB/km for f in kHz as [15.29]

$$10\log \mathrm{a}(f) = 0.11\frac{f^2}{1+f^2} + 44\frac{f^2}{4100+f^2} + 2.75\cdot 10^{-4} f^2 + 0.003 . \tag{15.2}$$

This formula is generally valid for frequencies above a few hundred Hz. The absorption coefficient is shown in Fig. 15.3. It increases rapidly with frequency, thus imposing a limit on the operational bandwidth that can be used for an acoustic link of a given distance.

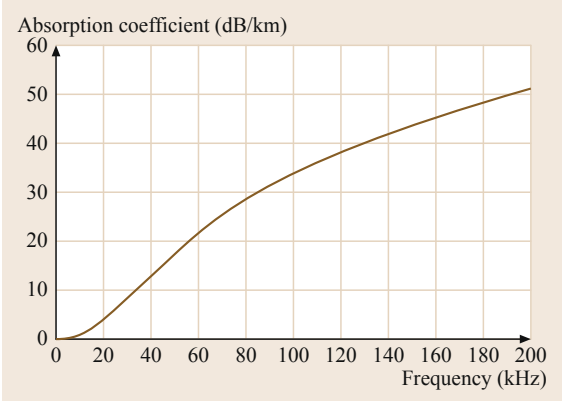


Fig. 15.3 Absorption coefficient, $10\log \mathrm{a}(f)$ in dB/km

Noise in an acoustic channel consists of ambient noise and site-specific noise. Ambient noise is always present in the background of the quiet deep sea. Site-specific noise, on the contrary, exists only in certain places. For example, ice cracking in polar regions creates acoustic noise, and so does shrimp snapping in tropical waters. Ambient noise comes from sources such as turbulence, distant shipping, and breaking waves, in addition to thermal noise. While this noise may be approximated as Gaussian, it is not white. The power spectral density (PSD) of the four noise components is given by the following empirical formulae in dB re μ Pa per Hz as a function of frequency in kHz

$$\begin{aligned} 10\log N_t(f) &= 17 - 30\log f \\ 10\log N_s(f) &= 40 + 20(s-0.5) + 26\log f \\ &\quad - 60\log(f+0.03) \\ 10\log N_w(f) &= 50 + 7.5 w^{1/2} + 20\log f \\ &\quad - 40\log(f+0.4) \\ 10\log N_{th}(f) &= -15 + 20\log f . \end{aligned} \tag{15.3}$$

Figure 15.4 shows the total PSD of the ambient noise for several values of the wind speed w (wind drives the surface waves which break and generate noise) and several levels of the distant shipping activity $s \in [0,1]$ (thousands of ships are present in the ocean at any time, and they generate distant noise, which is to be distinguished from the site-specific noise of vessels passing nearby). The noise PSD decays at a rate of approximately 18 dB/decade, as shown by the straight dashed line in Fig. 15.4. This line represents an approximate model for the noise PSD, $N(f) = N_0 \cdot (f/f_{\text{refx}})^{-\eta}$. The parameters N_0 and η can be fitted from the model, but also from the measurements taken at a particular site.

The attenuation, which grows with frequency, and the noise whose PSD decays with frequency, result in a signal-to-noise ratio (SNR) that varies over the signal

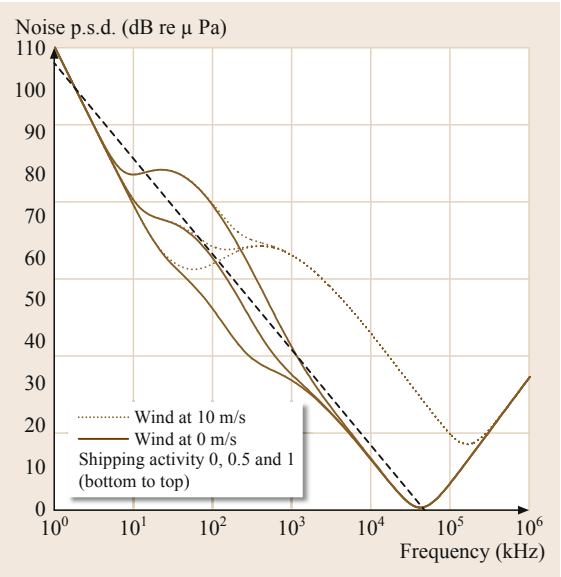


Fig. 15.4 Power spectral density of the ambient noise. The *dash-dotted* line shows an approximation $10\log N(f) = 50\,\text{dB re}\mu\text{Pa} - 8\log(f/1\,\text{kHz})$

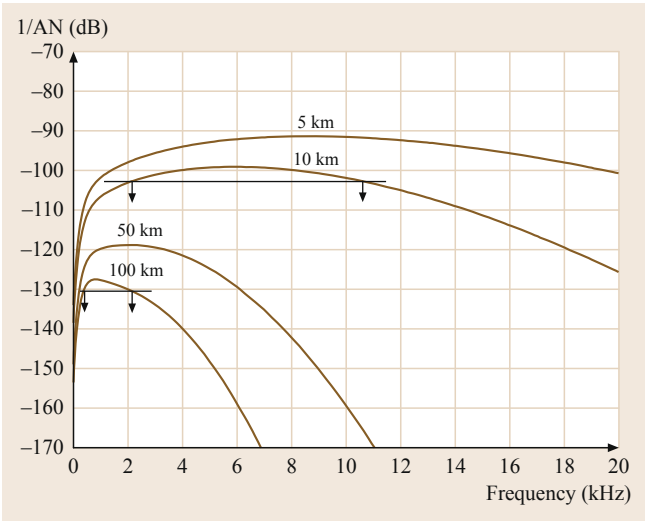


Fig. 15.5 Signal-to-noise ratio in an acoustic channel depends on the frequency and distance through the factor $1/A(l,f)N(f)$. This figure refers to the nominal SNR value that does not include variations induced by shadowing or multipath propagation

bandwidth. If one defines a narrow band of frequencies of width Δf around some frequency f, the SNR in this band can be expressed as

$$SNR(l,f) = \frac{S_l(f)\Delta f}{A(l,f)N(f)\Delta f}\,, \tag{15.4}$$

where $S_l(f)$ is the PSD of the transmitted signal. For any given distance, the narrowband SNR is thus a function of frequency, as shown in Fig. 15.5. From this figure it is apparent that the acoustic bandwidth depends on the transmission distance. In particular, the bandwidth and the power needed to achieve a prespecified SNR over some distance can be approximated as $B(l) = b \cdot l^{-\beta}$ and $P(l) = p \cdot l^{\psi}$, where the coefficients b, p, and the exponents $\beta \in (0,1)$, $\psi \geq 1$, depend on the target SNR, the parameters of the acoustic path loss, and the ambient noise [15.30]. The bandwidth is severely limited at longer distances: at 100 km, only about a kHz is available. At shorter distances, the bandwidth increases, but it will ultimately be limited by that of the transducer. The fact that the bandwidth is limited implies the need for bandwidth-efficient modulation methods if more than a bps/Hz is to be achieved over these channels.

Another important observation to be made is that the acoustic bandwidth B is often on the order of the center frequency f_c. This fact bears significant implications on the design of signal processing methods, as it prevents one from making the narrowband assumption, $B << f_c$, on which many radio communication principles are based. Respecting the wideband nature of the system is particularly important in multichannel (array) processing and in synchronization for mobile acoustic systems.

Finally, the fact that the acoustic bandwidth depends on the distance has important implications on the design of underwater networks. Specifically, it makes a strong case for multihopping, since dividing the total distance between a source and destination into multiple hops enables transmission at a higher bit rate over each (shorter) hop. Since multihopping also ensures lower total power consumption, its benefits are doubled from the viewpoint of energy-per-bit consumption on an acoustic channel [15.31]. Granted, by increasing the number of hops, the level of interference will increase, and packet collisions may become more likely, requiring, in turn, more re-transmissions. However, as shorter hops support higher bit rates, data packets containing a given *number* of bits will have shorter *duration* if the bit rate is appropriately increased, and the chances of collision will be reduced. This fact speaks further in favor of acoustic system design based on high-bandwidth, multihop links.

15.4.2 Multipath Propagation

While the basic propagation loss describes energy spreading and absorption, it does not take into account the specific system geometry and the resulting multipath propagation. Multipath formation in the ocean is

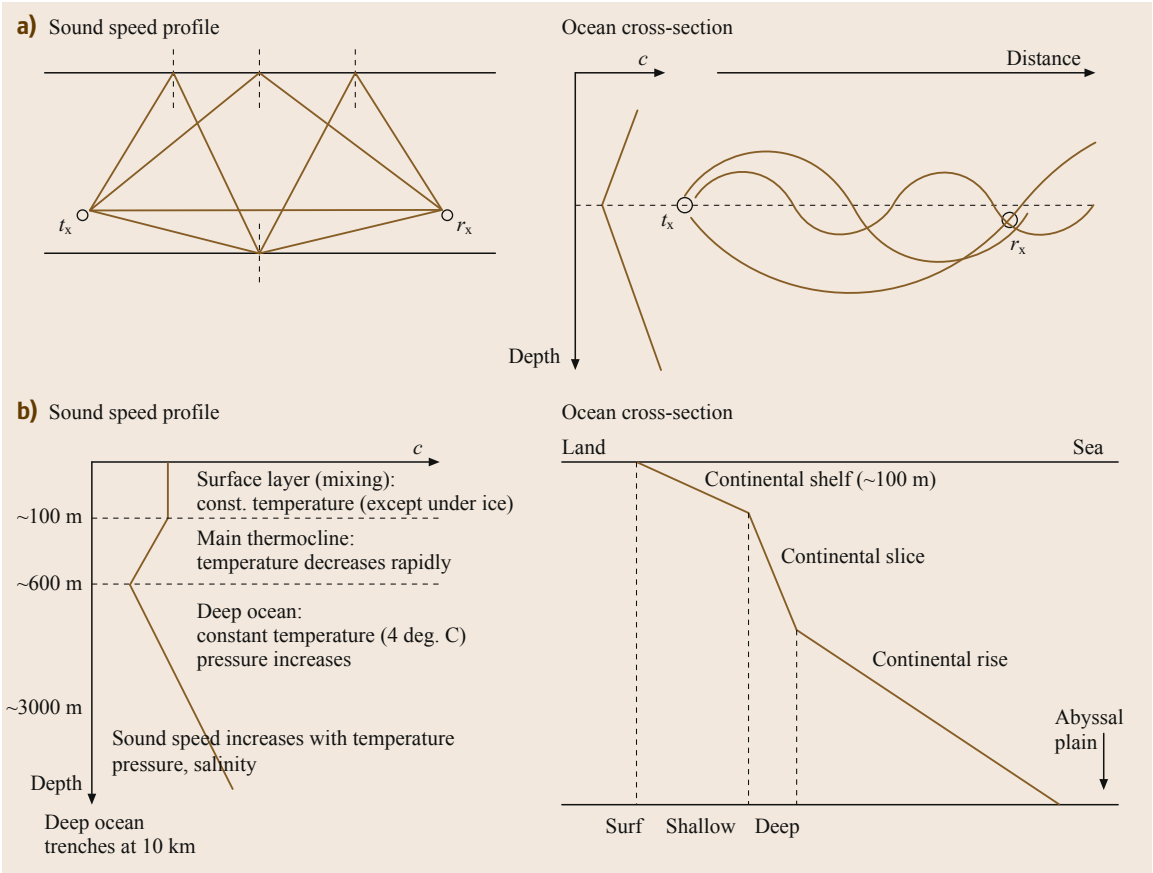


Fig. 15.6a,b Multipath formation in shallow and deep water **(a)**. Sound speed as a function of depth and the corresponding ocean cross-section **(b)**

governed by two effects: sound reflection at the surface, bottom and any objects, and sound refraction in the water. The latter is a consequence of sound speed variation with depth, which is mostly evident in deep water channels. Figure 15.6 illustrates the two mechanisms.

Sound speed depends on the temperature and pressure, which vary with depth, and a ray of sound always bends towards the region of lower propagation speed, obeying Snell's law. In shallow water, both the temperature and the pressure are constant, and so is the sound speed. As the depth increases, the temperature starts to decrease, but the weight of the water above is still not significant enough to considerably change the pressure. The sound speed thus decreases in the region called the main thermocline. After a certain depth, the temperature reaches the constant level of 4°C, and from there on the sound speed increases with pressure. When a source launches a beam of rays, each ray will follow a slightly different path, and a receiver placed at some distance will observe multiple signal arrivals. Note that a ray traveling over a longer path may do so at a higher speed, thus reaching the receiver before a direct, stronger ray. This phenomenon results in a nonminimum-phase channel response.

The first approach to modeling multipath propagation is a deterministic one, which provides an exact solution for the acoustic field strength in a given system geometry with a given sound speed profile. Because of computational complexity, approximate solutions are often used instead of the exact ones. An approximation that is suitable for frequencies of interest to acoustic communication systems is based on ray tracing. A widely-used package, Bellhop, is available online [15.32] (see also [15.33, 34]). Figure 15.7 illustrates a ray-trace obtained for a transmitter placed inside the circle to the left. Lighter colors in this figure indicate locations of higher received signal strength. If a receiver is placed at some distance away from the transmitter in this field, the signal strength will vary depending upon the *exact* location. In other words, two receivers (e.g., two circles to the right) placed at the *same* distance away from the transmitter, may experi-

ence propagation conditions that are quite different. In this example, shadowing is caused by ray bending in deep water. In shallow water with constant sound speed, signal strength can be calculated using simple geometrical considerations. Alternating constructive/destructive combining of multiple reflections will now form pockets of strong/weak signal reception. In either case, multipath effects make the signal strength location dependent, i. e., different from the value predicted by the basic propagation loss (15.1). If the transmitter or receiver were to move through this frozen-in-time acoustic field, the received signal strength would be perceived as varying. In practice, of course, neither the channel geometry nor the environmental parameters are frozen, and small disturbances cause the signal strength to vary in time even if there is no intentional motion of the transmitter or receiver.

The impulse response of an acoustic channel is thus influenced by the geometry of the channel and its reflection and refraction properties, which determine the number of significant propagation paths, their relative strengths and delays. Strictly speaking, there are infinitely many signal echoes, but those that have undergone multiple reflections and lost much of the energy can be discarded, leaving only a finite number of significant paths.

To put a mathematical channel model in perspective, let us denote by $\bar{l}_p$ the nominal length of the p-th propagation path, with $p = 0$ corresponding to a reference path (either the strongest, or the first significant arrival). In shallow water, where the sound speed c can be taken as constant, path delays can be obtained as $\bar{\tau}_p = \bar{l}_p / c$. The surface reflection coefficient equals -1 under ideal conditions, while bottom reflection coefficients depend on the type of bottom (hard, soft) and the grazing angle [15.35]. If we denote by Γ_p the cumulative reflection coefficient along the pth propagation path, and by $A(\bar{l}_p, f)$ the propagation loss associated with this path, then

$$\overline{H}_p(f) = \frac{\Gamma_p}{\sqrt{A(\bar{l}_p, f)}} \tag{15.5}$$

represents the nominal frequency response of the p-th path. Hence, *each path* of an acoustic channel acts as a low-pass filter. Figure 15.8 illustrates a reference path transfer function $\overline{H}_0(f) = 1/\sqrt{A(\bar{l}_0, f)}$.

For the distances and frequencies of interest to the majority of acoustic communication systems, the effect of path filtering is approximately the same for all paths, as the absorption coefficient does not vary much from its value at the reference distance and center frequency. Denoting this value by a_0, the transfer function of the

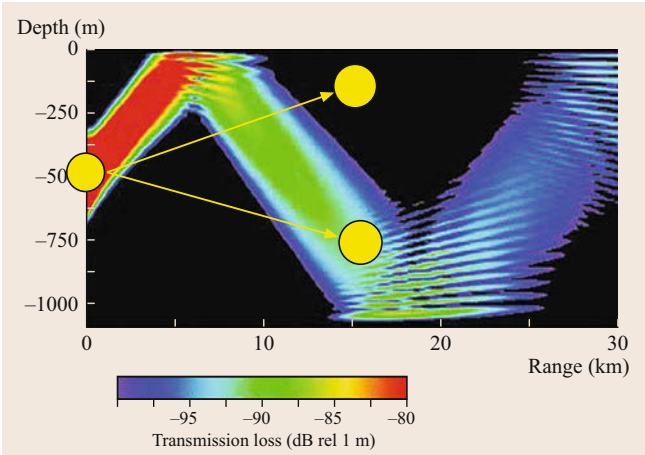


Fig. 15.7 A ray-trace shows areas of strong/weak signal reception. Sound reflection and refraction result in multipath propagation which favors some locations while placing others in a shadow

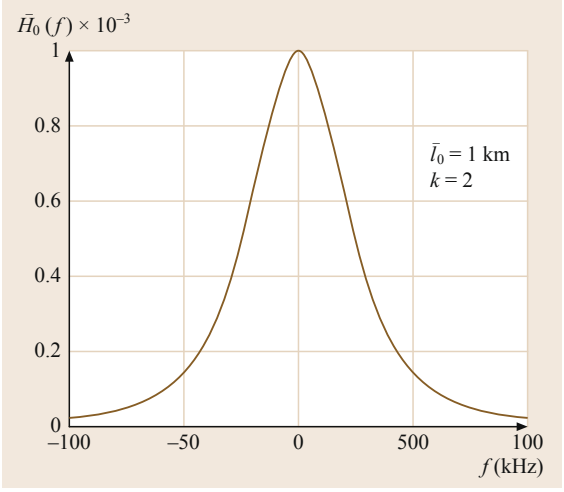


Fig. 15.8 Example of a reference path transfer function $\overline{H}_0(f)$

p-th path is modeled as

$$\overline{H}_p(f, t) = \overline{h}_p \cdot \overline{H}_0(f)\,, \tag{15.6}$$

where

$$\overline{h}_p = \frac{\Gamma_p}{\sqrt{(\bar{l}_p/\bar{l}_0)^k \mathsf{a}_0^{\bar{l}_p - \bar{l}_0}}} \tag{15.7}$$

is the real-valued (passband) path gain. Figure 15.9 shows an example of nominal multipath structure.

In an acoustic communication system, the channel can be observed only in a limited band of frequencies supported by the transducer. As a result, the observable response will be smeared, as illustrated in Fig. 15.10.

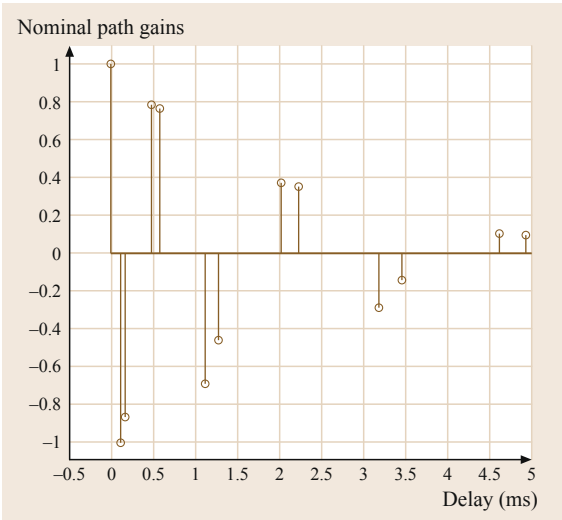


Fig. 15.9 Nominal path gains (15.7) for a channel with transmission distance $d = 1\,\text{km}$, water depth $h = 20\,\text{m}$, and transmitter/receiver height above the bottom $h_T = 10\,\text{m}$ and $h_R = 12\,\text{m}$. The surface reflection coefficient is -1; bottom reflection coefficients are calculated assuming speed of sound and density of $c = 1500\,\text{m/s}$, $\rho = 1000\,\text{kg/m}^3$ in water and $c_b = 1300\,\text{m/s}$, $\rho_b = 1800\,\text{kg/}m^3$ in the soft bottom. Those paths whose absolute value is above one tenth of the strongest path's are shown

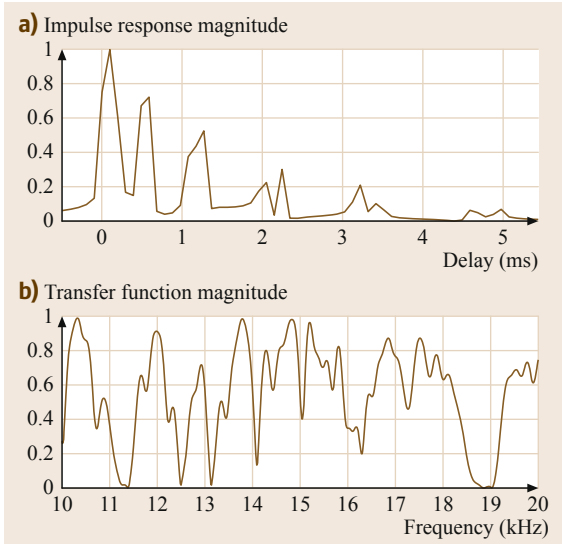


Fig. 15.10 **(a)** Impulse response (baseband) and **(b)** the transfer function for the nominal (time-invariant) channel geometry corresponding to the multipath structure of Fig. 15.9, as seen in $B = 10\,\text{kHz}$ of bandwidth centered at $f_c = 15\,\text{kHz}$. In a wideband acoustic system, multipath propagation creates discernible signal echoes and a frequency-selective channel distortion

It is important to make the distinction between the natural channel *paths* of the channel (Fig. 15.9), and the samples of the observable response (Fig. 15.10). The latter are often referred to as the channel *taps*. It is also important to note that the path gains can be very stable, while taps may exhibit fast variation. Such behavior is explained by the fact that each tap contains contributions from *all* the paths, wherein each path carries a pulse of finite duration and a different delay that varies in time. An example of tap variation is shown in Fig. 15.13, which we will discuss later when we address channel fading.

Multipath propagation creates a frequency-selective channel distortion, which is evident as the variation of the transfer function in Fig. 15.10. As a result, some frequencies of a wideband acoustic signal will be favored, while others will be severely attenuated. Channel equalization has to be employed to deal with this type of distortion.

The total delay spread between significant multipath arrivals or the *multipath spread* of the channel, T_{mp}, is an important figure of merit for the design of acoustic communication systems. Vertical links exhibit short multipath spreads, while horizontal channels exhibit multipath spreads that can range between a few milliseconds and hundreds of milliseconds. In single-carrier broadband systems, where the data symbols are short compared to the multipath spread, the meaningful figure of merit is the number of symbols that fit within the multipath spread. This number determines the extent of inter-symbol interference (ISI), which in turn dictates the length of the filters needed to equalize the channel. The fact that the multipath spread may be on the order of several milliseconds, or more, implies that the ISI may span tens or even a hundred symbol intervals; a situation very different from that typically found in radio systems, where ISI may involve a few symbols only. Multicarrier systems avoid this problem by transmitting in parallel on many carriers, each occupying a narrow subband. A meaningful figure of merit for these systems is the frequency coherence, $\Delta f_{\text{coh}} = 1/T_{\text{mp}}$. In a properly designed multicarrier system, carrier separation is kept well below this value.

The acoustic channel is often *sparse*, i. e., there are only several significant paths that populate the *total* and possibly long delay spread. This fact has an important implication on channel estimation and the associated equalization methods. Namely, while the entire multipath spread is represented by $L \approx T_{\text{mp}}B$ samples taken at intervals $T_s = 1/B$, fewer than L coefficients may suffice to represent the channel response. Ideally, only as many coefficients as there are propagation paths, $P < L$,

are needed. Channel modeling thus becomes an important aspect of signal processing, and sparsing has been investigated for decision-feedback equalization [15.36, 37], turbo equalization [15.38, 39], and multicarrier detection [15.40–42]. It is also important to note that although a greater bandwidth implies more ISI and more frequency selectivity, it also implies a better resolution in delay (less smearing in the observable channel response). Hence, although the attendant signal distortion is perceived as more severe, channel estimation will be more efficient if a proper sparse model is used, which may, in turn, lead to improved signal processing. In addition, signaling at a higher rate enables more frequent channel observations and, consequently, easier channel tracking [15.43].

15.4.3 Time Variability: Motion-Induced Doppler Distortion

There are two sources of the channel's time variability: inherent changes in the propagation medium, which contribute to signal fading, and those that occur because of the transmitter/receiver motion and contribute to the frequency shifting. The distinction between the two types of distortions is that the first one appears as random, while the motion-induced Doppler effects can be modeled in a deterministic manner and, therefore, compensated to some degree through synchronization.

Motion of the transmitter, receiver, or a reflection point along the signal path causes the path distances to vary with time. The resulting Doppler effect is evident as time compression/dilation of the signal, which causes frequency shifting and bandwidth spreading/shrinking. The magnitude of the Doppler effect is proportional to the ratio $a = v/c$ of the relative transmitter-receiver velocity to the speed of sound. Because the speed of sound is very low as compared to the speed of electromagnetic waves, motion-induced Doppler distortion of an acoustic signal can be extreme. AUVs move at speeds that are on the order of a few m/s, but even without intentional motion, underwater instruments are subject to drifting with waves, currents, and tides, which may occur at comparable velocities. In other words, there is always *some* motion present in the system, and a communication system has to be designed taking this fact into account. The only comparable situation in radio communications occurs in low Earth orbiting (LEO) satellite systems, where the relative velocity of satellites flying overhead is extremely high (the channel there, however, is not nearly as dispersive in delay). The major implication of motion-induced distortion is on the design of synchronization algorithms.

To model the Doppler distortion, let us focus on a single propagation path, assuming that the relative transmitter/receiver velocity v_p along this path stays constant over some interval of time. The path delay can then be modeled as

$$\overline{\tau}_p(t) = \overline{\tau}_p - a_p \cdot t \,, \tag{15.8}$$

where $a_p = v_p/c$ is the Doppler factor corresponding to the p-th path. If there is a single dominant component of velocity, these factors can be approximated as equal for all paths, $a_p = a, \forall p$; however, this is not the case in general [15.37]. To assess the effect of time-varying delay on the signal, let us focus on a particular signal component centered around frequency f_k in a narrow band $\Delta f \ll f_k$. The signal $s_k(t)$ and its equivalent baseband $u_k(t)$ are related by

$$s_k(t) = Re\{u_k(t)e^{i2\pi f_k t}\} \,. \tag{15.9}$$

Since the signal occupies only a narrow band of frequencies, its replica received over the p-th propagation path is given by

$$s_{k,p}(t) = \overline{h}_p \bar{H}_0(f_k) s_k(t - \overline{\tau}_p(t)) \,. \tag{15.10}$$

Defining the equivalent baseband component of the received signal via

$$s_{k,p}(t) = Re\{v_{k,p}(t)e^{i2\pi f_k t}\} \,, \tag{15.11}$$

we have the baseband relationship

$$v_{k,p}(t) = \overline{c}_{k,p} e^{i2\pi a_p f_k t} u_k(t + a_p t - \overline{\tau}_p) \,, \tag{15.12}$$

where

$$\overline{c}_{k,p} = \overline{h}_p \overline{H}_0(f_k) e^{-i2\pi f_k \overline{\tau}_p} \,. \tag{15.13}$$

The Doppler effect is evident in two factors: (i) frequency shifting by the amount $a_p f_k$ and (ii) time scaling by the factor $(1 + a_p)$. The situation of putting together all the signal components at frequencies f_0, f_1, etc., that span a wide bandwidth $B = K\Delta f$, and assuming only a single propagation path with the Doppler factor a is illustrated in Fig. 15.11. This figure offers an exaggerated view of the Doppler effect, but nonetheless one that illustrates the point: each spectral component in a wideband mobile acoustic system is shifted by a *different* amount, and the total bandwidth B is spread over $B(1 + a)$, i. e., a signal of duration T is observed at the receiver as having duration $T/(1 + a)$.

The way in which these distortions affect signal detection depends on the actual value of the factor a. For comparison, let us look at a highly mobile radio system: at 160 km/h (100 mph), $a = 1.5 \cdot 10^{-7}$. This value is low enough that compression/dilation can be neglected.

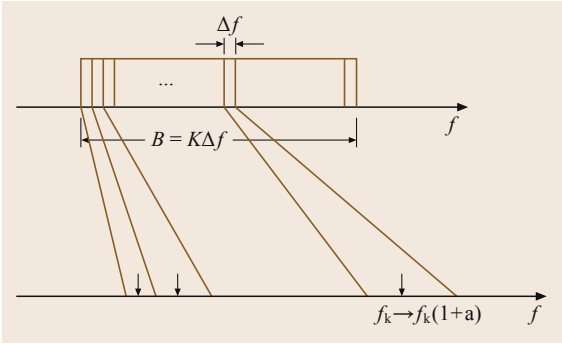


Fig. 15.11 Motion-induced Doppler shift is not uniform in a wideband acoustic system

In other words, there is no need to account for it explicitly in symbol synchronization. The error made in doing so is only 10^{-3} of a bit per 10 000 bit. In contrast to this situation, a stationary acoustic system may experience unintentional motion at 0.5 m/s (1 kt), which would account for $a = 3 \cdot 10^{-4}$. For an AUV moving at several m/s (submarines can move much faster), the factor a will be on the order of 10^{-3}, a value that cannot be ignored.

Nonnegligible motion-induced Doppler shifting and spreading thus emerge as another major factor that distinguishes an acoustic channel from the mobile radio channel, and dictates the need for explicit phase and delay synchronization in all but stationary systems [15.10]. In multicarrier systems, the Doppler effect creates a particularly severe distortion. Unlike in radio systems, where time compression/dilation is negligible and the Doppler shift appears equal for all frequencies within the signal bandwidth, in an acoustic system each frequency component may experience a markedly different Doppler shift, creating a nonuniform Doppler distortion across the signal bandwidth [15.44].

15.4.4 Time Variability: Random Effects (Fading)

Inherent channel changes range from the very large-scale, slow variations that occur seasonally (e.g., the change in sound speed profile from winter to summer) or daily (e.g., the change of water depth due to tides) to the small-scale, fast variations that are often caused by the rapid motion of the sea surface (waves) or the system itself. Many of these variations appear as random, and as such require an additional stage of modeling, namely statistical modeling.

The distinction between various scales is meaningful from the viewpoint of the duration of communication signals. While large-scale phenomena affect *average* signal power, causing it to vary over longer periods of time (many data packets), small-scale phenomena affect the *instantaneous* signal level, causing it to vary over shorter periods of time (few data packets, or bits of the same data packet). In light of system design, it is also useful to distinguish between those variations that are slow enough to admit a feedback by which the receiver can inform the transmitter of the change, and those that are too fast to be conveyed over the time it takes the signal to propagate from one end of the link to the other. In this sense, modeling of large-scale phenomena is meaningful for adaptive power control and modulation (which are performed at the transmitter in response to delayed feedback), while modeling of small-scale phenomena is important for adaptive signal processing at the receiver (channel estimation, equalization). It is also important to note that the apparently random channel variation can be a consequence of both the changes in the physical parameters of the environment (sound speed profile, surface scattering, internal waves) and the changes in the placement of the system within a given environment (transmitter/receiver displacement due to intentional motion or drifting with currents, etc.). Note also that these effects will differ across different environments, e.g., deep or shallow water, tropical, or ice-covered regions.

Large-Scale Effects

To assess the random variation of the channel, one can start by allowing the path lengths to deviate from the nominal values, so that

$$l_p = \bar{l}_p + \Delta l_p \,, \tag{15.14}$$

where Δl_p is regarded as a random displacement. This displacement models the large-scale variation which occurs because the channel geometry deviates from the nominal, either as the system is moved to a different location, or because the water depth has changed with the tide, etc. The deterministic, motion-induced delay variation which can be compensated through proper synchronization, is not considered as part of statistical channel analysis. The resulting transfer function of the channel, which is now *random* as well, is given by

$$H(f) = \overline{H}_0(f) \sum_p h_p e^{-i2\pi f \tau_p} \,, \tag{15.15}$$

where the path coefficients h_p can be calculated similarly as for the nominal case (15.7), using l_p instead of $\bar{l}_p$, and the delays can be calculated as $\tau_p = l_p/c - t_0$ in reference to some time $t_0 = \bar{l}_0/c$. The corresponding random path gains can also be approximated as [15.45]

$$h_p = \overline{h}_p e^{-\xi_p \Delta l_p/2} \,, \tag{15.16}$$

where

$$\xi_p = a_0 - 1 + k/\bar{l}_p \,, \tag{15.17}$$

and $\bar{h}_p$ is given by the expression (15.7). Note that one can also explicitly express the salient parameters h_p, τ_p as functions of time, i. e., $h_p(t)$, $\tau_p(t)$, if one is interested in their variation during a long-term deployment in the *same* area (deep/shallow water, etc.).

Small-Scale Effects

The above model accounts only for the variations induced by the path length displacements, i. e., it does not include the effects of scattering. Scattering, which causes micro-path dispersion along each propagation path, can be modeled through an additional factor that accompanies the path gain $h_p(t)$. An overall channel model is now obtained in the form

$$H(f,t) = \bar{H}_0(f) \sum_p \gamma_p(f,t) h_p(t) e^{-i2\pi f \tau_p(t)} \,. \tag{15.18}$$

In this model, the factors $h_p(t)$ account for the slower-changing process caused by deviations of the system geometry from the nominal, while the factors $\gamma_p(f,t)$ represent the faster-changing scattering process. In other words, during a typical communication transaction of several packets, the path gains can be considered fixed, $h_p(t) = h_p$, but the additional factors $\gamma_p(f,t)$ cannot. Assuming a relatively large number of micro-paths, these factors can be modeled as complex-valued Gaussian processes, with some mean and variance. In general, scattering in an acoustic channel depends on the signal frequency, making each macro-path experience a different, possibly independent filtering effect. Note, however, that the transfer function $H(f,t)$ will in general exhibit correlation in both frequency and time (as well as across the elements of a transmit/receive array).

Statistical Characterization

Because of the different dynamics of the two types of processes, statistical channel analyses are typically conducted in two forms: one that targets the locally-averaged received signal power, and another that targets the instantaneous channel response. Specifically, large-scale statistical modeling focuses on the channel gain defined as

$$G(t) = \frac{1}{B} \int_{f_c - B/2}^{f_c + B/2} E_\gamma \left\{ \left| H^2(f,t) \right| \right\} \mathrm{d}f \,, \tag{15.19}$$

where the expectation is taken over small-scale fading. In contrast, small-scale modeling focuses on the processes $\gamma_p(f,t)$, as well as the effects of random Doppler shifting.

A complete statistical model for either large or small-scale phenomena must specify the probability density function (PDF) and the power spectral density (PSD), or the correlation properties of the random process of interest. These functions can be assessed using an analytical approach that relies on a mathematical model such as the one outlined above, a repeated application of a propagation model such as the Bellhop ray tracer for a large number of slightly disturbed channel conditions, or an experimental data analysis. All approaches have been considered in the literature, which abounds in statistical models that were found to adequately match the channels observed in different locations, at different time scales, and in different frequency bands. As of this time, however, the jury is still out on establishing standard models that would concisely describe the statistics of *typical* underwater acoustic channels. Crucial to that effort is identification of the scales of stationarity for the different processes at hand.

As far as the large-scale modeling goes, recent experimental results offer some evidence in support of a log-normal model for the channel gain [15.45–47]. Figure 15.12 illustrates such a model. Several of these studies also find that the path gains, and consequently the overall large-scale channel gain, can be well modeled as auto-regressive processes, at least on the time scales of several seconds, which are meaningful for implementation of packet-level feedback schemes.

Experimental studies on small-scale fading have offered evidence of Ricean [15.46, 48, 49] as well as Rayleigh phenomena [15.50]. Figure 15.13 illustrates an experimental data set that exhibits traits of Ricean fading. Combining the effects of large and small-scale fading leads to a mixture of log-normal and Ricean distributions, which can be approximated in closed form by the compound K-distribution. Experimental evidence to this effect is provided in [15.51]. Less is known about modeling the time-correlation functions of the small-scale processes, but it is generally understood that coherence times on the order of 100 ms can be assumed for a general-purpose design. Frequency-correlation, as well as space-correlation, as they pertain to communications systems, also remain to be addressed in a systematic manner.

The importance of statistical channel modeling is twofold. On the one hand, it enables the design of practical adaptive modulation methods which rely on identifying those channel parameters that can be *predicted* via delayed receiver-transmitter feedback [15.52]. On the other hand, it will enable the development standardized simulation models, thus allowing performance analy-

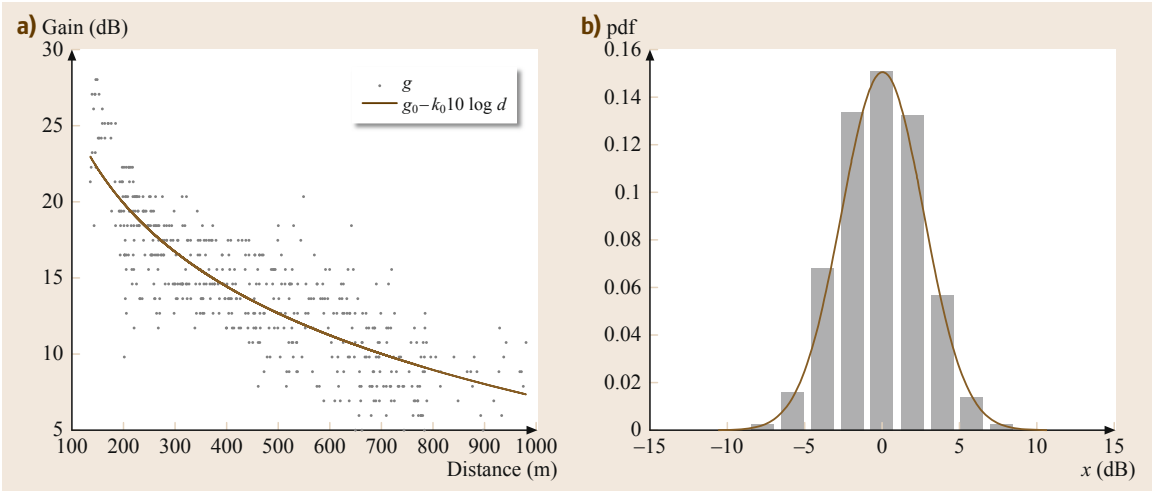


Fig. 15.12a,b During an experiment off the coast of Northern California, an acoustic signal occupying the 8–12 kHz band was transmitted between a fixed station and an AUV, and the strength of the signal received by the WHOI micro-modem was measured over a period of time during which the transmission distance changed. **(a)** shows the recorded signal strength, an estimate of the channel gain (15.19), as a function of distance (*dots*) and the accompanying trend which exhibits a log-distance dependence (*solid*). **(b)** shows the histogram of the difference between the recorded gain and the nominal trend, which is well described by a Gaussian distribution on the dB scale, i. e., a log-normal distribution on the linear scale

sis not only of signal processing methods, but also of network protocols and system-level design strategies, without the need for actual system deployment. Currently available simulators, such as the World Ocean Simulation System [15.53], use accurate ray tracing models, but they lack the statistical aspect of assessing the system performance. Statistical channel modeling was addressed recently in [15.45].

15.4.5 System Constraints

In addition to the fundamental limitations imposed by acoustic propagation, there are hardware constraints that affect the operation of acoustic modems. The most obvious of these constraints is the fact that acoustic transducers have their own bandwidth limitation, which constrains the available bandwidth beyond that offered by the channel. Additional constraints include power efficiency (transmitter amplifiers are often nonlinear, imposing a limit on the *peak* signal power), and the fact that acoustic modems operate in half-duplex fashion. These constraints do not only affect the physical link but all the layers of a network architecture. For example, the half-duplex mode of operation, combined with the low speed of sound, challenges the throughput efficiency of acoustic links that require automatic repeat request (ARQ) [15.54].

In an acoustic system, the power required for transmitting may be much greater than the power required for receiving. Transmission power depends on the distance, and its typical values are on the order of tens of Watts. (An acoustic signal propagates as a pressure wave, whose power is measured in Pascals, or commonly in dB relative to a micro Pascal. In seawater, 1 W of radiated acoustic power creates a sound field of intensity 172 dB re μPa 1 m away from the source.) In contrast, the power consumed by the receiver is much lower, with typical values ranging from about 100 mW for listening or low-complexity detection, to no more than a few Watts required to engage a sophisticated processor for high-rate signal detection. In hibernation mode, from which a modem can be woken on command, no more than 1 mW is required.

Underwater instruments are often battery-powered, and, hence, it is not simply the power, but the energy consumption that matters. This is less of an issue for mobile systems, where the power used for communication is a small fraction of the total power consumed for propulsion, but it is important for networks of fixed bottom-mounted nodes, where the overall network lifetime is the figure of merit. One way to save energy is by transmitting at a higher bit rate. For example, the WHOI modem [15.55] has two modes of operation: high rate at 5 kbps and low rate at 80 bps. This modem will require about 60 *times* less energy per bit (18 dB) in the high-rate mode. The receiver's energy consumption will also be lower, although it requires 3 W for detection of high-rate signals as opposed to 80 mW for detection of

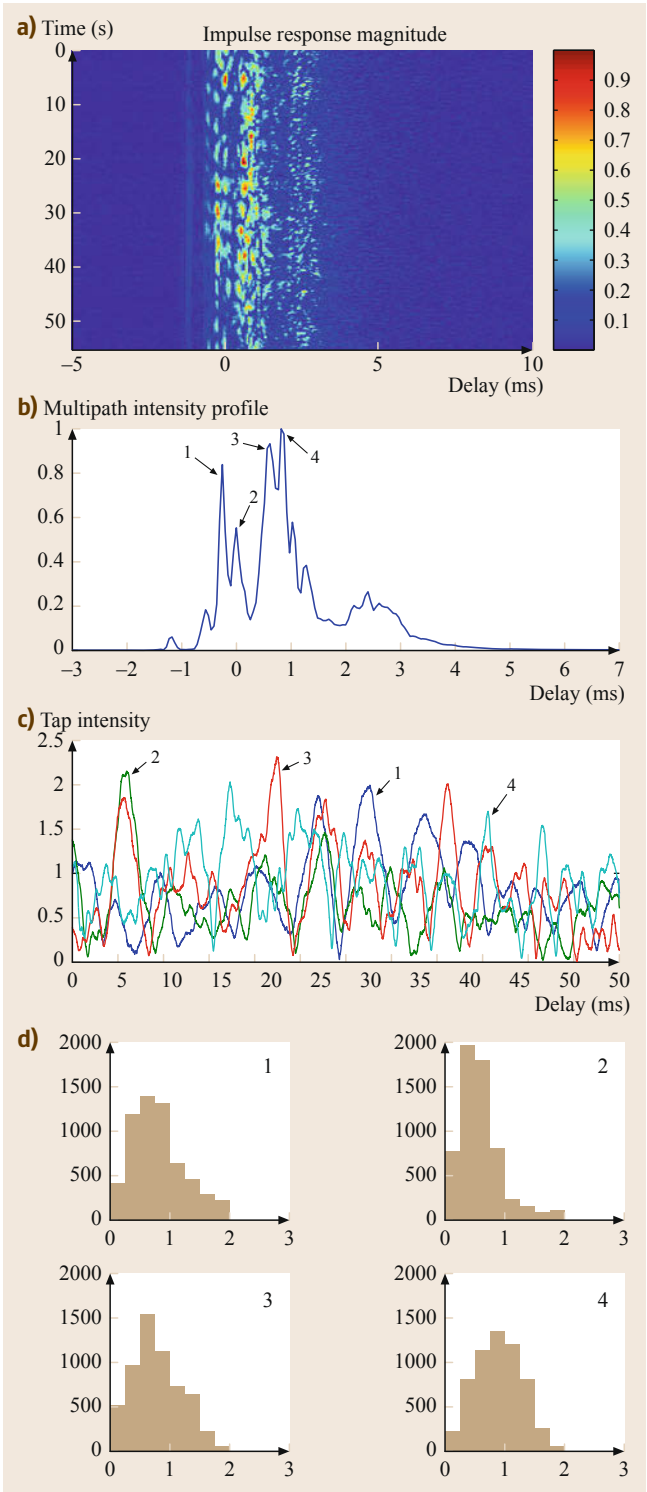


Fig. 15.13a–d During an experiment that took place in Narragansett Bay, near the coast of Rhode Island, the transmitter and the receiver were mounted on fixed tripods at 4 and 2 m above the bottom, at a distance of 1 km, with the channel depth ranging between 9 and 14 m. The sea condition was calm. A pseudo-random binary sequence of length 4095, modulated onto a 13 kHz carrier, was transmitted repeatedly at a rate of 10 kbps. The received signal was digitally down-converted, and correlated with the replica of the transmitted signal to yield an estimate of the channel response. Channel responses (magnitude) are shown in **(a)**. The multipath intensity profile **(b)** represents the average of magnitude-squared channel responses observed from 132 consecutive sequences. Shown also is the time-variation of several channel *taps* **(c,d)**. These figures reveal a channel consisting of multipath arrivals whose amplitude varies in time in a such a manner that there is no clearly defined strongest arrival. The phase (not shown) exhibits a random behavior around a constant slope, corresponding to a small but different Doppler shift on each of the arrivals, while the histograms of the tap magnitudes indicate Ricean fading

low-rate signals (the difference is about 2 dB). Another way to save the energy is by minimizing the number of retransmissions (we discuss the relevant protocols in Sect. 15.5.6). Finally, energy spent in idle listening over a longer period of time is not to be neglected. Having modems that are capable of hibernation, or instituting intentional sleeping schedules for those that are not, can help to control this type of power expenditure [15.56].

15.5 Point-to-Point Links: Signal Processing

15.5.1 Noncoherent Modulation/Detection

Noncoherent signaling is commonly used to relay data and to transmit command-and-control information at low bit rates. Frequency-hopped m-ary frequency-shift-key (FH-MFSK) modulation is widely accepted as a reliable albeit inefficient choice for such modems. FSK modulation uses distinct tonal or frequency-modulated pulses to map digital information. Frequency-hopping is a spread-spectrum technique, which alternates portions of the frequency band over fixed periods of time, so that successive FSK-modulated symbols use a different portion of the frequency spectrum, thus minimizing inter-symbol-interference (ISI) caused by reverberation. Many commercial suppliers offer several versions of FH-MFSK acoustic modems. In the case of a combination of N tones transmitted simultaneously, the symbol transmitted during the m-th signaling interval ($mT_s \le t < (m+1)T_s$) is

$$s_m(t) = A_m d(t) \sum_{n=1}^{N} g_{m,n} c_{m,n}(t), mT_s \le t < (m+1)T_s\,, \tag{15.20}$$

where, in the simplest case, $c_{m,n}(t)$ is a tone (either real or complex),

$$c_{m,n}(t) = e^{2\pi i f_{m,n}(t-mT_s)}, mT_s \le t < (m+1)T_s\,. \tag{15.21}$$

A_m is the peak acoustic pressure of symbol $s_m(t)$, $d(t)$ is the pulse shape (e.g., Gaussian), T_s is the pulse duration in seconds, $f_{m,n}$ is the carrier frequency of tone n for symbol $s_m(t)$. The binary information $b(m)$ is coded in symbol $s_m(t)$ as a combination $g_{m,n}$ of frequency bands centered at $f_{m,n}(n = 1, \ldots, N)$. $g_{m,n} = 1$ if the frequency band is used, otherwise $g_{m,n} = 0$ (at least one band is used for every symbol $s_m(t)$). If $s_m(t)$ uses only one of N bands and N is a power of 2, $\log_2(N)$ bits can be coded simultaneously within a single symbol. For example, if $N = 4$, $s_m(t)$ will contain $log_2(4) = 2$ bits of binary information ($b(m)$ =(0,0),(0,1),(1,0) or (1,1)). If $g_{1,n} = 1$ and $g_{2,n} = g_{3,n} = g_{4,n} = 0$, then $b_m = (0, 0)$. If $g_{2,n} = 1$ and $g_{1,n} = g_{3,n} = g_{4,n} = 0$, then $b_m = (0, 1)$, and so on. Figure 15.14 shows a section of an FH-MFSK modulated signal and the corresponding periodogram.

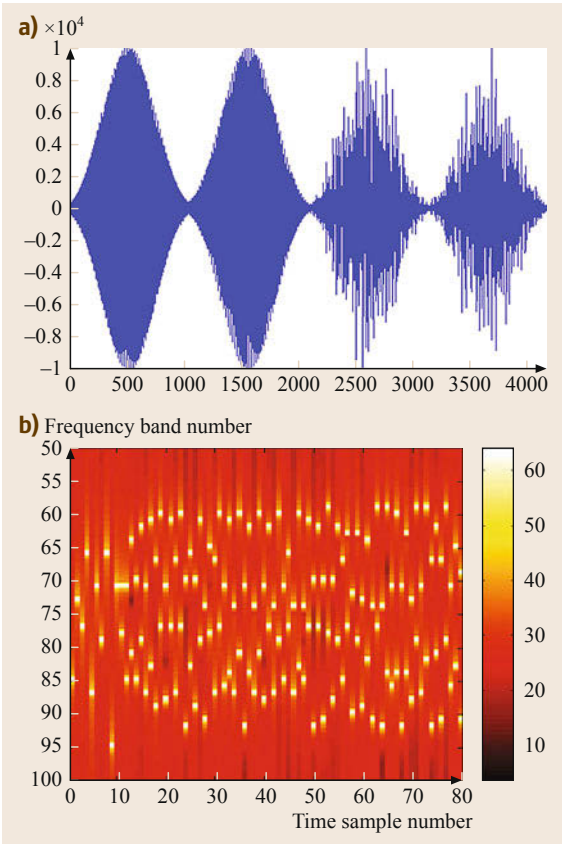


Fig. 15.14 (a) The first four symbols of a FH-MFSK sequence. (b) Spectrogram of a FH-MFSK sequence

At the receiver, the incoming message is first detected and synchronized using a short FH-MFSK modulated sequence. Following this, each data symbol is processed individually. First, the cross-correlation between the incoming faded symbol and a set of reference symbols $c_{m,n}(t)$ is computed,

$$R_{r_m,c_{m,n}}(\tau) = \int_{-T_s}^{T_s} r_m(t) c^*_{m,n}(t+\tau)\mathrm{d}t\,, \tag{15.22}$$

where * denotes the complex conjugate. For every received symbol $r_m(t)$, we retain the peak of correlation

within each band centered at $f_{m,n}$,

$$Z_{m,n} = \max_t \left\{ R_{r_m,c_{m,n}}(\tau); mT_s \le t < (m+1)T_s \right\} . \tag{15.23}$$

If we assume once more that $s_m(t)$ uses only one of N bands and N is a power of 2, we identify the frequency band that contains the most energy by comparing the coefficients $Z_{m,n}$,

$$n_{\max}(m) = \max_n \{Z_{m,n}; n = 1, \ldots, N\} . \tag{15.24}$$

Note that $n_{\max}$ is a function of the symbol index m. Under the same assumption, $g_{m,n} = 1$ if $n = n_{\max}$, otherwise $g_{m,n} = 0$. For example, if $n_{\max} = 1$, then $g_{m,1} = 1$ and $g_{m,2} = g_{m,3} = g_{m,4} = 0$. Therefore, $s_m(t)$ contains the couple of bits $b_m = (0, 0)$. If $n_{\max} = 2$, then $g_{m,2} = 1$ and $g_{m,1} = g_{m,3} = g_{m,4} = 0$, so that $s_m(t)$ contains the couple of bits $b_m = (0, 1)$, and so on.

15.5.2 Coherent Modulation/Detection

Coherent modulation methods include phase shift keying (PSK) and quadrature amplitude modulation (QAM). These methods offer bandwidth efficiency, i. e., the possibility to transmit more than 1 bit per second per Hertz of occupied bandwidth. However, because the information is encoded into the phase of the signal, precise knowledge of the received signal's frequency and phase is required in order to perform coherent detection. This fact presents a major challenge because an acoustic channel introduces a rather severe phase distortion on each of its multiple paths. A coherent receiver thus needs to perform phase synchronization together with channel equalization.

Coherent systems fall into two types: single-carrier and multicarrier systems. In single-carrier systems, a broadband information-bearing signal is directly modulated onto the carrier and transmitted over the channel. A typical high-rate acoustic signal occupies several kHz of bandwidth over which it experiences uneven channel distortion (Fig. 15.10). This distortion must be compensated at the receiver through the process of equalization. Multicarrier modulation bypasses this problem by converting the high-rate information stream into many parallel low-rate streams, which are then modulated onto separate carriers. The carriers are spaced closely enough such that the channel appears as frequency-flat in each narrow subband. After demodulation, each carrier's signal now only has to be weighted and phase-synchronized, i. e., a single-coefficient equalizer suffices per carrier. Each of these methods has its advantages and disadvantages when it comes to practical implementation: single-carrier systems are capable of faster channel tracking but they need high-maintenance equalizers; multicarrier systems are efficiently implemented using the fast Fourier transform (FFT), but they have high sensitivity to residual frequency offsets. In what follows, we review the basics of both types of systems as they apply to phase-distorted frequency-selective acoustic channels, and we outline the current research trends.

Single-Carrier Modulation/Detection

The problem of joint phase tracking and equalization in single-carrier broadband systems was addressed in [15.10]. The basic receiver structure [15.10] was extended to multichannel (array) configuration [15.57, 58] to enable diversity combining, which is often essential on acoustic channels. The resulting multichannel decision-feedback equalizer (DFE), integrated with a second-order phase-locked loop (PLL), provided the benchmark design for the high-speed acoustic modem [15.55]. It also became a de-facto standard for the development and comparison of future techniques, which focused on replacing the basic DFE with more sophisticated equalization and channel estimation schemes (e.g., turbo equalization and sparse channel estimation [15.39]), or inclusion of multiple transmitters, be it in the form of multiuser systems [15.59, 60] or multi-input multi-output (MIMO) structures for spatial multiplexing and/or diversity [15.38, 61].

The block diagram of a multichannel DFE is shown in Fig. 15.15. The receiver includes several stages: precombining (which may or may not be used, depending upon the array structure), phase correction, and multichannel equalization. The receiver is designed to process an incoming signal of the form

$$v(t) = \sum_n d(n) g_c(t - nT) e^{i\theta(t)} + w(t) , \tag{15.25}$$

where $d(n)$ are the data symbols transmitted at the rate of one symbol per T; $g_c(t) = g_T(t) * c(t) * g_R(t)$ is the composite impulse response of the transmitter, channel, and receiver (all in equivalent baseband with respect to the carrier frequency f_c); $\theta(t)$ is the carrier phase, and $w(t)$ is the noise. The channel response $c(t)$ and the phase $\theta(t)$ are unknown (and possibly different across the array elements), necessitating an adaptive receiver. Adaptive equalization has been considered in two forms: direct, which does not perform explicit channel estimation, and indirect, in which the channel is first estimated and then used to adjust the equalizer. In either form, the data symbols are estimated as

$$\hat{d}(n) = \sum_{p=1}^{P} \boldsymbol{a}_p^H(n) \boldsymbol{x}_p(n) e^{-i\hat{\theta}_p(n)} - \sum_{i>0} b_i^*(n) d(n-i) , \tag{15.26}$$

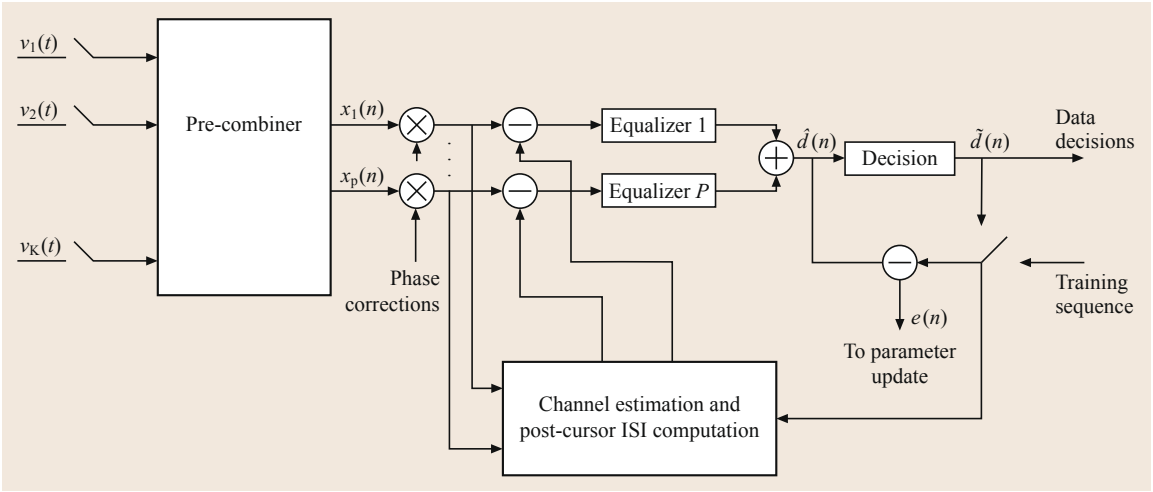


Fig. 15.15 Multichannel decision feedback equalizer

where the input vector $\boldsymbol{x}_p(n)$ to the p-th equalizer branch consists of the sampled precombiner output

$$x_p(t) = \sum_{k=1}^{K} c_{p,k}^{*} v^{(k)}(t) \, .$$

The samples are taken at the Nyquist rate, e.g., two per symbol interval for signals band-limited to $\pm 1/T$. The precombiner coefficients $c_{p,k}$, the feedforward equalizer vectors $\boldsymbol{a}_p$, the feedback equalizer coefficients b_i, and the phase estimates $\hat{\theta}_p$ are adjusted adaptively so as to minimize the mean squared error (MSE) in data detection. Each of these parameters is thus driven by the input signal and the error $e(n) = d(n) - \hat{d}(n)$. Adaptive operation begins with a training sequence of known data symbols $d(n)$ until convergence has been established, after which the symbols $d(n)$ in (15.26) are replaced by the decisions $\tilde{d}(n)$ made on the estimates $\hat{d}(n)$. The details of various algorithm variants can be found in [15.36, 57, 58].

Performance of the multichannel DFE is illustrated in Fig. 15.16. The excellent performance achieved by this receiver in many environments testifies to the benefits of careful algorithm structuring that caters to the acoustic channel. While a propagation-ignorant design may result in an unnecessarily large number of receiver parameters (many equalizer branches with long filters), which will, in turn, cause increased sensitivity to noise and numerical errors, results similar to those of Fig. 15.16 show that one may not need more equalizer branches than there are significant propagation paths; that each of these branches may not need an excessively long filter, and that the feedback filter may only need to activate a select subset of coefficients, although its

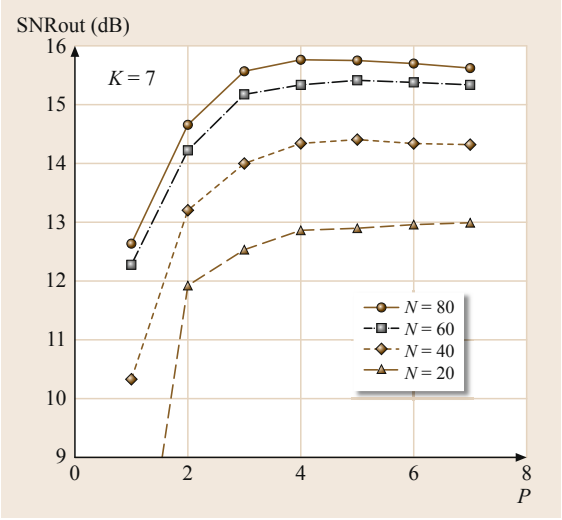


Fig. 15.16 Performance of the multichannel DFE on an experimental channel: output SNR (inversely proportional to the MSE) as a function of the number of equalizer channels P for a fixed number of input channels K. The parameter on the curves is the length N of each feedforward filter used. The total span of the feedback filter is 100 symbols. Real data used for this study correspond to quadrature amplitude modulation (QPSK) signals modulated at a rate of 2 kbps onto a carrier of 1 kHz, and transmitted acoustically over an 85 km long channel in about 50 m of water. The signals were received using a vertical array with inter-element spacing of 1 m. The signals were processed using a double recursive least squares (RLS) algorithm integrated with a second-order decision-directed PLL ►

total span must match that of the multipath. In other words, respecting the *physical* aspects of acoustic propagation in developing a concise channel representation is the key to successful signal processing. Identification of *significant* channel components has attracted due interest in the acoustic community over the past several years, and much attention has been devoted to the topic of sparse channel estimation. [15.37] gives an excellent primer.

Multicarrier Modulation/Detection

Multicarrier modulation in the form of orthogonal frequency division multiplexing (OFDM) has been adopted as a standard for many of the emerging wireless radio systems – wireless local area networks (WLAN), digital audio and video broadcast (DAB/DVB), and long-term evolution that supports the fourth generation (4G) and future generations of cellular systems. However, it has only recently come into the forefront of acoustic communications research, and at the moment, there appears to be little effort to implement this technology in a commercial modem (e.g., by the French Thales).

The appeal of OFDM lies in the computational efficiency of FFT-based processing, and in the fact that it easily scales to different bandwidths (unlike with single-carrier systems, where the equalizer length has to be adjusted in accordance with the bandwidth B because it determines the symbol duration and hence the extent of ISI, with OFDM it simply suffices to increase/decrease the number of carriers K, i. e., the size of the FFT, while keeping the same carrier separation $\Delta f = B/K$). Namely, the demodulated signal is simply represented as

$$y_k(n) = H_k(n)d_k(n) + z_k(n)\,, \qquad (15.27)$$

where $d_k(n)$ is the data symbol transmitted on the k-th carrier during the n-th OFDM block of duration $T = 1/\Delta f$; $H_k(n)$ is the channel transfer function $H(f,t)$ evaluated at the k-th carrier frequency at the time of the n-th block, and $z_k(n)$ is the noise. This signal admits a very simple detector, and extensions to multiple receiving elements are straightforward. In addition, by virtue of having a narrowband signal on each carrier, OFDM is easily conducive to MIMO processing [15.40, 62], adaptive modulation [15.63], and differentially coherent detection [15.64]. However, its sensitivity to frequency offset and time-variation of the channel demands special attention. Issues related to power efficiency also need to be kept in mind, as OFDM is sensitive to nonlinear distortions [15.65].

OFDM signal processing encompasses two stages: pre-FFT synchronization and post-FFT data detection. To account for motion-induced Doppler frequency shifting, which can amount to more than a full carrier spacing, front-end resampling is often necessary. A simple method for estimating the needed resampling rate is to measure the time between two synchronization preambles that frame several OFDM blocks and compare it to the expected frame duration [15.66].

Since the Doppler factor is relatively large to begin with (e.g., on the order of 10^{-3} for a relative velocity of 1.5 m/s), Doppler shifting that remains after initial resampling cannot be neglected. Two approaches have been pursued to address this issue: one is based on the assumption that residual Doppler shift is the same for all carriers [15.66], while another allows for nonuniform frequency shifting across the signal bandwidth [15.44].

Channel estimation for OFDM systems has been addressed in different forms: in one, each OFDM block is processed independently of the other blocks, thus allowing for the possibility that the channel changes completely from one block to another [15.41, 62], while another form exploits correlation between adjacent blocks [15.40, 67], which makes it advantageous on slowly varying channels. Similarly as in single-carrier systems, accurate channel estimation is the key to successful data detection in OFDM, and it benefits greatly from proper channel modeling to reduce the number of unknown parameters that need to be estimated. Methods for identification of sparse systems, such as matching pursuit (MP) and basis pursuit (BP), which improve upon traditional least squares (LS) estimation, were found to be beneficial and well suited to channel estimation in acoustic OFDM. These methods have been applied to both block-individual channel estimation that uses pilot carriers only [15.41], and to block-adaptive, decision-directed channel estimation [15.68].

The effect of time variability is a subtle one when it comes to acoustic OFDM systems: if the channel can change from one block to another, then perhaps it will change within each block. If this variation is nonnegligible, inter-carrier interference (ICI) will arise at the demodulator output, and carrier-by-carrier detection will no longer be optimal. ICI equalization then becomes necessary. Such a situation is somewhat counter-productive, as it destroys the original motivation of using OFDM for the simplicity of its implementation. Nonetheless, it may arise from pushing the limits of bandwidth efficiency: the more carriers in a given bandwidth, the better the bandwidth utilization, but the narrower the carrier spacing $\Delta f = 1/T$, the longer the OFDM block duration T, and the channel variation may become nonnegligible. The problem of ICI equalization is analogous to that of ISI equalization in single-carrier systems, except that the equalizer now operates across

carriers, and typically involves fewer interfering terms. Methods based on one-shot linear equalization of a full block of carriers [15.41, 69], as well as recursive linear or decision-feedback equalization [15.70], have been investigated. Further improvements are available from front-end (pre-FFT) filtering, which extracts the information about the time-varying channel before it has been lost in the process of FFT demodulation. A receiver that mimics the operation of an optimal, channel-matched filter was proposed in [15.71], while one that follows the analogy of fractionally-spaced equalization in the frequency domain was investigated in [15.72].

Virtually all acoustic OFDM systems addressed to date have focused on coherent detection and the attendant issues of channel estimation and Doppler tracking. However, a properly designed OFDM system (one in which there is no ICI) is well suited to differentially coherent detection as well. Differential encoding can be performed either across blocks or across carriers. In the latter case, differential coherence is naturally satisfied with tight carrier packing, which simultaneously supports bandwidth efficiency. Preliminary experimental results [15.64, 73] have demonstrated the benefits of differentially coherent OFDM, whose computational complexity is by far lower than that of any coherent system, and whose performance can surpass that of coherent detection when channel estimation fails.

15.5.3 Data Link Reliability

Channel coding or forward error coding (FEC) in a crucial feature of underwater acoustic modems. By introducing a redundancy part in the transmitted message, error correction codes allow for detection and correction of the bit errors caused by noise and inter-symbol interference. Generally, the ability to correct the errors is achieved by adding redundancy, which in turns reduces the effective data rate. Therefore, a proper coding scheme for a communication system is chosen based on the trade-off between the desired data bit rate and error rate performance.

Traditional FEC techniques are divided into convolutional codes (usually decoded with the ubiquitous Viterbi algorithm) and block codes (e.g., Bose-Chaudhuri-Hocquenghem or BCH, Reed–Solomon or RS) [15.74, 75]. Concatenated codes usually combine convolutional and block codes to improve the reliability of the acoustic modem even further. More recently, turbo codes [15.76, 77] and low-density parity-check (LDPC) codes [15.78] have become part of the field of telecommunication as they offer near Shannon error correction capability. Having inherently a low coding rate and a potentially large decoding latency, turbo codes have emerged as a good choice in communication systems where achieving a very low error rate performance is a top priority. A communication link designed for the transmission of the command-and-control messages to a remotely operated robot or vehicle is such a system. Since control messages do not contain a large number of bits that need to be transmitted in a short time, high bit rates are not required.

Even the most sophisticated error codes cannot ensure complete link reliability. To guarantee an error-free communication link (if it is physically possible), the acoustic modem must include a quality of service (QoS) feature, typically in the form of an automated retransmission query (ARQ) process implemented between two acoustic modems [15.54, 79]. This operation requires a communication protocol that includes acknowledgments and may dramatically reduce the overall data rate. The optimization of the QoS protocol is a key feature of modern acoustic modems. Acoustic modem users should be aware that commercially advertised throughputs (data rates) assume a single message transmission and do not include any retransmission.

15.5.4 Turbo Equalization

Turbo equalization is a very powerful method used to improve the removal of inter-symbol interference and additive noise. This technique can be implemented with single-carrier and multiple-carrier signaling. In particular, extensive research in combining turbo equalization and OFDM has taken place over the past decade. The simplest and best explanation is provided in the seminal work published in [15.80]. In short terms, turbo-equalization jointly optimizes adaptive equalization and channel decoding in an iterative process. The result is a dramatic improvement of transmission quality, even for high spectral efficiency modulations and time-varying fading channels. Modern underwater acoustic modems make growing use of this technique, which is powerful but processor intensive.

At the heart of this process is interleaving and error coding. The binary information at the source is carefully error-coded, modulated, and interleaved before it is pulse-shaped and transmitted through the water channel. In a traditional equalized receiver, the distorted signal is equalized, deinterleaved, converted to the binary domain, and the error-coded binary information is decoded and, hopefully, every binary error is corrected.

The obvious issue with this method is the fact that some error may remain. The turbo-equalizer goes two steps further in using this information. First of all, it modulates and interleaves the error-corrected binary information (which may still contain errors), thus cleaning the signal of some (or most of the) inter-symbol interference. The original signal is equalized

once again, but this time the *cleaned* signal is used as a new reference in the decision feedback portion of the adaptive equalizer. The second step is to repeat the entire operation again and again, so that the original signal is equalized against an ever *cleaner* reference signal.

The statistical properties of the acoustic channel, the type of interleaver, error coding, and decoding have a significant impact on turbo-equalizers. However, the same can be said of every equalized communication system, and overall this technique can lead to significant improvements in data transmission quality over a more traditional equalizer approach. Therefore, turbo-equalizers are becoming increasingly more common in underwater acoustic communication systems [15.38, 39, 69].

15.5.5 Adapting to the Environment

Adapting to the communication environment involves not only receiver-end channel estimation, but transmitter adaptation as well. This task is much more challenging in a time-varying environment which hinders the feedback from the receiver to the transmitter. In an acoustic channel, the situation is exacerbated by the long propagation delay which may make the feedback obsolete by the time it arrives. It is thus imperative to identify those environmental parameters that can withstand the feedback delay, and focus on the related transmitter functions.

In light of the large-scale and small-scale channel variations discussed in Sect. 15.4.4, adaptive mechanisms that have been considered include large-scale power control and small-scale waveform control. The first refers to adjusting the total transmitting power regardless of the modulation/detection method used, while the latter refers to adjusting the shape of the signal spectrum, i. e., distribution of the total available power over a given bandwidth. While the former targets power savings over extended periods of time (e.g., hours or days), the latter targets improved signal detection in a particular communication session (several data packets).

Neither type of adaptivity has been implemented in a commercial system, although research suggests significant benefits. Specifically, large-scale power control was shown to have significant benefits both in the sense of power savings [15.45] and in the sense of interference control in networked systems [15.81].

The work on small-scale adaptation has been versatile, including active time-reversal [15.82], single-mode excitation [15.83], and adaptive modulation [15.63]. Time reversal (or phase conjugation in the frequency domain) has in particular been extensively studied as an adaptive matched-filtering technique for acoustic communications [15.82], both in the sense of receiver adaptivity and transmitter adaptivity. It uses a time-reversed replica of a received signal waveform to implement a filter matched to that waveform, and can operate either passively or actively. Passive time-reversal resides at the receiver side only, where its role is to acquire a probe signal and use it to perform low-complexity front-end filtering prior to equalization or interference cancelation [15.84, 85]. In contrast, active time-reversal operates at the transmitter side, where its role is to time reverse the feedback signal and use it as the basic pulse (basic transmit waveform) that will best match the channel. By doing so, the transmitter, typically equipped with a large array, focuses its energy not only in time, but also in space ([15.82] and references therein). In repeated actions of this type, both ends of the link can focus their energy.

Adaptive modulation has been considered in the context of single-carrier MIMO systems [15.61], and, more recently, in the context of multicarrier systems [15.63], where adaptive power and/or rate control (adaptive bit loading) can be implemented easily by adjusting the amplitude and/or the modulation level of each carrier separately. The performance improvement available from these techniques is contingent on the quality of the channel state information that is fed back to the transmitter. Recent results [15.63], which report on the first experimental demonstration of this type, suggest the possibility to isolate the more slowly varying channel parameters (i. e., the predictable propagation path gains) from the more rapidly varying ones (phases) and use them to design an adaptive modulation system.

15.5.6 Networks

With the technological advances in acoustic modems, vehicles, and sensors, the concept of an underwater network of distributed autonomous nodes [15.14] is coming closer to realization. Experimental deployments are well underway, demonstrating the existing hardware capabilities in series of experimental deployments supported by the US Navy, such as the SeaWeb that uses telesonar modems [15.28], or deployments that use WHOI micro-modems [15.86]. Recently, efforts have begun to standardize acoustic network protocols with the goal of ensuring inter-operability between different modems. In particular, an international NATO-led effort is active in establishing JANUS [15.12], a standard that specifies data packet structures, acknowledgement (ACK) mechanisms, bit rates, etc., for use in given frequency band/transmissison distance configurations. Simultaneously, research is steadily advancing on the design of multiple-access methods, medium

access control (MAC) and routing protocols for underwater acoustic networks. The fundamental aspects of underwater acoustic networking have been highlighted in several survey articles [15.87–90]. At the moment, there is a solid base of custom-designed protocols that are suitable for poor-quality, interference-prone, delay-challenged acoustic channels, and are ready for testing in the field.

A full spectrum of network applications is not yet obvious, as integrated system capabilities depend on many aspects whose proof-of-concept, in turn, depends on in-situ testing. Unlike with signal processing, testing of networked systems cannot be performed off-line. Network simulation tools have thus started to emerge; an example is the World Ocean System Simulator (WOSS) [15.53], which integrates the Bellhop ray tracer with digital ocean maps and the NS-2 network simulator.

Due to a lack of *typical applications*, we tend to think about two types of networks: one with fixed, bottom-mounted nodes deployed for longer periods of time (e.g., sensor networks for environmental monitoring), and another with mobile nodes deployed on shorter missions (e.g., fleets of *cooperating* AUVs, where vehicles have the capability to respond to one another and make decisions, not only to be guided by supervisory commands from a central authority that amounts to *switch from mission A to mission B*). Interactions between mobile and fixed sensors are, of course, also of interest. Large-scale integration of autonomous systems includes the typical cross-layer network optimization at the boundary of two adjacent layers (i. e., routing and data link layer, or data link and physical layer). More importantly, it extends all the way from the application to the physical link, and includes external system functions such as localization and navigation (which themselves are challenging tasks as there is no global positioning or timing underwater).

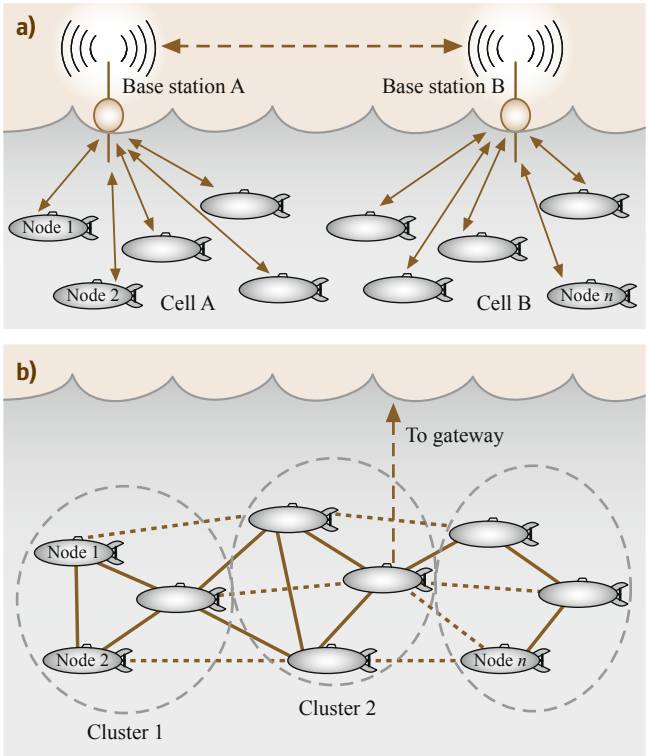


Fig. 15.17a,b Network topologies and the associated architectures come in various forms: in a centralized network of a cellular type, nodes communicate through a single-hop connection to the base station, while base stations are connected to a common infrastructure **(a)**; in a de-centralized network of an ad-hoc type, nodes communicate over multihop, peer-to-peer links **(b)**

The design of underwater networks is daunted by the constraints of acoustic propagation: limited, distance-dependent bandwidth, high bit (packet) error rate, and high channel latency. The fact that acoustic signals propagate at 1500 m/s, while bit rates supported by the (half-duplex) acoustic modems are on the order of a kbit/s, makes the propagation delay anything but negligible with respect to the packet size – otherwise a common assumption in the design of many channel access schemes. As a result, direct application of protocols developed for terrestrial sensor networks may incur a serious penalty in efficiency, requiring instead a careful re-design for acoustic applications. Energy efficiency is an important aspect in bottom-mounted sensor networks, since re-charging the batteries of submerged instruments is to be postponed for as long as possible. In contrast, mobile systems are already constrained to several-hour missions before they have to re-charge, and power usage for communications is not a major issue (unless self-propelled gliders are in question). It must also be kept in mind that underwater instruments (sensors, modems, vehicles) are neither cheap nor disposable. This fact may be the single most-important feature that distinguishes underwater networks from their terrestrial counterparts, fundamentally changing many network design principles that are usually taken for granted. Finally, the questions of network topology optimization and the associated capacity analysis remain open.

In terms of system topology and architecture, two extreme situations are illustrated in Fig. 15.17. In a centralized network, distributed nodes communicate through a base station in whose footprint they are. Coverage can be extended in this type of network by replicating the cells across space, similarly as in terrestrial cellular systems (except that the *infrastructure* that connects the base stations can consist of radio links). In a de-centralized network, the nodes communicate di-

rectly with each other. There is no central authority such as a base station, although the nodes can form clusters, and the end goal may as well be to transmit all the data to a common node (a surface sink). The network nodes can either have pre-established routes over which to send the packets (as would be the case in fixed bottom-mounted networks), or they can do this in an ad-hoc manner, learning the routes as they go (which would be the case in a mobile network). Between these two extremes lie various hybrid architectures.

15.5.7 Channel Sharing

Channel sharing is the major design issue for band-limited acoustic networks. Methods for channel sharing can be divided into two types: deterministic and random. In deterministic access, resources such as frequency bands or time slots are allocated to particular nodes. This type of multiple access is well suited to systems in which the nodes transmit continuously, since the a-priori allocated resources are then well utilized. If the nodes instead transmit in a bursty manner and not too often, random access may be the preferred choice. The nodes now transmit more-or-less at their own will and in the same frequency band, risking a loss if their packets collide at the receiver. Medium access control (MAC) protocols can then be used to orchestrate the nodes' behavior and reduce the chances of collision. Random and deterministic access can also be combined in a reservation-based channel sharing scheme. Nodes now first contend by transmitting short reservation packets in a random access fashion. The central authority chooses the winners, and deterministically allocates them fixed resources for the duration of the data transmission phase.

Deterministic Access

Deterministic access for acoustic networks has been considered in various forms – frequency, time, and code-division multiple access (FDMA, TDMA, CDMA) [15.91]. Typically, FDMA is considered inefficient on account of the fact that the already limited bandwidth will be wasted if allocated permanently to a node that does not transmit, while those nodes that transmit will be doing so at a lower bit rate, thereby increasing the energy consumption. This reasoning is justified in today's applications where the nodes typically do not stream continuously, but only report to the base station upon request. A polling mechanism is then put in place via time-division duplexing (TDD). Although this method may resemble TDMA (and is often called that) a network based on polling is not a true TDMA network, as it involves a two-way propagation delay for each transaction, whereas in a true TDMA network nodes schedule their transmissions so that they arrive back-to-back to the central node. In an acoustic setting with variable propagation delays, guard times are needed to ensure that there is no overlap of packets at the base station. Acoustic modems that are currently in use are well conducive to such an implementation [15.92].

An alternative that has been considered for small multiuser systems is CDMA based on direct-sequence spread-spectrum (DS-SS) modulation, which simultaneously provides the low probability of detection (LPD) needed for operation in hostile environments. It may be worth noting that the usual assumptions involved in the design of radio DS-SS systems do not hold in an acoustic channel, because it requires chip-rate (as opposed to symbol-rate) adaptive processing to take advantage of the available spreading gain [15.60]. Passive time reversal has also been used for multiuser detection [15.85].

While today's applications do not require more than a single cell, spatial frequency reuse across multiple cells offers the key to large area coverage with a limited bandwidth. Acoustic propagation, however, dictates cellular design principles that are more complex than those used in terrestrial networks, and makes the system capacity heavily dependent on the center frequency [15.93]. Specifically, moving to a higher frequency region than that dictated by simple SNR maximization (Fig. 15.5) improves the signal-to-interference ratio and yields a greater capacity.

Random Access

Random access has been used for acoustic networks in which the nodes transmit infrequently [15.28, 86]. This type of access is also suitable for networks with many nodes, so long as the their aggregate traffic does not exceed a certain limit. The simplest form of random access is the Aloha mechanism whereby each node transmits whenever it wants to, and if it becomes aware of a collision (via an acknowledgement/negative acknowledgement (ACK/NAK) feedback from the intended receiver), it re-transmits after a random back-off time. This simple mechanism can be augmented by carrier sensing multiple access (CSMA), which imposes an additional condition that a node can transmit only if it senses the channel to be free. Compared to radio systems where propagation delay is negligible compared to the packet duration, this listen-before-transmitting principle has fewer benefits in acoustic systems, where the packets propagate slowly, and the fact that none are overheard does not mean that some are not present in the channel. Multiple-access collision avoidance (MACA) introduces a hand-shaking procedure to secure the link using short control packets of

the request-to-send/clear-to-send type (RTS/CTS) prior to transmitting the longer data packets.

In the past years there have been major developments on the medium access control (MAC) layer, with focus on protocol design that is not ignorant of the high acoustic latency, but strives to overcome it in an efficient manner or even take advantage of it. Examples of such protocols include the distance-aware collision avoidance protocol (DACAP) [15.94], T-Lohi [15.95], propagation-delay-tolerant Aloha [15.96], and many others, e.g., [15.97, 98]. For example, DACAP is based on a standard control packet exchange to secure the link for data transmission, with the addition of a warning signal that interrupts an on-going transmission that is bound to fail. T-Lohi uses dedicated collision-avoidance tones to announce data transmission, thus enabling all those who overhear the tones to count the number of active nodes, which, in turn, is used to optimize channel access.

15.5.8 Routing and Cross-Layer Integration

Work on higher layer protocols has focused in particular on routing. Geographical routing, which exploits the knowledge of the nodes' locations to determine the best next relay so as to minimize the total energy consumption, was addressed in [15.99]. A distributed geographical routing protocol with integrated power control was proposed in [15.81]. Distributed routing for delay-sensitive and delay-insensitive applications was investigated in [15.100], while a distributed procedure for initial neighbor discovery was proposed in [15.101]. Transport control protocols have been considered to a much lesser extent, since end-to-end retransmission suffers greatly from long delays and high packet error rate.

Cross-layer design plays an important role in acoustic systems, both between adjacent network layers and between adjacent system functions. Examples of cross-layer design include frequency allocation and power control integration with MAC and routing [15.81]; topology/sleep control for energy-efficiency in fixed networks [15.56, 98, 102]; packet size selection for optimizing the throughput/energy performance of MAC protocols [15.103], and optimization of rateless packet coding (network coding) for use with half-duplex acoustic links [15.104]. An example of cross-*function* design is integration of localization/navigation with acoustic communications. The use of acoustic modems as traveling beacons that help to localize AUVs by measuring the relative propagation delays in a network is described in [15.92]. Such localization may, in turn, support geographical routing protocols in which location information is used to find the best routes through the network.

The issues of mobility and delay-tolerance have received particular attention. Channel access in a network composed of both fixed and mobile nodes was studied in [15.102], via a simulation analysis that compares the Aloha and the DACAP protocols. Both were found to provide satisfactory performance, which is further enhanced for mobile applications by simple modifications such as back-to-back repetitions. [15.105] investigated routing in a delay-tolerant network (DTN), comparing two classes of protocols, one based on spray-and-wait principle and the other based on resource allocation for intentional DTN. In conditions of high load, the latter was found to outperform the former as well as basic flooding in terms of packet delivery ratio regardless of mobility conditions, while the former showed better performance in terms of average delivery ratio. [15.106] considered a swarm of free-floating mobile sensor nodes that need to report their measurements to a central station. [15.107] considered location-based routing in a mobile network. To improve the performance, the nodes do not use the current knowledge of the destination's location, but instead form a prediction based on the history of motion. [15.108] also investigated location-based routing, which it enhanced by inclusion of a link quality metric into the decision-making process. Finally, [15.109] explored the idea of waiting for a favorable link to present itself before forwarding a packet. In other words, an AUV may not wish to waste its energy transmitting to a relay far away, but may instead want to wait for a better opportunity. Once such an opportunity presents itself, highly directional transmission is exploited to focus the energy in a desired direction.

These references offer a valuable first step towards understanding the network behavior in the presence of long acoustic delay and frequency-dependent attenuation. They identify vulnerable points of existing protocols and offer clever solutions to recover the performance. However, although there is a common understanding that mobility needs special care in an acoustic setting, the analysis is often conducted via simulation, using a time-invariant propagation model. This fact emphasizes the need for *statistical* channel models that will reflect temporal as well as spatial variability of acoustic channel and enable the development of more accurate simulation tools.

15.6 Future Trends

As underwater acoustic modem technology keeps evolving, new research trends appear. The most obvious trend is the development of message format and protocols optimized for a specific application. A good example is the development of the compact control language (CCL) for unmanned underwater vehicle (UUV) applications [15.23]. Another example is digital diver modems. In addition, data compression algorithms tailored for underwater acoustic modems have now been developed, as traditional compression routines are often not well suited to this type of communication system. The integration of underwater acoustic modems in larger system now includes the combination of multiple functions in order to reduce cost, size, and power consumption. For example, several underwater acoustic modems also include long-baseline and/or ultra-short baseline acoustic positioning capability [15.110, 111].

Some of the most innovative research in underwater communication now combines various modalities, such as optics (laser, lidar), and magnetic and electromagnetic signaling with broadband acoustics [15.112]. In particular, excellent work is taking place in combining short-range, high-bit-rate underwater communications with slower, longer-range underwater acoustic communications. Finally, one should not overlook the extensive research work taking place to develop better acoustic propagation models and statistical performance models [15.113]. For example, great efforts are being made in modeling and optimizing network operations using underwater acoustic modems, possibly in combination with surface radio communication devices. Also, integrated underwater acoustic communications and localization are increasingly being combined to improve data routing [15.114].

References

15.1 X. Che, I. Wells, G. Dickers, P. Kear, X. Gong: Re-evaluation of RF electromagnetic communication in underwater sensor networks, IEEE Commun. Mag. **48**(12), 143–151 (2010)

15.2 N. Farr, A. Bowen, J. Ware, C. Pontbriand, M. Tivey: An integrated, underwater optical /acoustic communications system, Proc. IEEE OCEANS'10 (2010) pp. 1–6

15.3 L. Freitag, S. Singh: Performance of micro-modem PSK signaling under variable conditions during the 2008 RACE and SPACE experiments, Proc. MTS/IEEE OCEANS'09 (2009) pp. 1–8

15.4 L. Freitag, M. Stojanovic: Basin-scale acoustic communication: a feasibility study using tomography m-sequences, Proc. MTS/IEEE OCEANS'01, Vol. 4 (2001) pp. 2256–2261

15.5 L. Freitag, M. Johnson, D. Frye: High-rate acoustic communications for ocean observatories-performance testing over a 3000 vertical path, Proc. MTS/IEEE OCEANS'00, Vol. 2 (2000) pp. 1443–1448

15.6 A. Quazi, W. Konrad: Underwater acoustic communications, IEEE Commun. Mag. **20**(2), 24–30 (1982)

15.7 A.B. Baggeroer, D.E. Koelsch, K. von der Heydt, J. Catipovic: DATS - A digital acoustic telemetry system for underwater communications, Proc. IEEE OCEANS'81 (1981) pp. 55–60

15.8 J. Catipovic, D. Brady, S. Etchemendy: Development of underwater acoustic modems and networks, Oceanography **6**(3), 112–119 (1993)

15.9 M.D. Green, J.A. Rice: Channel-tolerant FH-MFSK acoustic signaling for undersea communications and networks, IEEE-JOE **25**(1), 28–39 (2000)

15.10 M. Stojanovic, J.A. Catipovic, J.G. Proakis: Phase-coherent digital communications for underwater acoustic channels, IEEE-JOE **19**(1), 100–111 (1994)

15.11 S. Singh, S.E. Webster, L. Freitag, L.L. Whitcomb, K. Ball, J. Bailey, C. Taylor: Acoustic communication performance of the WHOI micro-modem in sea trials of the Nereus vehicle to 11,000 m depth, Proc. IEEE OCEANS'09 (2009) pp. 1–6

15.12 J.R. Potter, A. Berni, J. Alves, D. Merani, G. Zappa, R. Been: Underwater communications protocols and architecture developments at NURC, Proc. IEEE OCEANS'11 (2011) pp. 1–6

15.13 V. Tunnicliffe, C.R. Barnes, R. Dewey: Major advances in cabled ocean observatories (VENUS and NEPTUNE Canada) in coastal and deep sea settings, Proc. IEEE/OES US/EU-Balt. Int. Symp. (2008) pp. 1–7

15.14 T.B. Curtin, J.G. Bellingham, J. Catipovic, D. Webb: Autonomous oceanographic sampling networks, Oceanography **6**(3), 86–94 (1993)

15.15 Teledyne Benthos: https://teledynebenthos.com/product/acoustic_modems/910-series-atm-916

15.16 Woods Hole Oceanographic Institution: http://acomms.whoi.edu/micro-modem

15.17 LinkQuest Inc.: http://www.link-quest.com/html/uwm1000.htm

15.18 EvoLogics: http://www.evologics.de/en/products/acoustics/index.html

15.19 Sercel: http://www.sercel.com/products/Pages/mats3g.aspx

15.20 L-3 Oceania: http://www2.l-3com.com/oceania/products/uc_modem.htm

15.21 Tritech: http://www.tritech.co.uk/product/micron-data-modem

15.22 P.-P.J. Beaujean, E. Carlson: HERMES – A high bit-rate underwater acoustic modem operating at high-frequencies for ports and shallow wa-

ter applications, Mar. Technol. Soc. J. **43**(2), 21–32 (2009)
15.23 R.P. Stokey, L.E. Freitag, M.D. Grund: A compact control language for AUV acoustic communication, Proc. IEEE OCEANS'05 (2005) pp. 1133–1137
15.24 D.B. Kilfoyle, A.B. Baggeroer: The state of the art in underwater acoustic telemetry, IEEE-JOE **25**(1), 4–27 (2000)
15.25 R.J. Urick: *Principles of Underwater Sound*, 3rd edn. (McGraw-Hill, New York 1983)
15.26 H. Medwin, C.S. Clay: *Fundamentals of Acoustical Oceanography (Applications of Modern Acoustics)* (Academic Press, Waltham 1998)
15.27 M.B. Porter, V.K. McDonald, P.A. Baxley, J.A. Rice: SignalEx: Linking environmental acoustics with the signaling schemes, Proc. MTS/IEEE OCEANS'00 (2000) pp. 595–600
15.28 J.A. Rice, C.W. Ong: A discovery process for initializing underwater acoustic networks, Proc. SENSORCOMM (2010) pp. 408–415
15.29 L. Berkhovskikh, Y. Lysanov: *Fundamentals of Ocean Acoustics*, Springer Series in Electronics and Photonics, Vol. 8 (Springer, Berlin, Heidelberg 1982)
15.30 M. Stojanovic: On the relationship between capacity and distance in an underwater acoustic communication channel, Proc. WUWNet (2006) pp. 41–47
15.31 M. Stojanovic: Capacity of a relay acoustic channel, Proc. IEEE OCEANS'07 (2007) pp. 1–7
15.32 M.Porter: Ocean Acoustics Library, http://oalib.hlsresearch.com/ (2015)
15.33 M.B. Porter, H.P. Bucker: Gaussian beam tracing for computing ocean acoustic fields, J. Acoust. Soc. Am. **82**(4), 1349–1359 (1987)
15.34 M. Siderius, M.B. Porter: Modeling broadband ocean acoustic transmissions with time-varying sea surfaces, J. Acoust. Soc. Am. **124**, 137–150 (2008)
15.35 F.F. Jensen, W. Kuperman, M. Porter, H. Schmidt: *Computational Ocean Acoustics*, Modern Acoustics and Signal Processing (Springer, New York 2011)
15.36 M. Stojanovic: Efficient processing of acoustic signals for high rate information transmission over sparse underwater channels, Phys. Commun. **1**(2), 146–161 (2008)
15.37 W. Li, J.C. Preisig: Estimation of rapidly time-varying sparse channels, IEEE-JOE **32**(4), 927–939 (2007)
15.38 S. Roy, T.M. Duman, V.K. McDonald: Error rate improvement in underwater MIMO communications using sparse partial response equalization, IEEE-JOE **34**(2), 181–201 (2009)
15.39 J.W. Choi, T.J. Riedl, K. Kim, A.C. Singer, J.C. Preisig: Adaptive linear turbo equalization over doubly selective channels, IEEE-JOE **36**(4), 473–489 (2011)
15.40 M. Stojanovic: MIMO OFDM over underwater acoustic channels, Proc. 43rd Asilomar Conf. Signals Syst. Comput. (2009) pp. 605–609
15.41 C.R. Berger, S. Zhou, J.C. Preisig, P. Willett: Sparse channel estimation for multicarrier underwater acoustic communication: From subspace methods to compressed sensing, IEEE Trans. Sig. Process. **58**(3), 1708–1721 (2010)
15.42 T. Kang, R.A. Iltis: Iterative carrier frequency offset and channel estimation for underwater acoustic OFDM systems, IEEE J. Sel. Areas Commun. **26**(9), 1650–1661 (2008)
15.43 M.M. Stojanovic, J. Catipovic, J. Proakis: Performance of High-Rate Adaptive Equalization on a Shallow Water Acoustic Channel, J. Acoust. Soc. Am. (1996) pp. 2213–2219
15.44 M. Stojanovic: Low complexity OFDM detector for underwater acoustic channels, Proc. IEEE OCEANS'06 (2006) pp. 1–6
15.45 P. Qarabaqi, M. Stojanovic: Statistical characterization and computationally efficient modeling of a class of underwater acoustic communication channels, IEEE-JOE **38**(4), 701–717 (2013)
15.46 W.B. Yang, T.C. Yang: High-frequency channel characterization for M-ary frequency-shift-keying underwater acoustic communications, J. Acoust. Soc. Am. **120**(5), 2615–2626 (2006)
15.47 B. Tomasi, G. Zappa, K. McCoy, P. Casari, M. Zorzi: Experimental study of the space-time properties of acoustic channels for underwater communications, Proc. IEEE OCEANS'10 (2010) pp. 1–9
15.48 A. Radosevic, D. Fertonani, T.M. Duman, J.G. Proakis, M. Stojanovic: Capacity of MIMO systems in shallow water acoustic channels, Proc. 44th Asilomar Conf. Sig. Syst. Comput. (2010) pp. 2164–2168
15.49 F. Socheleau, C. Laot, J. Passerieux: Stochastic replay of non-WSSUS underwater acoustic communication channels recorded at sea, IEEE Trans. Signal Process. **59**(10), 4838–4849 (2011)
15.50 M. Chitre: A high-frequency warm shallow water acoustic communications channel model and measurements, J. Acoust. Soc. Am. **122**(5), 2580–2586 (2007)
15.51 J. Zhang, J. Cross, Y.R. Zheng: Statistical channel modeling of wireless shallow water acoustic communications from experiment data, Proc. MILCOM (2010) pp. 2412–2416
15.52 A. Radosevic, T.M. Duman, J.G. Proakis, M. Stojanovic: Channel prediction for adaptive modulation in underwater acoustic communications, Proc. IEEE OCEANS'11 (2011) pp. 1–5
15.53 F. Guerra, P. Casari, A. Berni, J. Potter, M. Zorzi: Performance evaluation of random and handshake-based channel access in collaborative mobile underwater networks, Proc. OCEANS'10 (2010) pp. 1–7
15.54 M. Stojanovic: Optimization of a data link protocol for an underwater acoustic channel, Proc. IEEE OCEANS'05, Vol. 1 (2005) pp. 68–73
15.55 L. Freitag, M. Grund, S. Singh, J. Partan, P. Koski, K. Ball: The WHOI micro-modem: An acoustic communications and navigation system for multiple platforms, Proc. MTS/IEEE OCEANS'05, Vol. 2 (2005) pp. 1086–1092

15.56 A. Harris, M. Stojanovic, M. Zorzi: Idle-time energy savings through wake-up modes in underwater acoustic networks, Elsevier J. Ad Hoc Netw. **7**, 770–777 (2009)

15.57 M.M. Stojanovic, J. Catipovic, J. Proakis: Adaptive multichannel combining and equalization for underwater acoustic communications, J. Acoust. Soc. Am. (1993) pp. 1621–1631

15.58 M. Stojanovic, J.G. Proakis, J. Catipovic: Reduced complexity spatial and temporal processing of underwater acoustic communication signals, J. Acoust. Soc. Am. **98**(2), 961–972 (1995)

15.59 M. Stojanovic, Z. Zvonar: Multichannel processing of broad-band multiuser communication signals in shallow water acoustic channels, IEEE-JOE **21**(2), 156–166 (1996)

15.60 M. Stojanovic, L. Freitag: Multichannel detection for wideband underwater acoustic CDMA communications, IEEE-JOE **31**(3), 685–695 (2006)

15.61 D.B. Kilfoyle, J.C. Preisig, A.B. Baggeroer: Spatial modulation experiments in the underwater acoustic channel, IEEE-JOE **30**(2), 406–415 (2005)

15.62 B. Li, J. Huang, S. Zhou, K. Ball, M. Stojanovic, L. Freitag, P. Willett: MIMO-OFDM for high-rate underwater acoustic communications, IEEE-JOE **34**(4), 634–644 (2009)

15.63 A. Radosevic, R. Ahmed, T.M. Duman, J.G. Proakis, M. Stojanovic: Adaptive OFDM modulation for underwater acoustic communications: Design considerations and experimental results, IEEE-JOE **39**(2), 1–14 (2013)

15.64 Y. Aval, M. Stojanovic: A method for differentially coherent multichannel processing of acoustic OFDM signals, Proc. IEEE Workshop Sensor Array Multichannel Signal Process. (SAM) (2012)

15.65 G. Rojo, M. Stojanovic: Peak-to-average power ratio (PAR) reduction for acoustic OFDM systems, MTS J. **44**(4), 30–41 (2010)

15.66 B. Li, S. Zhou, M. Stojanovic, L. Freitag, P. Willett: Multicarrier communication over underwater acoustic channels with nonuniform doppler shifts, IEEE-JOE **33**(2), 198–209 (2008)

15.67 P. Ceballos Carrascosa, M. Stojanovic: Adaptive channel estimation and data detection for underwater acoustic MIMO OFDM systems, IEEE-JOE **35**(3), 635–646 (2010)

15.68 A. Radosevic, T.M. Duman, J.G. Proakis, M. Stojanovic: Selective decision directed channel estimation for UWA OFDM systems, Proc. 49th Commun. Control Comput. (2011) pp. 647–653

15.69 J. Huang, S. Zhou, J. Huang, C.R. Berger, P. Willett: Progressive inter-carrier interference equalization for OFDM transmission over time-varying underwater acoustic channels, IEEE J. Sel. Top. Signal Process. **5**(8), 1524–1536 (2011)

15.70 K. Tu, D. Fertonani, T.M. Duman, M. Stojanovic, J.G. Proakis, P. Hursky: Mitigation of intercarrier interference for OFDM over time-varying underwater acoustic channels, IEEE-JOE **36**(2), 156–171 (2011)

15.71 S. Yerramalli, M. Stojanovic, U. Mitra: Partial FFT demodulation: A detection method for highly doppler distorted OFDM systems, IEEE Trans. Signal Process. **60**(11), 5906–5918 (2012)

15.72 Z. Wang, S. Zhou, G.B. Giannakis, C.R. Berger, J. Huang: Frequency-domain oversampling for zero-padded OFDM in underwater acoustic communications, IEEE-JOE **37**(1), 14–24 (2012)

15.73 Y.M. Aval, M. Stojanovic: Fractional FFT demodulation for differentially coherent detection of acoustic OFDM signals, Proc. 46th Conf. Signals Syst. Comput. (ASILOMAR) (2012) pp. 1525–1529

15.74 S. Lin, D.J. Costello: *Error Control Coding: Fundamentals and Applications* (Prentice Hall, Englewood Cliffs 1983)

15.75 B. Sklar: *Digital Communications* (Prentice Hall, Englewood Cliffs 1988)

15.76 C. Berrou, A. Glavieux: Near optimum error correcting coding and decoding: turbo-codes, IEEE Trans. Commun. **44**(10), 1261–1271 (1996)

15.77 M. Pajovic, P.-P. Beaujean: Turbo-coded frequency-hopped frequency division multiplexed signaling for underwater acoustic communications between 60 and 90 kHz in ports and very shallow waters, Proc. MTS/IEEE OCEANS'09 (2009) pp. 1–7

15.78 J. Huang, S. Zhou, P. Willett: Nonbinary LDPC coding for multicarrier underwater acoustic communication, IEEE J. Sel. Areas Commun. **26**(9), 1684–1696 (2008)

15.79 B. Tomasi, P. Casari, L. Badia, M. Zorzi: A study of incremental redundancy hybrid ARQ over Markov channel models derived from experimental data, Proc. WUWNet (2010)

15.80 C. Laot, A. Glavieux, J. Labat: Turbo equalization: Adaptive equalization and channel decoding jointly optimized, IEEE J. Sel. Areas Commun. **19**(9), 1744–1751 (2001)

15.81 J. Montana, M. Stojanovic, M. Zorzi: On joint frequency and power allocation in a cross-layer protocol for underwater acoustic networks, IEEE-JOE **35**(4), 936–947 (2010)

15.82 G.F. Edelmann, H.C. Song, S. Kim, W.S. Hodgkiss, W.A. Kuperman, T. Akal: Underwater acoustic communications using time reversal, IEEE-JOE **30**(4), 852–864 (2005)

15.83 J.R. Buck, J.C. Preisig, M. Johnson, J. Catipovic: Single-mode excitation in the shallow-water acoustic channel using feedback control, IEEE-JOE **22**(2), 281–291 (1997)

15.84 J.A. Flynn, J.A. Ritcey, D. Rouseff, W.L.J. Fox: Multichannel equalization by decision-directed passive phase conjugation: experimental results, IEEE-JOE **29**(3), 824–836 (2004)

15.85 S.E. Cho, H.C. Song, W.S. Hodgkiss: Successive interference cancellation for underwater acoustic communications, IEEE-JOE **36**(4), 490–501 (2011)

15.86 C. Petrioli, R. Petroccia, J. Shusta, L. Freitag: From underwater simulation to at-sea testing using the ns-2 network simulator, Proc. OCEANS'11 (2011) pp. 1–9

15.87 J. Heideman, M. Stojanovic, M. Zorzi: Underwater sensor networks: Applications, advances, and challenges, Phil. Trans. R. Soc. **370**, 157–175 (2012)

15.88 D. Pompili, I. Akyildiz: Overview of networking protocols for underwater wireless communications, Communications Magazine, IEEE **47**(1), 97–102 (2009)

15.89 J. Heidemann, W. Ye, J. Wills, A. Syed, Y. Li: Research challenges and applications for underwater sensor networking, Proc. IEEE WCNC, Vol. 1 (2006) pp. 228–235

15.90 M. Chitre, S. Shahabodeen, M. Stojanovic: Underwater acoustic communications and networking: Recent advances and future challenges, Mar. Technol. Soc. J. **42**(1), 103–116 (2008)

15.91 E.M. Sozer, M. Stojanovic, J.G. Proakis: Underwater acoustic networks, IEEE-JOE **25**(1), 72–83 (2000)

15.92 R. Eustice, H. Singh, L. Whitcomb: Synchronous-clock one-way-travel-time acoustic navigation for underwater vehicles, J. Field Robotics **28**, 121–136 (2011)

15.93 M. Stojanovic: Design and capacity analysis of cellular-type underwater acoustic networks, IEEE-JOE **33**(2), 171–181 (2008)

15.94 B. Peleato, M. Stojanovic: Distance aware collision avoidance protocol for ad-hoc underwater acoustic sensor networks, IEEE Commun. Lett. **11**(12), 1025–1027 (2007)

15.95 A. Syed, W. Ye, J. Heidemann: Comparison and evaluation of the T-Lohi MAC for underwater acoustic sensor networks, IEEE J. Sel. Areas Commun. **26**(9), 1731–1743 (2008)

15.96 J. Ahn, A. Syed, B. Krishnamachari, J. Heidemann: Design and analysis of a propagation delay tolerant ALOHA protocol for underwater networks, Ad Hoc Netw. J. **9**(1), 752–766 (2011)

15.97 N. Chirdchoo, W.-S. Soh, K. Chua: RIPT: A receiver-initiated reservation-based protocol for underwater acoustic networks, IEEE J. Sel. Areas Commun. **26**(9), 1744–1753 (2008)

15.98 M.K. Park, V. Rodoplu: UWAN-MAC: An energy-efficient MAC protocol for underwater acoustic wireless sensor networks, IEEE-JOE **32**(3), 710–720 (2007)

15.99 M. Zorzi, P. Casari, N. Baldo, A. Harris: Energy-efficient routing schemes for underwater acoustic networks, IEEE J. Sel. Areas Commun. **26**(9), 1754–1766 (2008)

15.100 D. Pompili, T. Melodia, I.F. Akyildiz: Distributed routing algorithms for underwater acoustic sensor networks, IEEE Trans. Wirel. Commun. **9**(9), 2934–2944 (2010)

15.101 A. Patil, M. Stojanovic: A node discovery protocol for ad hoc underwater acoustic networks, Wirel. Commun. Mob. Comput. **13**(3), 277–295 (2013)

15.102 F. Zorzi, M. Stojanovic, M. Zorzi: On the effects of node density and duty cycle on energy efficiency in underwater networks, Proc. IEEE OCEANS'10 (2010) pp. 1–6

15.103 S. Basagni, C. Petrioli, R. Petroccia, M. Stojanovic: Optimized packet size selection in underwater WSN communications, IEEE-JOE **37**(3), 321–337 (2012)

15.104 D.E. Lucani, M. Stojanovic, M. Medard: Random linear network coding for time division duplexing: When to stop talking and start listening, Proc. IEEE INFOCOM (2009) pp. 1800–1808

15.105 M.S. Rahim, P. Casari, F. Guerra, M. Zorzi: On the performance of delay tolerant routing protocols in underwater networks, Proc. IEEE OCEANS'11 (2011) pp. 1–7

15.106 L.F.M. Vieira, U. Lee, M. Gerla: Phero-trail: a bio-inspired location service for mobile underwater sensor networks, IEEE J. Sel. Areas Commun. **28**(4), 553–563 (2010)

15.107 N. Chirdchoo, W.-S. Soh, K.C. Chua: Sector-based routing with destination location prediction for underwater mobile networks, Proc. WAINA (2009) pp. 1148–1153

15.108 E.A. Carlson, P.-P. Beaujean, E. An: Location-aware routing protocol for underwater acoustic networks, Proc. IEEE OCEANS'06 (2006) pp. 1–6

15.109 B. Chen, D. Pompili: QUO VADIS: QoS-aware underwater optimization framework for inter-vehicle communication using acoustic directional transducers, Proc. IEEE SECON (2011) pp. 341–349

15.110 D. Green: Acoustic modems, navigation aids, and networks for undersea operations, Proc. IEEE OCEANS'10 (2010) pp. 1–6

15.111 M. Stojanovic, L. Freitag, J. Leonard, P. Newman: A network protocol for multiple AUV localization, Proc. MTS/IEEE OCEANS'02, Vol. 1 (2002) pp. 604–611

15.112 N.E. Farr, A.D. Bowen, J.D. Ware, C. Pontbriand, M.A. Tivey: An integrated, underwater optical /acoustic communications system, Proc. IEEE OCEANS'10 (2010) pp. 1–6

15.113 M. Stojanovic, J. Preisig: Underwater acoustic communication channels: Propagation models and statistical characterization, IEEE Commun. Mag. **47**(1), 84–89 (2009)

15.114 E.A. Carlson, P.P.J. Beaujean, E. An: Location-aware source routing (LASR) protocol for underwater acoustic networks of AUVs, J. Electr. Comput. Eng. **2012**, 1–18 (2012)

16. Autonomous Underwater Vehicle Docking

James G. Bellingham

This chapter describes the practical challenges of docking autonomous underwater vehicles (AUVs) in the ocean environment and approaches undertaken to create working docking systems. Early docking work was motivated by the desire to operate AUVs for multiple sorties without recovery or human servicing of the AUV. To accomplish this, the docking system provides a method for the AUV to find the dock, to physically attach, to recharge AUV batteries, to establish a communication link, to wait in a low power state for a new mission, and to undock. A variety of homing methods, dock configurations, power transfer approaches, and communications links have been employed to achieve these respective needs, and are described herein. While the original applications driving docking were oceanographic in nature, present interest is broader, driven by commercial and military applications as well. Docking has been demonstrated, but has not been deployed as an operational capability, in part because docking requires an infrastructure for the dock that provides physical mounting, a source of power, and communication connectivity. The advent of cabled and mooring-based observatories, and subsea infrastructure for oil and gas production, create a foundation supporting docking system deployment. Consequently, the prospects for operational docking systems seem promising.

16.1 **Technical Elements of Docking** 388
16.2 **AUV Characteristics** 390
16.2.1 Vehicle Configuration and Controllability 390
16.2.2 Vehicle Size 390
16.3 **Sensors For Homing** 391
16.3.1 Acoustic Homing 392
16.3.2 Optical Homing 393
16.3.3 Electromagnetic Homing Systems 394
16.3.4 Constraining Vertical Position 395
16.4 **Capture and Connection Mechanisms** 395
16.4.1 Directional Docks: Funnels 396
16.4.2 Omnidirectional Docks: Poles 397
16.4.3 Soft Docking Approaches 398
16.4.4 Other Docking Mechanisms 398
16.5 **Coupling Power and Communications** 399
16.5.1 Establishing a Communications Link 399
16.5.2 Power Transfer 400
16.6 **AUV Control Considerations** 400
16.6.1 Navigation, Currents, and Arriving at the Dock 400
16.6.2 Managing State: AUV Control 401
16.6.3 Managing State: Coordinating AUV and Dock 401
16.6.4 Detecting and Recovering from Failures 403
16.7 **Conclusions and Future Prospects** 403
References 404

Autonomous underwater vehicle (docking AUV) is the process whereby an AUV purposefully transitions from a state of free flight to being physically connected to another device or platform, called the dock. One driver for the development of AUV docking stems from the desire to decrease reliance of AUV operations on ships and thus enable persistent AUV operations. In effect, if the AUV batteries can be recharged, mission data retrieved, and a new mission initiated without recovering the AUV to a ship, then the potential exists to dramatically reduce operational costs and increase deployment duration. Examples of docks designed to support extended deployment include docking systems integrated on moorings [16.1] or attached to cabled observatories [16.2]. The advent of a growing number of scientific observatories directly enables such AUV docking concepts, a circumstance recognized early in the seafloor observatory planning process [16.3].

One of the earliest observing system concepts to employ docking as a key element was the autonomous ocean sampling network (AOSN) paradigm [16.4]. AOSN proposed the use of a heterogeneous mix of vehicles to synoptically characterize dynamic ocean processes. The system used a mix of buoyancy-driven

vehicles (*gliders*) and propeller-driven AUVs. Docking was invoked as a method to extend the endurance of the propeller driven vehicles and to collect and report observed data. It is worth noting that at the time of the AOSN paper, propeller driven AUVs had typical endurances of a fraction of a day, buoyancy-driven systems (gliders) did not exist yet, and satellite communication systems like Iridium were still in the future. The AOSN concept attracted significant effort, and a range of docking technologies were developed and demonstrated [16.5–8].

With growing acceptance of survey-class AUVs in the commercial sector, interest is growing in the use of AUVs for activities such as periodic inspection of subsea equipment installations [16.9] and the use of AUVs for maintenance and repair activities [16.10–12]. Many scenarios envision an AUV that is resident with the undersea equipment, thus creating a docking requirement. Another scenario uses an AUV to deliver a package, such as an ROV (remotely operated vehicle) [16.13, 14] to the equipment requiring servicing. In this scenario, the delivered ROV, once attached, can be tele-operated using connectivity through the dock and undersea infrastructure, allowing ROV operations without a ship.

The demands of launch and recovery often limit the operational use of AUVs. Consequently docking has been identified as an enabling capability to support off-board operation of AUVs from submarines [16.15, 16], autonomous surface vessels [16.17], other AUVs [16.18], ships [16.9], and under ice [16.19, 20]. Even in circumstances where humans are available to handle recovery, safety factors will limit deck operations as sea conditions deteriorate. Expanding the weather window for AUV operations can have important financial benefits for commercial operations, and can be enabling for military operations. Consequently, automating launch and recovery is attractive.

The diverse uses of docking systems lead to equally diverse design strategies. The steps include: navigating the AUV to the docking device vicinity, approaching the dock with the appropriate velocity and orientation, and achieving physical linkage of the AUV and the dock. In addition, some docking systems integrate communications and power transfer functions into their design, and in these cases the process of physical linkage is also associated with establishing communications and power links. Factors influencing the nature of the docking include: is the vehicle docking with a fixed or moving object? Will the dock provide power to the AUV, or vice versa? Will a human operator intervene in the process at any stage? Are there reasons to minimize the amount of hardware required for docking on either the AUV or the dock? Even this brief list of questions demonstrates how different applications might drive a docking system design in very different directions.

16.1 Technical Elements of Docking

To understand the elements of docking, it is useful to review an example docking scenario. Consider a dock attached to a cabled observatory that provides power and high-bandwidth communication to a docked vehicle. The dock itself is a funnel structure (Fig. 16.1) into which the torpedo-shaped vehicle (Fig. 16.2) will drive in the docking process. The docking process starts with the vehicle at a considerable distance, and following sequence:

- *En route*: the first phase of docking is for the vehicle to navigate to the vicinity of the dock so that it can directly sense the dock with onboard sensors. During the en route phase of operations, the vehicle typically navigates relative to the Earth.
- *Approach setup*: once the vehicle is in the general vicinity of the dock, it must position itself for an approach to the dock. For example, the vehicle might use an ultrashort baseline (USBL) acoustic system that can measure the direction and range to an acoustic transponder mounted on the dock to position itself in front of the dock cone, aligned for an approach.
- *Approach*: once the vehicle is positioned appropriately, it can make its approach to the dock. Approach techniques range from the simple strategy of simply pointing directly at the dock, to more sophisticated methodologies which attempt to compensate for currents to follow a specific trajectory.
- *Terminal homing*: the last segment of the final approach, as the vehicle closes to within a few vehicle lengths of the dock, is referred to as the terminal homing phase. The accuracy of the final approach to the docking funnel can be improved by using a terminal guidance sensor with very high update rate and high accuracy, for example, an optical homing system.
- *Capture*: a successful final approach terminates with the vehicle impacting inside the docking funnel. The vehicle will continue to thrust forward, driving the nose of the vehicle to the center of the

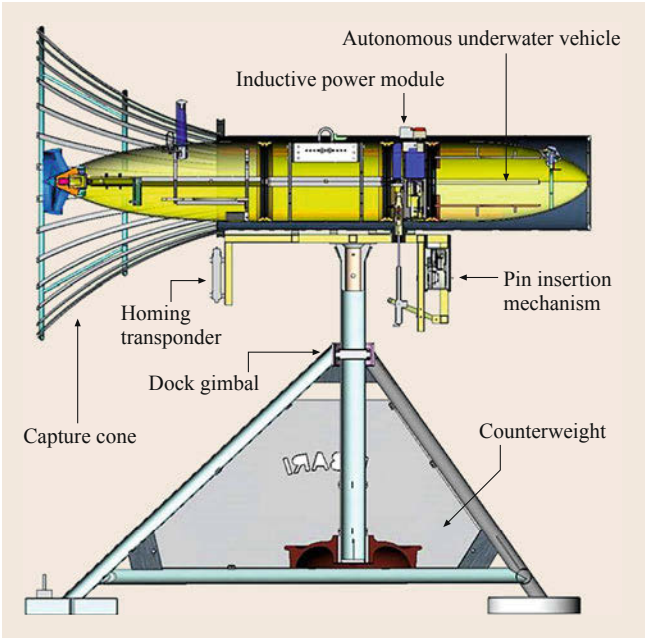


Fig. 16.1 A solid model of a docking station with an AUV captured within the dock (after [16.2]). This docking system is made to rest on the seafloor and be connected to a cabled observatory. The system includes a flared capture cone, which increases the capture aperture of the dock, and a cylindrical housing section, which encloses the docked AUV. A power connection is created by inserting a pin containing an inductive coil into the vehicle, enabling inductive power transfer. The pin also serves to restrain the vehicle within the dock. Communication is provided by an 802.11 ink, supporting short-range communication through seawater. The entire cone assembly is mounted on a gimbal and counterweighted so that the dock will self-level on deployment. Solid works drawing by Jon Erickson, MBARI ►

funnel, where the vehicle will be guided to a capture mechanism.

- *Missed approach*: a robust docking system must anticipate the chance that the final approach will not result in a successful capture. In this case, the vehicle must be capable of detecting that it has missed the dock and repositioning itself for another attempt. This is called executing a missed approach.
- *Connection*: following successful capture, power and communication connections can be made. These may involve actuation by either the vehicle or the dock to make electrical, inductive, or (in the case of communications) electromagnetic/optical connection.
- *Servicing*: once attached, the AUV can connect to the cabled observatory network through the dock. Its batteries can be recharged, data can be downloaded, new missions uploaded, software updated, and so on.
- *Undocking*: to start a new mission, the process must be reversed. Electrical and power connections must be broken. The vehicle capture mechanism must release the vehicle. The vehicle must extract itself from the dock. Once the vehicle is clear of the dock, it can begin its new mission.

The common elements of docking include sensors that allow the AUV to home on the dock, mechanisms for mechanically connecting the vehicle and the dock, and software controlling the overall process. In some cases, the docking system also includes one or more communications links between the vehicle and the dock, and power transfer systems to power and recharge the vehicle.

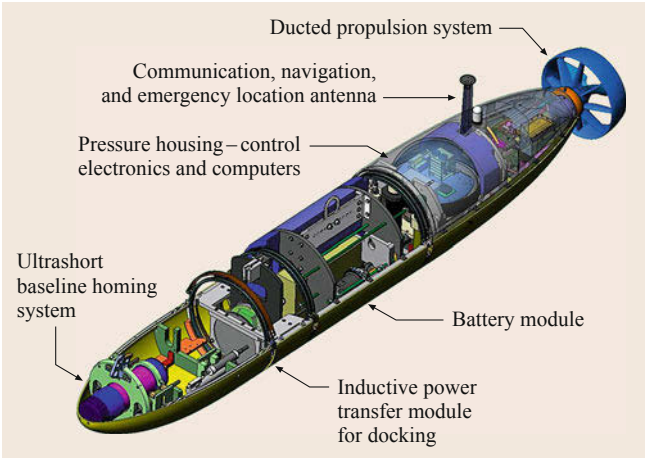


Fig. 16.2 Autonomous underwater vehicle (AUV) configured to dock with docking station of Fig. 16.1 (after [16.3]). Elements of the vehicle are labeled. Docking-related components include the ultrashort baseline homing system, mounted in the nose, and the inductive power module. Other features of the system facilitating docking include the absence of external appendages on the forward portion of the vehicle, which would hinder capture and insertion of the vehicle into the docking enclosure. This system is one of the Dorado vehicles (after [16.21]). Jon Erickson, MBARI ►

16.2 AUV Characteristics

The design and operational characteristics of a docking system are intimately connected to the characteristics of the AUV. Important parameters include the controllability of the platform, its external characteristics, its size, and its ability to withstand impacts. For example, some docking methodologies involve a slow and careful approach to the dock, while others achieve docking by running the vehicle into the dock at or close to its maximum speed. The reasons behind these different strategies largely revolve around the nature of the AUV.

16.2.1 Vehicle Configuration and Controllability

Perhaps the most common type of propeller-driven AUVs are designed like torpedoes; they are propelled by a single propeller, and their attitude controlled by control surfaces at the stern (Fig. 16.2). Such platforms are optimized for efficient movement through the water, but are unable to hover or operate at low speed. They are analogous to aircraft, which are controllable only as long as they move forward. For most torpedo-like vehicles, the minimum controllable speed is usually greater than 0.5 m/s, setting a lower bound on docking approach speed. Examples of torpedo-like vehicles used for docking include: Odyssey [16.22], REMUS [16.23], and Dorado [16.21].

Other vehicles are hover-capable, using multiple thrusters to allow control at low and zero speed (Fig. 16.2). For example, the Endurance AUV, used to explore an Antarctic lake [16.24], is an ellipsoidal vehicle with six thrusters, and is able to move independently in x, y, and z axes and rotate in yaw. While Endurance has little control of pitch or roll, it can pivot in place and exert fine control of its horizontal position while rising vertically. The Seaeye Sabertooth hybrid ROV-AUV [16.25] has vertical and lateral thru-hull thrusters, and forward thrusters at its stern. The SWIMMER AUV developed by IFREMER has an ROV-like configuration [16.13]. The important characteristic of these hover-capable vehicles is that they can make controlled, low speed approaches to a dock, and thus are amenable to precision docking. The ability to operate at low speeds is also attractive when the dock is located in the vicinity of other equipment that might be damaged by the impact of a fast-moving vehicle.

16.2.2 Vehicle Size

The size of the vehicle and dock has a substantial effect on design of the docking system, as size influences the sensors that can be accommodated, the impact velocities that can be sustained, and the forces the vehicle can generate. A small dock size brings obvious disadvantages from the perspective of accommodating the hardware associated with docking. Mechanisms for making mechanical connections between vehicle and dock, for power transfer and communications, all have minimum-size requirement. Further, sensor performance is most often a function of size as well. For example, the direction finding performance

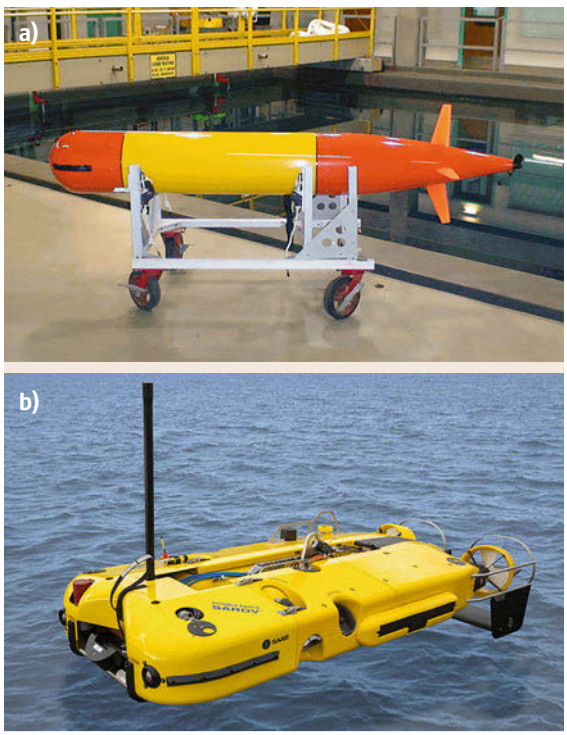


Fig. 16.2a,b Two different configurations of AUVs. **(a)** Torpedo-like AUVs are particularly common as the torpedo shape can support low drag and efficient propulsion, which is attractive for AUV operations in which covering long ranges is important. The vehicle depicted here is the *Tethys* AUV (after [16.26]). The flight characteristics of torpedo-like AUVs encourage docking solutions where the vehicle flies into the dock at some appreciable speed, usually greater than 1 m/s. **(b)** The *Double Eagle SAROV* is characteristic of AUVs that have large numbers of thrusters allowing the vehicles to pivot, and to control forward, lateral, and vertical motion at very low speeds. This allows docking strategies involving slow, precision approaches to docking stations. From SAAB Center Images, reproduced with permission (photos by J. Bellingham) ◄

of an acoustic array for homing is dictated by its aperture compared to the acoustic wavelength at the operating frequency, with smaller arrays having poorer directionality.

However, small size has benefits as well, for example, generally allowing impacts at higher speeds. Consider the docking of a torpedo-like vehicle weighing 200 kg, impacting a dock at 1.5 m/s. The dock and vehicle must absorb 225 J of kinetic energy, not counting added mass of entrained water. If the linear dimension of the system is scaled up by a factor of 2, the vehicle will weigh 1600 kg, and the dock must now dissipate eight times the energy. If, on the other hand, the approach speed can be dropped by half, then the quadratic dependence of kinetic energy on velocity drops the energy dissipated by a factor of 4. Consequently, as vehicle size increases, the speed of docking needs to decrease, and this, in turn, drives larger vehicles to multi-thruster configurations that can be controlled at lower speeds.

The vehicle's ability to apply useful force to the docking process using its propulsion system is likely also to be a function of size. Use of the vehicle's propulsion system to create a force that can complete physical elements of docking is attractive because it minimizes the number of additional actuators that are required. A useful metric, therefore, is the thrust to weight ratio for a given vehicle. Propulsion systems are typically sized to overcome vehicle drag and drag scales with the area of the vehicle. In contrast, mass scales as the vehicle volume. Consequently, if all other factors are held constant, then shrinking a vehicle by two in length will double the vehicle's thrust to mass ratio. Thus smaller vehicles have higher accelerations and are able to apply significantly more force as compared to their mass in the docking process.

16.3 Sensors For Homing

The homing sensor provides the AUV with the information necessary to find and connect to the dock. The ideal homing sensor would allow the AUV to reliably sense the relative location of the dock at any distance with high accuracy, and would update that information many times a second and with no latency. In practice, there is no such technology. Consequently, homing systems for docking are either a combination of different technical approaches with complementary characteristics, or have non-ideal characteristics that must be mitigated by other aspects of the docking system design.

The primary information provided by the homing sensor is the relative position of the two platforms. This provides the vehicle with the information required to approach the dock in a controlled manner. While the AUV presumably has some knowledge of its own location and that of the dock in Earth coordinates, this knowledge is unlikely to be sufficiently accurate for the AUV to achieve physical connection to the dock. Thus the approach and terminal homing phases of docking require accurate relative position information.

Drivers for the homing system key characteristics are as follows:

- *Acquisition distance*: how far away can the AUV acquire information on the relative location of the dock? A homing system that operates at a great distance greatly minimizes the chance that the AUV will not be able to find the dock and/or provides a larger margin for maneuvering the vehicle to position for final approach.
- *Accuracy*: how accurately can the vehicle measure the relative position of the dock? The importance of accuracy increases as the vehicle gets closer to the dock. The simplest homing sensors provide only left–right, up–down cues, without providing accurate relative bearing, elevation, and range. While these systems are adequate for head-on approaches to a homing target, they can be less attractive in more complex situations, for example, when there are currents or when the homing target is offset from the dock.
- *Environmental sensitivity*: many sensing modalities in the ocean have highly variable performance. The range of a particular optical sensor might vary from sub-meter to greater than 100 m, depending on water properties. Methods depending on acoustic signals are sensitive to acoustic conditions, including refraction due to sound speed variations and reflections from the sea surface and seafloor. The severity of performance variability is highly location dependent, and, consequently, the choice of homing system can be heavily influenced by the nature of the environment in which the dock will be deployed.
- *Update rate*: the more frequent the update rate, the more rapidly the vehicle can correct a non-optimal approach. For example, a vehicle traveling 1.5 m/s traverses 3 m between updates given an update rate of 2 s. Under ideal conditions, this might not be a problem; however any cross-currents, dock motion, or errors in the homing system measurement can cause a missed approach.

- *Latency*: some homing methodologies introduce significant latency to the relative position measurement. This increases the ambiguity of the current state of the vehicle, and delays the vehicle's ability to correct for disturbances. In many cases, latency is introduced by signal processing inside the sensor rather than the physics of the sensing technology.

An accurate navigation system onboard the AUV can be particularly useful for docking, allowing the vehicle to navigate between homing system updates, to compensate for latency, and to detect environmental influences like currents.

16.3.1 Acoustic Homing

Acoustic signals propagate over long distances under water and, consequently, are the basis of homing for a number of docking systems. Numerous acoustics navigation methodologies exist, and are often grouped by the size of the transducer array required to achieve a position fix, e.g., long baseline, short baseline, and ultrashort baseline navigation systems. The most commonly used configurations are transponding in nature, in that one system (usually the vehicle) emits an interrogation ping, and the other system (usually the dock) replies. Range is determined by measuring the round trip travel time of the acoustic signal. Practical homing systems used by AUVs have acquisition distances ranging from a fraction a kilometer to many kilometers. The performance depends on both the characteristics of the acoustic homing systems, principally its operational frequency, and the nature of the environment.

The physics of sound in the underwater environment creates a range of challenges. These include the speed of sound, sound refraction due to sound speed variations, interactions of sound with the seafloor and sea surface, ambient noise, and the absorption properties of seawater [16.27]. In complex environments, like shallow water, reverberation and the time varying nature of the ocean surface and interior may cause individual pulses in the received signal to fade or amplify from ping to ping. In signal processing terms, the channel impulse response function is both complicated and time varying. The net effect is that measured properties of a received ping, for example, time of arrival of the leading pulse, may appear to jump in a non-Gaussian way [16.28]. This is a well-known problem for undersea acoustic navigation systems, and a variety of methods have been developed, ranging from methods to reject the jumps [16.29] to other approaches that seek to use the complex nature of the arrival as a probe of the undersea environment sound speed environment [16.30]. Most often, however, the appearance of a multi-path as the separation between source and receiver increases, which becomes yet another factor limiting the useful range of the system.

The update rate and latency of an acoustic system homing are bounded by the finite speed of sound underwater. For the case when the AUV interrogates the dock, and the dock replies, the round-trip for a vehicle at 1.5 km is approximately 2 s. While the latency cannot be improved, the update rate can be increased by simply accelerating the update rate as the vehicle gets closer to the dock. After all, a fast update rate is really only necessary in the terminal stages of homing.

Ultra-Short Baseline Acoustic Systems

Ultra-short baseline (USBL) acoustic homing systems make a phase-sensitive measure of the difference in time of arrival of an acoustic pulse at several hydrophones [16.31]. The difference in time of arrival can be used, in turn, to compute the direction of propagation of the pulse. The phase-sensitive nature of USBL allows comparatively small hydrophone arrays, employing hydrophone separations less than half a wavelength of the pulse. The small size of the hydrophone array, for example, less than 7.5 cm for a 10 kHz signal, allows practical systems with arrays that could fit in the palm of a person's hand. By combining USBL with standard acoustic ranging methods, a vehicle carrying a small USBL can measure the distance and direction to a dock equipped with a transponder. Consequently, a wide number of AUV homing systems have used USBL as the primary homing sensor, including *Light* [16.32], *Singh* et al. [16.5], *Stokey* et al. [16.8], *de Goede* et al. [16.33], *Allen* et al. [16.34], and *McEwen* et al. [16.2].

USBL system integration on an AUV can require careful calibration. Commercial USBL arrays are usually characterized in acoustic facilities to quantify the as-built performance of the array. It is not unusual to require an additional calibration step once the USBL system is mounted on a vehicle, and even to require the addition of materials like acoustical absorbing foam to ensure adequate performance [16.35]. Sources of error include acoustic refraction by any enclosure used for the USBL array, as well as interfering acoustic reflections from vehicle components.

Short Baseline Systems

In contrast to ultra-short baseline systems, short baseline systems (SBL) dispense with the phase comparison and only determine arrival time difference. For monofrequency pulse detection systems the electronics are simple, most often an in-band energy detector of a hard-limited acoustic signal. However, the accuracy of such arrival-time schemes are not as good as phase-sensitive

methods, and, consequently, array sizes must be significantly larger to get usable accuracy. For this reason, homing methods using SBL methods have been implemented on docks rather than on vehicles.

An *inverted* SBL scheme was implemented by *Cowan* [16.36] in which three radially distributed hydrophones on a dock measure the azimuth of a pulse arriving from the vehicle. The resulting measurement is encoded in two reply pulses, separated in frequency and time. The time separation of the two pulses communicates the azimuth information. The attractive aspect of this system is that it minimizes the complexity of the system on the vehicle, reducing the job of the vehicle to simply measuring time-of-arrival of the two pulses. However, some small latency is introduced by the round-trip time. Latency is minimized both by making the system highly integrated and by operating it at short ranges. Cowan's system was intended to be used at ranges of approximately 300 m or less.

Another inverted SBL system was tested by *Smith* and *Kronen* [16.37] for application to docking. The commercial Desert Star Dive Tracker system introduced sufficient latency, even at short ranges; Smith concluded that the approach was unattractive for vehicles making a fast approach (a meter per second or more). However, the approach was deemed useful for slower moving vehicles.

Acoustic Imaging as a Docking Aid

Acoustics can be used to map undersea structures, thus providing a potential cue for docking. The SWIMMER vehicle, which is an AUV that delivers and connects an ROV to subsea infrastructure, used an acoustic imaging system for the final approach of the SWIMMER vehicle to a dock [16.38]. The preliminary tests used an imaging system consisting of two orthogonally mounted scanning sonars. The target consisted of the docking station of which the vehicle was provided with the CAD (computer-aided design) model. The imaging system compared the actual sonar return to a simulation of the sonar return, generated with an estimated position, the CAD model, and a sonar model, to generate updated positions. The system was demonstrated with a mockup of the docking system in shallow water at Cybemetix' Marseilles harbor facilities.

16.3.2 Optical Homing

The advanced state of light sources, sensors, imaging systems, image processing algorithms, computation, and associated technologies provides a strong foundation for a variety of optical homing methods. Optical systems can be roughly divided into some simple categories: systems that home on active lighting systems, systems that home on structured patterns, and systems capable of homing on arbitrary objects. The complexity of the different approaches varies tremendously. Homing on an active light source can be one of the simplest homing systems available, while software intensive vision-based homing methodologies require substantial computational capability, and even then may be limited in update rate.

The acquisition range of optical systems is, in general, greatly limited compared to acoustics in natural ocean environments. Attenuation of optical signals occurs due to scattering and absorption of light by both seawater itself, and the particulate and dissolved constituents of seawater [16.39]. The optical properties of natural seawater are highly wavelength dependent, with the best clarity coinciding roughly with those of human vision. The presence of particulates and dissolved chemicals greatly changes the absorption and scattering properties of water, and can increase attenuation many order of magnitude. As a consequence, light propagation in the ocean is highly location and depth-dependent. Proximity to coastal runoff, phytoplankton blooms, seafloor sediment, and other sources of particles and dissolved compounds can dramatically change optical properties and, therefore, the performance of optical homing systems.

Homing on Active Light Sources

Cowen et al. [16.7, 36] demonstrated an elegant optical homing system using a chopped light source as the homing target and a quadrant detector on the AUV to provide homing cues. The system provided right–left and up–down cues to the vehicle. The directional nature of both the light source and the detector meant that the vehicle needed to approach the dock towards the light source in order to detect the source. Despite water clarity so poor that the vehicle was not visible at a depth of 2 m, acquisition distances of the optical system were 10–15 m. The performance of the optical homing system in Buzzard Bay demonstrated acquisition distances of roughly 28 m. The author assisted with these tests and observed that when the Odyssey AUV acquired lock on the light source at sufficient distance (more than a few meters), it always hit the cone dock target.

Ambient illumination from the sun or other light sources poses a potential source of confusion when homing on the light source. The system described by *Cowen* discriminated against ambient light using two techniques. By modulating the light source at 40 Hz the system could reject the more slowly varying ambient light using a high-pass filter. The more spatially constant ambient light could also be rejected, as compared to the point source target, by using a differencing

scheme with quadrant detectors to reject uniform lighting [16.7]. In tests in San Diego Bay, when water clarity reportedly obscured the yellow Odyssey AUV submerged at depth of 2 m, the homing system was only confused when submerged very shallowly (about 1 m or less) and pointing towards a setting sun.

The use of multiple lights on the dock and a vision system on the AUV allows more sophisticated homing. With the appropriate configuration of lights, the vehicle can compute the vehicle position in dock coordinates. This is particularly useful if the dock is a cone or some other system that requires the vehicle to approach from a particular direction. *Hong* et al. [16.40] proposed a system using five lights arranged around the perimeter of a docking cone, and a sixth light in the same plane as the other five, but outside of the cone. This arrangement allowed the vehicle to compute its position and orientation relative to the dock when all dock lights were in the field of view of the AUV imaging systems.

The five-light system implemented by *Park* et al. was tested in a Korea Ocean Research and Development Institute (KORDI) engineering basin [16.41]. Park eliminated the sixth light, focusing on extracting a directional cue and a range measurement. The environment provided challenges such as interfering light sources, including reflection of the dock lights off the water surface. Docking was initiated with the ISiMI AUV [16.42] at a distance of 15 m pointing at the dock. The authors report successful docking without using range estimation, but also provide a range of suggestions for improving the image processing and control algorithms.

Murarka et al. [16.20] used light-source homing for recovery of their Endurance AUV through ice holes in the Antarctic. This system used a flashing light, reminiscent of *Cowen* et al. [16.7], to guide the AUV to the ice hole. However, the AUV was equipped with a vision system rather than a simple quadrant detector. The light source was suspended to the bottom of the ice hole, where it would be detected by an upward looking camera on the AUV. Endurance is capable of hovering and moving vertically. Once the flashing optical target was detected, Endurance navigated to the ice hole entrance, and then ascended through the 2 m-diameter, 3 m high ice hole. The vision base systems accomplished both approach to and ascent through the ice hole. The authors report using the algorithms for eight descents and ten ascents through the ice.

Homing on Patterns and Arbitrary Structures

Vision systems can also be used to home on passive targets. These targets may be patterned to have specific visual signatures useful to the vision processing algorithm. For example, *Maire* et al. [16.17] tested a vision-based system for homing on black and white striped poles. The system depended on ambient light for illumination, and was tested in an outside diving pool with the Starbug AUV. The vision processing algorithms were able to recognize and distinguish between two different poles by the differing width of the black and white stripes. The docking concept in this case consisted of landing on the bottom between the vertically oriented poles.

Kushnerik et al. [16.43] describe a vision-based docking system using a passive visual marker on the bottom as the homing target. They explored a number of potential marker patterns, finally selecting nested rectangles, as they provided orientation information. They report using the tunnel sea lion (TSL) autonomous vehicle to demonstrate detection, servoing, and landing on the markers using a down-looking camera in natural light at a water depth of 8 m in Peter the Great Bay.

16.3.3 Electromagnetic Homing Systems

Electromagnetic methods face a special challenge in the ocean environment due to the conductive nature of seawater. Not only are radio frequencies strongly attenuated, but the effects are frequency dependent. Below gigahertz frequencies, attenuation goes as the inverse of the root of the frequency. At a frequency of 1 MHz, the $1/e$ attenuation distance, also called the skin depth, is 0.3 m. However, at 1 kHz the skin depth increases to approximately 10 m. Consequently, homing systems using electromagnetic techniques must function at low frequencies. Although the speed of propagation of electromagnetic signals is greatly slowed in a conducting medium, it is still effectively instantaneous in the undersea environment. Thus the latency of an electromagnetic homing system is largely determined by the implementation of the homing system, rather than the physics of electromagnetic propagation in the ocean.

Magnetic Dipole Systems

Feezor et al. [16.6] developed an electromagnetic homing system for docking, using an oscillating magnetic dipole to generate a homing signal. Since magnetic fields are vector quantities, a sensor that fully resolves magnetic components can determine the direction of the magnetic field. By aligning the homing dipole with the docking system, the magnetic measurement, in effect, conveys the orientation of the dock. Thus the magnetic dipole system is attractive for docks that require a directional approach, but whose orientation might be unknown.

The oscillating nature of the magnetic homing signal creates a problem in that while the alignment of the magnetic field can be determined, its absolute direction

is ambiguous. To resolve this ambiguity, an additional dipole source was introduced to transmit a timing signal, allowing synchronization of the homing signal. The source coil for the timing dipole field was oriented vertically at the dock, so that a vehicle approaching in the plane of the dock would see a vertically oriented timing magnetic field. The homing dipole was oriented so that its axis aligned with the axis of the docking funnel. The vertical source coil generated a dipole timing field for the AUV oscillating at 1 kHz, while the homing dipole field oscillated at 2 kHz. The two signals were locked such that a frequency doubled signal generated by the timing field would be synchronized to the homing dipole field.

A three-axis magnetic sensor on an Odyssey IIb AUV was employed to detect and home on magnetic signals generated by the dipole fields. On the AUV, a vertical detection coil was used to measure the magnetic field generated by the vertical source dipole, following which a phased-lock-loop captured the phase information in the timing signal. The frequency-double output of the phase-locked-loop provided a reference for a lock-in amplifier style detection of the homing signal. The provision of a reference signal, generated from the timing signal, allowed determining both the amplitude and the direction of the homing field. Because the AUV had a three-axis sensor, the homing field could be completely resolved in the coordinate frame of the AUV. The AUV homing algorithm commanded the vehicle heading to align the vehicle with the homing field. In effect, the AUV followed the field lines into the dock. Tests of this system on Odyssey II AUVs in Buzzards Bay, Massachusetts, were highly successful with the caveat that the vehicle needed to approach to a range of approximately 30 m to detect the magnetic dipole fields.

16.3.4 Constraining Vertical Position

Homing can be reduced to a two-dimensional horizontal homing problem if the vehicle knows the vertical location of the dock. This can be particularly attractive when there is the potential that the homing signal may suffer from degradation in the vertical plane, for example, when the dock is near a reflecting surface such as the sea surface or seafloor. In general, the vehicle will have at least a rough knowledge of the dock depth, for example, that it is near the seafloor or in a particular depth range in the water column. However, to use depth as a cue for the terminal phase of homing requires the vehicle to know the dock depth to an accuracy better than the vertical capture aperture of the docking system.

Pressure can be used as a cue for vertical plane homing, effectively allowing homing at a constant depth, if the vehicle knows the depth of the dock. However, pressure as a cue can be complicated by factors such as tidal variations in water depth, changing atmospheric pressure at the surface, inaccuracies and hysteresis in pressure sensors, and changes in the density of the water column. For example, tidal variability can cause the depth of the dock to change as local sea level rises and fall if the dock is secured relative to the seafloor. Environmental factors can be mitigated by providing a method for the dock to communicate its pressure depth to the vehicle, but at the cost of added complexity.

Altitude is a particularly robust vertical plane cue if the dock is mounted in a way that keeps its height over the seafloor constant. For example, the docking system in Fig. 16.1 uses a rigid tripod, holding the dock a constant 3 m over the seafloor. For this system, the AUV uses its altimeter to maintain a constant 3 m altitude during the final homing phase, resulting in highly accurate homing in the vertical plane.

16.4 Capture and Connection Mechanisms

Capture is the process whereby the AUV becomes physically attached to the dock. Connection describes the activity of securing the vehicle and making any links, for example, for power or communications. Typically, the capture mechanism guides the vehicle into alignment, brings the vehicle into its final docked position, and latches the vehicle and dock together to ensure that the vehicle will stay connected. The connection mechanisms establish electrical or optical connection between the vehicle and the dock, and may secure the vehicle within an enclosure that protects the vehicle against the environment while docked.

Chief among the goals associated with design of the capture mechanism is to simplify the homing problem. Considerations include providing a large capture aperture in the horizontal and vertical planes, and reducing constraints on approach direction and attitude. This is simplest to understand in the context of a dock shaped like a funnel, into which the vehicle will attempt to drive. If the funnel is very small, then both homing sensors and vehicle control must be very accurate to ensure that the vehicle impacts the interior of the funnel. The funnel also requires that the vehicle approach the mouth of the funnel along the axis of the funnel. Thus, the funnel dock requires the vehicle to know the orientation of the funnel and be able to achieve an orientation to make a final approach along the funnel axis.

The concept of a capture aperture is a useful abstraction for comparing docking methodologies. This is best understood in the context of a simple model of a dock as an area in space which defines the region the vehicle must encounter during final approach for a successful capture. Typically, there are constraints on the alignment of the vehicle with the dock as well.

16.4.1 Directional Docks: Funnels

Funnel docks are designed to provide a large cross-section with which to capture the vehicle and guide it to a capture mechanism. In some cases, the funnel guides the vehicle to a cylindrical structure which encloses all or part of the AUV. In most cases, the vehicle drives into the dock under propulsion power. Examples of funnel docks include the early Odyssey docking systems [16.6, 7], the REMUS docking system [16.34], the Dorado docking system [16.2], and the UUV (unmanned underwater vehicle) Docking and Recharge System designed to accommodate multiple vehicles [16.18]. Cone docking systems are most often used with torpedo-like AUVs.

Attractive features of funnel-type docks include the potential for a large capture aperture. Further, funnels can be constructed to fully constrain the AUV after capture, for example by placing a cylindrical vehicle *garage* at the base of the funnel (Fig. 16.1). This simplifies the subsequent problem of establishing connections for power and communications. Further, if the dock completely envelops the AUV, then the AUV is better protected while it is docked, which can protect the vehicle against biofouling or other environmental hazards. Finally, the funnel dock can be constructed to support most of the docking functionality, minimizing the amount of additional hardware the vehicle must carry.

The funnel dock imposes constraints on the approach direction, as the funnel is fundamentally directional in nature. This creates the need that an approaching vehicle be aware of the docking orientation. In cases in which the orientation of the dock is fixed on deployment, the vehicle only needs to know the dock alignment. In other cases, the dock orientation is not fixed; for example, it may be allowed to weathervane to align with prevailing currents. If the orientation of the dock is variable, then the orientation of the dock must be communicated to the vehicle.

Most funnel docks are either bottom-mounted (*McEwen* et al. [16.2]) or designed to be connected to a large platform of some sort [16.1, 18]. In the first case, the orientation of the dock can be fixed during

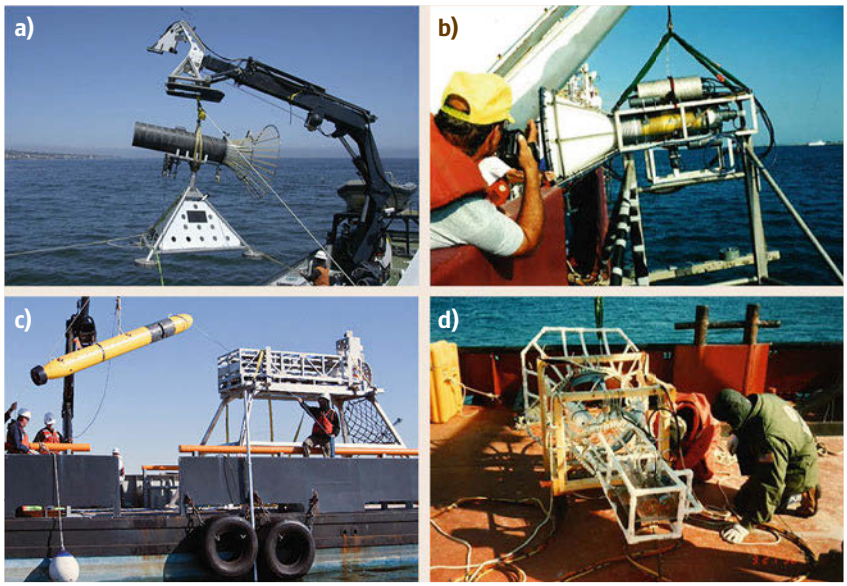


Fig. 16.4a–d Four cone-type docking stations. **(a)** Docking station depicted in Fig. 16.4 supporting the Dorado AUV docking (after [16.2]). The system is depicted as it was deployed on the MISO cabled observatory in Monterey Bay. **(b)** REMUS docking station described by *Stokey* et al. (after [16.8]). **(c)** A docking station built by Bluefin Robotics and Bettelle Memorial Foundation, in which a single cone can be aligned in front of a rectangular enclosure containing several vehicle garages (provided by Jeff Smith, Bluefin Robotics, reproduced with permission). Thus the system is unique in its ability to support multiple AUVs. **(d)** A docking system using a unique electromagnetic homing system, used with the Odyssey IIb AUVs, described in the text (after [16.6])

deployment. In the second case, the dock orientation is determined by the platform hosting the dock. Since almost all dock homing methodologies assume a constant or near constant depth approach, assuring the dock is level is an important consideration in design and deployment. For example, a bottom-mounted system might be placed on an uneven surface, causing the dock to be tilted. The Dorado dock [16.2] gimballed the dock on the support structure, using a large counterweight to hold the dock upright and level. Various funnel docks are depicted in Fig. 16.4.

16.4.2 Omnidirectional Docks: Poles

Pole docks are vertical structures, usually rigid poles or cables under tension, to which the vehicle attaches with some latching device. In effect, the vertical aperture of the dock is defined by the length of the docking segment of the pole, and the horizontal aperture is defined by the width of the capture mechanism on the AUV. The capture mechanism is often a fork on the nose of the AUV (Fig. 16.5) or tines which extend from the side of the vehicle that can capture the pole as the vehicle sweeps by. Pole docks are a solution when a dock must be integrated into a mooring. An early pole dock was developed and successfully demonstrated during the AOSN project [16.5].

A key feature of a pole dock is that an AUV can approach from any direction, eliminating the need to control the dock orientation in the horizontal plane or communicate the orientation to an AUV. This means, for example, that a vehicle can always approach into

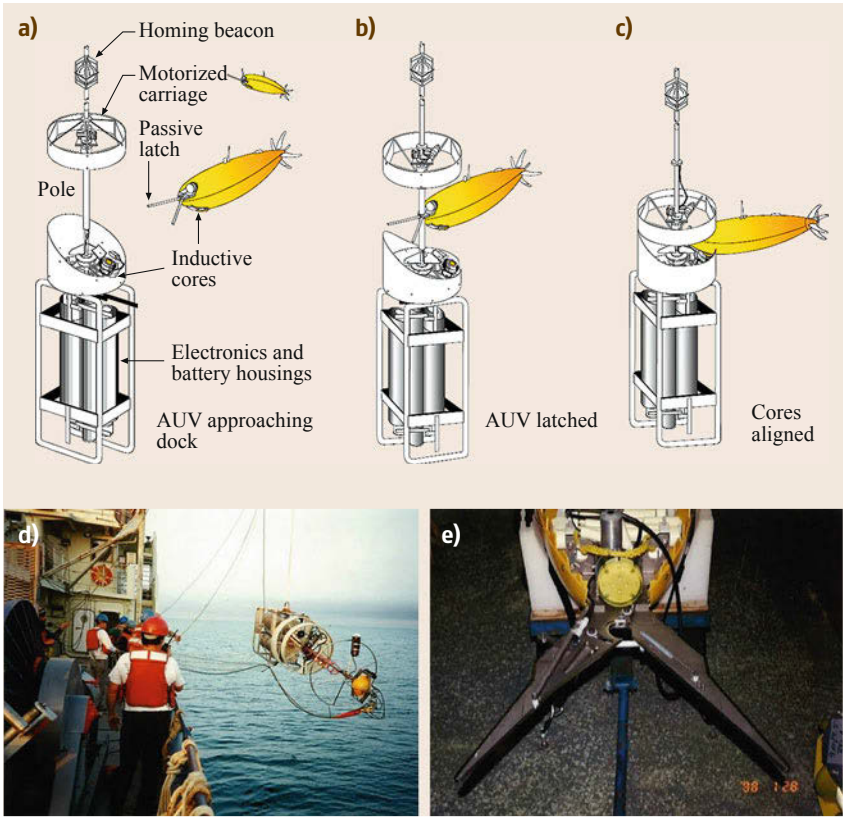


Fig. 16.5a–c This pole dock system (after [16.5]) was built to be an integral element of a deep-water mooring (after [16.1]) and used with an Odyssey IIb AUV. **(a–c)** The function of docking station is demonstrated, consisting of approach, capture, and engagement. The final step, where the vehicle is clamped by the motorized carriage, brings inductive cores on the AUV and the dock onto alignment, enabling both power and communication connectivity. This docking methodology has the advantage of allowing an approach from any horizontal direction (after *Singh* et al. [16.5]). **(d)** The docking station being deployed. **(e)** The front of the AUV, with the capture fork extending towards the photographer, and the ultrashort baseline homing array disk positioned just behind the above the docking fork. The top fairing of the vehicle is removed for the photo. Photos by the author J. Bellingham

the current, provided the vehicle can measure current direction. Pole docks can also support a larger vertical aperture fairly easily by simply making the docking segment of the pole long. This feature significantly simplifies homing in the vertical plane. However, pole docks have the disadvantage that the vehicle must carry some sort of mechanism for attaching itself to the pole, thus adding to vehicle size and weight, and possibly to drag.

Although omnidirectional in the horizontal plane, pole docks can be more complicated homing targets in the vertical plane, despite their large capture aperture. It is not simple to place a homing target like an acoustic beacon or a light source in the middle of a pole. Singh placed the homing beacon at the top of the docking pole. In such a configuration, the AUV must home on a location offset from the homing signal emitter. This complicates vertical plane homing, as a simple dead-ahead homing strategy does not work. Options include constraining vertical plane homing by pressure or altitude, or using a homing system that provides good off-axis navigation.

Ensuring a pole dock stays vertical can be difficult or impossible, depending on the structure hosting the dock. For example, if the pole is part of a mooring, then a current can introduce a tilt into the mooring and by extension to the dock. If the dock is not upright, the vertical offset between the pole and the homing target introduces a horizontal offset for approach directions perpendicular to the tilt. Further, a tilted dock can complicate engagement of the vehicle with the dock.

While pole docks simplify the AUV approach phase, they complicate making power and communications links between the dock and the vehicle. The ability of the vehicle to attach to the dock from any direction means that any communication or power connection to the dock must either be omnidirectional, or the vehicle and the dock must be aligned. In the case of the Woods Hole Oceanographic Institution (WHOI) dock, a mechanism nicknamed the *juicer* was used to clamp the vehicle into place and align the vehicle with the dock's communications and power transformer [16.5]. However, clamping the AUV to the pole also creates potential problems. The clamped AUV is potentially subject to rotating currents imposing a winding motion of the AUV around the mooring axis. One could allow the AUV to twist the entire mooring, or put some sort of electrical slip-ring somewhere between the mooring and the AUV, but such additions increase system complexity and compromise reliability.

16.4.3 Soft Docking Approaches

Soft docking methodologies use a low approach speed during the final approach, and thus require a vehicle equipped to be controllable at low and no speed. Motivations for soft docking include removing constraints on the nature of the docking structure, achieving greater control of alignment of the vehicle and dock during capture, and minimizing docking forces. For example, AUV servicing of subsea infrastructures may require that the vehicle attach itself to a subsea control panel [16.44] similar or identical to those designed for servicing by an ROV. Thus the AUV must be able to approach the panel in a similar manner to a ROV. Another motivation for soft docking is to allow simpler dock design. For example, some docking work has simply used a cradle on which the AUV *lands* [16.43].

Soft docking implies a vehicle with control over four or more degrees of freedom at zero and near zero speed. For example, ROVs can control surge, heave, sway, and yaw, but usually not roll or pitch. Thus, soft-docking AUVs are typically multi-thruster vehicles. Examples of vehicles used for soft docking work include the Australian Starfish and the SAAB Double Eagle SAROV in Fig. 16.2.

16.4.4 Other Docking Mechanisms

In the methods described above, the vehicle is the active element that achieves a mechanical connection between the vehicle and the dock. At most, the dock actuates mechanisms to engage to the vehicle after the initial connection has already been established. However, there is a class of approaches that involves the dock actively grappling the vehicle. Examples include a vehicle recovery scheme in which a towed vehicle homes and grapples with a free swimming AUV [16.45, 46]. An under-ice docking scheme [16.19, 47] involved a multi-step process whereby an AUV (the Explorer) autonomously homed to the vicinity of an ice-mounted dock and came to rest on the under-ice surface, where a small ROV attached lines used to pull the AUV into the dock.

Yet another approach is discussed by *Smith* and *Kronen* [16.37], who evaluated a tail-hook docking system suitable for bottom mounting and *full speed* approaches. The tail hook deployed by the AUV contained a puck at its end, which was captured by a slotted capture mechanism. The capture mechanism allowed entry of the puck from any of four directions. Like the cone and pole docks, the tail-hook concept in principle allows a vehicle to approach at full speed.

16.5 Coupling Power and Communications

As outlined previously, many docking scenarios are motivated by the desire to extend undersea deployment of an AUV. Initially such scenarios were oceanographically driven, however, as outlined in the introduction, commercial and military operational scenarios are incorporating docking as a way to extend deployment durations of AUVs. The need to recharge AUV batteries and to extract data and send new commands is common to almost every extended deployment scenario. Furthermore, when the dock is connected to the shore, the dock provides the physical portal through which an operator can interact with the vehicle. For example, docking tests of the Dorado AUV [16.2] used the Naval Postgraduate School (NPS) Monterey Inner Shelf Observatory (MISO) cabled observatory, allowing video monitoring of the docking and undocking process in real time from the shore. Once docked, the vehicle effectively became a part of the shore network, with all the advantages such connectivity brings.

Even when the dock is not connected to the network, there are strong motivations for enabling the vehicle to download data to the dock. In effect, the dock is usually the *safe* part of the system, so it is the logical location in which to save data. The alternative, in which the vehicle simply accumulates data, creates a situation where each successive operation of the vehicle puts at risk both the vehicle, and all previous data taken by the vehicle. At some point, the risks of operation will outweigh the potential benefit of additional data.

The ability to recharge an AUV's batteries at sea is a key feature of many docking systems. AUVs that carry power hungry payloads like mapping sonar are most often attended by a ship and recovered on a daily schedule for servicing. Recharging an AUV at a dock creates the opportunity to simultaneously support a power hungry payload and remain deployed for extended periods.

16.5.1 Establishing a Communications Link

Communication techniques that have been employed for connecting an AUV and a dock include optical, inductive, radio frequency, and direct electrical connection. Each of these technologies is driven by large commercial markets, and thus the technical foundation is constantly evolving. Factors influencing the choice of communication link include the amount of data that will be transferred; the nature of the engagement and capture mechanisms; the role of communications in the docking and undocking process; the characteristics of the dock environment; and the type of AUV being employed.

Bandwidth requirements for communication connectivity is largely driven by the intersection of data download and operational duty cycle needs. Communication between the AUV and the dock is employed for a variety of reasons, including:

- Downloading mission data from the vehicle to the dock
- Uploading mission scripts from the dock to the vehicle
- Commanding the vehicle to start a mission
- Modifying the vehicle software configuration.

Interacting with the vehicle to upload and start new missions and to modify vehicle software can be accomplished over comparatively low bandwidth connections. In contrast, high bandwidth sensors like multibeam sonar and sidescan sonar generate on the order of 300 kb/s. So if the vehicle can download data at 1 Mb/s, the vehicle must spend roughly 1 h downloading data for every 3 h of mission. Camera systems can generate even more data. Consequently, high-bandwidth connection between the vehicle and the dock can be critical if rapid turn-around of the vehicle between missions is important.

Optical communications was used by the Flying Plug docking system [16.36], inductive coupling by the Odyssey IIb [16.48], radio frequency communications by the Dorado [16.2], and direct electrical connection by the REMUS [16.8]. The advantage of both optical and inductive communications is that no electrical connection need be exposed to seawater or mated via an underwater connector. The advantage of the direct electrical system is that it is relatively simple and uses off-the-shelf hardware and software. The radio frequency (RF) communication system used by the Dorado employed off-the-shelf 802.11 hardware and provided 10 Mb/s communications through seawater at separations up to 6.25 cm.

The RF-based communication link used by the Dorado offered the particular attraction that the dock and the vehicle established a communication link while the vehicle was not yet fully docked. When the dock is cabled to shore, as in the Dorado tests that occurred at the MISO cabled observatory [16.49], the potential exists for human intervention in the docking process. For example, the Dorado tests could be monitored by a video camera on the dock, and when the dock established a communication link with the vehicle, that communication link by extension allowed operator interaction with the vehicle. With the appropriate interface to the vehicle, this allows the operator to take control of the vehicle and manually control docking should the need arise.

16.5.2 Power Transfer

The primary objective of many docking scenarios is to extend AUV deployment times by recharging AUV batteries. Consequently the ability to transfer power through the dock to the AUV is a key feature of many docks. Being able to recharge batteries quickly allows a higher operational tempo of the AUV, and thus the power capacity of the dock is a critical characteristic. The total energy capacity of a smaller AUV might be only a few kilowatt-hours, while larger 21″ diameter AUVs might have battery packs with capacities on the order of 10 kilowatt-hours or more. Other considerations include reliability, alignment, mating force, complexity, power capacity, efficiency, size and weight.

There are a number of ways to establish a power connection between the vehicle and the dock. These include underwater mateable electrical connectors, inductive electrical coupling, transfer of material, or mechanical coupling. Docking systems to date use the first two approaches, direct electrical connection or inductive coupling. In principle, transfer of material might include physically swapping batteries or fueling chemical reactants. These might be attractive for faster recharge of a vehicle, but have not been implemented. Mechanical coupling might be as simple as letting the vehicle's propeller windmill in a current, and using the resulting power generated by the spinning propulsion motor to charge the battery. No mechanical power transfer methods have been reported in the literature.

Wet mateable electrical connections are conceptually simple, but require careful alignment, significant mating forces, and can have a limited lifetime. A survey of six wet mateable connectors [16.50] reports mating forces ranging from approximately 7–55 N per circuit. Most connectors were rated for 100 mate cycles, with two commercial connectors claiming greater than a thousand cycles. A unique wet mateable connector described by *Stokey* et al. [16.8] takes advantage of the fact that only two connections are required. The connector establishes a two-contact electrical connection between the vehicle and the dock, but only insulates one of those connections from seawater. Electrical isolation of the power system from the vehicle hull prevents anodic dissolution of the hull.

Inductive power transfer systems have the advantage of no exposed electrical contacts and can be designed to be less sensitive to alignment, and have low mating force. Reliability is typically limited by the lifetime of the electronics associated with the inductive process, rather than the number of mate/demate cycles. Power capacity and size and weight are related, as efficient high power transfer requires a larger inductor, and the high-permeability materials that make high efficiency power coupling possible are quite dense. Systems developed by *Feezor* et al. [16.48] and *Granger* et al. [16.51] used inductive coupling for both communication and power transfer. The Florida Atlantic University (FAU) inductive power system developed by *Coulson* et al. [16.52] was used for the Dorado docking system.

Autonomous recharging requires a sophisticated battery management system that can maximize energy stored in the batteries while simultaneously minimizing battery failures [16.53]. Examples of battery failures include the destruction of the advanced swimmer delivery system by a battery fire. Lithium-ion battery fires on the Boeing 787 grounded that aircraft's commercial operations for months. In the undersea environment, where batteries might be in a pressure vessel, a battery fire is a catastrophic event.

16.6 AUV Control Considerations

In most AUV control systems, AUV control is segregated by the level of abstraction. At the lowest level, dynamic control systems command actuators to achieve desired orientation, speed, and position. At a higher level, algorithms attempt to accomplish a particular task by commanding the dynamic control system. Above the task level, software monitors the state of tasks, and manages sequencing of activities. While a variety of software architectures are used on operational AUVs, most conform to some semblance of this hierarchy. In principle, even more capable systems might include yet another level that could create plans to achieve user defined goals, and replan when circumstances change; however, in practice this level of autonomy has not been used in docking AUVs thus far.

16.6.1 Navigation, Currents, and Arriving at the Dock

Control algorithms for bringing an AUV to a dock have diverse levels of sophistication, depending on the demands of the docking process. Perhaps the simplest are *pure pursuit* algorithms, which simply point the vehicle directly at the homing target. These are appropriate for docking systems that can be ap-

proached from any direction, like pole docks. Slightly more sophisticated homing control is required when the dock must be approached from a particular direction, as is the case for a funnel-type dock. Control becomes more demanding for the directional approach when there is a significant cross-current. Even more complex homing is required in circumstances where there are obstacles, and soft docking is required.

Pure pursuit, in which the vehicle always points directly at the dock, is the simplest homing method. In the event of a cross-current, pure pursuit results in a hooked trajectory, in which the vehicle is *blown* to one side of the dock. This can result in the vehicle missing the dock, and also means that the final approach direction depends on the speed and direction of the current. This makes pure pursuit unattractive for docking systems that require the direction of approach to be controlled. A variation on pure pursuit is to estimate the current and use the current estimate to compute a heading correction that cancels the cross-current component (the current orthogonal to the direction to the dock). *Wu* et al. [16.54] described simulations of docking approaches in currents. This removes the *hook* from the approach trajectory. Two-phase approaches are possible as well, for example, using an initial line of sight approach and converting to controlled orientation in the terminal phase, as simulated by *Jantapremjit* and *Wilson* [16.55]. A fuzzy-behavioral approach was proposed by *Smith* et al. [16.56] and *White* et al. [16.57] that drove the vehicle towards a virtual funnel to arrive at a docking station.

The ability to measure the current environment in the vicinity of the dock can substantially improve docking skills. An AUV equipped with a Doppler velocity log will be able to measure current by computing the difference of dead reckoned speed and measured speed over bottom. For example, *Yakimenko* et al. [16.58] describe a trajectory generation approach that can incorporate currents in planning the approach to the dock. Docking methods that are omnidirectional in nature, like the pole dock, can use knowledge of the current to approach into the current so as to minimize cross-current and reduce approach speed. If the current cannot be measured directly by the AUV, an alternative approach is for the dock to be free to rotate and align with the current. If the AUV can sense or interrogate the dock orientation, the near-field docking trajectory will always be directly up-current, avoiding the necessity for a crab angle of approach to compensate for a cross-current. While this necessitates a method to communicate the dock orientation to the vehicle, some homing techniques like the electromagnetic homing method described by *Feezor* et al. [16.6] carry the information with the homing signal.

16.6.2 Managing State: AUV Control

From a control perspective, docking of the AUV can be understood as a sequence of tasks that culminate with the vehicle connected to and secured by the dock. Each of these tasks has specific sensing and control challenges, and in some cases actions by the vehicle must be coordinated with actions by the dock. Overall management of the process involves sequencing between tasks, possibly communicating with the dock, detecting failures, and invoking contingency activities when a failure occurs. To achieve the individual tasks, and to sequence between tasks, many operational systems employ some variant of behavior-based control [16.59, 60].

The building blocks of behavior-based control are the individual *behaviors*, which are control laws that can be combined to accomplish tasks, which in turn can be conditionally connected to achieve more complex missions or respond to contingencies. *Bellingham* and *Leonard* [16.59] describe a behavior-based architecture, illustrating an implementation of homing as used for under-ice homing and recovery. Docking work reported by *McEwen* et al. [16.2], *Feezor* et al. [16.6], and *Singh* et al. [16.5] employed the same behavior-based architecture. Successful docking depends on the vehicle carrying out a sequence of activities such as setting up for the approach, making an approach, and perhaps activating actuators upon contacting the dock to secure itself and make connections for communications and power. Figure 16.6 illustrates the state transition table for an AUV for both docking and undocking, as adapted from McEwen.

16.6.3 Managing State: Coordinating AUV and Dock

Docking involves the coordination of two distinct entities, the dock and the AUV. Consequently, successful control strategies for docking must address both platforms and contend with the reality that communications between undersea systems are unreliable. Some docking designs require significant interaction between the dock and the AUV, and consequently each platform must be aware of the state of the other system in order to execute the appropriate action. Docking systems that demand coordinated activities on the part of both the vehicle and the dock introduce failure modes associated with the communication and coordination process. Consequently, contingencies for failures must encompass the entire system.

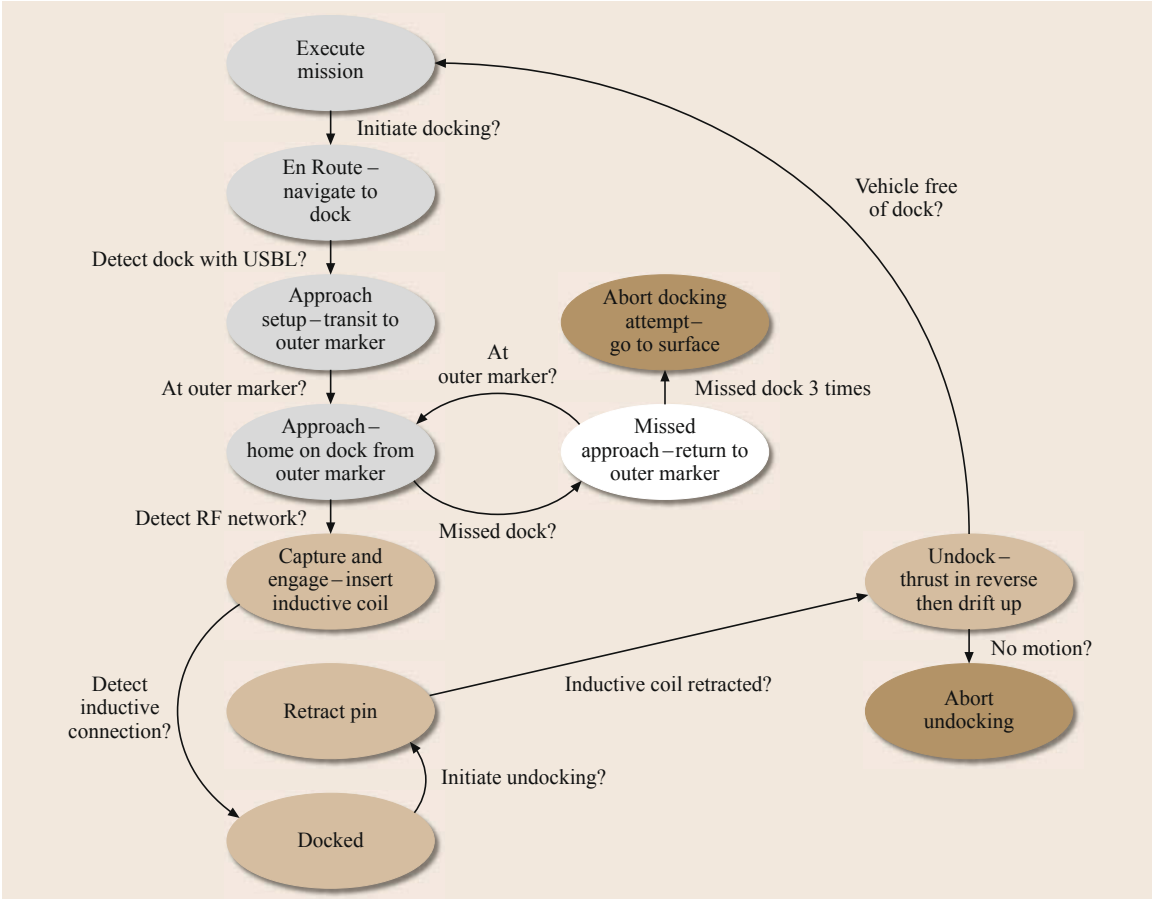


Fig. 16.6 State diagram for docking, undocking, and handling more common failures, derived from *McEwen* (after [16.2]). Starting from the *top* is the sequence whereby the AUV progressively navigates to a position in front of the dock (the *outer marker*) and docks with the dock. If the vehicle misses the dock in its final approach, a missed approach procedure is executed, which repositions the vehicle for another approach from the outer marker. Missing from this diagram is the dock logic, which must recognize that the vehicle is captured in the dock so that it can insert the inductive power pin into the vehicle. Logic for undocking the vehicle is also shown

Examples of AUV-dock interactions that must be managed during the docking process include the following:

- The pole dock (Fig. 16.5) described by *Singh* et al. [16.5] uses a carriage (the *juicer*) to clamp the vehicle in place after the vehicle has attached itself to the pole. Should the clamps activate without an attached vehicle, the dock effectively closes itself to the vehicle, precluding docking.
- The funnel dock described by *Stokey* et al. [16.8] guides the AUV to a cylindrical chamber, in which the vehicle is clamped for electrical connection.
- *McEwen* et al. [16.2] also described a funnel dock that guides the vehicle to a cylindrical chamber. In this system, the dock drives an inductor into the side of the AUV, pinning the vehicle in place and engaging the inductive power transfer system.

In the systems above, the dock must take certain actions at the right time. If the dock attempts to engage the vehicle too soon, they make it impossible for docking to succeed. For example, if the pole dock attempts to clamp the vehicle, it effectively blocks the docking pole, preventing the vehicle from attaching. Although the undocking process has not been discussed much, it should be clear that extracting the vehicle from the dock requires that the dock and the AUV proceed in synchrony as well.

Operator control of elements of the docking process is possible under some circumstances, as described earlier. For example, docks attached to a cabled ob-

servatory are likely to have a very high bandwidth connection with the shore. This provides the potential for operator monitoring and control of dock, and to the extent that the vehicle is in communication, monitoring and control of the vehicle as well.

16.6.4 Detecting and Recovering from Failures

Failures are inevitable with autonomous systems, particularly when deployed for extended periods in a hostile and unpredictable environment. The robustness of a system is thus a function of both its inherent ability to avoid failures and its ability to recover from failures when they do occur. Typically the first is achieved through overdesign, which is expensive, and the second remains a research topic. For most complex engineered systems in use today, robustness is achieved largely through experience; by having processes in place to learn from failures and to implement the lessons learned. The few docking systems that have been tested in the ocean, have been in use only briefly. Consequently, it seems likely that only the most common failure modes have been experienced.

The large number of discrete control functions involved with docking, and the sensitivity of many of these activities to environmental conditions or disturbances, makes failures difficult to predict. Consequently, the overall docking process must be robust to failure of individual steps. For example, a missed approach, where the vehicle attempts to home on the dock but misses, is a particular class of failure that is likely to occur many times. This might be caused by currents, poor sensing conditions, turbulence, or anomalies in vehicle performance. The appropriate response for the AUV may be quite different, depending on the root cause of the missed approach. This leads to both a classification problem and a contingency management problem. It is not enough to know that an approach has failed, because the reason for failure is important as well. However, to the degree that contingency responses are created for different failure modes, mission management becomes more complex. In effect, each contingency response creates the potential for unintended consequences that must be anticipated and mitigated. At present there is no good way to do this.

Handling of some simple failure modes is discussed by both *Singh* et al. [16.5] and *McEwen* et al. [16.2]. The state transition diagram in Fig. 16.6 describes the vehicle response to some simple failures. For example, the missed dock transition leads to a state of which the goal is repositioning the vehicle for another approach. If the vehicle misses the approach three times, it aborts the docking attempt and goes to the surface.

16.7 Conclusions and Future Prospects

Of the activities attempted with AUVs, docking is among the most challenging. The literature reflects this difficulty, with most papers testing some element of docking, and only a few reporting results from a deployed system in the ocean. Most tests focus on homing. Those that deploy and exercise a full docking suite, including battery recharge, data dump, and initiation of new vehicle missions, report results for comparatively brief deployments. Consequently, one must conclude that docking has not made the transition from demonstration to operations. This prompts the questions: why has docking not become an operational capability, and what are the conditions that might allow this event to occur?

In the early 1990s, when the AOSN paper first proposed AUV docking as an integral part of an ocean observing system, little attention was given the infrastructure required to support and power a docking system. However, docking does imply the existence of an infrastructure. For example, if the objective is to support ten AUV full missions from the dock, then the dock must be able to supply power levels equal to ten times the battery capability of the vehicle (assuming the vehicle missions uses the full battery capacity each mission). Early attempts to use docking to support field programs thus had to invest in their own infrastructure, such as moorings with satellite communications and large quantities of batteries [16.1]. This created a significant barrier to the adoption of AUV docking. Not only does a team have to master the most difficult aspect of AUV operations, it must also be capable of building from scratch the dock and the infrastructure supporting the dock.

The deployment of cabled observatories, which adapt subsea telecommunications technology to provide high bandwidth, low-latency communications and high power to nodes in the ocean interior, provide an ideal infrastructure to support docked AUVs. Early scientific cabled observatories include a system deployed off Japan in the late 1970s [16.61] and the LEO-15 system, installed a few kilometers off the coast of New Jersey in the 1990s [16.62]. Since then, a wide number of cabled systems have been installed, supporting *plug and play* instruments on the deep seafloor

connected by many hundreds of kilometers of cable to the internet [16.63, 64]. While one of the more ambitious cabled observatories efforts is funded in the United States (Ocean Observatories Initiative (OOI)), a number of countries are developing and operating cabled observatories [16.65]. It should be noted that AUVs provide a useful complement to cabled systems, the former being mobile and capable of deploying instruments to locations of interest, the latter being fixed in location.

AUV reliability is also an important factor driving adoption. Most AUV operations today are attended by a ship, or at least have ships in the vicinity. Should there be problems, the vehicle can ascend to the surface, where it can be recovered and serviced. A docked AUV is most likely far from rescue and must carry out a particularly complex sequence of activities for every mission. Consequently, operating a docked AUV entails some risk. While one might argue that the risk associated with docking is a necessary part of the docking development process, the ability of the vehicle to successfully complete repeated missions without human servicing is demanding. While buoyancy driven systems are routinely deployed for months at a time, propeller-driven AUVs are typically deployed on the order of a day. However, the advent of a class of very long endurance AUVs, operating up to 3 weeks at a time per deployment [16.26, 66] offers encouragement that AUV reliability can support extended dock deployments.

A final factor moderating docking adoption stems from the comparatively high cost of scientifically equipped AUVs. In effect, there is an opportunity cost associated with committing an AUV to a docked deployment. The owner of the AUV has the choice of keeping the AUV close at hand, or deploying a vehicle at a remote dock at sea. While the sea-deployed system is positioned to respond rapidly to opportunities in the vicinity of the dock, it is effectively unavailable for operations that are beyond operational range of the vehicle from the dock. When the AUV must support a variety of scientific goals, this creates an inherent conflict between users, which is simplest to handle by keeping the vehicle on shore where it is equally available to all potential users. This calculus breaks down, however, when the dock location is remote and unique. For example, science associated with a docked system in winter in the Arctic Ocean removes the AUV from ready access, but also places the vehicle in a location where it is uniquely capable of making observations, as the Arctic Ocean is all but inaccessible during winter. It is worth noting that such remote deployments place a high premium on autonomy, including adaptive capability such as event detection capabilities for triggering AUV missions.

In summary, while AUV docking has been experimented with since the 1990s, only recently has the needed combination of supporting infrastructure, vehicle reliability, and operational need emerged for docking to transition to an operational capability. In the science domain, the emergence of cabled and mooring-based observatories provides a framework that can both support and benefit from docked AUVs. In the commercial domain, the rising interest in inspection, repair, and maintenance (IRM) capabilities for offshore structures offers an application domain where the problems associated with infrastructure and the risk-reward calculation are tipping in favor of resident vehicles. The drivers for adoption of docking capability are growing, encouraging hope that docking will become an operational capability supporting a more permanent presence of mobile platforms in the ocean.

References

16.1 D.E.F. Kemp, W. Paul, D. Peters: Mooring developments for autonomous ocean-sampling networks, IEEE-JOE **26**(4), 477–486 (2001)

16.2 R.S. McEwen, B.W. Hobson, L. McBride, J.G. Bellingham: Docking control system for a 54-cm-diameter (21-in) AUV, IEEE-JOE **33**(4), 550–562 (2008)

16.3 National Research Council (US) Committee on Seafloor Observatories: *Illuminating the Hidden Planet: The Future of Seafloor Observatory Science* (National Academies Press, Washington 2000)

16.4 T.B. Curtin, J.G. Bellingham, J. Catipovic, D. Webb: Autonomous oceanographic sampling networks, Oceanography **6**(3), 86–94 (1993)

16.5 H. Singh, J.G. Bellingham, F. Hover, S. Lemer, B.A. Moran, K. von der Heydt, D. Yoerger: Docking for an autonomous ocean sampling network, IEEE-JOE **26**(4), 498–514 (2001)

16.6 M.D. Feezor, F.Y. Sorrell, P.R. Blankinship, J.G. Bellingham: Autonomous underwater vehicle homing/docking via electromagnetic guidance, IEEE-JOE **26**(4), 515–521 (2001)

16.7 S. Cowen, S. Briest, J. Dombrowski: Underwater docking of autonomous undersea vehicles using optical terminal guidance, Proc. IEEE OCEANS'97, Vol. 2 (1997) pp. 1143–1147

16.8 R. Stokey, B. Allen, T. Austin, R. Goldsborough, N. Forrester, M. Purcell, C. von Alt: Enabling technologies for REMUS docking: An integral component of an autonomous ocean-sampling network, IEEE-JOE **26**(4), 487–497 (2001)

16.9 J. Jamieson, L. Wilson, M. Arredondo, J. Evans, K. Hamilton, C. Sotzing: Autonomous inspection vehicle: A new dimension in life of field operations, Proc. Offshore Technol. Conf. (2012), OTC-23365
16.10 L. Brignone, M. Perrier, C. Viala: A fully autonomous docking strategy for intervention AUVs, Proc. IEEE OCEANS'07 (2007) pp. 1–6
16.11 S. Krupinski, F. Maurelli, G. Grenon, Y. Petillot: Investigation of autonomous docking strategies for robotic operation on intervention panels, Proc. IEEE OCEANS'08 (2008) pp. 1–10
16.12 J. Jacobson, P. Cohen, A. Nasr, A.J. Schroeder, G. Kusinski: Laying the groundwork for AUV standards for deepwater fields, Mar. Technol. Soc. J. **47**(3), 13–18 (2013)
16.13 Y. Chardard, T. Copros: Swimmer: Final sea demonstration of this innovative hybrid AUV/ROV system, Proc. IEEE Underw. Technol. (2002) pp. 17–23
16.14 J. Evans, P. Redmond, C. Plakas, K. Hamilton, D. Lane: Autonomous docking for intervention-AUVs using sonar and video-based real-time 3-D pose estimation, Proc. IEEE OCEANS'03, Vol. 4 (2003) pp. 2201–2210
16.15 M.O. Piggott: *USS Scranton Completes Successful UUV Test* (Navy Newsstand, Washington 2006), NNS 060309-10
16.16 M.S. Stewart, J. Pavlos: A means to networked persistent undersea surveillance, Proc. Submar. Technol. Symp. (2006)
16.17 F.D. Maire, D. Prasser, M. Dunbabin, M. Dawson: A vision based target detection system for docking of an autonomous underwater vehicle, Proc. Australas. Conf. Robotics Autom. (2009)
16.18 D. Pyle, R. Granger, B. Geoghegan, R. Lindman, J. Smith: Leveraging a large UUV platform with a docking station to enable forward basing and persistence for light weight AUVs, Proc. IEEE OCEANS'12 (2012) pp. 1–8
16.19 P. King, R. Lewis, D. Mouland, D. Walker: CATCHY: An AUV ice dock, Proc. IEEE OCEANS'09 (2009) pp. 1–6
16.20 A. Murarka, G. Kuhlmann, S. Gulati, M. Sridharan, C. Flesher, W.C. Stone: Vision-based frozen surface egress: A docking algorithm for the endurance AUV, Proc. 15th UUST (2009)
16.21 M. Sibenac, W.J. Kirkwood, R. McEwen, F. Shane, R. Henthorn, D. Gashler, H. Thomas: Modular AUV for routine deep water science operations, Proc. IEEE OCEANS'02, Vol. 1 (2002) pp. 167–172
16.22 J.G. Bellingham, C.A. Goudey, T.R. Consi, J.W. Bales, D.K. Atwood, J.J. Leonard, C. Chryssostomidis: A second generation survey AUV, Proc. AUV'94 (1994) pp. 148–155
16.23 B. Allen, R. Stokey, T. Austin, N. Forrester, R. Goldsborough, M. Purcell, C. von Alt: REMUS: A small, low cost AUV; system description, field trials and performance results, Proc. IEEE OCEANS'97, Vol. 2 (1997) pp. 994–1000
16.24 W. Stone, B. Hogan, C. Flesher, S. Gulati, K. Richmond, A. Murarka, G. Kuhlman: Design and deployment of a four-degrees-of-freedom hovering autonomous underwater vehicle for sub-ice exploration and mapping, J. Eng. Marit. Environ. **224**(4), 341–361 (2010)
16.25 B. Johansson, J. Siesjo, M. Furuholmen: Seaeye Sabertooth a hybrid AUV/ROV offshore system, Proc. IEEE OCEANS'10 (2010) pp. 1–3
16.26 J.G. Bellingham, Y. Zhang, J.E. Kerwin, J. Erikson, B. Hobson, B. Kieft, M. Godin: Efficient propulsion for the Tethys long-range autonomous underwater vehicle, Proc. IEEE/OES AUV'10 (2010) pp. 1–7
16.27 R.J. Urick: *Principles of Underwater Sound* (McGraw-Hill, New York 1983)
16.28 D.R. Yoerger, M. Jakuba, A.M. Bradley, B. Bingham: Techniques for deep sea near bottom survey using an autonomous underwater vehicle, Int. J. Robotics Res. **26**(1), 41–54 (2007)
16.29 J. Vaganay, J.J. Leonard, J.G. Bellingham: Outlier rejection for autonomous acoustic navigation, Proc. IEEE ICRA'96, Vol. 3 (1996) pp. 2174–2181
16.30 M. Deffenbaugh, H. Schmidt, J.G. Bellingham: Acoustic positioning in a fading multipath environment, Proc. IEEE OCEANS'96, Vol. 2 (1996) pp. 596–600
16.31 K. Vickery: Acoustic positioning systems. A practical overview of current systems, Proc. IEEE AUV'98 (1998) pp. 5–17
16.32 R.D. Light, J. Morison: The autonomous conductivity-temperature vehicle: First in the seashuttle family of autonomous underwater vehicle's for scientific payloads, Proc. IEEE OCEANS'89, Vol. 3 (1989) pp. 793–798
16.33 G.M. de Goede, D. Norris: Recovering unmanned undersea vehicles with a homing and docking sonar, Proc. IEEE OCEANS'05 (2005) pp. 2789–2794
16.34 B. Allen, T. Austin, N. Forrester, R. Goldsborough, A. Kukulya, G. Packard, M. Purcell, R. Stokey: Autonomous docking demonstrations with enhanced REMUS technology, Proc. IEEE OCEANS'06 (2006) pp. 1–6
16.35 M.J. Stanway: Private communication, February 2013
16.36 S. Cowen: *Flying plug: A small UUV designed for submarine data connectivity*, Tech. Rep. NCCOSC Ser., Vol. 796 (Space and Naval Warfare Systems Center, San Diego 1997)
16.37 S.M. Smith, D. Kronen: Experimental results of an inexpensive short baseline acoustic positioning system for AUV navigation, Proc. IEEE OCEANS'97, Vol. 1 (1997) pp. 714–720
16.38 J.C. Evans, K.M. Keller, J.S. Smith, P. Marty, O.V. Rigaud: Docking techniques and evaluation trials of the swimmer AUV: An autonomous deployment AUV for work-class ROVs, Proc. IEEE OCEANS'01, Vol. 1 (2001) pp. 520–528
16.39 D. Stramski, E. Boss, D. Bogucki, K.J. Voss: The role of seawater constituents in light backscattering in the ocean, Prog. Oceanogr. **61**(1), 27–56 (2004)
16.40 Y. Hong, J. Kim, P. Lee, B. Jeon, K. Oh, J. Oh: Development of the homing and docking algorithm for AUV, Proc. 13th Int. Offshore Polar Eng. Conf. (2003) pp. 205–212
16.41 J. Park, B. Jun, P. Lee, J. Oh: Experiments on vision guided docking of an autonomous underwater

vehicle using one camera, Ocean Eng. **36**(1), 48–61 (2009)
16.42 B.-H. Jun, J.-Y. Park, F.-Y. Lee, P.-M. Lee, C.-M. Lee, K. Kim, Y.-K. Lim, J.-H. Oh: Development of the AUV ISiMI and a free running test in an ocean engineering basis, Ocean Eng. **36**(1), 1–14 (2009)
16.43 A.A. Kushnerik, A.V. Vorontsov, A.P. Scherbatyuk: Small AUV docking algorithms near dock unit based on visual data, Proc. IEEE OCEANS'09 (2009) pp. 1–9
16.44 T. Palmer, D. Ribas, P. Ridao, A. Mallios: Vision based localization system for AUV docking on subsea intervention panels, Proc. IEEE OCEANS'09 (2009) pp. 1–10
16.45 R.M. Stenson, D.J. Braun, L.A. Hamme, C.D. Mailey: Underwater unmanned vehicle recovery system and method, US Patent 7854569 B1 (2010), issued December 21
16.46 T.F. Tureaud, D.N. Dietz, S.J. Hills, D.E. Humphreys, A.V. Roup: Docking Apparatuses and Method, US Patent 8141369 B1 (2012)
16.47 C. Kaminski, T. Crees, J. Ferguson, A. Forrest, J. Williams, D. Hopkin, G. Heard: 12 days under ice – An historic AUV deployment in the Canadian high arctic, Proc. IEEE/OES AUV'10 (2010) pp. 1–11
16.48 M.D. Feezor, F.Y. Sorrell, P.R. Blankinship: An interface system for autonomous undersea vehicles, IEEE-JOE **26**(4), 522–525 (2001)
16.49 K.M. Tjoa: *The Bottom Boundary Layer Under Shoaling Inner Shelf Solitons*, M.Sc. Thesis (Naval Postgraduate School, Monterey 2003)
16.50 H. Painter, J. Flynn: Current and future wetmate connector technology developments for scientific seabed observatory applications, Proc. IEEE OCEANS'06 (2006) pp. 1–6
16.51 R.P. Granger, C.M. Baer, N.H. Gabriel, J.J. Labosky, T.C. Galford: Non-contact wet mateable connectors for power and data transmission, Proc. MTS/IEEE OCEANS'13 (2013) pp. 1–4
16.52 R. Coulson, J. Lambiotte, E. An: A modular docking system for 12.75-inch class AUVs, Sea Technol., 49–54 (2005)
16.53 A.M. Bradley, M.D. Feezor, H. Singh, F.Y. Sorrell: Power systems for autonomous underwater vehicles, IEEE-JOE **26**(4), 526–538 (2001)
16.54 L. Wu, Y. Li, S. Su, P. Yan, Y. Qin: Hydrodynamic analysis of AUV underwater docking with a cone-shaped dock under ocean currents, IEEE-JOE **85**, 110–126 (2014)
16.55 P. Jantapremjit, P.A. Wilson: Control and guidance for homing and docking tasks using an autonomous underwater vehicle, Proc. IROS'07 (2007) pp. 3672–3677
16.56 S.M. Smith, G.J.S. Rae, D.T. Anderson: Applications of fuzzy logic to the control of an autonomous underwater vehicle, Proc. IEEE 2nd Int. Conf. Fuzzy Syst. (1993) pp. 1099–1106
16.57 K.A. White, S.M. Smith, K. Ganesan, D. Kronen, G.J.S. Rae, R.M. Langenbach: Performance results of a fuzzy behavioral altitude flight controller and rendezvous and docking of an autonomous underwater vehicles with fuzzy control, Proc. AUV'96 (1996) pp. 117–124
16.58 O.A. Yakimenko, D.P. Horner, D.G. Pratt: AUV rendezvous trajectories generation for underwater recovery, Proc. 16th Med. Conf. Control Autom. (2008) pp. 1192–1197
16.59 J.G. Bellingham, J.J. Leonard: Task configuration with layered control, Proc. IARP 2nd Mob. Robots Subsea Environ. (1994) p. 193
16.60 M.R. Benjamin, H. Schmidt, P.M. Newman, J.J. Leonard: Nested autonomy for unmanned marine vehicles with MOOS-IvP, J. Field Robotics **27**(6), 834–875 (2010)
16.61 H. Mikada, K. Asakawa: Development of Japanese scientific cable technology, Proc. IEEE OCEANS'08 (2008) pp. 1–4
16.62 N.C. Forrester, R.P. Stokey, C. von Alt, B.G. Allen, R.G. Goldsborough, M.J. Purcell, T.C. Austin: The LEO-15 long-term ecosystem observatory: Design and installation, Proc. IEEE OCEANS'97, Vol. 2 (1997) pp. 1082–1088
16.63 Monterey Bay Aquarium Research Institute: Monterey Accelerated Research System (MARS), http://www.mbari.org/mars/
16.64 C.R. Barnes, M.M.R. Best, F.R. Johnson, B. Pirenne: Final installation and initial operation of the world's first regional cabled ocean observatory (NEPTUNE Canada), Canadian Meterological and Oceanographic Society Bulletin **38**(3), 89–96 (2010)
16.65 Y. Kaneda: Advanced ocean floor network system for mega thrust earthquakes and tsunamis, Proc. IEEE UT'11 SSC'11 (2011) pp. 1–6
16.66 B.W. Hobson, J.G. Bellingham, B. Kieft, R. McEwen, M. Godin, Y. Zhang: Tethys-class long range AUVs-extending the endurance of propeller-driven cruising AUVs from days to weeks, Proc. IEEE/OES AUV'12 (2012) pp. 1–8

17. Underwater Vehicle Manipulators

Tae Won Kim, Giacomo Marani, Junku Yuh

This chapter covers current commercial developments and major research activities for underwater vehicle-manipulator systems. It briefly describes the dynamics of underwater vehicle-manipulator systems to show the level of its complexity. The teleoperation of the underwater vehicle-manipulator system and sensor-based underwater vehicle manipulator controls, such as force control and visual servoing are discussed. Furthermore, it describes coordinated motion controls of underwater vehicle and manipulator systems and presents the results of recent developments in autonomous manipulation for autonomous underwater vehicles.

17.1 Underwater Vehicles for Intervention Missions 407
17.2 Dynamics of Underwater Vehicle Manipulators 409
17.3 Teleoperation of Underwater Vehicle Manipulators 410
17.4 Sensor-Based Manipulator Control 411
17.5 Coordinated Motion Control of Underwater Vehicle-Manipulator Systems 413
17.6 Underwater Autonomous Manipulation 414
17.6.1 Phase 1: Undock from the Pier and Navigate to a Search Area 417
17.6.2 Phase 2: Search for the Submerged Platform 417
17.6.3 Phase 3: Navigate and Dive Toward the Platform 417
17.6.4 Phase 4: Hover (Station Keeping)..... 417
17.6.5 Phase 5: Hook a Recovery Tool to the Target Object (Autonomous Manipulation)..... 418
17.6.6 Phase 6: Return to the Pier 418
17.7 Conclusions 419
References 419

17.1 Underwater Vehicles for Intervention Missions

As about 70% of our planet is covered with water, the importance of the ocean cannot be overemphasized. A great deal of effort has been put into acquiring living and nonliving resources in the ocean. Most marine activities have worked on the ocean surface or shallow water, and deep-sea research activities have been limited due to economic and technological reasons. However, recent advancements in deep-sea technologies such as unmanned underwater vehicles (UUVs), have made it more feasible to exploit deep-sea resources, such as mining for seabed minerals, collecting marine organisms for pharmaceutical applications, and extracting oil and gas. There are basically two types of UUVs: remotely operated vehicles (ROVs) and autonomous underwater vehicles (AUVs) [17.1–3]. While many of them are just fly-by-type vehicles for survey and monitoring, some of them are equipped with manipulators for intervention missions. While there have been some recent developments and breakthroughs in underwater manipulator technologies for AUVs (so called *autonomous manipulation*), most commercial UUVs with manipulators are ROVs. A typical work-class ROV for intervention missions has a 7 degrees-of-freedom (DOF) manipulator and a 5 DOF grabber. The manipulator is used for interactive manipulation by a human operator on the supporting ship, whereas the grabber is used to maintain a relative position of ROV to the target object by grabbing a target structure in the presence of an underwater current or wave. Most heavy-duty underwater manipulators are designed for ROVs and are actuated by hydraulic drivers. They are also designed for teleoperation rather than autonomous manipulation. Table 17.1 lists samples of heavy duty underwater manipulators currently available in the market.

A typical work-class ROV for intervention missions requires two individual operators: one steering the vehi-

Table 17.1 Samples of commercially available underwater manipulators

Company name	Eca robotics	International submarine engineering	Kraft telerobotics	FMC technologies schilling robotics	Western Space and Marine
Model	Arm 7H	Magnum-7	Predator	Titan 4	The ARM
DOF	6 plus gripper	6 plus gripper	6 plus gripper	6 plus gripper	6 plus gripper
Power source	Hydraulic	Hydraulic: max 1000 psi/19 lpm (5 gpm)	Hydraulic: 103 bar (1500 psi) ≈ 207 bar (3000 psi), 19 lpm (5 gpm)	Hydraulic:103 bar (1500 psi) ≈ 207 bar (3000 psi), 19 lpm (5 gpm)	Hydraulic: max 204 bar (3000 psi), 7.6 lpm (2 gpm)
Material	Titanium or aluminum	Aluminum with stainless steel fittings	Anodized aluminum and stainless steel	Titanium	
Max grip force	490 N (110 lbf)	2009 N (450 lbf)	1334 N (300 lbf)	4092 N (920 lbf)	1467 N (330 lbf)
Wrist torque	108 Nm (80 ft lbs)	108 Nm (80 ft lbs)	135 Nm (100 ft lbs)	170 Nm (125 ft lbs)	
Actuator	Hydraulic cylinder	Hydraulic cylinder	Hydraulic cylinder	Hydraulic cylinder	Hydraulic cylinder
Max reach	1700 mm (66″)	1524 mm (60″)	2019 mm (79.50″)	1920 mm (75.7″)	1700 mm (66″) from azimuth axis to fingers
Max lift capacity		454 kg (1000 lb)	227 kg (500 lb)	454 kg (1000 lb)	45.4 kg (100 lb)
Max lift at full ext.	90 kg (198 lb)	295 kg (650 lb)	91 kg (200 lb)	122 kg (270 lb)	29.5 kg (65 lb)
Working depth (standard)	7500 msw (24 600 fsw)	6000 msw (19 700 fsw)	3000 msw (9800 fsw)	4000 msw (13 100 fsw)	Unlimited
Working depth (extended)	Unlimited	11 000 msw (36 000 fsw)	6500 msw (21 000 fsw)	7000 msw (22 967 fsw)	Unlimited
Homepage	www.eca-robotics.com	www.ise.bc.ca	www.krafttelerobotics.com	www.fmctechnologies.com/Schillingrobotics.aspx	www.wsminc.com
Data source	Eca robotics [17.4]	International submarine engineering [17.5]	Kraft teleRobotics [17.6]	FMC technologies [17.7]	Western space and marine [17.8]

cle and the other operating the manipulator. One of the main tasks of the vehicle operator is to keep the ROV as steady as possible (station-keeping) since the manipulator is attached to the vehicle, and the vehicle motion can disturb the manipulator and greatly affects the accuracy of its end-effector interacting with the environment. The accuracy of the end-effector would be greatly improved if the manipulator controller could compensate for any disturbance due to the motion of the vehicle. However, most commercial ROV systems do not use vehicle motion compensation control for the manipulator that is directly teleoperated by a human operator. Due to the increase in demand for more dexterous and precise underwater manipulation, many researchers have performed advanced studies for dynamics and control of underwater vehicle manipulators. Considering the vehicle and the attached manipulator as one multibody system, several researchers have studied coordinated motion control of the vehicle-manipulator system to improve the accuracy of the end-effector. The coordinated motion control automatically compensates for any disturbance due to the motion of the vehicle and also takes advantage of the redundancy of the overall vehicle-manipulator system. Many underwater tasks require the position accuracy of the end-effector of the order of a centimeter or subcentimeter, whereas most sensors for the vehicle position have accuracy of the order of a meter. Therefore, in order to effectively implement the coordinated motion control in practice especially in autonomous mode, the sensor accuracy of the vehicle position must be improved. In the past, very few AUV systems for intervention missions have been reported and most of them were used as a testbed for developing autonomous manipulation or as a working model for research.

17.2 Dynamics of Underwater Vehicle Manipulators

The analysis of the dynamics of underwater vehicle manipulators is much more complex than that of regular manipulators on the factory floor. For example, it is difficult to accurately introduce the effects of the added mass and the added moment of inertia in the modeling process for multiple links of the manipulator. The frictional force and the drag force due to the velocity of the manipulator itself, waves, and currents are also complex in modeling. The modeling and control of underwater manipulators were studied by several researchers, including *Kato* and *Lane* [17.9], *Dunningan* et al. [17.10], *Levesque* and *Richard* [17.11], *Mahesh* et al. [17.12], *McMillan* et al. [17.13], *McLain* et al. [17.14], *Tarn* et al. [17.15], *Ioi* and *Itoh* [17.16], *Schjølberg* and *Fossen* [17.17], *Leabourne* and *Rock* [17.18], *Lee* and *Choi* [17.19], and *Ryu* et al. [17.20]. In this section, it is not our intention to derive details of the full dynamic equations of the underwater vehicle manipulators. Instead, we would like to point out several specific elements that differentiate underwater vehicle manipulators from regular manipulators.

The hydrodynamic forces acting on a rigid body moving in fluid can be found in *Patel* [17.21] and *Troesch* and *Kim* [17.22]. Similar to the vehicle main body, each link of the manipulator will be influenced by hydrodynamic forces including added mass, buoyancy acting at the center of buoyancy (COB) of the link, the fluid acceleration force resulting from the acceleration of the fluid itself, and the fluid drag forces exerted on the link.

As the profile drag forces are dominant for slowly moving objects, the drag force exerting on an infinitesimal element of the ith link and the total drag force on the link can be expressed as

$$\mathrm{d}D_{fi} = \frac{1}{2}\rho C_{\mathrm{D}} b_i v_{r_i} |v_{r_i}| \mathrm{d}x \,, \tag{17.1}$$

$$D_{fi} = \frac{1}{2}\rho \int_0^l C_{\mathrm{D}} b_i v_{r_i} |v_{r_i}| \mathrm{d}x \,, \tag{17.2}$$

where C_{D} is the drag coefficient, ρ is the fluid density, v_{r_i} is the relative velocity of the element to the fluid flow velocity, $b_i \mathrm{d}x$ is a projected area of the element, and l is the link length.

Considering the effects of the hydrodynamic variables above and assuming that the vehicle is stationary during the manipulation, the dynamic model of an underwater manipulator having a series of links and joints can be represented by the following equations

$$\begin{aligned}\mathbf{M}_m(\boldsymbol{q})\ddot{\boldsymbol{q}} + \boldsymbol{C}_m(\boldsymbol{q},\dot{\boldsymbol{q}}) + \boldsymbol{D}_m(\boldsymbol{q},\dot{\boldsymbol{q}}) + \boldsymbol{F}_m(\dot{\boldsymbol{q}}) \\ + \boldsymbol{G}_m(\boldsymbol{q}) + \boldsymbol{\tau}_d = \boldsymbol{\tau}_m \,,\end{aligned} \tag{17.3}$$

where $q \in \boldsymbol{R}^{n\times 1}$ is a joint angle vector, $\boldsymbol{\tau}_m \in \boldsymbol{R}^{n\times 1}$ is a joint torque vector, $\mathbf{M}_{\boldsymbol{m}}(\boldsymbol{q}) \in \boldsymbol{R}^{n\times n}$ an inertia matrix including added mass terms, $\boldsymbol{C}_{\boldsymbol{m}}(\boldsymbol{q},\dot{\boldsymbol{q}}) \in \boldsymbol{R}^{n\times 1}$ is a nonlinear vector arising from centrifugal and Coriolis effects including added mass terms, $\boldsymbol{D}_{\boldsymbol{m}}(\boldsymbol{q},\dot{\boldsymbol{q}}) \in \boldsymbol{R}^{n\times 1}$ is a nonlinear vector due to hydrodynamic forces such as drag forces, $\boldsymbol{F}_{\boldsymbol{m}}(\dot{\boldsymbol{q}}) \in \boldsymbol{R}^{n\times 1}$ is a nonlinear vector due to friction at the manipulator joints, $\boldsymbol{G}_m(\boldsymbol{q}) \in \boldsymbol{R}^{n\times 1}$ is a nonlinear vector due to gravity and buoyancy forces, and $\boldsymbol{\tau}_d \in \boldsymbol{R}^{n\times 1}$ is a vector of unknown signals due to unmodeled dynamics or external disturbances such as current.

During the manipulation, the end-effector or tool of the manipulator could be in contact with the environment, and forces and moments would occur at the end-effector. In this situation, (17.3) must be modified by adding the reaction torque to the end-effector force on the right side of (17.2).

$$\boldsymbol{\tau}_r = \mathbf{J}_m^T(\boldsymbol{q}) f_e \,, \tag{17.4}$$

where is $\mathbf{J}_m \in \boldsymbol{R}^{6\times n}$ the manipulator Jacobian matrix, $\boldsymbol{f}_e \in \boldsymbol{R}^{6\times 1}$ a vector of forces and torques at the end-effector.

The dynamics of the underwater vehicle manipulator is quite complicated as shown in (17.2), highly nonlinear, and involves coupled equations with unknown parameters and disturbances. It is almost impossible to accurately model the dynamic equations and to operate the manipulator at a nominal speed using a conventional controller. However, when the manipulator moves very slowly in the friendly environment of no or minimal current or wave, the dynamics of the manipulator could be represented by the simplified linear version of (17.3) since the effect of velocity-dependent terms in (17.3) becomes negligible. In fact, during the actual operation in the field, most commercial underwater manipulators move very slowly (much less than 1 rad s^{-1}) for safety and other reasons, and they use conventional joint controllers for the master–slave teleoperation.

17.3 Teleoperation of Underwater Vehicle Manipulators

A typical setup for teleoperation involves the operator sitting in the control room located on the mother ship and holding a master arm that is a miniature of the actual underwater vehicle manipulator (slave arm) attached to the ROV. The operator controls the slave arm by moving the master arm whose motion is followed by the slave arm. During the operation, the operator relies on the visual information of the work site in the form of a series of 2-D video images that are captured by cameras on the vehicle's main body and the slave arm and then transmitted to the control room monitors by an underwater cable. Not only the view of the work site is limited but also the visual information especially for the deep-sea operation is often delayed. If the vehicle is not holding or sitting on an underwater structure, the ROV operator must try to control the vehicle for station-keeping during the manipulation. However, even if there is no current or wave during the station-keeping, the vehicle moves like free-floating within the accuracy of the vehicle position sensors in the order of a meter or so. Therefore, it is difficult to achieve the accuracy at the end-effector since underwater vehicle manipulators are attached to the vehicles that are constantly moving, unlike industrial manipulators whose bases are fixed on the factory floor. As *Hildebrandt* et al. [17.23] state, many simple tasks on the ground such as plugging become very difficult to perform by the underwater vehicle manipulator. Underwater plugging tasks, done on a trial-error basis since the required precision could not be achieved by the ROV manipulator system, take hours to complete. Therefore, operator fatigue often becomes a critical issue in teleoperating the underwater vehicle manipulator.

Several researchers have addressed the issues mentioned above and have improved the accuracy of the manipulator in teleoperation. Among them include *Sayers* et al. [17.24], *LaPierre* et al. [17.25], *Zhang* et al. [17.26], *Bowen* et al. [17.27], *Hildebrandt* et al. [17.23], *Hildebrandt* et al. [17.28], and *Shim* et al. [17.29]. *Zhang* et al. [17.26] developed a virtual model-based teleoperating system and experimentally showed a good performance of moving target tracking with up to a two second time delay. They used on-line parameter estimation and showed that the manipulator could be continuously controlled with a low bandwidth network. *Hildebrandt* et al. [17.23] implemented precise manipulator position control by compensating for errors due to the ROV movement. The overall scheme was divided into:

1. Prediction of vehicle movement
2. Calculation of manipulator compensation movement
3. Compensating motion with the manipulator.

The prediction of vehicle movement relied on a position sensor in the vehicle and a mathematical model of the vehicle. The manipulator compensation movement was calculated by the difference between the current position and the predicted position. The compensation position in the Cartesian space was then transformed into a joint space position to be given to the manipulator. They showed that the ROV movement compensation algorithm considerably increased the position accuracy of the end-effector of the commercial hydraulic underwater manipulator, Schilling Orion 7P (Fig. 17.1). Notably, *Hildebrandt* et al. [17.28] and *Shim* et al. [17.29] have developed additional computer control systems to the top of a commercial underwater vehicle manipulator system, Schilling's Orion

Fig. 17.1 Schilling's Orion 7P (courtesy of FMC technologies schilling robotics)

7P, and presented experimental results. *Hildebrandt* et al. [17.28] added two new controllers (a second position controller and an adaptive speed controller) to the top of Orion's position controller to improve the joint position accuracy at the level of less than 1° without changing the original hardware or the respective hardware-near software/firmware. The second position controller has the desired subdegree precision and the adaptive speed controller makes it possible to have very slow but precise movements. Experiments were conducted for the plugging of a Gisma Series 80 connector. The plugging procedure consists of three steps: the manipulator's end-effector with the plug is positioned in front of the receptacle within an accuracy of 1 mm and 1°; the plug is precisely aligned with the receptacle until the top of both parts touch; and then the plugging process is completed by having a linear movement of 100 mm. Results show that their control approach can improve accuracy and plugging was successfully done in 50% of all trials. *Shim* et al. [17.29] developed a new master system that allows the operator to control the slave arm in workspace. Instead of using the master arm, a joystick is used for a velocity input of the end-effector in workspace and the new master system's computer computes the desired joint angles by solving forward and inverse kinematics of the slave arm for the desired velocity of the end-effector. The computed desired joint angles are then sent to the slave arm controller. The master system also offers a graphic model of the slave arm, which follows its actual motion based on the joint signals. In addition to the limited video images of the slave arm at the work site, this graphic model view of the slave arm's motion helps the operator.

It is worthwhile to note the Woods Hole Oceanographic Institute (WHOI)'s Hybrid-ROV, also known as Nereus [17.27], which operates in two different modes. For broad area survey, the vehicle can operate untethered as an autonomous underwater vehicle capable of exploring and mapping the sea floor with sonars and cameras. For close-up imaging and sampling, Nereus can be converted to operates at sea as a tethered ROV, using a lightweight, micro-thin, fiber-optic tether. Nereus uses its manipulator by teleoperation via the fiber-optic tether while the vehicle could be considered an AUV. In May 2009, Nereus explored the Mariana Trench, being the first vehicle to explore the Mariana Trench since 1998.

17.4 Sensor-Based Manipulator Control

Various sensors are used for manipulator control. Among them, force and vision sensors are discussed in this section. They are widely used for manipulator control in terrestrial or space environments [17.30–33]. As demand on the high performance of the underwater manipulation is increased especially for applications in deeper water like subsea oil/gas industry, archaeological or scientific studies, force and visual-based manipulator control received more attention. However, their use for underwater manipulator control has been limited due to environmental conditions. The performance of a force sensor would be influenced by the change of the operating depth, whereas the performance of an optical vision sensor would be dependent on the turbidity of the water. In this section, the current use of force and visual feedback for underwater manipulator control is briefly reviewed and the future direction toward autonomous manipulation is discussed.

Force feedback contains information about contact force between a manipulator's end-effector and an underwater environment. This information helps an underwater robot or operator determine how much power is needed to perform a given task such as handling a fragile object. A few force control methods that have been implemented to control underwater manipulators (are/include) bilateral force feedback [17.6], back-electromotive force (EMF)-based motor force feedback [17.34], and direct force control with underwater force–torque (FT) sensor [17.35].

The bilateral force feedback control has been developed for master–slave teleoperation [17.30, 36, 37]. Its

basic concept is to keep the joint angles of both master and slave arms the same so that the operator feels the contact force from any changes in (master/slave) arm positions. To achieve a 1 : 1 direct mapping of both arms' angular positions, a motor-driven active master arm is needed for duplicating the slave arm's position. Kraft Telerobotics [17.6] has commercialized a bilateral force control-based deep-sea underwater manipulator for ROVs such as WHOI's Hybrid ROV Nereus.

Back-EMF feedback is another way of (measuring/assessing/mapping) the force on the underwater manipulator [17.34]. It uses back-EMF voltage of electric motor in the manipulator to calculate reaction force. Since the back-EMF voltage is proportional to the motor speed and has relatively low noise [17.38], it can be used to measure the motor speed and position without a direct sensor. An obstruction of the robotic arm or an object inside the gripper could be detected by calculating a feedback force on each motor from the command voltage and back-EMF voltage. This method can be implemented only on electric motor-driven manipulators. *Kerdels* et al. [17.34] showed the performance of the back-EMF-based force control of the gripper by grabbing a light bulb with a 3 DOF Sub-Atlantic 123 DC-manipulator (rated for a depth of 300 m).

An FT sensor-based force control scheme for autonomous underwater manipulation was developed by *Marani* et al. [17.35]. Actual contact forces and torques on the manipulator are directly measured by the underwater FT sensor and used to control the manipulator. However, as mentioned above, accuracy and precision of most FT sensors in the market are heavily influenced by the operating pressure. Developing FT sensors for deepwater applications would be useful.

Visual feedback control schemes use a noncontact sensor like a low-light charge-coupled device (CCD) camera. However, the performance of visual feedback control depends on environmental conditions such as light and visibility. One good example of using a vision sensor for robot manipulator control is visual servoing. It has been used to control a robot manipulator with visual information, mostly with relative pose (position and orientation) between a camera and an object in a hostile environment [17.32, 33, 39–42]. With the visual servoing concept, the target object can be kept within the camera field of view (FOV), so that a manipulator maintains its relative pose with respect to a static or moving object.

It is well known that visual servoing has two types of control methods: position-based and image-based ones [17.43]. The position-based visual servoing is used to calculate the relative pose of an object with respect to the camera, which is used to control the manipulator. In this method, the camera calibration is very critical to get accurate relative pose. The image-based method uses image features such as size or orientation of the object in the camera space to calculate control commands making the current features match desired ones that were precalculated. While it is free from the camera calibration, it requires control commands mapping from image features to real world control space since it does not use the position information.

One commercial position-based visual servoing for underwater manipulator control was recently developed by FMC Technologies Schilling Robotics for autonomous hot stab insertion [17.44]. The hot stab insertion task is an underwater version of traditional peg-in-hole task with a manipulator, and is more difficult than the traditional one due to very tight (almost zero) mechanical clearance between a hot stab plug and a receptacle (port) and the unstable base of the manipulator. In addition, the main body of the vehicle that is the base of the manipulator is continuously moving due to underwater current and buoyancy unless the vehicle is sitting on the ocean floor or fixed to an underwater structure. Due to the aforementioned difficulties, it usually takes 30 min to 1 h for well-experienced operators to complete the task. Furthermore, failure occurs often, causing serious damage to the plug or the port. The visual servoing helps making the operation much faster than the human operator, completing the task before it fails to keep the target object within the camera FOV. The visual servoing is more efficient than the human operator in completing the operation. To overcome drawbacks in using an optical camera in turbid water, a high-frequency acoustic imaging sonar camera has been developed [17.45, 46]. Acoustic imaging is a valid alternative for a human operator to get information about the object, but it requires further research to make a computer automatically recognize the object [17.47–49]. A sensor fusion scheme for autonomous underwater manipulation was developed for the SAUVIM project, using various sensors such as an imaging sonar, a dual-frequency identification sonar (DIDSON), optical cameras, and an underwater FT sensor [17.50, 51]. Details are described in Sect. 17.6.

17.5 Coordinated Motion Control of Underwater Vehicle–Manipulator Systems

The vehicle's movements affect its manipulator's and vice versa, unless the vehicle is completely stationary during the manipulation by holding or sitting on an underwater structure. The effect of the manipulator's motion on a large underwater vehicle main body is negligible, compared to the effect of the vehicle's motion on the accuracy of the manipulator's end-effector. For a small vehicle, the coupled effect of the movements from the vehicle main body and the manipulator is substantial and must be considered in the overall control system design. This issue can be addressed by the coordinated motion control of the vehicle and the manipulator.

The overall underwater vehicle-manipulator system could be seen as a multibody system with redundant degrees-of-freedom. The redundancy can be used to assign additional motion without impeding the end-effector's performance, and the motions of the vehicle and the manipulator can be coordinated by redundancy resolution of the overall system with optimizing an index. Therefore, the coordinated motion control of the vehicle-manipulator system could be advantageous by having additional degrees-of-freedom in manipulation. The coordinated motion control system should have a planning module capable of generating trajectories for the vehicle and the manipulator for given tasks and a control module capable of generating control signals for the vehicle and the manipulator to accurately track the trajectories in the presence of coupled effects and uncertainties.

Several researchers have studied modeling and coordinated motion control of underwater vehicle-manipulator systems and showed that the coordinated motion control approach could improve the accuracy of the end-effector [17.12–15, 17–20, 43, 52–67]. Some of these previous studies are described below.

Schjølberg and *Fossen* [17.17] derived the dynamic equations of an underwater vehicle-manipulator system, and applied feedback linearization for transforming the nonlinear system dynamics into a linear system for controller design. *Antonelli* [17.52] proposed an adaptive control scheme that keeps the advantage of model-based adaptive control while limiting the computational load. The control algorithm was designed as a modular structure by applying the virtual decomposition approach [17.68]. *Han* et al. [17.59] proposed a performance index for redundancy resolution. Redundant DOFs of the vehicle and manipulator were used to optimize the index by applying the gradient projection method. In addition, the inverse optimal nonlinear H-infinity control was developed to guarantee tracking performance under uncertainty caused by modeling or identification errors, or unmodeled dynamics.

One must note that most commercial UUVs with manipulators are heavy-duty large ROVs. They do not have coordinated motion controllers for the vehicle and manipulator systems and their manipulator controllers do not automatically compensate for errors due to the vehicle motion. It is partly because the current state of the art in positioning sensors such as long baseline (LBL) does not offer the accuracy required for a high performance of the manipulator, especially its end-effector. The vehicle and the manipulator may also operate in different bandwidths, mainly due to the difference in size and dynamic characteristics. The actual position of the vehicle remains within the accuracy of the vehicle position sensor. For example, the accuracy of the vehicle position sensor of the order of a meter could mean that the actual vehicle keeps moving within a meter and yet the sensor reads the same position value. Since the vehicle is the base of the manipulator, the required accuracy of the end-effector cannot be achieved without accurate sensory information about the vehicle position. However, recent developments in integrated navigation sensor technologies show very promising results for more accurate position control of the vehicle, which would improve the precision and accuracy of the end-effector. *Panish* and *Taylor* [17.69] experimentally demonstrated that both the Kearfott T-24 and the IXSEA PHINS III [17.70] provide exceptional navigation accuracies, exceeding the published specifications–position drift of less than 0.1% of distance travelled, CEP that refers to the circular error probability, or a circle about a mean value which includes 50% of the population. A drift of 0.1% of distance travelled CEP means that the position drift of less than 1 m for travelling 1 km between global positioning system (GPS) fixes is expected for 50% of the time. Such developments could make the practical use of the coordinated motion control possible.

17.6 Underwater Autonomous Manipulation

Autonomous manipulation on a moving base, such as terrestrial mobile robots, humanoids, and underwater robotic vehicles, is a very challenging task in the area of robotics in general, especially in unstructured environments, such as underwater. It refers to the capability of a robot system that performs intervention tasks requiring physical contact with unstructured environments without continuous human supervision. Unlike industrial manipulators having fixed bases on the floor, autonomous manipulation requires a system capable of assessing a situation, including self-calibration based on sensory information, and executing or revising a course of manipulating action without continuous human intervention. Therefore, developing a system capable of fully autonomous manipulation would be a great achievement and make a substantial impact on a variety of application areas with significant economical, societal, and scientific importance [17.71].

Let us look at one scenario of cutting an underwater cable. With given information about the cable's location and shape, the vehicle has to navigate to the location, identify the cable, position itself, and cut the cable. It may sound like a very simple task. However, it would not be so simple if it has to be done in autonomous mode. Even in teleoperation mode, it would not be an easy task without the coordinated motion control if the vehicle is floating in water. In fact, the low bandwidth and significant time delay inherent in acoustic subsea communications represent a considerable obstacle to remotely operate a manipulation system, making it impossible for remote controllers to react to problems in a timely manner. Nevertheless, robots for autonomous underwater intervention would pave the way for a different range of new operations, such as deep-ocean and under-ice exploration, tasks in hazardous areas, tasks in natural or man-made disastrous regions, automated searches, surveillance missions, to name a few.

In the past, many researchers have studied advanced control of AUV itself as reported by *Carreras* et al. [17.72, 73], *Zhao* and *Yuh* [17.74], *Hanai* et al. [17.75], *Kim* and *Yuh* [17.76], *Lee* et al. [17.77], *Yuh* and *Nie* [17.78], and *Yang* et al. [17.79], but only a few AUVs with manipulators were introduced. OTTER is an AUV equipped with a single degree-of-freedom arm, which was designed to be used as a testbed for autonomous technologies at the Stanford Aerospace Robotics Lab. in 1996. It is a hovering-capable underwater vehicle, which operates in a test tank at the Monterey Bay Aquarium Research Institute (MBARI). Various research topics were studied with OTTER, including precise, high-speed underwater manipulators, task-level programming with an advanced finite-state machine paradigm, underwater video navigation, station-keeping, tracking, and automatic objects retrieval [17.80]. Another intervention-AUV (I-AUV), namely Autonomous Light Intervention Vehicle (ALIVE), was developed by Cybernetix in 2003. The aim of the EU-funded ALIVE project was to develop an intervention-I-AUV capable of docking to a subsea structure, which has not been specifically modified for AUV use. A description of the ALIVE vehicle was given by *Evans* et al. [17.81]. The most recent effort in developing an AUV for intervention missions is the TRIDENT FP7 EU research project funded by the European Commission for 3 years, since March 2010. The consortium is made up of eight partners: Universitat Jaume-I, University of Girona, University of the Balearic Islands, University of Bologna, University of Genova, Heriot-Watt University and Instituto Superior Tecnico, and company Graal Tech. The long-term objective of the project is the design and implementation of a new methodology enabling multipurpose underwater intervention missions with a very high autonomy level [17.82–84]. A team of two cooperative heterogeneous robots with complementary skills, an autonomous surface craft (ASC) and an intervention-AUV (I-AUV) endowed with a dexterous manipulator, are used to perform underwater manipulation tasks. The proposed methodology is based on two steps. First, the I-AUV cooperates with the ASC performing path to survey an area of interest, gathering opto/acoustic data from the seafloor. During this phase, the I-AUV performs accurate path following and terrain tracking, to maximize bottom coverage and data quality. The motion of the ASC is coordinated with that of the I-AUV to achieve precise USBL (ultra short base line) positioning and reliable acoustic communications. After the survey, a multimodal two-and-a-half-dimensional (2.5-D) elevation map of the environment is built consisting on a photomosaic rendered over an accurate bathymetric map, both being jointly and globally optimized. At the second step, I-AUV is launched to search for an object of interest manually selected by the user over the constructed map. When the object (the target of the intervention) is found, the I-AUV switches to free floating navigation mode. An optimization algorithm is used to decide how to move the arm, the vehicle, or both together in order to satisfy a set of constraints, such as keeping the object within the camera FOV, keeping a desired altitude, and avoiding arm singularities. The manipulation of the object takes place through a dexterous three-fingered hand attached to a redundant robot arm (7 DOF) and assisted with proper percep-

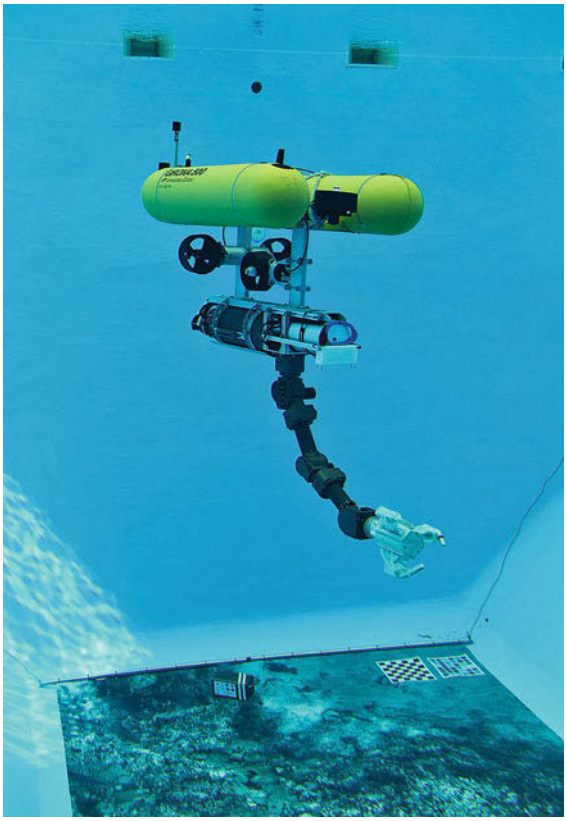


Fig. 17.2 Girona 500 I-AUV

tion (a stereo pair). The intervention task is seen as a semiautonomous process where the target is manually selected over the map built during the survey phase, but then it is automatically recognized and manipulated by the robot in a completely autonomous way. Girona 500 I-AUV is shown in Fig. 17.2.

The key technology in underwater intervention performed with autonomous vehicles is autonomous manipulation. In the literature, no such system fully functional for autonomous manipulation was reported until recent developments with a semi-autonomous underwater vehicle for intervention missions (SAUVIM) [17.50, 51]. Therefore, this section describes underwater autonomous manipulation based on the results of SAUVIM.

Autonomous underwater intervention is the primary objective of the SAUVIM research team. SAUVIM (Fig. 17.3) funded by the US Office of Naval Research was jointly developed by the Autonomous Systems Laboratory (ASL) of the University of Hawaii, Marine autonomous systems engineering (MASE), Inc. in Hawaii, and Naval undersea warfare center division(NUWC) Newport in Rhode Island. The first fully autonomous manipulation in an unstructured ocean environment was demonstrated with SAUVIM at Snug Harbor, Honolulu, Hawaii on January 20, 2010.

Fig. 17.3 The SAUVIM vehicle (Sand Island, Hawaii, 2008)

SAUVIM was built with the primary goal of performing autonomous underwater intervention. Navigation and hovering movements are precisely actuated with eight thrusters located around the center of mass (COM). The sensor navigation equipment includes a photonic inertial navigation system (PHINS) unit from IXSEA, a differential GPS, a Doppler velocity log (DVL), and depth sensor. The above navigation sensor system is capable of providing a stable and precise feedback of the vehicle position, velocity and acceleration on all the six degrees of freedom. The stability and precision of the vehicle, SAUVIM, during station-keeping was tested by experiments with a six DOF model-based dynamic controller, active on position and orientation. The inertial navigation system (INS) was aided only by the DVL, since the GPS antenna was, in this case, submerged. The vehicle was able to maintain the target position, over a 15 min period, with a standard deviation of the order of 0.01 m, for the translational part. This was confirmed by the manipulator camera output, which was looking toward an Earth-fixed target. During the experiment, only a slow change was noticed in the vertical position, which was due to the tide activity. In fact, while the target was fixed with respect to the earth, the INS used a depth sensor to correct errors in the Z-coordinate.

SAUVIM is equipped with a 7 DOF robotic manipulator, MARIS 7080 (Fig. 17.4). The arm, unlike the classical hydraulic technology in use for ROVs, is actuated by electromechanical components, to meet low-power requirements and high accuracy in manipulation tasks. Each joint is actuated by a brushless motor with a reduction unit (harmonic drive). A force/torque sensor, installed between the wrist and the gripper, senses the amount of the force and torque acting on the gripper. Designed for underwater applications at extreme depths, it is internally compensated with appropriate oil. The sensor devices of SAUVIM are the most critical components of a generic intervention mission, since at the task-execution level the system must be capable of acting and reacting to the environment with the extensive use of sensor data processing. SAUVIM is equipped with a dual frequency identification sonar camera DIDSON (Fig. 17.5), a digital multifrequency scanning sonar, several video cameras with image-processing computational unit and a special ultrasonic device for tracking the position of a generic target in 6 DOF.

Fig. 17.4 MARIS 7080 underwater manipulator (ASL, Hawaii, 2004)

Finally, the hardware architecture is composed of several on-board computers and peripherals for sensor data acquisition, enabling the following capabilities: for the vehicle, 6 DOF dynamic control, 6 DOF position feedback, extended Kalman filter (EKF) for parameters identification, target identification with DIDSON, and ability to correct absolute position with the DIDSON; for the manipulator, advanced multipriority task space control, collision avoidance, workspace optimization in conjunction with the navigation control, and optical vision with 6 DOF target position recognition.

Autonomous manipulation on SAUVIM is described by illustrating different phases during its final demonstration conducted on January 20, 2010. It consists of the following six phases that would be a major part of the typical procedure for underwater recovery tasks:

- Phase 1: Undock from the pier and navigate to a search area
- Phase 2: Search for the submerged platform
- Phase 3: Navigate and dive toward the platform
- Phase 4: Hover (station keeping)
- Phase 5: Hook a recovery tool to the target object (autonomous manipulation)
- Phase 6: Return to the pier.

Fig. 17.5 The DIDSON sonar (courtesy Sound Metrics)

Each phase was executed with no human intervention. The following information was known in advance and given to the SAUVIM computer:

1. Shape of the platform
2. Shape of the target to recover
3. Search area for the platform
4. Directional information about the target's location on the platform (frontal area of the platform).

17.6.1 Phase 1: Undock from the Pier and Navigate to a Search Area

The surface navigation, in this undocking phase, was aided by the differential global positioning system (DGPS), with an accuracy of about 2–3 m. This accuracy is sufficient enough for SAUVIM to move to a target area where to begin the search for the platform (Fig. 17.6). After detecting the platform by the scanning sonar, SAUVIM aligned itself to the platform and became ready to dive toward it.

17.6.2 Phase 2: Search for the Submerged Platform

After arriving at the search area, SAUVIM started searching for the main platform using the DIDSON sonar (Fig. 17.7) with acoustic lenses that operates at two frequencies, 1 and 1.8 MHz, with an operative range up to 40 m [17.85]. Target localization in the mid-range using the DIDSON sonar is still a challenging task under development. Very few studies on target identification and localization with the DIDSON have been presented in the literature. A preliminary study on target identification has been developed by *Yu* et al. [17.47] and *Yu* [17.48] while a possible use of the DIDSON imagery for SLAM has been investigated by *Walter* et al. [17.86]. The SAUVIM approach to determining the pose estimation with the DISDON sonar was presented by *Marani* and *Choi* [17.87]. Here, a recursive model estimator was implemented: at a fixed frequency, a model of the target, as ideally seen from the DIDSON, was built according to object dimension (a-priori information) and DIDSON orientation with help of the inertial navigation system and the pan/tilt modules on-board SAUVIM vehicle. Maximum (white color) acoustic reflection from the object surface, no reflection (black color) from the object edge and average reflection (gray) from the background are the object model hypothesis. Then, for the target identification, a bi-dimensional convolution-based approach was implemented using high-performance libraries.

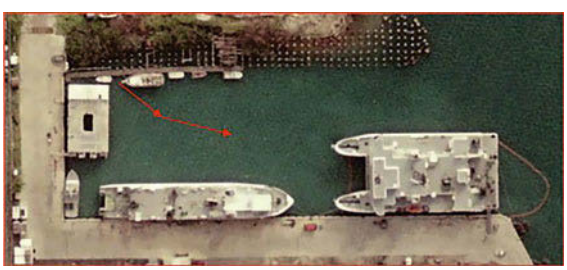


Fig. 17.6 Experimental area (SNUG harbour, Hawaii, courtesy Google Earth)

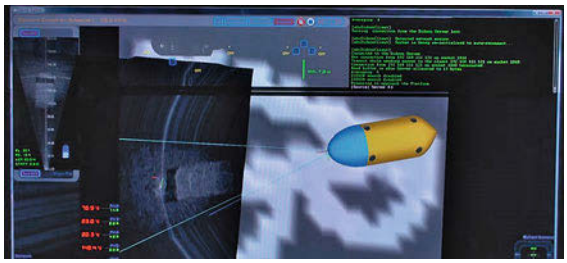


Fig. 17.7 Search of the underwater platform

17.6.3 Phase 3: Navigate and Dive Toward the Platform

After the platform was located in phase 2, the vehicle initiated diving toward the platform. Using the PHINS corrected by the DVL, SAUVIM performed dead reckoning in order to position itself just 30 cm above the platform. Here, the vehicle used the ballast for diving, enabling the COB identification when beyond a depth of 1 m.

In order to practically quantify the standard deviation of the vehicle position over the platform, a predefined experiment of positioning the vehicle in the same hovering configuration was repeated several times. In this experiment, the relative Cartesian coordinates of the platform with respect to the vehicle were computed using DIDSON imagery taken at a distance of 10 m from the target area. Successively, the vehicle performed dead-reckoning navigation from the target area to the hovering configuration. This experiment confirmed an excellent repeatability, and the global standard deviation of the position error was confined within a few centimeters in the successive trials.

17.6.4 Phase 4: Hover (Station Keeping)

SAUVIM is equipped with an advanced navigation control system capable of precise station-keeping (hovering). This is accomplished with a 6 DOF dynamic controller and an EKF for the real-time identification of the COB.

This is another important capability necessary to optimally set a roll and pitch configuration, as presented by *Marani* et al. [17.50]. In fact, optimal positioning for an I-AUV should minimize the energy consumption and improve the stability in orientation. In general, during an intervention task, the vehicle is maintained in a hovering configuration, thus requiring a 6 DOF control of the vehicle positioning. The choice of roll and pitch, if done arbitrarily, can severely impact the power efficiency of the vehicle. This is especially significant in heavy systems, since the COB may not be necessarily aligned over the COM. An EKF was used to identify the location of the COB relative to the COM, making it possible to compute the working orientation that maintains the COB vertically aligned above the COM. The EKF was implemented online to detect any change in the COB caused by various reasons, such as ballast operations. This approach helped achieve very precise position control of the order of subcentimeter in translation.

17.6.5 Phase 5: Hook a Recovery Tool to the Target Object (Autonomous Manipulation)

The manipulator started searching for the target object by using its optical camera mounted on the manipulator for short-range optical vision. Once the target object and its location (6 DOF) were detected, the arm started a tracking mode. After successfully locking on the target object, it physically connected the recovery tool to the target object. One important feature during this phase is the workspace optimization. The vehicle is set to adjust its position in real time in order to optimize the manipulation. In this approach, the entire vehicle-manipulator system is regarded as a unique open chain with multi-degrees of freedom joint. In particular, the vehicle position is modeled as a 6 DOF (free) joint (Fig. 17.8).

SAUVIM uses the concept of task reconstruction [17.88–90] along with the appropriate choice of primary and secondary tasks, which makes it possible for SAUVIM to autonomously reposition itself to increase the manipulability only if needed. The choice of the tasks was set in the following order of priority:

- Task 1 (3 DOF): Cartesian position of *Link 8* (the end-effector)
- Task 2 (3 DOF): Orientation of *Link 8* (the end-effector)
- Task 3 (6 DOF): Generalized velocity of *Link 1* (the vehicle main body).

Tasks 1 and 2 were set to track the relative position and orientation of the target object measured by using the optical camera system, whereas the generalized velocity of the vehicle (task 3) was set to zero.

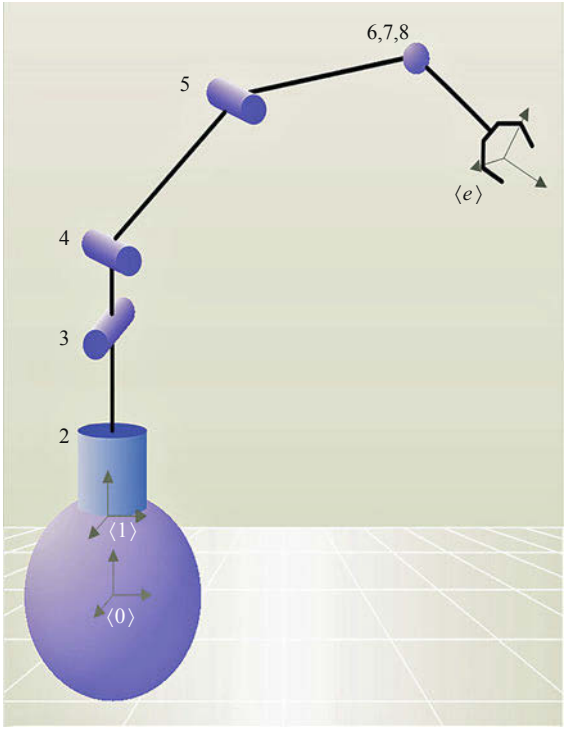


Fig. 17.8 Modeling of the SAUVIM vehicle-manipulator system

17.6.6 Phase 6: Return to the Pier

Finally, SAUVIM returned to the docking pier. Precise docking was performed using feature-based relative navigation. After surfacing, SAUVIM went to the central area and looked for the submerged platform. Then, using the relative position of the docking pier with respect to this reference, SAUVIM navigated to its home. The measured standard deviation of the final position was about 10 cm, which is accurate enough for precise-docking.

From the above six phases, we have implicitly identified the complexity of problems encountered in autonomous underwater interventions. Among those, the following are most noticeable:

- *Target area navigation*. Without the help of advanced underwater imaging technology, it would have been impossible to autonomously locate and navigate toward the target area. The above is one of the first attempts to use the DIDSON sonar as multi-DOF pose sensor to guide autonomous navigation.
- *Vehicle positioning*. While in classical ROV operation, it is the responsibility of the operator to position the ROV in the target area and an autonomous system must be capable of performing some kind

of workspace optimization. For SAUVIM, a large effort has been devoted to develop automatic optimization algorithms capable (as seen in phase 5) to precisely correct the hovering position in order to maximize the manipulability of the arm during its operations.

- *Arm control system*. Many problems commonly encountered in robotics such as kinematic singularities, collisions, and joint limit and motor saturation, are no longer handled by the robot operator. An autonomous robotic control system must be able to intelligently assess all the above situations in order to allow a successful completion of the task when it is, at least theoretically, possible. This was another area in which SAUVIM research was devoted for.
- *Future development* in autonomous underwater manipulation should be able to provide robust improvements of all of the above issues. It is important to increase the general confidence of all the organizations toward such systems, and to initiate a gradual transition from human-teleoperated ROVs to a new level of autonomy of intervention AUVs.

17.7 Conclusions

This chapter presented basic materials about underwater vehicle manipulators that are essential for underwater intervention missions, such as underwater sampling, construction, and recovery. Most commercial underwater vehicle manipulators are hydraulic-driven and used for ROVs. They are teleoperated by human operators using the master–slave configuration, which often causes operator fatigue since most underwater manipulator operations are done on a trial-and-error basis due to the vehicle movement and delayed sensory feedback. In order to minimize the presence of human operators and maximize the advantage of vehicle autonomy, the coordinated motion control of the vehicle and manipulator system and autonomous manipulation are key technologies for future underwater vehicle manipulators. The US Navy funded the underwater vehicle, SAUVIM, which has experimentally demonstrated autonomous manipulation. Following its success, more underwater vehicles having the capability of autonomous manipulation are expected to develop in the near future.

Among some exemplar underwater intervention tasks are object recovery/rescue, and maintenance/repairing of underwater facilities. Therefore, one may consider a task like *Black box recovery-grasping an object of interest and placing it at a desired location* for benchmarks in future development. One recent development after SAUVIM is the TRIDENT project in Europe. They performed an autonomous underwater intervention task, *Black box recovery* in Port de Sóller's harbor (Mallorca) in October 2012 [17.91].

References

17.1 J. Yuh: Modeling and control of underwater robotic vehicles, IEEE Trans. Syst. Man. Cybern. **20**(6), 1475–1483 (1990)

17.2 J. Yuh: Design and control of autonomous underwater robots: A survey, Int. J. Auton. Robots **8**, 7–24 (2000)

17.3 J. Yuh, G. Marani, D.R. Blidberg: Applications of marine robotic vehicles, Intell. Serv. Robotics **4**(4), 221–231 (2011)

17.4 eca Robotics: ARM 7H 7 Function HYDRAULIC Manipulator ARM (2011), http://www.eca-robotics.com/ftp/ecatalogue/486/ARM_7H_GB.pdf

17.5 International Submarine Engineering: Magnum-7 A rugged 7-function subsea manipulator (2012), http://www.ise.bc.ca/manips/pdf/Magnum-7-data-sheet-v2.pdf

17.6 Kraft TeleRobotics: Predator Force Feedback Manipulator (2005), http://krafttelerobotics.com/products/PDFs/Predator_PDF.pdf

17.7 FMC Technologies: Schilling Robotics TITAN 4 (2012), http://www.fmctechnologies.com/en/SchillingRobotics/Technologies/Schilling-TITAN-4.aspx

17.8 Western Space and Marine (WSM): The ARM-Standard Specifications (2012), http://www.wsminc.com/TheArm/Specifications.html

17.9 N. Kato, D. Lane: Coordinated control of multiple manipulators in underwater robots, Proc. IEEE ICRA'96 (1996) pp. 2505–2510

17.10 M.W. Dunningan, D.M. Lane, A.C. Clegg, I. Edwards: Hybrid position/force control of a hydraulic underwater manipulator, Proc. Control Theory Appl., Vol. 143 (1996) pp. 145–151

17.11 B. L'evesque, M.J. Richard: Dynamic analysis of a manipulator in a fluid environment, Int. J. Robotics Res. **13**(2), 221–231 (1994)

17.12 H. Mahesh, J. Yuh, R. Lakshmi: A coordinated control of an underwater vehicle and robotic manipulator, J. Robotic Syst. **8**(2), 339–370 (1991)

17.13 S. McMillan, D. Orin, R. McGhee: Efficient dynamic simulation of an underwater vehicle with a robotic manipulator, IEEE Trans. Syst. Man Cybern. **25**(8), 1194–1206 (1995)

17.14 T. McLain, S. Rock, M. Lee: Experiments in the coordinated control of an underwater arm/vehicle system, Auton. Robots **3**, 213–232 (1996)

17.15 T.J. Tarn, G.A. Shoults, S.P. Yang: Dynamic model of an underwater vehicle with a robotic manipulator using Kane's method, Auton. Robots **3**(2/3), 269–283 (1996)

17.16 K. Ioi, K. Itoh: Modeling and simulation of an underwater manipulator, Adv. Robotics **4**, 303–317 (1990)

17.17 I. Schjølberg, T. Fossen: Modeling and control of underwater vehicle-manipulator systems, Proc. 3rd Conf. Mar. Craft Manouvering Control (1994) pp. 45–57

17.18 K.N. Leabourne, S.M. Rock: Model development of an underwater manipulator for coordinated arm-vehicle control, Proc. IEEE OCEANS'98, Vol. 2 (1998) pp. 941–946

17.19 M. Lee, H. Choi: A robust neural control for underwater robot manipulators, IEEE Trans. Neural Netw. **11**(6), 1465–1470 (2000)

17.20 J.H. Ryu, D.S. Kwon, P.M. Lee: Control of underwater manipulators mounted on a ROV using base force information, Proc. IEEE ICRA'01 (2001) pp. 3238–3243

17.21 M. Patel: Dynamics of Offshore Structures (Butterworths Scientific, Sevenoaks, 1989)

17.22 W. Troesch, S.K. Kim: Hydrodynamic forces acting on cylinders oscillating at small amplitudes, J. Fluids Struct. **5**, 113–126 (1991)

17.23 M. Hildebrandt, L. Christensen, J. Kerdels, J. Albiez, F. Kirchner: Realtime motion compensation for ROV-based tele-operated underwater manipulators, Proc. of IEEE OCEANS'09 (2009) pp. 1–6

17.24 C.P. Sayers, R.P. Paul, L.L. Whitcomb, D.R. Yoerger: Teleprogramming for subsea teleoperation using acoustic communication, IEEE-JOE **23**(1), 60–71 (1998)

17.25 L. LaPierre, P. Fraisse, N. M'Sirdi: Hybrid position/force control of a ROV with a manipulator, Proc. IEEE OCEANS'98, Vol. 2 (1998) pp. 931–935

17.26 P. Zhang, K. Tanaka, E. Shimizu, M. Ito: A teleoperating system for underwater manipulator to practice a time limit task, Proc. IEEE ISIC (2003) pp. 467–471

17.27 A.D. Bowen, D.R. Yoerger, C. Taylor, R. McCabe, J. Howland, D. Gomez-Ibanez, J.C. Kinsey, M. Heintz, G. McDonald, D.B. Peters, B. Fletcher, C. Young, J. Buescher, L.L. Whitcomb, S.C. Martin, S.E. Webster, M.V. Jakuba: The nereus hybrid underwater robotic vehicle for global ocean science operations to 11,000m depth, Proc. IEEE OCEANS'08 (2008) pp. 1–10

17.28 M. Hildebrandt, J. Kerdels, J. Albiez, F. Kirchner: A multi-layered controller approach for high precision end-effector control of hydraulic underwater manipulator systems, Proc. MTS/IEEE OCEANS'09 (2009) pp. 1–5

17.29 H. Shim, B. Jun, P. Lee, H. Baek, J. Lee: Workspace control system of underwater tele-operated manipulators on an ROV, IEEE-JOE **37**, 1036–1047 (2010)

17.30 R.J. Anderson, M.W. Spong: Bilateral control of tele-operators with time delay, IEEE Trans. Autom. Control **34**(5), 494–501 (1989)

17.31 B. Sciliano, L. Villani: *Robot Force Control*, The Springer International Series Engineering and Computer Science, Vol. 540 (Springer, New York 2000)

17.32 S. Hutchinson, G. Hager, I.C. Corke: A tutorial on visual servo control, IEEE Trans. Robotics Autom. **12**(5), 1465–1470 (1996)

17.33 F. Chaumette, S. Hutchinson: Visual servo control, Part 1: Basic approaches, IEEE Robotics Autom. Mag. **13**(4), 82–90 (2006)

17.34 J. Kerdels, J. Albiez, F. Kirchner: Sensorless computer control of an underwater DC manipulator, Proc. of MTS/IEEE OCEANS'08 (2008) pp. 1–5

17.35 G. Marani, S.K. Choi, J. Yuh: Underwater autonomous manipulation for intervention missions AUVs, IEEE-JOE **36**(1), 15–23 (2009)

17.36 H. Shao, K. Nomani: Bilateral control of tele-hand system with neuro-fuzzy scheme, Int. J. Indust. Robot **33**(2), 216–227 (2006)

17.37 N. Chopra, M.W. Spong, R. Lozano: Synchronization of bilateral teleoperators with time delay, Automatica **44**, 2142–2148 (2008)

17.38 B. Drury: *Control Techniques Drives and Controls Handbook*, The Institute of Electrical Engineers (Athenaeum, Gateshead 2001)

17.39 F. Chaumette, P. Rives, B. Espiau: Positioning of a robot with respect to an object, tracking it and estimating its velocity by visual servoing, Proc. IEEE ICRA'91, Vol. 3 (1991) pp. 2248–2253

17.40 F. Chaumette, E. Malis: 2 1/2 d visual servoing: A possible solution to improve image-based and position-based visual servoings, Proc. IEEE ICRA'00, Vol. 3 (2000) pp. 630–635

17.41 D. Dementhon, L. Davis: Model-based object pose in 25 lines of codes, Int. J. Comput. Vis. **15**, 123–141 (1995)

17.42 B. Espiau, F. Chaumette, P. Rives: A new approach to visual servoing in robotics, IEEE Trans. Robotics Autom. **8**(2), 313–326 (1992)

17.43 A.C. Sanderson, L.E. Weiss: Adaptive visual servo control of robots. In: *Robot Vision, International Trends in Manufacturing Technology*, ed. by A. Pugh (Springer, Berlin 1983) pp. 107–116

17.44 T. Ranstrom: *Visual Servoing: Summary Report 130-1694* (Schilling Robotics, Davis, 2011)

17.45 Sound Metrics Corp.: DIDSON Sonars (2013), http://www.soundmetrics.com/Products/DIDSON-Sonars

17.46 Teledyne BlueView: High resolution multibeam imaging sonar (2013), http://www.blueview.com/

17.47 S.C. Yu, T.W. Kim, G. Marani, S.K. Choi: Real-time 3-D sonar image recognition for underwater vehicles, Proc. UT'07 & SSC'07 (2007) pp. 142–146

17.48 S.C. Yu: Development of real-time acoustic image recognition system using by autonomous marine vehicle, IEEE-JOE **35**(1), 90–105 (2008)

17.49 T.W. Kim, S.C. Yu, J. Yuh: Neural network-based underwater image classification for AUV, Proc. 17th IFAC World Congress (2008) pp. 15991–15995

17.50 G. Marani, J. Yuh, S.K. Choi: Real-time center of buoyancy identification for optimal hovering in autonomous underwater intervention, Intell. Serv. Robotics **3**(3), 175–182 (2010)

17.51 Robaid: SAUVIM robot completed its first fully autonomous mission (2010), http://www.robaid.

com/robotics/sauvim-robot-completed-its-first-fully-autonomous-mission.htm
17.52 G. Antonelli: *Underwater Robots Motion and Force Control of Vehicle-Manipulator Systems* (Springer, Berlin 2006)
17.53 G. Antonelli, F. Caccavale, S. Chiaverini: Adaptive tracking control of underwater vehicle-manipulator systems based on the virtual decomposition approach, Proc. IEEE ICRA'04, Vol. 20 (2004) pp. 594–602
17.54 G. Antonelli, S. Chiaverini: Singularity-free regulation of underwater vehicle-manipulator systems, Proc. Am. Control Conf., Vol. 1 (1998) pp. 399–403
17.55 G. Antonelli, S. Chiaverini: Task-priority redundancy resolution for underwater vehicle-manipulator systems, Proc. IEEE ICRA'98 (1998) pp. 768–773
17.56 C.C. de Wit, O. Diaz, M. Perrier: Nonlinear control of an underwater vehicle/manipulator with composite dynamics, IEEE Trans. Control Syst. Technol. **8**(6), 948–960 (2000)
17.57 M.W. Dunnigan, G. Russell: Evaluation and reduction of the dynamic coupling between a manipulator and an underwater vehicle, IEEE-JOE **23**(2), 260–273 (1998)
17.58 T. Fossen: Adaptive macro–micro control of nonlinear underwater robotic systems, Proc. ICRA, Vol. 1 (1991) pp. 1569–1572
17.59 J. Han, J. Park, W. Chung: Robust coordinated motion control of an underwater vehicle-manipulator system with minimizing restoring moments, IEEE-JOE **38**, 1197–1206 (2011)
17.60 M. Ishitsuka, K. Ishii: Development and control of an underwater manipulator for AUV, Proc. UT'07 & SSC'07 (2007) pp. 337–342
17.61 J. Kim, W. Chung, J. Yuh: Dynamic analysis and two-time scale control for underwater vehicle-manipulator systems, Proc. IROS'03, Vol. 1 (2003) pp. 577–582
17.62 F. Lizarralde, J. Wen, L. Hsu: Quaternion-based coordinated control of a subsea mobile manipulator with only position measurements, Proc. 34th IEEE CDC, Vol. 4 (1995) pp. 3996–4001
17.63 N. Sarkar, T.K. Podder: Coordinated motion planning and control of AUV-manipulator system subject to drag optimization, IEEE-JOE **26**, 228–239 (2001)
17.64 S. Soylu, F. Firmani, B.J. Buckham, R.P. Podhorodeski: Comprehensive underwater vehicle-manipulator system teleoperation, Proc. IEEE OCEANS'10 (2010) pp. 1–8
17.65 Y. Sun, C. Cheah: Adaptive setpoint control of underwater vehicle-manipulator systems, IEEE Conf. Robotics Autom. Mechatron., Vol. 1 (2004) pp. 434–439
17.66 N. Sarkar, J. Yuh, T. Podder: Adaptive control of underwater vehicle-manipulator systems subject to joint limits, Proc. IROS'99, Vol. 1 (1999) pp. 142–147
17.67 J. Han, W.K. Chung: Coordinated motion control of underwater vehiclemanipulator system with minimizing restoring moments, Proc. IROS ICRA'08, Vol. 1 (2008) pp. 3158–3163
17.68 W.H. Zhu, Y.G. Xi, Z.J. Zhang, Z. Bien, J. De Schutter: Virtual decomposition-based control for generalized high-dimensional robotic systems with complicated structure, IEEE Trans. Robotics Autom. **13**(2), 411–436 (1997)
17.69 R. Panish, M. Taylor: Achieving high navigation accuracy using inertial navigation systems in autonomous underwater vehicles, Proc. MTS/IEEE OCEANS'11 (2011) pp. 1–7
17.70 IXSEA: PHINS User Guide (2007)
17.71 O. Brock, R. Grupen: Final Report of NSF/NASA Workshop on Autonomous Mobile Manipulation (AMM), NSF/NASA Workshop on AMM (2005) pp. 27–29
17.72 M. Carreras, J. Yuh, J. Batlle, P. Ridao: A behavior-based scheme using reinforcement learning for autonomous underwater vehicles, IEEE-JOE **30**, 416–427 (2005)
17.73 M. Carreras, J. Yuh, J. Batlle, P. Ridao: Application of SONQL for real-time learning of robot behaviors, Int. J. Robotics Auton. Syst. **55**(8), 628–642 (2007)
17.74 S. Zhao, J. Yuh: Experimental study on advanced underwater robot control, IEEE Trans. Robotics Autom. **21**(4), 695–703 (2005)
17.75 A. Hanai, H.T. Choi, S.K. Choi, J. Yuh: Experimental study on fine motion control of underwater robots, Int. J. Adv. Robotics **18**(10), 963–978 (2004)
17.76 T.W. Kim, J. Yuh: Application of on-line neuro-fuzzy controller to AUVs, Int J. Inf. Sci. **145**(1/2), 169–182 (2002)
17.77 C.S.G. Lee, J.S. Wang, J. Yuh: Self-adaptive neuro-fuzzy systems for autonomous underwater vehicle control, Int. J. Adv. Robotics **15**(5), 589–608 (2001)
17.78 J. Yuh, J. Nie: Application of non-regressor based adaptive control to underwater robots: Experiment, Int. J. Comput. Electr. Eng. **26**(2), 169–179 (2000)
17.79 K.C. Yang, J. Yuh, S.K. Choi: Fault-tolerant system design of an autonomous underwater vehicle – ODIN: An experimental study, Int. J. Syst. Sci. **30**(9), 1011–1019 (1999)
17.80 H.H. Wang, S.M. Rock, M.J. Lees: Experiments in automatic retrieval of underwater objects with an AUV, Proc. IEEE OCEANS'95, Vol. 1 (1995) pp. 366–373
17.81 J. Evans, P. Redmond, C. Plakas, K. Hamilton, D. Lane: Autonomous docking for intervention-AUVs using sonar and video-based real-time 3-D pose estimation, Proc. IEEE OCEANS'03, Vol. 4 (2003) pp. 2201–2210
17.82 M. Prats: Reconfigurable AUV for intervention missions: A case study on underwater object recovery, Intell. Serv. Robotics **5**(1), 19–31 (2012)
17.83 D. Ribas, N. Palomeras, P. Ridao, M. Carreras, A. Mallios: Girona 500 AUV, from survey to intervention, IEEE/ASME Trans. Mechatron. **17**(1), 46–53 (2012)
17.84 P. Sanz, R. Ridao, G. Oliver, P. Casalino, C. Insaurralde, C. Silvestre, C. Melchiorri, A. Turetta: TRIDENT: Recent improvements about autonomous underwater intervention missions, Proc. IFAC (NGCUV'12), Vol. 3 (Porto, Portugal 2012) pp. 355–360

17.85 E. Belcher, W. Hanot, J. Burch: Dual-frequency identification sonar (DIDSON), Proc. Int. Symp. Underw. Technol. (2002) pp. 187–192
17.86 M. Walter, F. Hover, J. Leonard: SLAM for ship hull inspection using exactly sparse extended information filters, Proc. IEEE ICRA'08 (2008) pp. 1463–1470
17.87 G. Marani, S.K. Choi: Underwater target localization, IEEE Trans. Robotics Autom. **17**(1), 64–70 (2010)
17.88 J. Kim, G. Marani, W. Chung, J. Yuh: A general singularity avoidance framework for robot manipulators: Task reconstruction method, IEEE ICRA'04 (2004) pp. 4809–4814
17.89 G. Marani, J. Kim, J. Yuh, W.K. Chung: Algorithmic singularities avoidance in task-priority based controller for redundant manipulators, Proc. IROS'03, Vol. 4 (2003) pp. 3570–3574
17.90 G. Marani, J. Yuh, S.K. Choi: Autonomous manipulation for an intervention AUV. In: *Guidance and Control of Unmanned Marine Vehicles*, Vol. 69, ed. by G. Roberts, R. Sutton (IET, Stevenaze 2006)
17.91 TRIDENT Team: TRIDENT newsletter (2012) http://www.irs.uji.es/trident/files/2nd-TRIDENT-SCHOOL-Newsletter-Oct2012.pdf

18. Non-Acoustic Sensors

David Fries, William Kirkwood

The focus of the chapter is to provide a cross section of progress in the area of non-acoustic sensors for underwater mobile platforms and to provide a framework for structuring the subject of oceanographic non-acoustic sensors. State-of-the-art non-acoustic sensor systems are a summation of five areas: advanced analytical detectors, distributed computing, network technology, automation and power systems. The current state of sensor technology is partitioned and reviewed using a chemical, biological and physical sensing systems approach. A commentary on future directions for the three classes of sensor systems is also provided. Sampling technology with programmable fluidic processing systems capable of sampling with offline or online detectors is additionally presented to augment real-time non-acoustic sensor technology. The packaging of these types of sensor is briefly presented. Applications of ocean engineered AUV-based (AUV: autonomous underwater vehicles) sensing and sampling systems are also mentioned throughout the chapter to highlight the emergence of this class of ocean technology.

18.1 Non-Acoustic Ocean Sensors: Sourcing and Classification 423
18.2 Classical Non-Acoustic Ocean Sensors .. 424
18.3 Chemical Sensor Systems 426
18.4 Biological Sensor Systems 429
18.5 Physical Sensor Systems 432
18.6 AUV-Based Physical Sensors – Horizons 435
18.7 AUV-Chemistry Sensors – Horizons 435
18.8 AUV-Based Biological Sensors – Horizons 436
18.9 Autonomous Sampling Systems – Extending Real-Time AUV Sensors 436
18.10 Non-Acoustic Sensor Packaging 436
18.11 The Essential Need for Sensors 437
References 437

Scientists and engineers concerned with ocean changes that result from process dynamics work to investigate and design solutions for measuring the environmental systems observed. Process dynamic measurements and spatial-temporal observational abilities of the ocean historically are enhanced by the developmental evolution of physical, chemical, and biological sensor technologies. Historically, we would argue that field research activities in the ocean beyond diver depths, with the exception of sample collection, has been dominated by the research vessel paradigm using acoustics.

Predominantly the acoustic sensors for ocean science came from the development of sonars that have been focused on acoustic sensing for navigation as well as communications. Acoustic sensors are a primary instrument in the ocean observations and ongoing advances are well documented in the acoustic engineering happened rapidly over 50 years and literature. Fundamental advancements in acoustics continue, albeit incrementally, in comparison to the rapidly advancing methodologies in chemical, biological, and specialty physical sensing. Consequently, the focus of this chapter is to provide a cross-section of progress in this latter class and to provide a framework for structuring the subject of oceanographic non-acoustic sensors.

18.1 Non-Acoustic Ocean Sensors: Sourcing and Classification

The key to outlining the state and composition of non-acoustic ocean sensors is to consider the advancements of sensor technologies from larger sensor application spaces. Adapting technologies for ocean

Table 18.1 Sensor *space* based on stimulus to identify existing sensor system types

Stimulus	Applications
Acoustic	Wave amplitude, phase, polarization, spectrum, wave velocity
Biological	Bio-species (types, conc. states)
Chemical	Compositions (identity, conc. states)
Electric	Charge, current, potential, voltage, electric field (amplitude, phase, polarization, spectrum), conductivity, permittivity
Magnetic	Magnetic field (amplitude, phase, polarization, spectrum), magnetic flux, permeability
Mechanical	Positions (linear, angular), acceleration, stress, pressure, strain, mass, density, speed of flow, rate of mass transport shape, architecture, stiffness, viscosity, other
Radiation	Type, energy, intensity, other
Thermal	Temperature, flux, specific heat, thermal conductivity
Optical	Wave amplitude, phase, polarization, spectrum, wave velocity, refractive index, emissivity, reflectivity

applications from the terrestrial and space fields has greatly contributed to what is now considered part of an oceanographer's toolbox. For example, advanced analytical chemical systems, microelectronics, computers, and satellites are now commonly included as available ocean technologies. Innovations have and will continue to be brought over from the medical, military, and water treatment fields as they develop aquatic and parameter measurement sensors. Grouping the current state of sensor technology within the chemical, biological, and physical classes is adequate for a snapshot of the current state; however, in general, any of the sensors can be lumped into either radiation, fields, or mass transfer-based systems.

As a starting point, the treatment of non-acoustic ocean sensor solutions using a function classification scheme based on stimulus may be a reasonable starting point to categorize and organize existing sensors. Table 18.1 provides a summary of the stimuli and the composition of applications space. The generic sensor space categorization by stimulus leads to more specific non-acoustic sensor systems that most are familiar with, for example, non-acoustic current/flow meters, optical systems, thermistors, dissolved oxygen, plankton samplers, fluorometers, radiometers, magnetic current/flow meters, seismometers, mass spectrometers, and laser Raman spectrometers.

State-of-the-art non-acoustic sensor systems are a summation of five areas: advanced analytical detectors, distributed computing, network technology, automation and power systems. All of these areas are critical towards the advancement of more capable ocean sensing and observatories. However, new analytical functions are our focus and the basis for organizing the non-acoustic sensors space. Consider, for example, the question as to what scale the sensor system of interest needs to operate at. Spatial and temporal elements of the environmental continuum versus the spatial and temporal continuum of the sensor system have to be contemplated when looking to derive appropriate sensor-AUV selections.

18.2 Classical Non-Acoustic Ocean Sensors

Acoustics have been a dominant technology in ocean studies such as mapping and surveys. However, many ocean spatial and temporal process phenomena require other sensory modalities to properly access the space, time, and composition. In the case of particle characterization for size, form, and temporal dynamics, acoustic systems provide sensing abilities for longer temporal and spatial scales. In contrast, non-acoustic ocean optical systems provide similar time–space coverage but also extend to both longer and shorter characterization time scales. Depending on the platform selected for sensor deployment, optical systems can also extend into different spatial scales (Fig. 18.1).

Traditional non-acoustic sensors have matured primarily in the physical sensing area. The measurement of temperature is a great historical example of this maturity process, and it is still prevalent today. Examples of temperature sensing include the thermistor, thermistor string, and electronic reversing temperature sensor. Temperature measurement systems of all types, both general and high precision, are considered mature technologies.

Depth or water column pressure is another common measurement. Examples of pressure measurement systems include quartz pressure sensors, solid state diaphragms designed as submersible pressure transducers, such as piezoresistive sensors with an isolated stainless steel diaphragm in a stainless steel or titanium package. Their uses include wave and tide gauges, tsunami detection, ocean current surveys, and seabed subsidence.

Salinity is widely perceived as a fundamental variable in the regulation of natural marine systems. It is a key parameter in determining seawater physical

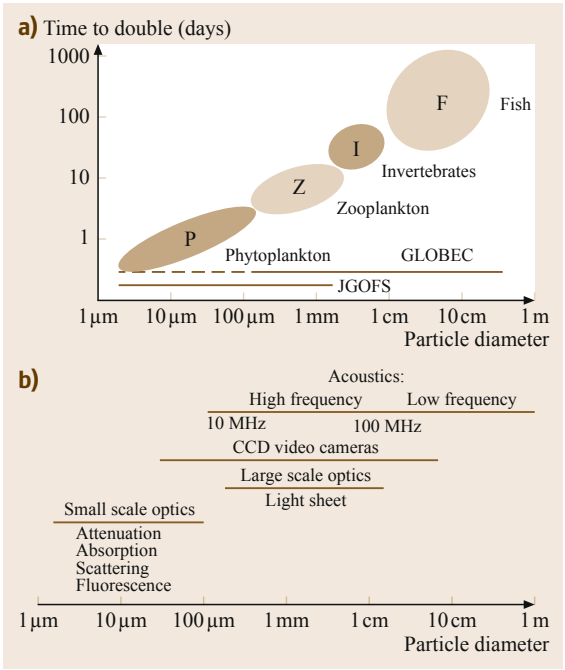


Fig. 18.1a,b Comparison of acoustic sensory range versus non-acoustic optical sensing for particle analysis and associated ocean function or phenomena. **(a)** Range of ocean *particles* and the timescale to functionally double relevant to acoustic and optical sensing modes, **(b)** Partitioning of the light and acoustic based measurements relative to spatial scale

characteristics such as density, conservative element concentrations, and solubility of gases [18.1]. In addition, as a conservative water mass tracer, it is critical to descriptive modeling of oceanic circulation, mixing, and climate processes. Salinity is an important ecological factor as it influences functional and structural properties of some organisms through changes in total osmotic concentration, relative proportions of solutes, coefficients of absorption, and saturation of dissolved gases, density, and viscosity [18.2].

Because of the global significance of salinity in marine systems, oceanographic researchers are continually exploring new tools, such as instrumenting animals, in order to study the physical, chemical, and biological structure of oceans [18.3]. Salinity is determined by electrical conductivity, temperature, and depth (CTD) instrumentation [18.4]. Electromagnetic inductive cell type conductivity and a platinum thermoresistance temperature are common sensors in CTD. There are high precision and miniaturized versions that are low cost with high precision for in-situ applications [18.5]. Additionally, towed-CTD instrument chains exist for accurate two-dimensional data sets of temperature and salinity in the upper layer of the ocean.

Fluid motion measurements are core instrumentation in classical ocean sensor systems. Currents, tides, and turbulence characterizations are needed for real-time forecasting and physical oceanographic model updates. Measurements are made from various platforms such as moorings, ships of opportunity, and autonomous underwater vehicles (AUV). Examples other than acoustic Doppler current profilers (ADCPs) include drag meters, electromagnetic current sensors, mechanical styles based on rotors, propellers, and electrode pairs using time-of-flight techniques. Wave and tide sensors provide additional information by recording wave speed, height, and direction using technologies like quartz pressure transducers, electrostatic capacitance, micro-machined silicon pressure chips, and semiconductor strain gauges, plus tri-axial accelerometers with three orthogonal rate gyros.

Seawater subsurface ocean optical properties comprise a history of rich instrumentation development to perform in-situ light measurements. Optical instrumentation allows the temporal and spatial surveying of light intensity, wavelength, reflectance, radiance, UV radiation, dissolved solids, bioluminescence, backscatter coefficients, beam attenuation, fluorescence, light transmittance, ocean color, chlorophyll concentration, absorption, angular distribution, and photosynthetically active radiation (PAR) irradiance sensors for the study of marine photosynthesis.

The distribution of optical energy in the upper ocean and ocean variable imaging are the drivers pushing advancement of ocean optical sensors for oceanography. Optical radiation sensors come in numerous configurations and target a variety of optical parameters. Systems based on absorbance, scattering, fluorescence, irradiance/radiance, and transmittance are all in operation in oceanography. Additionally, most of these optical instruments have been demonstrated on AUVs and other mobile platforms. Mobile spatial mapping has been accomplished using transmissometers, fluorometers, and radiometers.

18.3 Chemical Sensor Systems

Describing the state of chemical sensor systems is an enormous task and could easily be the basis for an entire volume dedicated to the subject. In this chapter we provide examples of technologies that represent aspects of chemical instrumentation and hopefully will provide a snapshot of the chemical instrumentation migration to maturity within the environment and mobility using AUVs. In general, key drivers for evolution of chemical instrumentation come from the ability of chemical sensors to provide trace detection and spatial/temporal chemical dynamics in water and within chemical ecology. Previous literature reviews have covered and summarized in-situ chemical sensors into two volumes on chemical sensor systems for aquatic sciences and chemical and biological sensors for time-series focused research [18.6, 7]. These reviews describe a variety of continuous flow analyzers, electrochemical sensors, and optical sensors for in-situ measurements of dissolved chemical species.

Chemical sensors need to have excellent detection limits to be desirable. Attributes such as precision, selectivity, response time, a large dynamic concentration range, low power consumption, robustness, and less variation of instrumental response with temperature and pressure are the most advantageous. Also, instruments that stay free of fouling problems are highly attractive to end users. It is also important to note that some chemical sensors are deployed in several different oceanic environments for the detection of targeted chemical species, whereas others are useful in only one environment because of various configuration limitations. There are a number of chemical sensors that can be used as stand alone in-situ sensors, whereas a greater number involve water being brought from the environment into an analytical cell or chamber via tubing, either aboard ship or onboard a mobile platform such an AUV. Multi-element sensors that measure several chemical species from the same water sample are an area that researchers are pushing for further development. These multi-function sensors will increase the capability of AUVs (for example) to characterize spatial ocean chemical composition and diversity.

At present, most sensors need further refinement to perform measurements at very high acquisition rates without a loss of precision or sensitivity. Fundamental inorganic compounds and ions (oxygen, basic nutrients, hydrogen sulfide, pH, carbon dioxide, NOx, SOx, Fe, Mn, and trace metals) are good candidates for continual AUV-based integration. Nitrate, a key nutrient, as a sensor system has seen major advancement in the last few years. Nitrate sensors are a current example of instrumentation that will see increased usage for nutrient mapping, particularly in AUV coastal surveys. Current commercial nitrate sensors can be accessed as either by the use of automated wet chemistry or ultraviolet (UV) spectroscopy performed in situ. Optical UV techniques function to detection levels of $\approx 0.5\,\mu$M, but for levels below that wet chemistry is required at present. UV nitrate sensors are, therefore, suitable for use in the ocean surface except for some rare areas. For work below the micromolar level there are wet chemical systems that can achieve nanomolar levels of detection [18.6]. At the time of writing this we are not aware of any having been deployed on an AUV. Additionally, wet chemical systems are evolving to measure ammonia, nitrite, orthophosphate, silicates, and iron at desired levels and acquisition rates for waters and sediments. We expect that these instruments will in time be in a format that is tractable for insertion on AUV platforms.

The study of boundaries and interface environments are important in ocean dynamics. Boundary layers include atmosphere-to-water, water-to-sediment/soil, as well as the boundaries between fluid masses with sharply different characteristics. These different contexts all require the measurement of fluxes. Additionally, deterioration of water quality, ocean acidification, and unknown processes have resulted in a tremendous decline of ocean communities [18.8]. Quantifying the fluxes of chemicals is of great importance to understand these local, regional, or global biogeochemical processes. These include fluxes of key carbon species such as CO_2 and CH_4, nitrogen species such as NH_4^+, N_2O, and NO_3^-, many important metals and trace molecular species, sulfur, and other measurements of CO_2 system parameters (pH, CO_2 fugacity (f_{CO_2}), total dissolved inorganic carbon (CT), total alkalinity (TA), and carbonate ion concentration [CO_3^{2-}]). An illustrative system is the spectrophotometric elemental analysis system (SEAS) [18.9]. Differently from the single-purpose instruments available on the market, SEAS offers a platform for CO_2 parameters, nutrients, and trace metals. In its third generation, the in-situ chemical sensor can be configured for a wide variety of chemical measurements, including trace metals, nutrients, and inorganic carbon system variables in different sample environments and deployment modes. Automated wet chemical analyzers such as these will see increased use in AUVs and possibly in combinations with AUV swarms to provide informative data on flux in different oceanic contexts.

Oxygen chemical sensors [18.10, 11] are in high demand, especially for AUV mapping studies of anoxic environments. Standard oxygen sensors such as oxygen electrodes, galvanic O_2 micro-sensors, and Clark

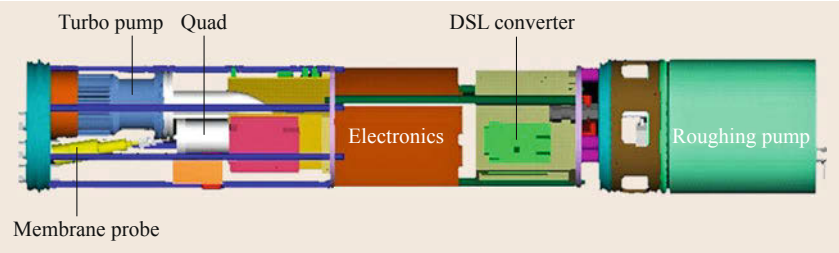


Fig. 18.2 Underwater mass spectrometer consisting of three primary subsystems: a high-pressure membrane inlet with a small-volume seawater pumping system; a quadrupole mass spectrometer and oil-less vacuum pumping system; and underwater housing. The membrane inlet assembly consists of a circular sheet or tube of polydimethylsiloxane (PDMS) membrane material

polarographic sensors are frequently being replaced by the rise of optode-based O_2 systems. Optode devices have lowered the cost of making oxygen measurements while they have also extended reliability and maintainability. Related to these types of sensors are membrane-covered amperometric sensors, a renewable reagent fiber optic sensor for measuring partial pressures, and membrane transport of analytes into reactive solutions for conductance detection. Pure optical systems such as non-dispersive infrared absorption and cavity ring down laser systems have emerged especially in the area of CO_2 sensing. Underwater methane sensors based on IR optical absorption and mass spectrometry have gone through rapid ascension driven by the need for assessment of their role in climate studies. We expect these systems will migrate over to mobile vehicle platforms from the initial demonstrations or be optimized further for more numerous placements.

Underwater mass spectrometer technology has been adapted from its lab-based origins to AUV-based field deployments [18.12]. For the detection of organic pollutants and other toxic chemicals, this evolving technology provides non-parallel analytical performance and allows the possibility of access to the entire chemical space. The technology (Fig. 18.2) can detect organic compounds of small molecular mass (< 300 Da) and large molecular mass (> 300 Da). Small molecules include methane, xenobiotics (e.g., halogenated hydrocarbons), and toxins. Large molecular mass compounds, which have been largely ignored, include dissolved organic compounds (DOC), dissolved organic nitrogen compounds (DON) and can in time be characterized with the evolution of the technology.

The initial demonstration of an underwater mass spectrometry on AUVs was membrane inlet mass spectrometry (MIMS) [18.12]. This same design has also proven to be an effective tool for measuring complex mixtures of dissolved gases (volatile organic compounds (VOCs), CO_2, etc.) in non-AUV applications also. MIMS has also been used to study metabolite flux and dissolved gases in surface waters and continuously aboard ships [18.13, 14]. The recent adaptation of MIMS to in-situ environmental analyses demonstrates the utility of such an instrument operating while underway at sea, allowing continual monitoring of many gas species in real time. This approach allows potent analytical monitoring of spatially explicit biogeochemical changes.

The possibility of accessing the entire chemical space is a unique feature of mass spectrometry and the rise in mass spectrometry in the lab over 2 decades will mostly likely be replicated in the field/ocean analytical sensor space. Awaiting the expansion of mass spectrometry in the ocean is the increase in demonstrations of ion sources beyond the membrane inlet method that has been used to date by all underwater mass spectrometers (quad, ion trap, cycloidal). Atmospheric ion source developments for liquid chromatography mass spectrometry (LC-MS), which are numerous in lab settings suggest new source designs that will take a liquid sample and apply vacuum/gaseous-based mass analysis. The use of mass spectrometry on AUVs will most certainly see a rise in use in the coming decade, with the need for oil and gas field pipeline monitoring and exploration activities using the MS/AUV combination. The rise in chemical ecology and the need to understand non-point source chemical pollution and natural products will also drive the need for spatial chemical mapping.

Laser Raman spectroscopy is another class of spectrometer for ocean science. Raman spectroscopy has the frequently desired capability of being non-destructive. Raman spectroscopy also has the ability to discern structural information many other systems lose when analyzing the sample. For example, Raman spectroscopy easily differentiates calcite from aragonite, which are the same calcium carbonate material but have different lattice structures. Raman spectroscopy uses a monochromatic laser to irradiate a spot on the target causing vibrational modes which return a portion of the photons returned are frequency shifted. The original

light frequency is filtered out and the resulting spectra collected is rich with information other sensors and instrument systems cannot provide. In addition, Raman spectroscopy looks at a spectrum of chemical compounds simultaneously, as long as they are within the laser beams' spot size. Raman systems are also capable of scanning a prescribed area or watching a reaction occur over time. Therefore, Raman spectroscopy is a very strong instrument for use in situ.

Natural gas hydrates occur within sediments and outcrops around the world; hydrates are an excellent target for Raman investigation [18.15]. It is helpful that hydrates, composed of 85% water and 15% gas, are ice-like and semi-transparent. The excitation laser can thus be focused within the specimen without the exact positioning strictures required for a fully opaque surface. Raman instruments have been evaluated for detection limits on a number of gas species. Samples of 1 atm CO_2, 0.7 atm N_2, 0.2 atm O_2, or 0.01 atm H_2O vapor contained in an Erlenmeyer flask were successfully tested using the excitation laser focused through the wall of a glass vessel. One experiment collected 64 acquisitions of 10 s each and determined the following detection limits: 0.2 bar CO_2, 0.1 bar N_2, 0.1 bar O_2, 0.01 bar CH_4, and 0.01 bar H_2O vapor [18.16]. The results observed are consistent with the well-documented Raman scattering cross-sections for CO_2, N_2, and O_2, where the scattering cross-section for CH_4 is almost an order of magnitude greater [18.17]. The detection limit for CH_4 is, therefore, about 10× greater than that for CO_2, N_2, and O_2. In addition, although seawater itself is of little interest for direct spectroscopy, one species peak is usually dominant, that this that of SO_4^{2-} at a near constant 28 mM concentration. In the normal oxygenated ocean, SO_4^{2-} varies with salinity; therefore, it is possible to use this prominent peak to provide a quantitative marker. Acquiring a seawater spectrum at depth serves as a valuable quantitative system check [18.18].

Current trends indicate a high likelihood that autonomous vehicles may in the near term be capable of successfully hosting instruments such as a Raman spectrometer. When suitable, autonomous target locating and sensing strategies can be devised; the detection of a chemical signature from gas plumes emitting from the seafloor are one potential example. Although autonomous operation is commonplace in industry, there are current technical barriers that inhibit long-term at sea operations for Raman spectroscopy in an autonomous mode, namely a suitable cost effective power source. In remotely operated vehicle (ROV) operations Raman spectroscopy has no limitations with the abundance of power, nominally 10 W continuous with today's technology.

In-situ fluorometers are another class of AUV compatible sensors that see almost routine use on AUVs and will see increased use over time (Fig. 18.3). Spectral sensors such as fluorometers have been developed for chlorophyll a in algal and cyanobacterial populations, largely related to trophic support (biomass) and primary productivity. The configurations of fluorometers have evolved over time into more simultaneous spectral wavelengths, extending the range of sensing wavelengths while reducing the packaging size and costs of the systems. This has largely been enabled with the enormous growth in solid-state opto-electronic devices, such as high power light-emitting diode (LEDs). One example of fluorometers moving beyond physiological flourometers is the application as hydrocarbon spill monitors. AUV deployments using UV flash lamp configured fluorometers have taken place, but with the rise of low cost UV LEDs in combination with market growth of lower cost and smaller form factors this class of oil sensor is becoming readily available from industrial sources. Continual improvement of antifouling systems for optical sensors is also actively under development, and we expect to see improvements that will be critical for long duration AUV studies.

Trace metal sensors are a type of chemical sensor suitable to perform in-situ metal analysis. Several analytical approaches exist, such as fractionation. Spectroscopic techniques such as laser-induced breakdown (LIBS) use emission spectrometry and are good candidates for future AUV inclusion. However, LIBS may need some form of chemical separation, such as preextraction to augment the utility. Traditional techniques use the electrochemical properties of metals and are divided into two major analytical classes, voltammetric and potentiometric systems. These electrochemical techniques have been used over the past 3 decades. Low power requirements with small form factor electronic packaging make electrochemistry a compatible chemi-

Fig. 18.3 Field fluorometer system using a xenon pulsed UV lamp in an underwater housing. The unit has detection capabilities for hydrocarbons (e.g., carbazole) in sub-ppb levels

cal sensor system for AUVs. Direct reduction methods and adsorptive styles, such as anodic stripping voltammetry illustrate the analytical performance and size of these sensors [18.19]. Potentiometric sensors are also applicable to AUV deployment, however the individual capability of a system is coupled to the performance of the ion selective coating that separates the sample from the medium. Further development of new ion-selective coatings is critical for the expansion of the chemical *vision* of these electrochemical metal species sensors for AUVs.

Automated wet chemical analysis has a proven history in aquatic observations, in particular for metals relevant to micronutrient activity. Continuous flow injection instrumentation has been used for profiling and surface ion detection mapping of specific metals (Fe, Co, Mn, Cu, Zn). As was mentioned prior, wet chemical systems have the ability to be tuned to different chemical species and in many ways can be viewed as *hardware processors* that are *programmed* to different chemicals based on the reaction chemistry used. This applies to electrochemistry applications and beyond. Insertion on AUV platforms will become more prevalent over time as form factors shrink and the cost of components is reduced.

Many chemical instrument systems benefit from the incorporation of microelectronics, which results in significant miniaturization. The realm of AUVs should substantially benefit from developments occurring in these miniaturized engineered instrument systems. An example of this is the water-stable organic transistor, which has already been developed and applied to chemical sensors [18.20]. The development of low-cost, organic thin-film transistor (OTFTs) devices capable of converting an analyte binding event to an easily read electrical signal is already leading to new chemical specific devices with reduced size and cost. OTFT sensors have been demonstrated in aqueous solutions with concentrations as low as ppb for trinitrobenzene, methylphosphonic acid, and other chemicals. This work indicates reliable OTFT operation in aqueous media, hence opening new possibilities of chemical sensing with OTFTs chemical sensors and on mobile platforms. OTFTs are ideal for inexpensive, chemical sensors because of their compatibility with flexible, large-area substrates, simple processing, and highly tunable active layer materials.

Nanomaterials also show promise in aquatic chemical analysis mainly due to their highly tunable size and shape-dependent chemical and physical properties [18.21]. Although a broad definition, nanomaterials are structured components materials that have at least one dimension less than 100 nm. Furthermore, they exhibit unique surface chemistry, thermal stability, high surface area, and large pore volume per unit mass that can be exploited for ocean chemical sensor fabrication. Nanoscale materials can be implemented as scaffolds for chemosensing. Nanomaterials have proven to be useful for the fabrication of sensors and have compositions including metals, metal oxides, chalcogenides, and polymers. Their structures range from nanoparticles, nanorods, and nanowires to nanoporous structures and core-shells. A few nanomaterials have been functionally altered to work with various kinds of ligands and active groups that produce sensitive and selective sensors for specific analytes. The ever expanding group of nanocomponents has the potential to allow chemical sensor designers to create highly selective chemoreception systems able to function within the complex background of ocean waters. Migration of this technological toolbox into ocean chemical sensor systems will emerge over time as they are implemented in industrial and medical applications and adapted to ocean research.

18.4 Biological Sensor Systems

Similar to the chemical research space, biosensor research and monitoring is fragmented in such a way that at best we can provide a broad overview of important AUV amenable sensors. The spectrum of sensors available, however, can address biogeochemical cycles, primary and secondary production, species distribution, animal movements, habitats, and pollutants. It is impossible to discuss all relevant systems or to provide appropriate detail for all of those discussed in this chapter. Therefore, this is an introduction to the sensor technologies currently in process or soon to be, and what variables are quantified.

The advancement of instruments has frequently preceded many major discoveries in the field of science and technology. The microscope is a classic example of the influence of scientific instrumentation on sensing within the biological sciences. Efforts in engineering ocean instrumentation and developing the associated sensors must take into consideration the inherent environmental limitations encountered in underwater applications. Environmental issues include the corrosive nature of seawater, bio-fouling, limited energy resources, pressure tolerant enclosures, data transfer reliability and the dynamic nature of the ocean itself.

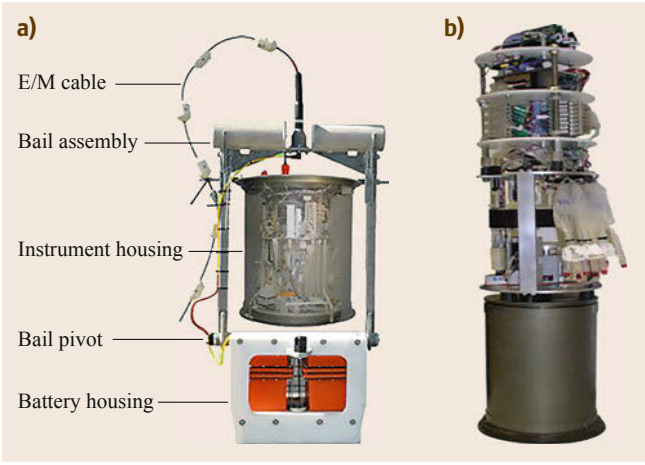


Fig. 18.4 (a) Environmental sampler processor (ESP) and (b) autonomous microbial genosensor (AMG), automated wet chemical analyzers that collect sample, homogenize, and filter the lysate and apply DNA/RNA probe arrays or amplification probes for nucleic acid detection and in the ESP case provide protein/toxin arrays for detection (images courtesy of MBARI and University of South Florida, respectively)

Biological sensor technology for underwater can be divided in one perspective as acoustic and non-acoustic based sensing. The major problem of acoustic-based sensors is that they are very prone to noise. Since this chapter is focused on non-acoustic sensors, we shall ignore the issues associated with those sensor technologies. With regard to non-acoustic sensors, they are not generalized instruments for the most part and must be tailor made for the specific application.

Biological, non-acoustic sensors can be further broken down into sensors for the visible (optical, imaging, etc.) and those for the microscopic (microbial, genetic, microscopy, etc.). In the microscopic realm sensors onboard AUVs are needed in order to measure biological reactions and physiological function in variable ecosystems. Biological sensors are, perhaps, the least developed of the sensor types but can provide key information on the function and structure/composition of biologically-influenced ecosystems in real time. As with chemical and physical sensor types, the signals produced by biological sensors need to be processed in a manner that allows for the future to perform adaptive sensing. To accomplish this they will need to be capable of long-term deployment and they should be suitable for integration with other sensing devices in an effort to provide a complete characterization of ecosystem structure over time. Although some aspects of sensor deployment may vary when implemented across different media (air, water, sediment), the general needs assessment for AUV-based sensor development is considered to be similar for all sensor types. The only real distinction between sensor types being the method to access and process a sample may change. These sensors need to be developed for monitoring all levels of biota, from microbiological to the highest levels in the eukaryotic domain.

Interdisciplinary research between engineering and biological cultures in many cases has spurred the development of sensors for AUVs and the ability to contribute to bio/eco function determination, including prediction. A working knowledge of modern molecular biological techniques synthesized with engineering approaches to sensor development is required to advance the biosensor ability of AUVs. In addition, bioinformatic data management and aquatic environmental engineering will contribute to these developments.

Areas where AUVs equipped with the latest biosensing technology can make inroads using environmental genomics and sensors for microbial viability (environmental viability of pathogens and indicators) to detect total microbial communities and identify emerging microbial problems before they create ecosystem deterioration. AUVs can help identify toxins in the environment, both toxin production and pathways. In many emerging biosensor systems computational biology and bioinformatics will be needed to manage the tremendous data streams generated from genomic and genetic sensors.

Fully operational genomic sensors (Fig. 18.4) on AUVs are still awaiting definitive demonstrations. Improvements for field deployment of AUV capable gene sensing systems are under development. These biosensor developments will benefit from an emerging complementary bio-analytical approach, namely, automated sampling onboard AUVs with post shore analysis in the lab using the highly developed genomics analytical systems. Using automated real-time sampling followed by near real-time analysis may be the first viable approach to significantly identify community structure through phylogenetic fingerprinting methods and to ground truth future biosensors.

Currently unattended genetics sensors, which can be classified as wet chemical analyzers, are the leading candidates for transfer onto AUVs. Both the environmental sampler processor (ESP) [18.22] and autonomous microbial genosensor (AMG) [18.23] are leading candidates for AUV deployments capable of species specific detection. The ESP and AMG are portable water laboratory systems that provide onsite biological analysis of water samples. These electromechanical fluidics systems are designed to collect and concentrate discrete water samples, autonomously apply molecular probes to identify microorganisms, including their DNA or protein products. Data gen-

erated are available by remote transmission and autonomous analysis in real time for adaptive sampling. Assay detection chemistries are available for many biological targets. These assays enable the sensor system to be tunable or programmable for different targeted species. This capability could be integrated into AUVs outfitted with them to act as a control function in the vehicle's control loop. Aquatic microbial detection technologies combined with AUV marine technology will provide near real-time sensing and detection capabilities to protect shellfish, beaches, coastal resources, groundwater, and drinking water sources. Polymerase chain reaction (PCR) amplification technology is the enabling technology being incorporated into the AUV compatible genetic sensor to bring down power requirements and scale while extending deployment time and the number of samples analyzed.

One aspect of these devices is that they can be configured with a variety of genetic probe arrays [18.24, 25] at one time. The probes are on small filters that have matching genetic markers which capture their genetic match when the assay of material is passed over them. Using a system of false positives and photosensitive chemistry, the identity and relative concentration can be captured in situ and in near real time. The probes can also be fixed and stored for later analysis in the lab.

Mass spectrometry and molecular biology-type systems [18.25] are demonstrating detection that indicates the presence of proteins and toxins. Real-time proteomic-based sensors that measure expression patterns in biota proteins using automated methods is a complex technical problem. Proteins are the functional elements in cells that are amenable to significant post-translational modifications not captured by functional genomic sensors. As such, proteomic-based sensors are expected to yield information that better relates to functionality at the macro-scale ecosystem. Real-time AUV sampling combined with a post-analysis approach may benefit the elucidation of the proteomics spatial-temporal dynamics in the immediate future. Similarly, metabolomic-based sensors that measure metabolic byproducts produced by functional biota will emerge on AUVs with the growth of mass spectrometry and the advancement of automated sampling methods. Detection and monitoring of known metabolic byproducts, or metabolic fingerprints, generated in response to perturbations within the environment will be highly informative to researchers as well as the general public.

Optical sensors, both spectral and imaging (Fig. 18.5) (beyond fluorometry and intrinsic properties ocean optical systems), are natural candidates for AUV missions. Optical sensors that measure biotic

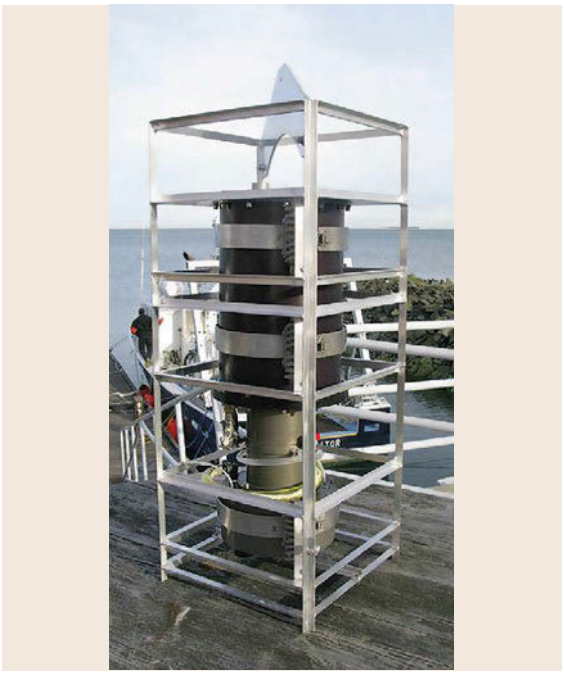


Fig. 18.5 Submersible FlowCAM acquires digital images of each particle passing through its flow cell along with up to 26 different measurements for each particle. Statistical pattern matching techniques enable automatic classification of taxonomic groups from libraries of known particle images. Particle size measurement range from 10–600 μm, deployment modes include moorings, as vertical profilers and on unmanned vehicle with a matched AUV housing (courtesy of MBARI)

morphology, and the shape and look of the organism, can be used to relate to biotic function and abundance measurements. An illustrative system for the spectral sensing pathway to measure microbial organisms is the optical phytoplankton discriminator, also known as BreveBuster [18.26]. BreveBuster was conceived and designed the using optical absorbance characteristics of particles in the water to assess the fraction of phytoplankton biomass made up by *Karenia brevis* (a red tide pathogen). An existing pre-production version of the instrument tolerates water pressure to 200 psi for deployments on AUVs and towed vehicles. The output parameter, the similarity index, represents the degree of similarity between the optical absorbance of particles in the sample under test and that of a target organism.

Flow cytometry is a biological particle imaging and counting system for morphological and abundance measurements. Flow cytometers developed from earlier efforts to count cells and particles in liquid suspension as they passed through a sensing region of the flow chamber. A primary example is the Coulter counter,

which was developed in the mid 1950s for counting and sizing red blood cells. In modern flow cytometers, cells are typically analyzed at rates up to a few thousand per second. Sensors measure cellular features and the corresponding signals are processed. These instruments have revolutionized automated cell analyses in that measurements are made at high speed. Miniaturization of the systems is rapidly occurring and shipboard, flow-through versions are now commercially available. The analysis of the intrinsic fluorescence profiles of individual marine algae can be used in general classification of organisms based on cell size and fluorescence properties.

Commercially available ocean deployable systems exist, including FlowCAM from Fluid Imaging Technologies. These instruments preform continuous imaging particle analysis, monitoring phytoplankton, zooplankton, and fluid borne particulates. The FlowCam is available in submersible models which can measure particles from 1 μm to 3 mm in size, which make it a good candidate for adoption onto AUVs. The FlowCytobot is a separate system that uses diode-pumped 532 nm lasers and can measure scattering and fluorescence of bio-particles. The FlowCytobot is currently designed for operation in coastal observatories with a connection to the shore for power. A next generation system has evolved, which is an imaging FlowCytobot version for imaging with a higher sampling throughput. The work to move these systems into more portable formats, capable of being carried on vehicles, is occurring outside the traditional ocean engineering field. Currently, lab-based engineering developments of optofluidic flow cytometers are underway primarily within the medical or clinical diagnostics arena. The field of optofluidics is comprised of optics and microfluidics working together to create novel functionalities within a small chip. These chips hold great promise for lab-on-a-chip flow cytometry and for the further miniaturization of the flow cytometer [18.27]. The Microflow cytometer has been developed on a chip for the characterization of phytoplankton. The Microflow cytometer measures distinct side scatter and fluorescence properties of various microorganisms. This device proved sensitive enough to detect and characterize picoplankton with diameters $\approx 1\,\mu m$ and larger phytoplankton up to 80 μm in length. The wide range in size discrimination coupled with the detection of intrinsic fluorescent pigments suggests that this Microflow cytometer should be able to distinguish different populations of phytoplankton as a miniaturized system on unmanned underwater vehicles.

It is hoped that highlighting these systems indicates the emergence of biological sensor technologies on mobile platforms. A large number of microbial biosensors have been developed for environmental, food, and biomedical applications, which we can expect to transition over to systems engineered for the ocean. The fabrication of microbial biosensors based on amperometry, potentiometry, conductometry, voltammetry, microbial fuel cell, fluorescence, bioluminescence, and colorimetry, is occurring and will be applied in the field. Sensors that provide reliable quantitative information related to growth rates are coming with the expansion of low cost, high performance embedded microscopy systems. Detection and data presentation of kinetic information allows direct and indirect indicators for quantitative predictions for the ecosystem being monitored. Sensors such as spectral optical instruments continue to evolve and help detect the initiation and extent of primary production. These spectral instruments are successful in connecting environmental perturbations and variations with primary production events in real time. The increased number and quality of underwater mobile imaging systems (beyond acoustics only) is helping track levels of predation to tease out changes in population numbers. These predation data give insight into the ecosystem's structural dynamics and helps researchers to better understand the impacts of environmental changes. For many of the biosensor designs outlined in this section the utilization of micro electro-mechanical (MEMS)-based underwater sensor system designs should be a major part of the future scope for research. MEMs-based sensors will optimize the cost of materials and energy in particular. One other potential dimension of underwater biosensor development is the bio-inspired system approach, which uses the existing biological system around us. The oceans in particular are a rich depository of ideas and solutions ready for adaptation to our man-made systems. Bio-inspired sensor systems, like bio-inspired AUVs, are not just about trying to imitate the living system but to synergize the currently available technology with nature to optimize any given sensor design.

18.5 Physical Sensor Systems

Physical sensor technologies for energy, heat, water, salt, and particulate phases in general are more mature than either chemical or biological sensor technologies. A large industrial base exists for the manufacture of

temperature, pressure, and physical property sensors. Consequently, there are more examples for these types of sensors than those in other areas [18.28]. Classical depth or pressure measurements in the ocean (highlighted earlier in this section) are a direct example of the progression of physical sensors. These developments as state-of-the-art changes are more evolutionary and usually are not revolutionary steps in progress. Water column measurements are of utmost importance to better understand ocean processes by setting the normal background conditions and, for instance, are required to perform tsunami wave detection. A specific example is the work on the reinforced piezoresistive pressure sensor, which employed double diaphragms with a Wheatstone bridge. The double diaphragm design achieved higher sensitivity, a wider operating range, and better temperature compensation for ocean depth measurements compared to conventional single diaphragm piezoresistive designs [18.29].

For AUV mounted current metering, one cannot avoid the strength of acoustics to deliver flow information. Acoustic current Doppler profilers (ACDPs) are available currently for roving current observations. Modern ADCPs are also employed as acoustic Doppler velocimeters (ADVs) looking at over ground velocities, and when using a specific configuration the water velocities can be estimated while the vehicle is flying mid water. The advancement of AUV adaptive oceanographic sensor networks, combined with 2-D and 3-D imaging techniques, should be able to assist in mapping the fluxes across scales in environmental volumes and interfaces. For example, atmospheric turbulence sensing requires 3-D flow field measurements that are within 10 cm of the air–water interface; the sensors must be environmentally robust for persistent observations and be accurate to $0.1\,\mathrm{cm\,s^{-1}}$. Water column sensing requires 3-D flow field levels quantified within 5% throughout the water column to better understand the turbulent transport and flux of gases.

Transport in the water column includes chemicals as well as biomass in water and at the water–air interface. Combining optical and acoustics sensors on mobile platforms may provide some of the functionality needed for simultaneously measuring ocean chemistry and biology across a variety of scales. As physical sensors have advanced in moored ocean observatories over the last decade, one should see opportunities to transfer the sensing functions onto AUVs. Modifications will be needed in areas of computation, power, and packaging to accommodate the unique features of mobile platforms.

Instruments to quantify the relationship between turbulence in water coupled with turbulence in the air directly above would provide major changes in physical oceanographic observations. This requires surface sensors for fluxes through the sea surface interface. To accomplish these turbulence goals, new capabilities for AUVs that advance mapping of 3-D flow field (transport, fluid velocity) systems are one focus, as well as microturbulence sensors. Standardization for field physical sensing devices will be required on AUVs, just as it has been on observatories and drifters. Continued miniaturization in physical sensors, electronics, and packaging will be driven by new unmanned system designs. A new area of rising research interest is the discrimination of nanoparticles and aerosols, manmade and naturally occurring. Particle size distribution and shape sensors with near real-time response for mobile surveys are areas in need of development for AUV-based event response. Sensor systems must be able to accurately establish the sensor/AUV orientation within the environment. Physical location and altitude is the largest uncertainty in mobile hydrological observations as one example. As submersible robots become increasingly available, knowledge of their exact position in four dimensions will become more and more critical. An error of alignment within ± 0.1 degrees or 0.5 cm will substantially degrade the utility of bio-geochemical observations.

Multi-parameter systems are commercially available (DO, pH, ORP, chlorophyll, rhodamine, turbidity, conductivity, temperature, salinity, specific conductivity, depth) and inclusion of these is becoming the backbone of AUV deployments; we expect this to expand in the future. Mini bio-optics packages: chlorophyll, beam transmittance, angular scattering, conductivity, temperature, depth etc., are available and further expand the ocean optics and bio-optical data streams. Maps generated with these data parameters will provide richer physio-chemical information sets to real-time circulation models and fisheries management. Greater optical sensing capability can be expected in both the spectral and imaging realms. Using imaging technologies from micrometer to the meter scale is and will further improve details on particles of benthic biota and fish [18.30, 31].

Several instruments for optical detection are already available; this is not a complete list and new developments are underway. A number of existing instruments have already been successfully adapted to AUVs. Sensors for bioluminescence, particle counting and imaging are available off the shelf from a number of manufacturers.

The underwater bioluminescence assessment tool (UBAT) from Wet Labs provides measurements of mechanically stimulated bioluminescence of continuously flowing samples for potential bioluminescence in units of photon per second per liter. UBAT is a small

bioluminescence sensor specifically designed for deployment on multiple platforms, including shipboard profilers, AUVs, and long-term deployment on moorings. The photomultiplier tube (PMT) and flow meter that together comprise the core of the UBAT system are calibrated to the appropriate National Institute of Standards and Technology (NIST) specifications. The current UBAT system operates to depths of 600 m at present and uses mechanical stimulation, not unlike the wake of a ship, necessary to obtain maximum bioluminescence potential from bioluminescent organisms entrained in the flow. Stated sample rates of 60 Hz are specified with a data output rate of 1 Hz. Laser in-situ scattering and transmissometry (LISST) instrumentation include a series of submersible devices that are configured for a number of applications, including AUVs. All employ the small-angle forward scattering principle, also called low-angle laser scattering or laser diffraction. Sequoia Scientific is a manufacturer of several LISST laser sediment/particle/bubble/droplet sensing instruments, which can be purchased currently. Some of these instruments deliver the optical volume scattering function (VSF) of interest to optical oceanographers. One variant of the LISST instrumentation is the LISST-hologram (LISST-HOLO). The optics end of the LISST-HOLO consists of a 5 cm optical path using a red (658 nm) laser beam to traverse the optical path, overfilling a charge-coupled device (CCD) array, thereby creating a hologram. This instrument is particularly well suited to the measurement of large, complex biological particles. The LISST-HOLO creates in-situ images of particles, using in-line digital holography to overcome the depth of field issues other particle imaging systems are limited by. The digitally recorded holograms yield images of particles from 25 μm to 2.5 mm throughout the entire sample volume (≈ 2 ml). The sampling frequency is adjustable, with a maximum sampling frequency of 0.2 Hz or a hologram every 5 s. Others are considering use of the LISST-HOLO to augment sensing capabilities in interdisciplinary research applications using AUVs.

Laser optical particle counters (Fig. 18.6) (LOPCs) are another variation of particle profiling instruments. LOPC instruments can be purchased as off-the-shelf equipment configured for installation on an AUV from Odim Brooke Ocean. The operational principle of the LOPC is a combination of laser diode and line generator producing a divergent linear beam ≈ 1 mm in width. The beam is focused through a cylindrical lens producing a parallel beam subsequently reflected by a mirror and directed through an air–water interface window into the sampling volume. At a selected distance from the window, a mirrored prism is used to redirect the beam back to the window on a parallel path directly below the emerging beam. The light beam and optics of the LOPC are designed to address the probability of two or more particles occurring in the beam simultaneously and being counted as one particle. The reduction in coincidence probability is essentially the ratio of the two smallest detection areas. The LOPC utilizes a laser diode line generator producing a narrow beam that is broader in height that impinges on a linear array photodiode. The cross-sectional effective area of the beam results in a reduction of coincidence by a factor of 80×. This effectively increases the LOPC operational capability in large particle concentrations. The LOPCs high speed processing and improved plane provide detection counts at higher resolutions and higher concentrations with lower coincidence. The resolution achievable with the LOPC is a significant improvement over existing technologies, allowing an improved lower detection limit. The design of the LOPC instrument allows it to add on components and optional equipment.

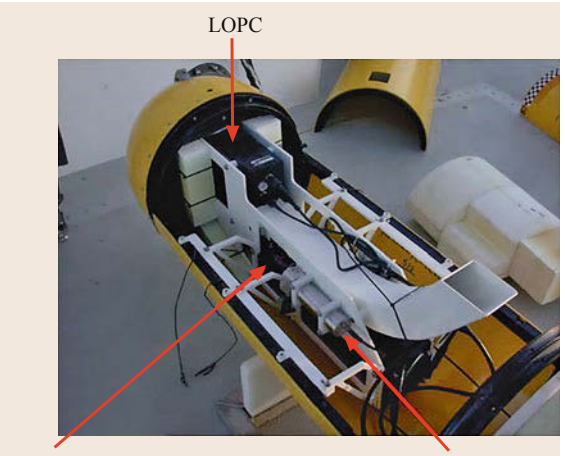


Fig. 18.6 AUV configured with laser optical particle counter and real-time flow imaging and classification (REFLICS) sensor concurrent with the embedded CTD system. Simultaneous deployment of both optical instruments permits geospatially defined particle quantitative and morphological characterizations (image courtesy of MBARI)

An accurate analysis of a large dynamic system like our oceans requires spatially fine and temporally matched data collection methods. The real-time flow imaging and classification (REFLICS) instrument offers higher resolution but at a lower frame rate than the LISST instruments. REFLICS is a method developed to estimate fish stock size from pelagic (marine) fish egg abundance taking point samples. REFLICS enhances

fish egg sampling by obtaining continuous, accurate information on fish egg abundance in the area of interest. REFLICS images the dynamic flow with a progressive-scan area camera and a synchronized strobe that provides backlighting. The strobe illumination allows the camera to acquire a sharp (no motion blur) image of the flow. Digitization and processing occur onboard using an image processing computer. REFLICS is capable of working as an integrated system built up with a number of modules, frequently a CTD, working together to achieve the objective of imaging, detecting, and classifying. REFLICS is also multi-platform capable and has been installed on AUVs. REFLICS is proving to be a cost-effective tool for oceanographers and marine ecologists who need high-resolution real-time data on fish egg distribution and abundance. REFLICS can be improved in several ways, such as integrating a classifier, to be more robust.

18.6 AUV-Based Physical Sensors – Horizons

Historically, physical sensor technology has matured at a faster pace than either chemical or biological sensor technologies. A large industrial base exists for the manufacture of temperature, pressure, and physical property sensors, and consequently there are more examples of these types of sensors on AUVs. However, current AUV-based physical sensor technology does not provide 3-D parametric physical information for water, sediment, and groundwater over spatial scales ranging from the micro to kilometer scale volumes. One particular topic in need of new sensors is fluid transport and fluid velocity for better characterization and understanding of the phenomenon within the associated fields of research. Specifically, there is a need for improved 1-D (one-dimensional) and 3-D flow field information in water and sediments. Additionally, there is also a need to improve the capability of AUV sensors to provide the proper orientation of the sensor in the environment; micro-turbulence measurements are an example of the need for further advancement in this area. Current developments in low power optical systems within the ultra-violet (UV) spectrum for particle detection and characterization are the type of technologies that may be an area of opportunity for the import of innovation in AUV oceanographic research. Discrimination of particles and nanoparticles, including the discrimination of non-biological and biological composition with particle size distribution and shape sensors, in particularly with near real-time response is an example of what is further expected to emerge in AUV sensors systems over time.

18.7 AUV-Chemistry Sensors – Horizons

The categorical space of chemistry is characterized by high level segmentation into classes of compounds and is further fragmented into a vast repository of individual compounds which elude detection by specific chemical sensors. Thus, unlike the physical sensor space, the chemical sensor space is highly fragmented and requires many different instrument configurations to access the variety of challenges associated with the field. A major factor in the lag of chemical sensor maturation relative to physical sensors is a continual issue. Consequently, there will be a continual call from the ocean research community to develop AUV-based chemical sensors for a wide range of compounds in the chemical space for organic, inorganic and biochemical samples within all environmental media, i. e., atmospheric aerosols, sediments and sub-surface waters in and around fresh and marine water systems. To date, there are few sensors that are capable enough to measure multiple chemicals at high acquisition rates, with suitable precision and sensitivity. There is a current need for sensors that can directly detect organic compounds of varying molecular mass along with the structural context, such as lattice cages. Small molecules include methane, halogenated hydrocarbons, and toxins. CO_2 sensors, either optical or chemical, and mass spectrometry are an example of sensing systems that have garnered substantial attention recently and have been placed on AUVs. Sensors having multi-element and/or chemical speciation capability at similar spatial and temporal resolutions will be highly desirable to elucidate environmental processes in 3-D. Multi-sensor packages for simultaneous physical and chemical parameters are essential to future AUV oceanographic research. It is reasonable to believe miniaturization will occur over time with widespread use driving investment from manufacturers. Sensors that can serve as chemical proxies for biological processes are also expected to emerge over time and to help the adoption rate of chem sensors on AUVs.

18.8 AUV-Based Biological Sensors – Horizons

Biological sensors to date have been developed for productivity measurements, species-specific detection, and organism mobility, among other targets. In order to measure biologic reactions and physiological function in ocean ecosystems, additional biosensors will need to emerge. Biological sensors can provide key real-time information on the function, structure, and composition of biologically influenced ocean ecosystems. AUV biosensors will evolve for monitoring all levels of biota, from microbiological (microscopy and imaging) to apex levels in the eukaryotic domain (passive acoustics tags). Future and existing technologies that need to be improved for AUV field deployment include: genomic-based sensors that identify community structure through phylogenetic fingerprinting methods or biological function by detecting gene expression and transcription; and proteomic-based sensors that measure protein expression patterns in biota. Metabolomic-based sensors that measure functional biota metabolic byproducts are a topic area ready for additional AUV-sensor systems development. Similarly to chemical sensors, the bio-space is largely fragmented and will require some revisions for specialized applications. Sensors that measure ocean biotic morphology, the shape and look of the organism in question, as well as sensors that provide reliable quantitative information related to growth rates are opportunities for transport of solutions from terrestrial life science microscopy innovations to the AUV sensing arena. New AUV sensor formats that detect the initiation and extent of primary production, as well as detection of predation levels and additional pathogens, will yield meaningful solutions in time. These would include invasive species and stress indicators along with toxicological responses.

18.9 Autonomous Sampling Systems – Extending Real-Time AUV Sensors

Higher content information of the ecosystems biotic layer is needed and desirable [18.31–33], along with simultaneous chemical and physical measurements. For those biological and chemical measurements that are not feasible by in-situ sensors (for example, isotope tracer measurements), modular high-resolution collection devices provide an approach to supply these difficult measurements. Configurable automated sampling/archiving modules on AUVs provide a solution that combines adaptive field sampling with subsequent analysis of acquired samples on established laboratory instruments using established assays. This sampling-AUV methodology enables complex chem/bio/physical measurements that compliment current and emerging AUV sensors.

Sampling technology will yield programmable systems capable of sampling with offline or online detectors while providing correlated high content information. The methodologies and automated systems can be targeted towards selected targets but with simple modification should be able to capture a wide range of intended chemical/biochemical/particulate material. One example of an adaptive sampling strategy would be the control of high end analytical fluidic sample processors triggered by a co-resident sensor that is commonly available – e.g., a nitrate sensor triggers sampling for isotopic analysis, or a bio-optical or physical sensor triggers the initiation of a confirmatory gene sampler.

This approach is a technology solution for persistent (temporal) and long-range (spatial) AUV derived automated/mobile/adaptive screening and sampling. It can be used for establishing baseline, responding, monitoring, and controlling ecosystems. This single technology can provide a range of adaptive and intelligent support for characterizing indicator chemistry, microbiology, and micro-particulate matter of an ecosystem, before, during, or after some time period. The sampling-AUV is based on three advanced technology components brought together into a unified approach:

1. Autonomous underwater vehicle
2. Adaptive real-time sensing module
3. Robotic fluidic sampler/archival unit.

Such a sensor-triggered robotic sampling system can either augment ship-based surveys or completely eliminate the need for expensive research vessel time while opening the access to the sea further to AUV technology.

18.10 Non-Acoustic Sensor Packaging

Non-acoustic sensors for ocean environment can be more highly affected by ocean deployment conditions than acoustic sensors (Table 18.2). Harsh environment salts, corrosion, vibration-motion, and biofouling will

Table 18.2 Listing of environmental problems and associated solutions for packaging of non-acoustic sensors

Marine environment problem	Approach/solutions
Rapid temperature change	Temperature isolation and compensation
Corrosive medium	Use of ceramics, special plastics, regenerable surfaces
High pressure	Spherical packages: pressure compensation
Bio-fouling	Acid reservoirs and pumps, Cu plating, regenerable surfaces
Energy source-volume limited	Seawater-Al cells for electrical energy
Ingestion	Inert outer casing, no strings or probes
Limited sample concentration	Bio-amplification, plate-strip-detect
Limited radio frequencies (RF) or optical communications range	Acoustic (piezo-based) communication
Limited reagent reservoir	On-board synthesis, microchem. reactors
Particulate concentrations/clogging	Bio-filtering/expendable samplers

affect non-acoustic sensors to a greater degree. Space efficiency is mandated for high packing density of mechanisms and minimal reagent storage, thus 3-D design is important. Heterogeneous materials are typically required for the complex systems: metals, ceramics, polymers, glasses, and composites and synergistic interferences such as electrical-RF-noise, motors, and light sources all become complications in non-acoustic sensor design. Lastly, the internal volume of the pressure vessels that house non-acoustic sensors in many instances also require some gases or solids to maintain a fixed air quality inside the pressure vessels.

18.11 The Essential Need for Sensors

Sensor-AUV-based observations combine the estimation of spatial-temporal distributions and trends for targeted environmental variables. From the sampling perspective, the introduction of new sensor types enables new observational vistas and truer representation of the water column dynamics. Sensors can further help influence the choice of sensing locations and dynamic redeployment of AUV monitoring resources in response to new observations and predictions. Consequently, sensors guide the selection and repositioning of vehicles to improve the estimation of spatially distributed fields under resource and environmental constraints.

Historically, acoustic sensors and non-acoustic optical sensing have been the primary and secondary development areas for sensor-AUV systems beyond physical sensors such as temperature and salinity. Greater space–time coverage and (chem-bio-physical) composition knowledge of the ocean is needed and new non-acoustic sensor technologies will be necessary to address this need.

The state of the ocean faces complex stressors and challenges to its valuable resources [18.34]. These include deterioration of water column and surface water quality, the emergence and increased intensity and duration of events such as harmful algal blooms and hurricanes, microbiological contamination of beaches, loss of reef habitats, and chemical contamination resulting from water usage and coastal runoff. There is a need to understand the processes that are responsible for these changes and to provide potential strategies that might mitigate these problems. Ocean-engineered AUV-based sensing and sampling systems can assist in the effort [18.35]. Automated mobile sensing and mapping of the oceans will have to circumvent the unique challenges related to the heterogeneity of this domain. Further non-acoustic sensors, automatic analyzers, and in-situ monitoring systems (and the associated deployment apparatus) are in an ongoing state of development, field deployment, and validation. However, AUVs augmented with acoustic and non-acoustic chemical/biological and physical sensors continue to be the leading candidate in platform technology to quantify and understand the true composition of the ocean's and its sustainability.

References

18.1 M. Pilson: *An Introduction to the Chemistry of the Sea* (Prentice-Hall, Upper Saddle River 1998)

18.2 O. Kinne: The effects of temperature and salinity on marine and brackish water animals, Oceanogr. Mar. Biol. Annu. Rev. **2**, 281–339 (1964)

18.3 G.W. Boehlert, D.P. Costa, D.E. Crocker, P. Green, T. O'Brian, S. Levitus, B.J. Le Boeuf: Autonomous pinniped environmental samplers: Using instrumented animals as oceanographic data collectors, J. Atmos. Ocean. Technol. **18**, 1882–1893 (2001)

18.4 A. Bradshaw, K. Schleicher: Electrical conductivity of seawater, IEEE-JOE **5**(1), 50–62 (1980)

18.5 H. Broadbent, S. Ivanov, D. Fries: Development of underwater PCB MEMS conductivity and temperature sensors for environmental measurements, J. Micromech. Microeng. **17**, 722–729 (2007)

18.6 K.L. Daly, R.H. Byrne, A.G. Dickson, S.M. Gallagher, M.J. Perry, M.K. Tivey: Chemical and biological sensors for time-series research: Current status and new directions, Mar. Technol. Soc. J. **38**, 121–143 (2004)

18.7 G. Mills, G. Fones: A review of in situ methods and sensors for monitoring the marine environment, Sens. Rev. **32**(1), 17–28 (2012)

18.8 P.M. Vitousek, H.A. Mooney, J. Lubchenco, J.M. Melillo: Human domination of Earth's ecosystems, Science **277**(5325), 494–499 (1997)

18.9 L.R. Adornato, E.A. Kaltenbacher, D.R. Greenhow, R.H. Byrne: High-resolution in situ analysis of nitrate and phosphate in the oligotrophic ocean, Environmental science technology **41**(11), 4045–4052 (2007)

18.10 J. Buffle, G. Horvai (Eds.): *In situ monitoring of aquatic Analysis and Speciation*, Analytical and Physical Chemistry of Environmental Systems IUPAC systems-Chemical, Vol. 6 (Wiley, Chichester 2000)

18.11 M. Varney (Ed.): *Chemical Sensors in Oceanography* (Gordon and Breach, Amsterdam 2000)

18.12 D.P. Fries, R.T. Short, L.L. Langebrake, R.J.W. Patten: In water field analytical chemistry: Mass spectrometers, robots, numerical models for wide area chemical plume localization, Field Anal. Chem. Technol. **5**(3), 121–130 (2001)

18.13 J. Kaiser, M.K. Reuer, B. Barnett, M.L. Bender: Marine productivity estimates from continuous O_2/Ar ratio measurements by membrane inlet mass spectrometry, Geophys. Res. Lett. **32**, L19605 (2005)

18.14 S.D. Wankel: New constraints on methane fluxes and rates of anaerobic methane oxidation in a Gulf of Mexico brine pool via in situ mass spectrometry, Deep Sea Res. II **57**(21–23), 2022–2029 (2010)

18.15 K.C. Hester, R.M. Dunk, S.N. White, P.G. Brewer, E.T. Peltzer, E.D. Sloan: Gas hydrate measurements at hydrate ridge using Raman spectroscopy, Geochim. Cosmochim. Acta **71**, 2947–2959 (2007)

18.16 S.N. White, P.G. Brewer, E.T. Peltzer: Determination of gas bubble fractionation rates in the deep ocean by laser Raman spectroscopy, Mar. Chem. **99**, 12–23 (2005)

18.17 X. Zhang, K.C. Hester, W. Ussler, P.M. Walz, E.T. Peltzer, P.G. Brewer: In situ Raman-based measurements of high dissolved methane concentrations in hydrate-rich ocean sediments, Geophys. Res. Lett. **38**(8), L08605 (2011), doi:10.1029/2011GL047141

18.18 R.M. Dunk, E.T. Peltzer, P. Walz, P.G. Brewer: Seeing a deep ocean CO_2 enrichment experiment in a new light: Laser Raman detection of dissolved CO_2 in seawater, Environ. Sci. and Technol. **39**, 9630–9636 (2005)

18.19 D.B. Nuzzio, E.R. Zettler, A. Aguilera, L.A. Amaral-Zettler: A direct in situ fingerprinting method for acid rock drainage using voltammetric techniques with a single renewable gold microelectrode, Sci. Total Environ. **409**(10), 1984–1989 (2011)

18.20 M. Roberts, S. Mannsfeld, N. Queralto, C. Reese, J. Locklin, W. Knoll, Z. Bao: Water-stable organic transistors and their application in chemical and biological sensors, Proc. Natl. Acad. Sci. USA **105**(34), 12134–12139 (2008)

18.21 T. Asefa, C.T. Duncan, K.K. Sharma: Recent advances in nanostructured chemosensors and biosensors, Analyst **134**(10), 1980–1990 (2009)

18.22 C. Scholin, G. Doucette, S. Jensen, B. Roman, D. Pargett, R. III Marin, C. Preston, W. Jones, J.C. Feldman, A. Everlove, N. Harris, E. Alvarado, J. Massion, D. Birch, R. Greenfield, C. Vrijenhoek, K. Jones: Remote detection of marine microbes, small invertebrates, harmful algae and biotoxins using the Environmental Sample Processor (ESP), Oceanography **22**, 158–167 (2009)

18.23 D. Fries, J. Paul, M. Smith, A. Farmer, E. Casper, J. Wilson: The autonomous microbial genosensor – An in situ sensor for marine microbe detection, Microsc. Microanal. **13**, 514–515 (2007)

18.24 C.A. Scholin, G.J. Doucette, A.D. Cambella: Prospects for developing automated systems for in situ detection of harmful algae and their toxins. In: *Real-time Coastal Observing Systems for Marine Ecosystem Dynamics and Harmful Algal Blooms*, Oceanographic Methodology, ed. by M. Babin, C. Roesler, J. Cullen (UNESCO Publishing, Paris 2008) pp. 413–462

18.25 L. Su, W. Jia, C. Hou, Y. Lei: Microbial biosensors: A review, Biosens. Bioelectron. **26**(5), 1788–1799 (2011)

18.26 K.G. Sellner, G.J. Doucette, G.J. Kirkpatrick: Harmful algal blooms: Causes, impacts and detection, J. ind. microbiol. biotechnol. **30**(7), 383–406 (2003)

18.27 S.H. Cho, J.M. Godin, C.H. Chen, W. Qiao, H. Lee, Y.H. Lo: Review article: Recent advancements in optofluidic flow cytometer, Biomicrofluidics **4**(4), 043001 (2010)

18.28 M. Luther: Status of sensors for physical oceanography measurements, Mar. Technol. Soc. J. **42**(1), 84–92 (2008)

18.29 S. Aravamudhan: Reinforced piezoresistive pressure sensor for ocean depth measurements, Sens. Actuators A. Physical. **142**(1), 111–117 (2009)

18.30 R. Holman, J. Stanley, T. Ozkan-Haller: Applying video sensor networks to nearshore environment monitoring, IEEE Pervasive Comput. **2**(4), 14–21 (2003)

18.31 D.C. Little, J. Konrad, P. Ishwar: A wireless video sensor network for autonomous coastal sensing, Proc. Conf. Coast. Environ. Sens. Netw. (2007)

18.32 H.T. Odum, E.C. Odum: *Modeling for All Scales* (Academic, San Diego 2000)
18.33 T.M. Straub, D.P. Chandler: Towards a unified system for detecting waterborne pathogens, J. Microbiol. Methods **53**(2), 185–197 (2003)
18.34 M.L. Parry, O.F. Canziani, J.P. Palutikof, P.J. van der Linden, C.E. Hanson (Eds.): *Climate Change 2007: Impacts, Adaptation and Vulnerability* (Cambridge University Press, Cambridge 2007)
18.35 B.M. Howe, Y. Chao, P. Arabshahi, S. Member, S. Roy, T. Mcginnis, A. Gray: A smart sensor web for ocean observation: System design, architecture, and performance, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **3**(4), 507–521 (2010)

19. Cooperative Vehicle Environmental Monitoring

Naomi Ehrich Leonard

This chapter reviews cooperative control of autonomous vehicles for environmental monitoring with a focus on methodologies that have been designed, deployed, and proven to provide efficient, reliable, and sustained monitoring of the uncertain and inhospitable ocean environment. Vehicles that communicate their state or measure the relative state of other vehicles in the team can cooperate by using feedback control to coordinate their motion as a mobile, reconfigurable sensor array, responding efficiently to changing signals, scales, and conditions in the environment. In a variety of contexts, a vehicle team with judiciously designed cooperative control can outperform the same team with each vehicle controlled independently. For example, cooperative control methodologies have been developed to improve the richness of information in the data that the vehicles collect, their accuracy in feature detection and tracking, and the robustness of their decisions to uncertainty and failures. The chapter begins with a survey of early work on ocean sampling and environmental monitoring, cooperative control, and collective motion. The theory, methodology, and field deployment are then highlighted for two projects on cooperative vehicle monitoring in the coastal ocean that demonstrated the applicability and associated performance advantages of cooperative control. The chapter concludes with a presentation of more recent developments as well as future directions in cooperative vehicle environmental monitoring.

19.1 **Motivation** 441
19.2 **Background and History** 443
19.3 **Advances in Cooperative Vehicle Ocean Monitoring** 445
19.3.1 Cooperative Gliders in AOSN II 445
19.3.2 Cooperative Gliders in ASAP 448
19.4 **Recent Developments and Future Directions** 452
References 454

19.1 Motivation

Over the last decade, methodologies for automated cooperative control of robotic vehicles have been designed, deployed, and proven to provide efficient, reliable, and sustained monitoring of the uncertain and inhospitable ocean environment. Unprecedented data sets have been collected from deployments of cooperative vehicles in the field, and both real-time and post-deployment analyses have led to new understanding of the environment. This first decade of success in cooperative vehicle environmental monitoring sets the stage for new opportunities and future gain, especially as the development of cooperative control methodologies can continue to leverage ongoing technological and scientific advances in underwater communication and sensing, energy and computational efficiency, vehicle size, speed, maneuverability and cost, and ocean modeling and prediction.

Indeed, the demonstrated potential of cooperative vehicle control has led to increased demand for fleets of autonomous underwater vehicles (AUVs) for use in measuring ocean physics, biology, chemistry, and geology to improve understanding of natural dynamics and human-influenced changes in the marine environment. Further, methodologies for cooperative control of robotic vehicles in the ocean are readily adaptable to applications on land, in the air and in space; likewise, there is much to be learned from developments in these other domains. The recent explosion in research on networks and complex systems, including investigation of mechanisms that explain a *collective intelligence* exhibited by animal aggregations on the move, are also being leveraged to advance design of cooperative vehicle dynamics.

For environmental monitoring to be successful, physical, chemical, and biological variables must be

measured across a range of spatial and temporal scales; in the ocean the monitoring strategy must also contend with a harsh, three-dimensional physical space that is highly uncertain and dynamic. Small spatial and temporal scales associated with the measured variables typically make a stationary sensor array impractical because a very large number of sensors would be needed to get sufficient resolution in space and/or time. An array of mobile sensors, however, may be very well suited to such a challenge since mobility can be exploited to dynamically distribute fewer sensors according to the spatial and temporal scales.

The underlying principle of cooperative control of vehicles for environmental monitoring leverages mobility of sensors and uses an interacting dynamic among the individual sensors to yield a collective behavior that performs better than the sum of the parts. If the vehicles can communicate their state or measure the relative state of others in the team, then they can cooperate and the cooperative vehicle dynamics can provide coordinated motion of the team as a whole. The resulting vehicle network functions as a dynamically reconfigurable sensor array with a capability for high performance in environmental monitoring not available at the level of individuals. High performance has been demonstrated with cooperative vehicle groups in the ocean in terms of richness of information in measurements, accuracy in feature detection and tracking, and robustness of decisions to uncertainty and failures.

Methodologies for systematic generation of feedback control laws that yield provable collective dynamics have been critical to the successful design of high-performing cooperative vehicle networks. Feedback control refers here to the automated changes that each vehicle makes in response to its measurements of the sampled fields, the relative state of other vehicles in the network, and any additional available measured or computed signals.

Consider, for example, the task of tracking high-density phytoplankton patches in the ocean with a team of AUVs that carry optical sensors for measuring phytoplankton concentration. Suppose also that when any two vehicles in the team are not separated by too great a distance (call them neighbors), they can measure or communicate to each other their relative position and/or velocity. Then, each vehicle can apply a feedback control law, at its sampling or communication frequency, that moves it in the direction of a combination of its best estimate of:

1. The direction of increasing concentration
2. The direction toward its neighbors that are farther than a prescribed separation distance
3. The direction away from its neighbors that are closer than a prescribed minimal separation distance

In ideal conditions, the vehicles will move as a regularly spaced array up the phytoplankton concentration gradient; in real conditions, feedback will provide robustness to noise, uncertainty, and disturbance within bounds. An augmentation to this feedback law to further reduce error due to noisy measurements has each vehicle compute a local estimate of the optimal vehicle array resolution for gradient climbing accuracy and then adapt the prescribed separation distances between neighbors to achieve this resolution.

Consider, as another example, the task of providing dynamic sampling coverage of the changing phytoplankton patches over a fixed region with this same team of AUVs. The goal is to enable the vehicles to efficiently sample the patches across the region so that the data can be used to map the phytoplankton patches with minimal mapping error. Each vehicle can apply a feedback control law in response to where its neighbors are collecting data and in accordance with priors on spatial and temporal scales associated with the phytoplankton patches. The feedback control law moves it toward a location that is easily accessible, is away from others, and has not been recently sampled. In this case in ideal conditions, the vehicle network will cooperatively perform dynamic sampling coverage of the patches over the region; in real conditions, feedback will provide robustness to some noise, uncertainty, and disturbance. An augmentation to this feedback law has each vehicle compute a local estimate of changing spatial and temporal scales to update its priors and adapt how far from other vehicles to move and how frequently to re-sample previously sampled locations.

Feedback has also been used to *close the loop* between cooperating vehicle networks and advanced ocean models when data collected by the vehicle network can be made available for assimilation into the ocean models and the ocean model predictions can be made available to one or more of the vehicles. In the first example of tracking high-density phytoplankton patches, the feedback with ocean models allows individual vehicles to modify their gradient climb based on predictions of high-density locations. In the second example of dynamic sampling coverage in a fixed region, the feedback with ocean models provides enhanced estimates of uncertainty so that individuals could bias their motion toward sampling locations with the greatest possibility of minimizing uncertainty in the mapping. Forecasts of ocean currents are also useful for navigation.

Cooperative feedback control makes possible a vehicle network that is autonomous, versatile, and robust to noise and uncertainty. Further, when each vehicle has the same feedback law, the vehicle network has an added robustness to vehicle failures or additions since an ordering of the vehicles is unnecessary and in particular the system does not depend on any special individuals. Other promising opportunities have been explored with a heterogeneous group of vehicles. In these cases, feedback can be used for cooperative subtask allocation or coordinated complementary actions: for example, slow-moving autonomous underwater gliders can provide coverage and fast-moving propelled underwater vehicles can be allocated to relay information and to move to *hot spots* at great speed. Similarly, unmanned aerial vehicles (UAVs) can provide large-scale mapping of sea-surface fields and AUVs can complement with in-depth feature tracking below the surface. In this chapter, we describe two projects in cooperative vehicle environmental monitoring with extensive field deployment in the coastal ocean that demonstrated for the first time at large scale and over several weeks the applicability and associated performance advantages of cooperative control methodologies for mobile sensor networks in the ocean. The first of these was the autonomous ocean sampling network (AOSN) II project [19.1, 2] with its field experiment in Monterey Bay, CA, over the month of August 2003 and the second was the adaptive sampling and prediction (ASAP) project with its field experiment in Monterey Bay, CA, over the month of August 2006 [19.3].

For AOSN-II, a methodology was designed [19.4] and demonstrated [19.5] that featured small networks of autonomous underwater gliders cooperating to compute temperature gradients and track cold upwelled water. The gliders were also used to demonstrate autonomous coordination of their motion to sample along the path of drifters and thus increase measurement density along fronts. For ASAP, a methodology was designed [19.6] and demonstrated [19.3] that featured a network of 10 gliders (of two types) to optimize dynamic sampling coverage, minimizing uncertainty for estimation of temperature, salinity, and currents in a large coastal region just north of Monterey Bay. In both field experiments, three numerical ocean models ran in near real time, assimilating data collected by the gliders and providing estimates and predictions that were used in the gliders' adaptive motion planning. Additional real-time data were provided by a research aircraft, satellite imagery, high-frequency radar, moorings, drifters, and propelled vehicles. In the ASAP experiment, the cooperative behavior of a network of six gliders ran autonomously without failure for almost 24 days straight.

An important factor in the success of the AOSN-II and the ASAP projects was the strong multidisciplinary collaboration among researchers with expertise in ocean science, vehicle dynamics, and control and dynamical systems theory. The methodologies developed drew inspiration from earlier work in ocean sampling and environmental monitoring, cooperative control, and collective motion; we present background and history on these subjects in Sect. 19.2. The AOSN-II and ASAP programs in cooperative vehicle ocean monitoring, from theory through full-scale ocean deployment, are reviewed in Sect. 19.3. More recent developments and future directions in cooperative vehicle environmental monitoring are described in Sect. 19.4.

19.2 Background and History

The autonomous oceanographic sampling network (AOSN) was introduced in 1993 by *Curtin* et al. [19.7] as an approach for dynamic measurement of the ocean environment and resolution of spatial and temporal gradients in the sampled fields. At that time most oceanographic data were collected from satellites, ships with towed underwater profilers, and arrays of moorings and floats. The AOSN concept was to deploy AUVs to take measurements that would complement those from distributed point sensors such as moorings and from remote sensors such as satellites, and enable adaptive sampling to improve forecast skill; the AOSN system would operate successfully with the use of acoustic and radio modems for communication and docking stations for recharging AUV batteries. It was anticipated that control would be critical for the AUVs to accomplish complex missions in the presence of uncertainties and real-world constraints, and that coordinated control of the multiple vehicles would lead to system efficiency and endurance. *Bellingham* described in [19.8] how nested approaches would allow sampling the ocean over a range of spatial and temporal scales.

The vision of an integrated ocean monitoring system was made possible with the development of small, relatively inexpensive AUVs. Propelled AUVs such as the autonomous benthic explorer (ABE) [19.9], the Odyssey [19.10], and the REMUS [19.11] were designed to provide maneuverability and speed. Buoyancy-driven autonomous underwater gliders, including the Slocum [19.12], Spray [19.13]

and Seaglider [19.14], were designed to provide endurance [19.15, 16]. ABE was used to map the sea bottom and to search out and study deep-sea hydrothermal vent sites and volcanoes [19.17–20]. The Odyssey vehicle was used in experiments under the ice in the Arctic [19.21]. REMUS was equipped with optical sensors and used to measure bioluminescence [19.22, 23]. Seagliders were sent on 5-month-long missions to measure physics, biology and chemistry off the coast of Washington and in the Labrador Sea [19.15]. Slocum gliders were introduced as an integral part of the long-term ecosystem observatory (LEO), an integrated observatory off the coast of New Jersey [19.24–26].

An early example of a control architecture design for multiple AUVs was described in 1987 by *Albus* and *Blidberg* in [19.27]: the architecture was designed to enable two AUVs to perform cooperative search, approach, and mapping using cooperative maneuvers such as *fly-formation*, *circle-split-and-rendezvous*, and *leader-follower*. In [19.28] the objective was to use multiple AUVs as an imaging system. Virtual chains of AUVs were considered by *Triantafyllou* and *Streitlien* in [19.29]; a technology for one vehicle to track another in a chain-like fashion using an ultrashort baseline (USBL) acoustic tracking system was demonstrated by *Singh* et al. [19.30] in Buzzards Bay off Woods Hole, MA in March 1996. In June 1996, two Odyssey vehicles were used along with an acoustic tomography network for mapping in the Haro Strait region of British Columbia [19.31]. *Stilwell* and *Bishop* presented a decentralized control framework for a cooperative platoon of AUVs in [19.32]. Formation flying to map salinity fronts was tested in the North River in North Carolina using the Ranger micro-AUV [19.33].

A 1997 survey by *Cao* et al. [19.34] described cooperative mobile robotics as a still emerging field rich with opportunities. The robotics community became heavily engaged in the late 1980s first with a focus on simulation [19.35, 36] and then physical implementations [19.37, 38]. Collective behavior in animal groups was an early inspiration [19.39], with a behavioral-based approach becoming popular in the 1990s [19.40]. However, according to [19.34] in 1997, *few applications of cooperative robotics* had been reported, and supporting theory was *still in its formative stages*.

Interest in very large-scale stationary sensor networks surged with advances in wireless communication technology and microsensors, and environmental monitoring was an early, important application driver for development of network architectures and algorithms in this context [19.41–44]. Energy considerations were used to justify the use of large numbers of stationary sensors over mobile sensors. However, the balance tipped the other way in the case of undersea sensing due to the challenges of undersea communication and the emergence of relatively inexpensive, high-endurance vehicles such as the autonomous underwater gliders.

Bretherton et al. [19.45] in the 1970s applied the technique of objective analysis (OA) [19.46], which uses classical linear estimation theory to compute objective maps, to address the problem of deployment design for an array of stationary or passively drifting sensors in the ocean where the aim was to provide coverage and minimize uncertainty in the estimates made from the data collected. Adaptive ocean sampling, as in [19.47], built on this concept to consider the design of a trajectory for a mobile sensor platform, complementing another sensor platform moving along a predetermined track, in order to minimize uncertainty in an ocean forecasting model. A performance metric that accounts for both spatial and temporal sampling requirements was derived by *Wilcox* et al. [19.48]; it was used to evaluate oceanographic survey performance with AUVs in [19.49]. A methodology for control of multiple sensor platforms based on information theory was presented in [19.50] and sampling strategies driven by distributed parameter estimation were described in [19.51, 52].

Motivated by the many potential applications and the rich theoretical possibilities, researchers in the control theory community began a significant effort in the early 2000s to use systems theoretic approaches to design and study cooperative control. Artificial potentials presented an attractive methodological basis for cooperative control of network formations [19.53–58] both because convergence and performance could be proved using Lyapunov stability theory (see early work on robot navigation and obstacle avoidance [19.59, 60]) and because control laws derived from artificial potentials resembled the distributed, cohesive, and repulsive forces used to model animals that move together [19.61, 62].

Artificial potential methods were also used to design cooperative gradient climbing strategies for a group of vehicles that could each only take a scalar measurement at a time of the field of interest (e.g., ocean temperature) [19.63]. Capitalizing on this idea and building on the methods of [19.55, 56], *Ögren* et al. [19.4] developed a provable methodology to control the shape of the formation as well as the rotation, translation, and expansion of the formation [19.64]; this was used to design control strategies for a network of vehicles to adaptively climb gradients in the sampled field and thus robustly find peaks (Sect. 19.3.1 below reviews the implementation of this methodology in the field). These ideas were extended further by *Zhang* and *Leonard* [19.65, 66] to design provable control laws for cooperative level set tracking, whereby small

vehicle groups could cooperate to generate contour plots of noisy, unknown fields, adjusting their formation shape to provide optimal filtering of their noisy measurements. Related work addressed environmental boundary tracking [19.67, 68], coverage control [19.69, 70], target tracking [19.71, 72], and maximization of information [19.50].

Researchers in control also took a strong interest in the dynamics of consensus within a network; the topology of the sensing and communication interconnections among agents was encoded using graphs, and the convergence of consensus dynamics was proved with approaches that exploited graph theory allowing for time-varying communication graphs and time-delayed communications [19.73–77]. Consensus in the positions of agents was used to address a variety of other problems including formations [19.78] and rendezvous [19.79].

In most of this consensus literature the dynamics are linear, and yet the problem of consensus on direction of motion that mobile robots must make is nonlinear, since the space of directions in the plane is a circle (and not a line). Consensus on the circle is called synchronization, and it has been studied extensively in the context of coupled phase oscillators used to model a variety of interconnected periodic processes in biology and physics (firefly flashing and neuron firing) [19.80, 81]. *Justh* and *Krishnaprasad* developed a geometric framework to design steering control laws to coordinate the motion of vehicles in [19.82]. This approach was generalized in the work of *Sepulchre* et al. [19.83, 84] using a model that extends coupled oscillator dynamics, in which the phase of each oscillator represents the direction of motion of a vehicle, to include the spatial dimensions, which represent the positions of the vehicles. These works provided a systematic methodology for designing provable, distributed control laws that stabilize motion patterns in the plane ([19.85–87] for 3-D); each vehicle uses a feedback law that depends only on what limited measurements it can make, and the controlled system can cope with a time-varying communication network and with real-time changes in the number of vehicles in the group. Because the methodology is systematic and robust and because distributed control of vehicle motion patterns is central to environmental monitoring, the methodology of [19.83, 84] was developed into an adaptive sampling methodology for mobile sensor networks in the ocean [19.6] (Sect. 19.3.2 below reviews the implementation of this methodology in the field).

19.3 Advances in Cooperative Vehicle Ocean Monitoring

The AOSN II and ASAP projects were driven by an interest in developing sustainable, portable, adaptive ocean observing and prediction systems for use in coastal environments. The projects used cooperating AUVs carrying sensors to measure the physics and biology in the ocean together with advanced ocean models in an effort to improve the ability to observe and predict ocean dynamics. A central focus was on reliable, efficient and adaptive coordinated control strategies for mobile sensor platforms to collect data of high value. Both the AOSN II and ASAP experiments were designed to bring together new techniques in sensing, forecasting and coordinated control; see [19.88] for a summary of goals and progress. The 2003 AOSN II experiment brought these techniques together for the first time, yielding an unprecedented data set. The 2006 ASAP experiment fully integrated these techniques to even greater benefit, demonstrating their potential in a versatile and high-performing adaptive coastal ocean observing and prediction system. The methodologies derived, integrated, and demonstrated are adaptable to a wide variety of environmental monitoring problems and settings.

19.3.1 Cooperative Gliders in AOSN II

In summer 2003, a multidisciplinary research group as part of the AOSN II project produced an unprecedented in-situ observational capability for studying upwelling features in Monterey Bay over the course of a month-long field experiment [19.1, 2]. A highlight was the simultaneous deployment of more than a dozen, sensor-equipped, autonomous underwater gliders [19.15], including 5 Spray gliders (Scripps Institution of Oceanography, La Jolla, CA) and up to 10 Slocum gliders (Woods Hole Oceanographic Institution (WHOI), Woods Hole, MA). Autonomous underwater gliders are high endurance, buoyancy-driven vehicles that move up and down in the ocean by controlling their net buoyancy using pumping systems. Their fixed wings and tail give them lift, which helps them make forward progress by following sawtooth-shaped trajectories. To control their attitude, gliders actively redistribute internal mass. The Slocum uses a rudder to control heading, and the Spray shifts mass to the side to roll, bank, and turn. During the field experiment, the gliders were operated to achieve a fixed velocity relative to the flow. Their effec-

tive forward speed was of the same order as the stronger currents in and around Monterey Bay: $\approx 25\,\mathrm{cm\,s^{-1}}$ in the case of Spray and $35\,\mathrm{cm\,s^{-1}}$ in the case of Slocum. Thus, when the currents were too strong, the gliders did not make progress if they were moving against the currents.

As part of the field experiment, sea trials were run with groups of three Slocum gliders controlled into triangular formations [19.5]; see also [19.6] for a survey. Feature-tracking capabilities of the glider formations were demonstrated under the challenging conditions of limited communication and limited feedback as well as a strong flow field. Two sea trials tested strategies for cooperative motion control and cooperative gradient estimation for the gliders at relatively small scales in the region, i. e., on the order of 3 km. In a third sea trial, a Slocum glider was used to track a Lagrangian drifter in real time and collect data in a volume surrounding the path of the drifter. This demonstrated the utility of a glider, and the possibilities for a network of gliders, to track Lagrangian particle features such as a water mass encompassing an algal bloom [19.5].

The Slocum gliders were operated to 200 m depth and were deployed far enough from the coast to avoid shallower water. Each glider surfaced every couple of hours, although the gliders did not surface synchronously. At the surface, each glider got a global positioning system (GPS) fix for navigation, and using Iridium satellite and ethernet, sent back to shore the data it had collected and received updated mission commands from the shore computer. Since the gliders were not equipped with underwater communication, the communication with the shore computers was the only means for (indirect) communication between gliders.

The strategy for coordinated motion control was based on the *virtual body and artificial potential* (VBAP) methodology for control of multiple vehicles described in [19.4]. VBAP is a general coordinated control strategy that stabilizes the translation, rotation, and expansion of a formation of autonomous vehicles; it is especially well suited to missions that require a changeable mobile sensor array such as gradient climbing in a scalar, sampled field, where expansion and contraction of the network modify the resolution of the array. The virtual body refers to a collection of moving reference points, each with dynamics that are computed centrally and made available to the autonomous vehicles. Spring-like control forces for the vehicles, and the virtual body, are derived from artificial potentials between the vehicles and the virtual body; they are designed to stabilize the dynamics of the vehicles and the virtual body into a formation. In the computation of its control law, each vehicle uses a measurement of the relative position of neighboring vehicles and nearby reference points on the virtual body.

The local gradients of a sampled scalar field can be estimated on-board or centrally, if the samples taken on-board the vehicles can be communicated among them or to a central computer. The group will control its motion in the direction of the gradient if the virtual body dynamics are designed to move it in the direction of the gradient; for example, if the dynamics of the virtual body move it toward the coldest water, as determined from an estimated temperature gradient, the vehicle group will move toward the coldest water as well since it moves with the virtual body in formation. In the VBAP methodology, stability and convergence of the vehicle formation is guaranteed with a feedback control on the speed of the virtual body.

Ideal assumptions, including continuous communication and feedback for the autonomous vehicles, were used to prove the control theory and algorithms described in [19.4]. Thus, in order to make the control methodology applicable to the conditions of the cooperative control sea trials of 2003, the VBAP control methodology was modified in a few key ways. For example, the algorithms were integrated with the on-board glider waypoint tracking routine and adjusted to accommodate the constant speed of the gliders, the high-speed ocean currents, the asynchronous surfacings of the gliders, as well as other latencies [19.89]. The approach was later systematized as part of the ASAP project in a fully automated software suite, the glider coordinated control system (GCCS) [19.90]. The GCCS automates the decentralized coordinated control methodology of [19.6] for adaptive sampling motion patterns.

The main idea behind the integration of VBAP with waypoint control was to use VBAP to produce waypoint lists, corresponding to coordinated glider trajectories, that the gliders would then follow using their on-board waypoint control [19.5, 89]. Accordingly, VBAP was run in a *planning* mode using a simulation of the gliders with initial conditions defined by the gliders' most recent GPS fixes and average flow measurements. The VBAP-planned trajectories were discretized into waypoint lists. Each Slocum glider uploaded its waypoint list when it surfaced, and followed those waypoints for the subsequent 2 h until its next surfacing.

For the Slocum vehicles, a waypoint is prescribed as a vertical cylinder in the ocean since it refers to a position in the horizontal plane and a radius that sets how close the vehicle should come to the waypoint position. Thus, when a sequence of waypoints is prescribed, the glider follows the waypoints by passing through each of the corresponding cylinders in the prescribed sequence. To follow a sequence of waypoints, each glider uses its on-board low-level heading control which depends

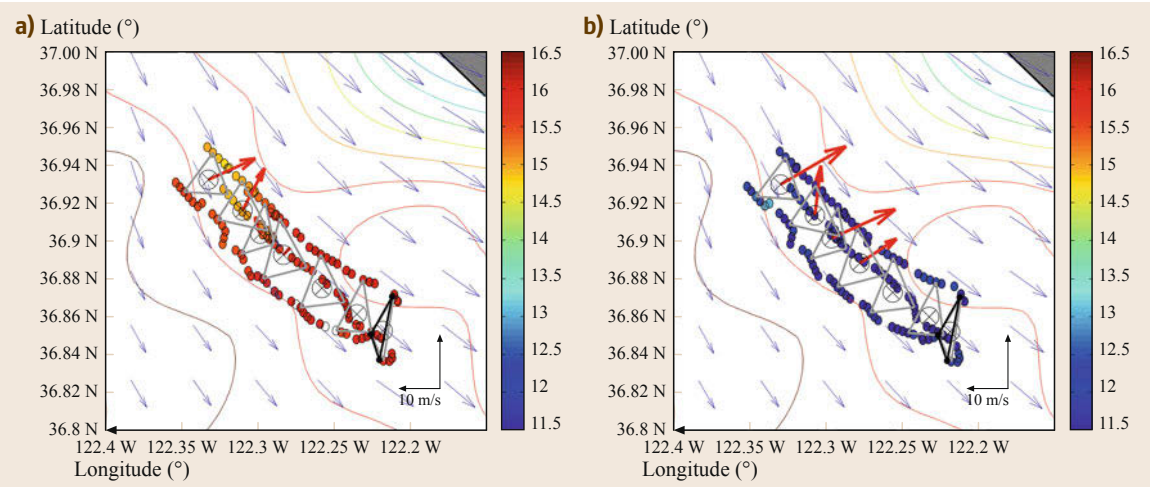


Fig. 19.1a,b From [19.2]. Three gliders moving northwest in triangular formation (*gray lines*) from 18:00 GMT August 6 into late morning August 7, 2003. *Colored dots* indicate the temperature in degree Celsius (see color scale) at **(a)** 5 m depth and **(b)** 30 m depth. *Black circles and lines* show initial position and formation, and *open circles with a cross inside triangles* illustrate the path of the center of the triangle formation. *Red arrows* correspond to estimate of the negative gradient of temperature in the horizontal plane along the path of the center of the triangular formation. The *color contours* indicate sea-surface temperature in degree Celsius (see color scale) as sampled by the Naval postgraduate school's TWIN OTTER aircraft

on its own heading (measured on-board) and a deduced reckoning estimate of its own position [19.91]. The deduced reckoning position is computed by integrating an estimate of horizontal speed using the most recent GPS fix as the initial condition. The horizontal speed is estimated from depth and vertical speed, which are estimated from on-board pressure measurements. The method also uses the glider's estimate of average flow, computed from the difference on the surface between its GPS and its deduced reckoned position.

The first sea trial in which three Slocum gliders coordinated their motion in an equilateral triangle formation was run over a period of 16 h on August 6–7, 2003, with asynchronous 2-hourly surfacings. The distance between gliders was prescribed to be 3 km and the formation prescribed to move along a linear path heading northwest to measure the incipient upwelling front. In the first half of the sea trial, there was no prescription on the orientation of the formation so that it could most efficiently maintain array resolution and follow its path. In the second half of the sea trial, to test the orientation control feature of the methodology, the orientation was prescribed such that one edge of the triangle would always be normal to the path of the center of mass of the group. The stable tracking of prescribed glider array resolution and linear path made it possible for in-situ estimates of gradients to be computed in near real time from the gliders' scalar measurements. The results suggest that the gliders could successfully be programmed to autonomously follow their estimate of the gradient if so desired.

Figure 19.1 shows a sequence of snapshots of the triangular glider formation over the August 6–7, 2003 sea trial. Temperature measurements are shown on 19.1a at 10 m depth and on 19.1b at 30 m depth. As shown, the three vehicles stayed in formation moving along the desired linear path despite relatively high-speed currents. The red arrows on the plot show a few example glider estimates of the negative gradient of temperature. These vectors point in the direction of the cold water, as verified from independent temperature measurements. The resolution corresponding to 3 km between gliders led to remarkably smooth gradient estimates over time.

Three gliders again coordinated their motion in an equilateral triangle formation in a second sea trial on August 16–17, 2003. This time the distance between gliders was prescribed to start at 6 km and then contract to 3 km. This was meant to demonstrate the expansion and contraction feature of the methodology and test the effect of the different glider array resolutions on the gradient estimates. The formation was prescribed to move along a zigzag path heading southwest across the upwelling front. Despite facing currents with magnitude as high as the Slocum's effective speed of 35 cm s^{-1}, the glider formation moved and contracted remarkably well. The results suggest that the gliders could successfully be programmed to autonomously adapt their

formation size in response to changing scales in the sampled field.

In the August 6–7 sea trial, the coordinated glider network measured the front close to its inception, while in the August 16–17 sea trial, the network measured the front after it had been advected further to the south across the mouth of Monterey Bay [19.2]. As a result, the data collected from these two sea trials added new insight into the evolving vertical structure of the upwelling plume. Notably, the gradient estimates from the glider networks would not have been possible using conventional profiling floats or drifters. Since frontal dynamics are typically nonlinear, precise gradient estimates are critical in forecasting frontogenesis and evolving instabilities.

In the third sea trial on August 23, 2003, a single glider followed a surface drifter in real time, making zigzags across and below its projected path. This demonstrated yet another opportunity for environmental monitoring with coordinated vehicles, namely that a glider or glider formation could collect scalar samples and thus estimate gradients both across and along tracer paths.

19.3.2 Cooperative Gliders in ASAP

In summer 2006, a multidisciplinary research group as part of the ASAP project performed an unprecedented field experiment, building on the successful efforts of the AOSN II project. The ASAP project demonstrated a full-scale adaptive ocean sampling network featuring a coordinated network of gliders controlled autonomously over the course of a month to efficiently sample a $22 \times 40\,\mathrm{km}^2$ and up to more than 1000 m-deep region of coastal ocean just northwest of Monterey Bay as shown in Fig. 19.2 [19.3]. The coordinated sampling of the gliders was integrated with an assortment of additional mobile and stationary sensing platforms, three real-time numerical ocean models, numerical optimization and prediction tools, a virtual control room, and a participating team of scientists.

When gliders move without taking into account, for example, through feedback, the relative position or motion of other gliders in the group, they are susceptible to strong currents driving them into clumps. In the AOSN II experiment current-driven clumping was observed and shown to lead to sensor redundancy with negative impact on sampling performance [19.6]. This motivated the development of coordinating feedback control laws that enforce dynamic distribution of vehicles to enhance sampling performance. The methodology proposed and justified in [19.6] provides this coordinated feedback control with two components: (1) the design of coordinated motion patterns for high-performance sampling and (2) the design of feedback control laws that systematically and automatically stabilize vehicles onto the desired coordinated patterns. The methodology of [19.6] was implemented in the 2006 ASAP experiment, which provided a proof of concept for the applicability of the methodology to the field.

The methodology proposed in [19.83, 84] was used for the second component, the systematic generation of feedback control laws for stable coordination of a network of autonomous vehicles to a family of motion patterns. The patterns, which consist of vehicles moving on a finite set of closed curves, are distinguished by a small number of parameters that encode synchrony. For example, two vehicles that move in parallel around two separate curves have synchronized heading directions, while two vehicles that move around the same curve but always on opposite sides of the curve have anti-synchronized heading directions. The corresponding feedback laws the vehicles use are likewise distinguished by control gains that depend on the same small number of synchrony parameters.

The control laws are distributed, which means that each vehicle applies its own control law that depends on its own measurements. Furthermore, the control laws are reactive, they do not require a prescription of where each vehicle should be as a function of time, but rather each vehicle moves in response to the relative position and direction of its neighbors. Each vehicle is constantly adjusting what it does to keep moving, to stay close to its assigned curve, and importantly to maintain the desired spacing with respect to the other vehicles, as encoded by the synchrony parameters. Because the responsive behavior of each individual can be defined as a function of the state of a small number of other vehicles, independent of the total number of vehicles, the control methodology is scalable. And because there are no leaders or special individuals in the network, the methodology is robust to vehicle failure.

The GCCS software infrastructure described in [19.90] and tested in [19.92] was used to implement the methodology in the field. In the ASAP experiment, it was observed that vehicles maintain their prescribed relative spacing in the presence of strong currents by moving off their assigned curve as needed: when a vehicle was slowed down by a strong opposing flow field, it cut inside its curve to make up distance, while its neighbor on the other side of the curve that was sped up by the strong current cut outside the curve to avoid overtaking the slower vehicle and compromising spacing.

There are several advantages to designing the coordinated motion patterns independently from the design of control laws to stabilize vehicles to those motion patterns. First, the patterns can be independently chosen to

optimize a sampling performance metric. Second, the pattern can be chosen for minimal performance sensitivity to disturbances in vehicle motion. Additionally, the pattern can be chosen to account for design requirements and constraints, such as avoiding or focusing on certain regions, leveraging information on the direction of strong currents so vehicles move with them rather than against them, and accommodating additions or removals of vehicles. Human-in-the-loop supervisory control, which can be critical for highly complex settings, can be fairly easily integrated when it is warranted. In the ASAP experiment, a team of scientists made supervisory decisions based on visualizations of observational data, modeling output, system performance, and availability of vehicles. A method was in place to translate these decisions into formal adaptations of the desired motion patterns, which could be refined using numerical optimization tools. To implement an adaptation, an intermittent, discrete change in the pattern was input to the GCCS and the vehicle network responded accordingly.

The sampling metric used to design motion patterns [19.6] is computed from the mapping error of the data assimilation scheme known as OA [19.45, 46]. OA provides a linear statistical estimation of a sampled field, and the mapping error measures the residual uncertainty. OA mapping error is a sampling performance metric since reduced uncertainty implies better measurement coverage. The mapping error at a given position and time is the error variance at that position and time. It can be computed from an empirically derived model of the covariance of fluctuations of the sampled field about its mean and from where and when data are taken. The OA mapping error is plotted in gray scale in Fig. 19.2b on the sampling domain. The sampling metric is computed as the negative log of the integral of the mapping error over the sampling region. In the ASAP experiment, the mapping error was computed in real time so that humans making adaptation decision could evaluate sampling performance.

An examination of the oceanographic and atmospheric conditions during the ASAP experiment using data and model output is described in [19.93]. The oceanographic focus of the ASAP experiment was the [19.3]

> *three-dimensional dynamics of the coastal upwelling frontal zone in Monterey Bay and the processes governing the heat budget of the* $22 \times 40\,\mathrm{km}^2$ *control volume during periods of upwelling-favorable winds and wind relaxations.*

An objective for the coordinated glider sampling was to be responsive to the dynamics of intermittent upwelling events.

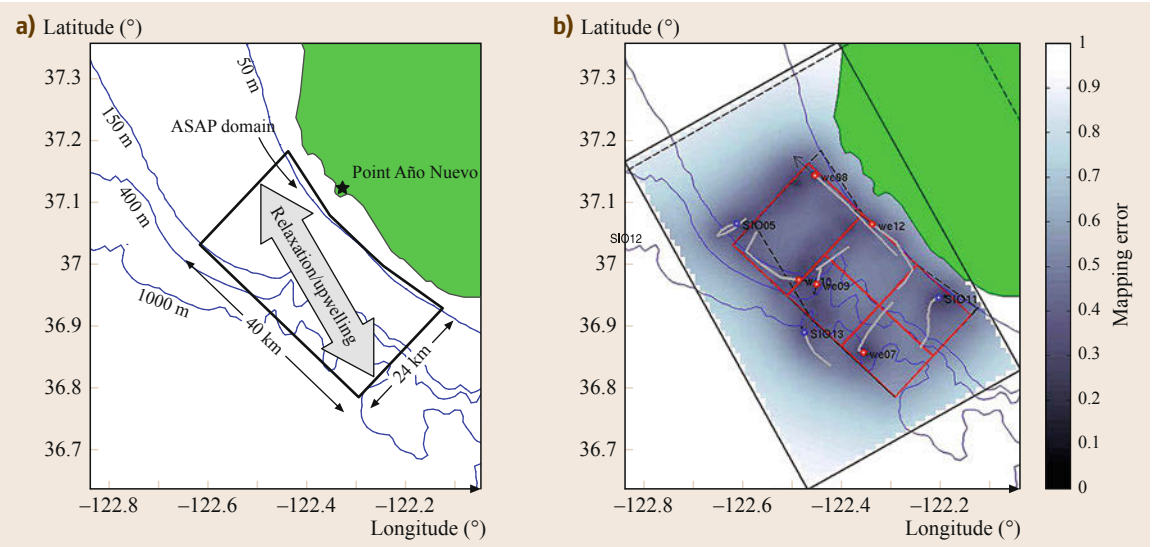


Fig. 19.2a,b From [19.3]. **(a)** Region of glider fleet operations in the 2006 ASAP field experiment, just northwest of Monterey Bay, California. The summertime ocean circulation in Monterey Bay oscillates between upwelling and relaxation. During an upwelling event, cold water often surfaces just north of the bay, near Point Año Nuevo and tends to flow southward across the mouth of the bay. During relaxation, poleward surface flow crosses the mouth of the bay past Point Año Nuevo. **(b)** Objective analysis mapping error plotted in *gray scale* on the ASAP sampling domain for July 30, 2006 at 23:30 GMT (Greenwich mean time); see text for details on the mapping error. Eight gliders are shown; their positions are indicated with *red* (Slocum) and *blue circles* (Spray)

Data were collected during the experiment from other sources as well as the gliders. These include a Naval postgraduate school research aircraft, satellite imagery, high-frequency radar, and several moorings, drifters deployed by the Monterey Bay Aquarium Research Institute (MBARI) and other ships and vehicles outside the control volume. Three different high-resolution ocean models regularly assimilated data: the Harvard Ocean Prediction System (HOPS) [19.94], the Jet Propulsion Laboratory implementation of the Regional Oceanic Modeling System (JPL/ROMS) [19.95] and the Navy Coastal Ocean Model/Innovative Coastal Ocean Observing Network (NCOM/ICON) [19.96]. Each model produced daily updated ocean predictions of temperature, salinity, and velocity. A central data server at MBARI was used to run a virtual control room (VCR) and to make all observational data and model outputs available in near real time. The VCR, developed for the 2006 ASAP field experiment, made it possible for participants to remain at their home institutions throughout the experiment but still be fully informed and connected team members [19.97]. There were a number of different panels on the VCR including those for team decision making and voting.

Virtual pilot experiments were run in advance of the field experiment to get experience with the coordinated control and adaptive sampling implementation. The virtual experiments were run as if they were real field experiments except for the replacement of real vehicles in the real ocean with simulated vehicles moving in the currents of a virtual ocean defined by a HOPS re-analysis of Monterey Bay in 2003. The simulation mode of the GCCS was designed to allow for virtual experiments with control of gliders, communication paths, and data flow exactly the same as what was used in the 2006 field experiment [19.90, 98].

For the ASAP experiment, the Slocum gliders were allocated to mapping the interior volume using automated coordinated sampling defined by motion patterns on a finite set of closed curves, with properties between measured paths inferred using interpolation. The automated feedback control laws for the Slocums were implemented with the GCCS. The Spray gliders were allocated to mapping the periphery of the volume. The boundary was divided up into segments and each Spray glider was assigned to move in an oscillatory manner along a segment of the boundary. A separate control law was implemented for this oscillatory behavior so the gliders were well distributed. The experiment started with a default coordinated motion pattern, and as the environment and operating conditions changed, the coordinated motion pattern was re-designed and updated.

The input file to the GCCS that defines a coordinated motion pattern is called a *glider coordinated trajectory* (GCT). As an example, GCT #2 used for the Slocum gliders in the ASAP experiment is illustrated in Fig. 19.3a. A GCT defines the curves that serve as tracks for the gliders as well as the synchronization of the motion of the gliders on and across the different curves. For example, the GCT #2 shown in Fig. 19.3a defines a pattern in which a pair of gliders, denoted with red circles, moves around the red curve with maximal intervehicle spacing and is synchronized with another pair of gliders, denoted with green circles, that moves around the green curve with maximal intervehicle spacing. The glider planner status panel of the VCR on July 30 at 23:10 GMT, when GCT #2 was active, is shown in Fig. 19.3b. The glider planner panel for OA mapping error at roughly the same time is shown in Fig. 19.4.

Pseudoelliptical curves were selected for the Slocums since they had nearly straight long sides. The curves were oriented to ensure that the gliders would repeatedly cross over the shelf break, each time sampling a cross-section of the dynamic ocean processes that propagate parallel to the shelf break. The shelf break refers to the end of the continental shelf characterized by a markedly increased slope toward the deep ocean bottom [19.3]:

> *By constructing a time sequence of cross-section plots, it would then be possible to reconstruct, identify and monitor ocean processes even before assimilating the glider profile data into an advanced ocean model.*

The distribution (synchronization) of the gliders relative to one another around the curves, as well as the dimensions and position of the curves, were selected to maximize the sampling performance metric. An on-line optimization tool was available for locally optimizing any candidate motion pattern. Additionally, candidate coordinated motion patterns were often pretested using the GCCS in simulation mode using one or more of the forecast ocean fields. Because the simulations of gliders moving in the forecast ocean could be run in faster than real time, it was possible to obtain predictions of glider performance in the predicted real ocean.

To implement an adaptation to sampling plans, a new GCT was prepared to replace the existing GCT. This was initiated manually by briefly interrupting the GCCS, swapping the new GCT file for the old one, and then re-starting the GCCS. Over 24 days of the ASAP experiment, 14 different GCTs were used to adapt the Slocum glider plan (Fig. 19.4a). Some of the adaptations were made in response to changes in the ocean involving strong and highly variable flow conditions. Other adaptations were made in response to changes in scientific objectives, for example, to add sampling over the head of the canyon and to chase an eddy moving off-

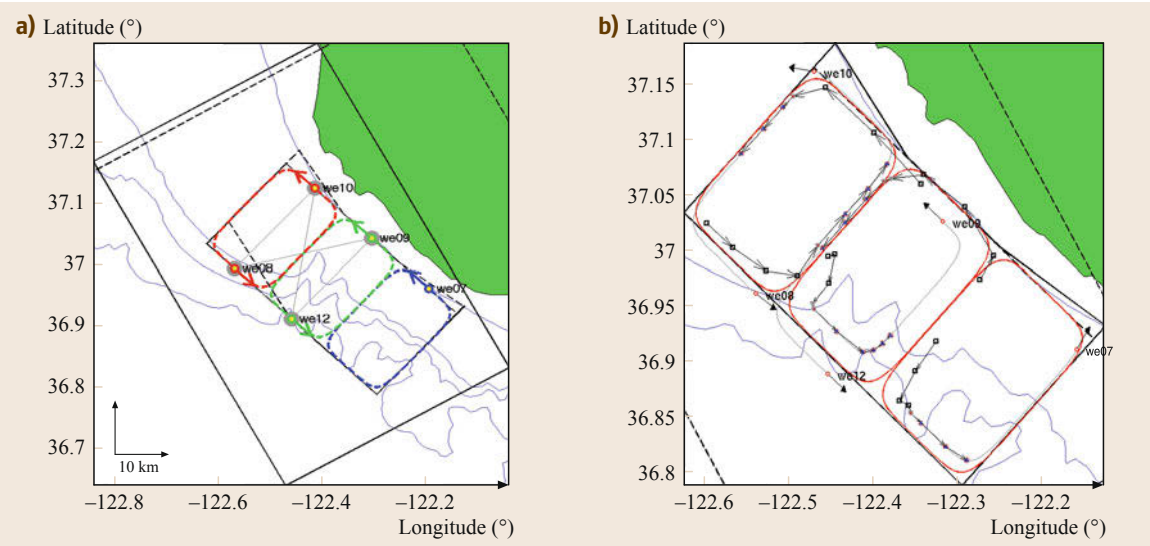


Fig. 19.3a,b From [19.3]. **(a)** The coordinated pattern for four of five Slocum gliders already in the water at the end of July 2006, with the pair we08 and we10 to move on opposites of the north track (*red*), the pair we09 and we12 on opposites of the middle track (*green*) and the two pairs synchronized on their respective tracks. Glider we07 should move independently around the south track. The *dashed lines* show the superelliptical tracks, the *circles* show a snapshot of the glider positions. The *thin gray lines* show the feedback interconnection topology for coordination and the *arrows* show prescribed direction of rotation. **(b)** A snapshot of one of the VCR panels, which was updated every minute. It presents, for each glider, surfacings over the previous 12 h (*black squares*), waypoints expected to be reached before the next surfacing (*gray triangle*), next predicted surfacing (*gray circle with red fill*), new waypoints over the next 6 h (*blue triangles inside red circles*), and planned position in 24 h (*hollow red circle*). Each glider is identified with a label at the planned position in 24 h

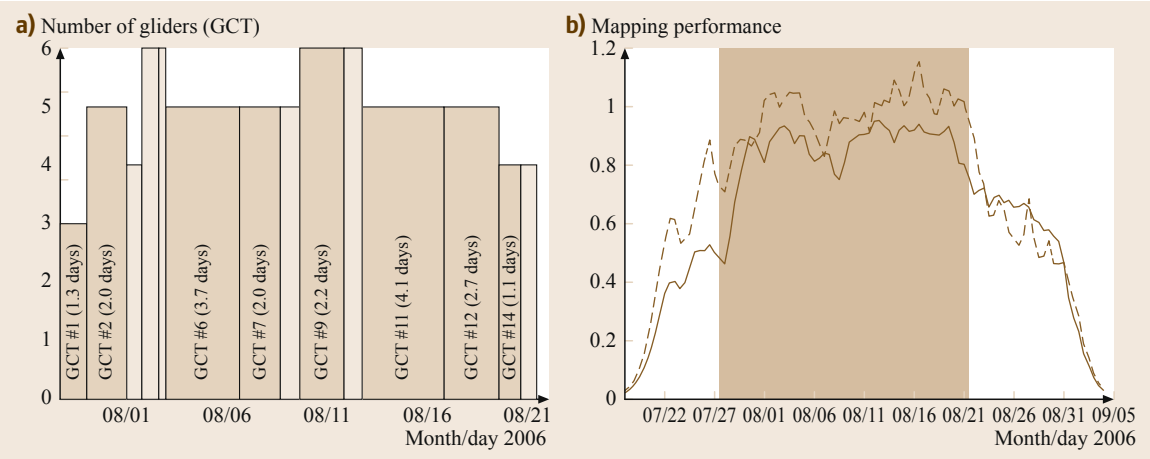


Fig. 19.4a,b From [19.3]. **(a)** Time of GCTs for Slocum gliders. **(b)** Slocum glider sampling performance over the interior of the sampling region (*solid line*) and along the boundary (*dashed line*). The portion of the plot that is *shaded* corresponds to the period of time during which the GCCS was actively steering the Slocum gliders

shore. The influence of these adaptations are reflected in the Slocum glider sampling performance, which is plotted as a function of time in Fig. 19.4b. In particular, poor coordination of gliders resulted in a decline in sampling performance. On August 6, during GCT #6, the sampling performance experienced its steepest decline as a result of flow conditions impairing coordination. Recovery of performance after a subsequent adaptation

of motion pattern demonstrated the positive impact of coordinated control on sampling performance. Details on the ocean conditions, adaptations, and performance of the gliders during the field experiment are described in [19.3].

The methodology of [19.6] was successfully implemented in the challenging coastal ocean environment of Monterey Bay, CA in August 2006, demonstrating a new capability for ocean sampling. The implementation points to the feasibility and versatility of the method for adaptation of motion patterns and integrated human decision making to address a complex multi-robot sensing task. The glider network coordination was autonomous and sustained with glider motion patterns adapted to meet the requirements of the changing ocean sampling mission and the changing dynamic state of the ocean. The methodology can be applied to collaborative robotic sensing in other domains.

19.4 Recent Developments and Future Directions

Recent years have seen an acceleration of theoretical developments and field testing of cooperative vehicle environmental monitoring; these have both motivated and leveraged advances in sensors, energy efficient and versatile mobile sensing platforms, communications, environmental modeling and estimation techniques, and control theory for cooperative exploration by networked mobile agents. It has become clear that robotic vehicles are especially useful for monitoring in remote and hostile environments and that their potential is much enhanced when they perform in cooperative teams [19.99]. A recent survey of robotics for environmental monitoring, including a discussion of cooperative systems, is provided by *Dunbabin* and *Marques* [19.100]. *Redfield* surveys works on cooperative underwater vehicles in [19.101]. Many of the recent advances are opening up further opportunities and avenues for continued research.

Further, new infrastructure developments, such as the cabled ocean observatories, provide significant resources that could potentially enhance the versatility of cooperative autonomous vehicles, most particularly with respect to data, communication, and power. For example, the NEPTUNE regional cabled ocean observatory boasts continuous high-power and high-bandwidth data transfer in real time between a large expanse of the ocean sea floor and the shore. Without having to surface, any autonomous vehicle could make use of sea floor nodes to send its data back to shore, or to share its location and data with other vehicles in the team, which would improve coordination among vehicles. A vehicle could also use the infrastructure to calibrate its location for its on-board deduced reckoning and to leverage the data collected in the cabled observed for its on-board (and thus collective) decision-making. Further, if docking stations were available, autonomous vehicles could recharge their batteries. Leveraging cabled observatories in this way would be most advantageous if dynamic sampling coverage were needed to complement the existing coverage from the sea floor nodes. Otherwise, there would be a tradeoff for the autonomous vehicles between moving into areas already well covered and exploring new areas where no cabled observatory exists.

Examples of recent field-tested methodologies for cooperative vehicle environmental monitoring include a decentralized strategy for coordinated harbor patrol using the theory of Gaussian processes implemented on three AUVs in Lisbon harbor by *Marino* et al. [19.102]. *Schofield* et al. [19.103] describe a number of field tests including a test of remote coordination of an array of acoustically networked AUVs and the coordinated sampling of underwater gliders and the space-based Hyperion imager flying on the Earth Observing-1 spacecraft. Using decentralized data fusion and control, two UAVs were used to demonstrate cooperative localization of ground-based features by *Cole* et al. [19.104]. *Techy* et al. [19.105] implemented a strategy for coordination based on speed modulation to synchronize two autonomous UAVs for tracking long-distance movement of plant pathogens above crop fields. *Maczka* et al. [19.106] demonstrated an efficient method for cooperative navigation of underwater vehicles from time-synchronized acoustic data transmissions. *Hollinger* et al. [19.107] demonstrated on a single AUV in the Southern California Bight a probabilistic planner that uses uncertainty in ocean current prediction based on an interpolation variance. *Merino* et al. [19.108] presented a cooperative perception system for multiple UAVs with different kinds of sensors and showed experimental results of forest fire detection with cooperating UAVs. *Alvarez* et al. [19.109] described methodology that estimates volumetric distribution of the geostrophic current field from glider measurements merged with satellite altimetry data; this methodology was validated using data collected from three Slocum gliders and one Spray glider moving along predefined paths during a field experiment in August 2010 in a coastal region of the Ligurian Sea. *Alvarez* and *Mourre* [19.110] examined optimal sampling strategies for a single underwater glider sampling in the presence of a mooring.

Lekien et al. [19.111] presented a method that uses Lagrangian coherent structures to coordinate vehicles robustly in the presence of very strong currents. Other methods to coordinate gliders and AUVs to maximize information in the data collected, taking explicit account of challenging ocean currents have been studied, [19.112–116].

Strategies for coordinated sampling that optimize information-based metrics have also been further explored [19.117–123].

Advances have been made in decentralized cooperative control strategies that improve or leverage the communication network structure of mobile robotic teams. These include new algorithms that use graph theoretic approaches [19.124] for computing, maintaining or maximizing connectivity [19.125–127], controllability [19.128], and robustness of coordinated motion to uncertainty [19.129]. Techniques from algebraic topology have also been applied to problems in multivehicle sensing [19.130]. Advances in cooperative routing and motion planning for multiple autonomous vehicles have been extensive, see, for example [19.131, 132].

Another source of inspiration for cooperative control design comes from mechanisms of collective behavior in animal groups such as fish schools and bird flocks. In these animal groups, remarkable collective behaviors result without centralized direction from relatively simple individuals who sense and respond to their local environment, including the relative position, heading or speed of *neighbors* in the group [19.133–136]. Mathematical models have been used to explain individual decision-making and interactions that lead to high-performing group behaviors [19.137–141]. These models can potentially be used to design provable decision-making feedback laws for individual robotic vehicles so that robotic teams inherit some of the critical group-level properties observed in nature, notably the ability of the group to forage efficiently (for information) despite individual-level limitations on sensing and communication and significant uncertainty in the environment. *Torney* et al. [19.142] showed how animal groups could apply a performance-dependent interaction to efficiently move to the source of a chemical gradient in a turbulent environment, and *Wu* et al. [19.143] turned this into a provable algorithm for efficient cooperative search in a noisy distributed field. In [19.144], *Young* et al. applied system-theoretic techniques to understand interactions in starling flocks that yield robustness of consensus to uncertainty, and used these in [19.129] to design decentralized feedback laws that enable networks of vehicles to improve their robustness to uncertainty.

Design of dynamics of decision-making teams of robots and humans is yet another important direction of research that has the potential to impact the success of cooperative vehicles in complex tasks. Humans are capable of intelligent and adaptable decision-making in response to reasoning about real-world information in real time, and robotic systems are capable of significant computational speed and memory. Challenging problems in complex settings, for example, with multiple scales and significant uncertainty, can be well served by solutions in which humans and robotic systems participate in complementary ways. Benefits and possible pitfalls of such human and robot collaboration were evidenced in the AOSN II and ASAP experiments. Indeed, with the data visualization tools, computational aids, and communication mechanisms accessible through the GCCS and ASAP's virtual control room, the ASAP field experiment demonstrated collaboration between a distributed team of humans and an automated group of underwater robots to perform adaptive ocean sampling in an uncertain environment with multiple objectives subject to a variety of safety and operational constraints. While the autonomous vehicles moved continuously in coordinated patterns, the human participants contributed in ways the automated system could not by making rapid decisions in response to critical environmental changes (sudden excessively strong currents or unanticipated shallow water) and operational failures. The human team also made important longer term decisions in response to observed or predicted indications of change in the environment, such as new features or locations of interest, or decline in performance with respect to one or more performance metric. And yet, opportunities may have been missed when the human participants had difficulty assessing the likelihood of risk in proposed scenarios or balancing the many competing objectives. Likewise, human supervisory strategies may have been less than optimal given that the autonomous robots had no opportunity to provide direct feedback to their human supervisors.

In this vein, progress has been reported on humans and robots in exploration [19.145–149] and on humans working together in search tasks [19.150–153]. Progress has also been made on algorithm development using what cognitive scientists understand about human decision-making. For example, decision-making algorithms with provable performance in search tasks were derived in [19.154], drawing on research on the heuristics humans use to trade off between exploiting well-known alternatives and exploring uncertain but possibly better alternatives. It has been shown in [19.155] that humans use their ability to learn correlation structure when it exists among decision alternatives, as in the case of alternatives representing different locations in a spatially distributed resource. The knowledge

of correlation structure has been shown to be critical to enhancing performance in the search algorithms of [19.154]. This suggests approaches to systematize means for robots to benefit from human input, for example, by using observations of human choices to estimate correlation structure and updating decision-making strategies accordingly. Further work that builds on research on human decision-making and behavior may allow derivation of engineering models and provable strategy design for well-integrated human–robot teams in complex missions such as cooperative vehicles for environmental monitoring.

References

19.1 S.H.D. Haddock, D.M. Fratantoni: Introduction to the autonomous ocean sampling network (AOSN-II) program, Deep-Sea Res. II **56**, 61 (2009)

19.2 S.R. Ramp, R.E. Davis, N.E. Leonard, I. Shulman, Y. Chao, A.R. Robinson, J. Marsden, P.F.J. Lermusiaux, D.M. Fratantoni, J.D. Paduan, F.P. Chavez, F.L. Bahr, S. Liang, W. Leslie, Z. Li: Preparing to predict: The second autonomous ocean sampling network (AOSN-II) experiment in Monterey Bay, Deep-Sea Res. II **56**, 68–86 (2009)

19.3 N.E. Leonard, D.A. Paley, R.E. Davis, D.M. Fratantoni, F. Lekien, F. Zhang: Coordinated control of an underwater glider fleet in an adaptive ocean sampling field experiment in Monterey Bay, J. Field Robotics **27**(6), 718–740 (2010)

19.4 P. Ögren, E. Fiorelli, N.E. Leonard: Cooperative control of mobile sensor networks: Adaptive gradient climbing in a distributed environment, IEEE Trans. Autom. Control **49**, 1292–1302 (2004)

19.5 E. Fiorelli, N.E. Leonard, P. Bhatta, D. Paley, R. Bachmayer, D.M. Fratantoni: Multi-AUV control and adaptive sampling in Monterey Bay, IEEE-JOE **31**(4), 935–948 (2006)

19.6 N.E. Leonard, D.A. Paley, F. Lekien, R. Sepulchre, D.M. Fratantoni, R.E. Davis: Collective motion, sensor networks, and ocean sampling, Proc. IEEE **95**, 48–74 (2007)

19.7 T.B. Curtin, J.G. Bellingham, J. Catipovic, D. Webb: Autonomous oceanographic sampling networks, Oceanography **6**(3), 86–94 (1993)

19.8 J.G. Bellingham: New oceanographic uses of autonomous underwater vehicles, Mar. Technol. Soc. J. **31**(3), 34–47 (1997)

19.9 D.R. Yoerger, A.M. Bradley, B.B. Walden: The autonomous benthic explorer (ABE): An AUV optimized for deep seafloor studies, Proc. 7th UUST (Autonomous Undersea Systems Institute, Lee 1991) pp. 60–70

19.10 J.G. Bellingham, C.A. Goudey, T.R. Consi, C. Chryssostomidis: A small, long-range vehicle for deep ocean exploration, Proc. Int. Offshore Polar Eng. Conf., Vol. 2 (International Society of Offshore and Polar Engineers, Mountain View 1992) pp. 461–467

19.11 C. Von Alt, B. Allen, T. Austin, R. Stokey: Remote environmental measuring units, Proc. AUV (1994) pp. 13–19

19.12 D.C. Webb, P.J. Simonetti, C.P. Jones: SLOCUM: An underwater glider propelled by environmental energy, IEEE-JOE **26**(4), 447–452 (2001)

19.13 J. Sherman, R.E. Davis, W.B. Owens, J. Valdes: The autonomous underwater glider Spray, IEEE-JOE **26**(4), 437–446 (2001)

19.14 C.C. Eriksen, T.J. Osse, R.D. Light, T. Wen, T.W. Lehman, P.L. Sabin, J.W. Ballard, A.M. Chiodi: Seaglider: A long-range autonomous underwater vehicle for oceanographic research, IEEE-JOE **26**(4), 424–436 (1994)

19.15 D.L. Rudnick, R.E. Davis, C.C. Eriksen, D.M. Fratantoni, M.J. Perry: Underwater gliders for ocean research, Mar. Technol. Soc. J. **38**(1), 48–59 (2004)

19.16 C.C. Eriksen, M.J. Perry: The nurturing of Seagliders by the National Oceanographic Partnership Program, Oceanography **22**(2), 146–157 (2009)

19.17 D.R. Yoerger, A. Bradley, B. Walden, H. Singh, R. Bachmayer: Surveying a subsea lava flow using the autonomous benthic explorer (ABE), J. Jpn. Soc. Mar. Surveys Technol. **9**(1), 43–60 (1997)

19.18 M.A. Tivey, H.P. Johnson, A.M. Bradley, D.R. Yoerger: Thickness of a submarine lava flow determined from near-bottom magnetic field mapping by autonomous underwater vehicle, Geophys. Res. Lett. **25**(6), 805–808 (1998)

19.19 M.V. Jakuba, D.R. Yoerger: High-resolution multibeam sonar mapping with the autonomous benthic explorer (ABE), Proc. 13th UUST (Autonomous Undersea Systems Institute, Lee 2003)

19.20 C.R. German, D.P. Connelly, R.D. Prien, D.R. Yoerger, M. Jakuba, A. Bradley, T. Shank, K.I. Nakamura, C. Langmuir, L.M. Parsons: New techniques for hydrothermal plume investigaton by AUV, Geophys. Res. Abstracts **7**, 04361 (2005)

19.21 J.G. Bellingham, J.J. Leonard, J. Vaganay, C.A. Goudey, D.K. Atwood, T.R. Consi, J.W. Bales, H. Schmidt, C. Chryssostomidis: AUV operations in the Arctic, Proc. 8th UUST (Autonomous Undersea Systems Institute, Lee 1993)

19.22 S. Blackwell, J. Case, S. Glenn, J. Kohut, M. Moline, M. Purcell, O. Schofield, C. Von Alt: A new AUV platform for studying near shore bioluminescence structure. In: *Bioluminesence and Chemiluminescence: Progress and Current Applications*, ed. by L.J. Kricka, P.E. Stanley (World Scientific, Singapore 2002) pp. 197–200

19.23 M.A. Moline, P. Bissett, S. Blackwell, J. Mueller, J. Sevadjian, C. Trees, R. Zaneveld: An autonomous vehicle approach for quantifying bioluminescence in ports and harbors, Proc. SPIE **5780**, 81–87 (2005)

19.24 J.F. Grassle, S.M. Glenn, C. Von Alt: Ocean observing systems for marine habitats, Proc. MTS OCC (Marine Technology Society, Washington 1998) pp. 567–570

19.25 S.M. Glenn, T.D. Dickey, B. Parker, W. Boicourt: Long-term real-time coastal ocean observation networks, Oceanography **13**(1), 24–34 (2000)

19.26 O. Schofield, T. Bergmann, P. Bissett, J.F. Grassle, D.B. Haidvogel, J. Kohut, S.M. Glenn: The long-term ecosystem observatory: An integrated coastal observatory, IEEE-JOE **27**, 146–154 (2002)

19.27 J.S. Albus, D.R. Blidberg: Control system architecture for multiple autonomous undersea vehicles (MAUV), Proc. 5th UUST (Autonomous Undersea Systems Institute, Lee 1987)

19.28 R.M. Turner, J.S. Fox, E.H. Turner, D.H. Blidberg: Multiple autonomous vehicle imagin systems (MAVIS), Proc. 6th UUST (Autonomous Undersea Systems Institute, Lee 1988) pp. 526–536

19.29 M.S. Triantafyllou, K. Streitlien: Distributed control of multiple AUV's forming effective chains, Proc. 7th UUST (Autonomous Undersea Systems Institute, Lee 1991) pp. 499–518

19.30 H. Singh, J. Catipovic, R. Eastwood, L. Freitag, H. Henriksen, F. Hover, D. Yoerger, J.G. Bellingham, B.A. Moran: An integrated approach to multiple AUV communications, navigation, and docking, Proc. IEEE OCEANS'96 (1996) pp. 59–64

19.31 H. Schmidt, J.G. Bellingham, M. Johnson, D. Herold, D.M. Farmer, R. Pawlowicz: Real-time frontal mapping with AUVs in a coast environment, Proc. IEEE OCEANS'96, Vol. 3 (1996) pp. 1094–1098

19.32 D.J. Stilwell, B.E. Bishop: Platoons of underwater vehicles: Communication, feedback and decentralized control, IEEE Control Syst. Mag. **20**(6), 45–52 (2000)

19.33 B. Schulz, B. Hobson, M. Kemp, J. Meyer: Field results of multi-UUV missions using Ranger micro-UUVs, Proc. IEEE OCEANS'03, Vol. 2 (2003) pp. 956–961

19.34 Y.U. Cao, A.S. Fukunaga, A.B. Kahng: Cooperative mobile robotics: Antecedents and directions, Auton. Robots **4**, 1–23 (1997)

19.35 T. Fukuda, S. Nakagawa: A dynamically reconfigurable robotic system, Proc. IEEE ICRA'87, Vol. 3 (1987) pp. 588–595

19.36 G. Beni: The concept of cellular robotic system, Proc. IEEE ISIC (1988) pp. 57–62

19.37 M.J. Mataric: *Distributed approaches to behavior control, Proc. SPIE* **1828** 1992) pp. 373–382

19.38 L.E. Parker: Adaptive action selection for cooperative agent teams, 2nd Int. Conf. Simul. of Adapt. Behav. (MIT, Cambridge 1992) pp. 442–450

19.39 R. Beckers, O.E. Holland, J.L. Deneubourg: From local actions to global tasks: Stigmergy and collective robotics. In: *Prerational Intelligence: Studies in Cognitive Systems*, Vol. 26, ed. by H. Crise, J. Dean, H. Ritter (Springer, Dordrecht 2000) pp. 1008–1022

19.40 M.J. Mataric: Designing emergent behaviors: From local interactions to collective intelligence, 2nd Int. Conf. Simul. of Adapt. Behav. (MIT, Cambridge 1992) pp. 432–441

19.41 D. Estrin, R. Govindan, J. Heidemann, S. Kumar: Next century challenges: Scalable coordination in sensor networks, Proc. ACM/IEEE MobiCom (1999) pp. 263–270

19.42 A. Cerpa, J. Elson, D. Estrin, L. Girod, M. Hamilton, J. Zhao: Habitat monitoring: Application driver for wireless communications technology, ACM SIGCOMM Workshop Data Commun. Latin America Caribb., Vol. 31 (2001) pp. 20–41

19.43 I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci: Wireless sensor networks: A survey, Comput. Netw. **38**, 393–422 (2002)

19.44 R. Cardell-Oliver, K. Smetten, M. Kranz, K. Mayer: Field testing a wireless sensor network for reactive environmental monitoring, Proc. IEEE ISSNIP (2004) pp. 14–17

19.45 F.P. Bretherton, R.E. Davis, C.B. Fandry: A technique for objective analysis and design of oceanographic experiments applied to MODE-73, Deep-Sea Res. **23**(7), 559–582 (1976)

19.46 A.R. Robinson, L.S. Gandin: Objective analysis of meteorological fields, Q. J. R. Meteorol. Soc. **92**(393), 447 (1966)

19.47 A.R. Robinson, S.M. Glenn: Adaptive sampling for ocean forecasting, Nav. Res. Rev. **51**(2), 28–38 (1999)

19.48 J.S. Willcox, J.G. Bellingham, Y. Zhang, A.B. Baggeroer: Performance metrics for oceanographic surveys with autonomous underwater vehicles, IEEE-JOE **26**(4), 711–725 (2001)

19.49 J.G. Bellingham, Y. Zhang: Observing processes that vary in time and space with heterogeneous mobile networks, Proc. Int. Workshop Underw. Robotics Sustain. Manag. Mar. Ecosyst. Environ. Monit. (Plus-Pisa University Press, Pisa 2005) pp. 9–16

19.50 B. Grocholsky: Information-Theoretic Control of Multiple Sensor Platforms, Ph.D. Thesis (University of Sydney, Sydney 2002)

19.51 S. Kumar, J.H. Seinfeld: Optimal location of measurements for distributed parameter estimation, IEEE Trans. Autom. Control **23**(4), 690–698 (1978)

19.52 D. Uciński: Optimal sensor location for parameter estimation of distributed processes, Int. J. Control **73**(13), 1235–1248 (2000)

19.53 C.R. McInnes: Potential function methods for autonomous spacecraft guidance and control, Adv. Astronaut. Sci. **90**, 2093–2109 (1996)

19.54 P.S. Krishnaprasad: Relative equilibria and stability of rings of satellites, Proc. 39th IEEE CDC (2000) pp. 1285–1288

19.55 N.E. Leonard, E. Fiorelli: Virtual leaders, artificial potentials and coordinated control of groups, Proc. 40th IEEE CDC (2001) pp. 2968–2973

19.56 P. Ögren, M. Egerstedt, X. Hu: A control Lyapunov function approach to multi-agent coordination, IEEE Trans. Robotics Autom. **18**, 847–851 (2002)

19.57 R. Olfati-Saber, R.M. Murray: Distributed cooperative control of multiple vehicle formations us-

ing structural potential functions, Proc. 15th IFAC World Congr. (2002) pp. 1–7
19.58 H. Tanner, A. Jadbabaie, G.J. Pappas: Stable flocking of mobile agents, part I: Fixed topology, Proc. 42nd IEEE CDC (2003) pp. 2010–2015
19.59 O. Khatib: Real time obstacle avoidance for manipulators and mobile robots, Int. J. Robotics Res. **5**, 90–99 (1986)
19.60 E. Rimon, D.E. Koditschek: Exact robot navigation using artificial potential functions, IEEE Trans. Robotics Autom. **8**, 501–518 (1992)
19.61 V. Gazi, K. Passino: Stability analysis of social foraging swarms, Proc. 41st IEEE CDC (2002) pp. 2848–2853
19.62 A. Mogilner, L. Edelstein-Keshet, L. Bent, A. Spiros: Mutual interactions, potentials, and individual distance in a social aggregation, J. Math. Biol. **47**, 353–389 (2003)
19.63 R. Bachmayer, N.E. Leonard: Vehicle networks for gradient descent in a sampled environment, Proc. 41st IEEE CDC (2002) pp. 112–117
19.64 B.J. Young, R.W. Beard, J.M. Kelsey: A control scheme for improving multi-vehicle formation maneuvers, Proc. IEEE Am. Control Conf. (2001) pp. 704–709
19.65 F. Zhang, N.E. Leonard: Generating contour plots using multiple sensor platforms, Proc. IEEE SIS (2005) pp. 1–8
19.66 F. Zhang, N.E. Leonard: Cooperative filters and control for cooperative exploration, IEEE Trans. Autom. Control **55**(3), 650–663 (2010)
19.67 F. Zhang, E. Justh, P.S. Krishnaprasad: Boundary following using gyroscopic control, Proc. 43rd IEEE CDC (2004) pp. 5204–5209
19.68 A.L. Bertozzi, M. Kemp, D. Marthaler: Determining environmental boundaries: Asynchronous communication and physical scales. In: *Cooperative Control: A Post-Workshop Volume*, Lecture Notes in Control and Information Sciences, Vol. 309, ed. by V. Kumar, N. Leonard, A.S. Morse (Springer, Berlin, Heidelberg 2005) pp. 25–42
19.69 H. Choset: Coverage for robotics: A survey of recent results, Ann. Math. Artif. Intell. **31**, 113–126 (2001)
19.70 J. Cortes, S. Martinez, T. Karatas, F. Bullo: Coverage control for mobile sensing networks, IEEE Trans. Robotics Autom. **20**, 243–255 (2004)
19.71 B. Sinopoli, C. Sharp, L. Schenato, S. Schaffert, S.S. Sastry: Distributed control applications within sensor networks, Proc. IEEE **91**, 1235–1246 (2003)
19.72 S. Martinez, F. Bullo: Optimal sensor placement and motion coordination for target tracking, Automatica **42**(4), 661–668 (2006)
19.73 J.N. Tsitsiklis, D.P. Bertsekas, M. Athans: Distributed asynchronous deterministic and stochastic gradient optimization algorithms, IEEE Trans. Autom. Control **31**(9), 803–812 (1986)
19.74 A. Jadbabaie, J. Lin, A.S. Morse: Coordination of groups of mobile autonomous agents using nearest neighbor rules, IEEE Trans. Autom. Control **48**, 988–1001 (2002)
19.75 R. Olfati-Saber, R. Murray: Consensus problems in networks of agents with switching topology and time-delays, IEEE Trans. Autom. Control **49**(9), 1520–1533 (2004)
19.76 L. Moreau: Stability of multiagent systems with time-dependent communication links, IEEE Trans. Autom. Control **50**(2), 169–182 (2005)
19.77 W. Ren, R.W. Beard: Consensus seeking in multiagent systems under dynamically changing interaction topologies, IEEE Trans. Autom. Control **50**(5), 655–661 (2005)
19.78 J.A. Fax, R.M. Murray: Information flow and cooperative control of vehicle formations, IEEE Trans. Autom. Control **49**(9), 1465–1476 (2004)
19.79 Z. Lin, M. Broucke, B. Francis: Local control strategies for groups of mobile autonomous agents, IEEE Trans. Autom. Control **49**(4), 622–629 (2004)
19.80 Y. Kuramoto: *Chemical Oscillations, Waves, and Turbulence* (Springer, Berlin, Heidelberg 1984)
19.81 S.H. Strogatz: From Kuramoto to Crawford: Exploring the onset of synchronization in populations of coupled oscillators, Physica D **143**, 1–20 (2000)
19.82 E.W. Justh, P.S. Krishnaprasad: Equilibria and steering laws for planar formations, Syst. Control Lett. **52**(1), 25–38 (2004)
19.83 R. Sepulchre, D.A. Paley, N.E. Leonard: Stabilization of planar collective motion: All-to-all communication, IEEE Trans. Autom. Control **52**(5), 811–824 (2007)
19.84 R. Sepulchre, D.A. Paley, N.E. Leonard: Stabilization of planar collective motion with limited communication, IEEE Trans. Autom. Control **53**(3), 706–719 (2008)
19.85 L. Scardovi, R. Sepulchre, N.E. Leonard: Stabilization of three dimensional collective motion, Commun. Inf. Syst. **8**(4), 473–500 (2008)
19.86 T.R. Smith, H. Hanssmann, N.E. Leonard: Orientation control of multiple underwater vehicles, Proc. 40th IEEE CDC (2001) pp. 4598–4603
19.87 S. Nair, N.E. Leonard: Stable synchronization of rigid body networks, Netw. Heterog. Media **2**(4), 595–624 (2007)
19.88 T.B. Curtin, J.G. Bellingham: Progress toward autonomous ocean sampling networks, Deep Sea Res, II **56**, 62–67 (2009)
19.89 E. Fiorelli, P. Bhatta, N.E. Leonard, I. Shulman: Adaptive sampling using feedback control of an autonomous underwater glider fleet, Proc. 13th UUST (Autonomous Undersea Systems Institute, Lee 2003)
19.90 D. Paley, F. Zhang, N.E. Leonard: Cooperative control for ocean sampling: The glider coordinated control system, IEEE Trans. Control Syst. Technol. **16**(4), 735–744 (2008)
19.91 J. Graver: Underwater Glider: Dynamics, Control and Design, Ph.D. Thesis (Princeton University, Princeton 2005)
19.92 F. Zhang, D.M. Fratantoni, D.A. Paley, J.M. Lund, N.E. Leonard: Control of coordinated patterns for ocean sampling, Int. J. Control **80**(7), 1186–1199 (2007)
19.93 S.R. Ramp, P.F.J. Lermusiaux, I. Shulman, Y. Chao, R.E. Wolf, R.L. Bahr: Oceanographic and atmo-

spheric conditions on the continental shelf north of the Monterey Bay during August 2006, Dyn. Atmos. Oceans **52**, 192–223 (2011)
19.94 A.R. Robinson: Forecasting and simulating coastal ocean processes and variabilities with the Harvard Ocean Prediction System. In: *Coastal Ocean Prediction*, ed. by C.N.K. Mooers (American Geophysical Union, Washington 1999)
19.95 A.F. Shchepetkin, J.C. McWilliams: The regional oceanic modeling system: A split-explicit, free-surface, topography-following-coordinate ocean model, Ocean Model. **9**, 347–404 (2005)
19.96 I. Shulman, C.-R. Wu, J.K. Lewis, J.D. Paduan, L.K. Rosenfeld, J.D. Kindle, S.R. Ramp, C.A. Collins: High resolution modeling and data assimilation in the Monterey Bay area, Cont. Shelf Res. **22**(8), 1129–1151 (2002)
19.97 M.A. Godin, J.G. Bellingham, K. Rajan, N. Leonard, Y. Chao: A collaborative portal for ocean observatories, Proc. IEEE OCEANS'06 (2006)
19.98 D.A. Paley: Cooperative Control of Collective Motion for Ocean Sampling with Autonomous Vehicles, Ph.D. Thesis (Princeton University, Princeton 2007)
19.99 J.G. Bellingham, K. Rajan: Robotics in remote and hostile environments, Science **318**(5853), 1098–1102 (2007)
19.100 M. Dunbabin, L. Marques: Robots for environmental monitoring: Significant advancements and applications, IEEE Robotics Autom. Mag. **19**(1), 24–39 (2012)
19.101 S. Redfield: Cooperation between underwater vehicles. In: *Marine Robot Autonomy*, ed. by M.L. Seto (Springer, New York 2013) pp. 257–286
19.102 A. Marino, G. Antonelli, A.P. Aguiar, A. Pascoal: A new approach to multi-robot harbour patrolling: Theory and experiments, Proc. IROS (2012) pp. 1760–1765
19.103 O. Schofield, S. Glenn, J. Orcutt, M. Arrott, M. Meisinger, A. Gangopadhyay, W. Brown, R. Signell, M. Moline, Y. Chao, S. Chien, D. Thompson, A. Balasuriya, P.F.J. Lermusiaux, M. Oliver: Automated sensor networks to advance ocean science, Eos Trans. Am. Geophys. Union **91**(39), 345–346 (2010)
19.104 D.T. Cole, P.R. Gaktogan, S. Sukkarieh: System development and demonstration of a cooperative UAV team for mapping and tracking, Int. J. Robotics Res. **29**(11), 1371–1399 (2010)
19.105 L. Techy, D.G. Schmale, C.A. Woolsey: Coordinated aerobiological sampling of a plant pathogen in the lower atmospher using two autonomous unmanned aerial vehicles, J. Field Robotics **27**(3), 335–343 (2010)
19.106 D.K. Maczka, A.S. Gadre, D.J. Stilwell: Implementation of a cooperative navigation algorithm on a platoon of autonomous underwater vehicles, Proc. IEEE OCEANS'07 (2007) pp. 1–6
19.107 G.A. Hollinger, A.A. Pereira, G.S. Sukhatme: Learning uncertainty models for reliable operation of autonomous underwater vehicles, Proc. IEEE ICRA'13 (2013) pp. 5593–5599
19.108 L. Merino, F. Caballero, J.R. Martinez de Dios, J. Ferruz, A. Ollero: A cooperative perception system for multiple UAVs: Application to automatic detection of forest fires, J. Field Robotics **23**, 165–184 (2006)
19.109 A. Alvarez, J. Chiggiato, K. Schroeder: Mapping sub-surface geostrophic currents from altimetry and a fleet of gliders, Deep-Sea Res. I **74**, 115–129 (2012)
19.110 A. Alvarez, B. Mourre: Optimum sampling designs for a glider-mooring observing network, J. Atmos. Ocean. Technol. **29**(4), 601–612 (2012)
19.111 F. Lekien, L. Mortier, P. Testor: Glider coordinated control and Lagrangian coherent structures, Proc. IFAC Workshop Navig. Guid. Control Underw. Veh. (2008) pp. 125–130
19.112 K.M. Lynch, I.B. Schwartz, P. Yang, R.A. Freeman: Decentralized environmental modeling by mobile sensor networks, IEEE Trans. Robotics **24**(3), 710–724 (2008)
19.113 K.A.C. Baumgartner, S. Ferrari, T.A. Wettergren: Robust deployment of dynamic sensor networks for cooperative track detection, IEEE Sens. J. **9**(9), 1029–1048 (2009)
19.114 A. Munafò, E. Simetti, A. Turetta, A. Caiti, G. Casalino: Autonomous underwater vehicle teams for adaptive ocean sampling: A data-driven approach, Ocean Dyn. **61**(11), 1981–1984 (2011)
19.115 X. Liang, W. Wu, D. Chang, F. Zhang: Real-time modelling of tidal current for navigating underwater glider sensing networks, Procedia Comput. Sci. **10**, 1121–1126 (2012)
19.116 R.E. Davis, N.E. Leonard, D.M. Fratantoni: Routing strategies for underwater gliders, Deep-Sea Res. II **56**, 173–187 (2009)
19.117 S. de Bruin, D. Ballari, A.K. Bregt: Where and when should sensors move? Sampling using the expected value of information, Sensors **12**(12), 16274–16290 (2012)
19.118 R. Graham, J. Cortes: Adaptive information collection by robotic sensor networks for spatial estimation, IEEE Trans. Autom. Control **57**(6), 1404–1419 (2012)
19.119 F. Lekien, N.E. Leonard: Nonuniform coverage and cartograms, SIAM J. Control Optim. **48**(1), 351–372 (2009)
19.120 C.H. Caicedo-Nunez, N.E. Leonard: Symmetric coverage of dynamic mapping error for mobile sensor networks, Proc. IEEE Am. Control Conf. (2011) pp. 4661–4666
19.121 D.A. Paley, N. Sydney: Optimal sampling of non-stationary spatiotemporal fields using a mobile sensor network, Proc. IFAC Workshop Distrib. Estim. Control Netw. Syst. (2012)
19.122 R.N. Smith, M. Schwager, S.L. Smith, B.H. Jones, D. Rus, G.S. Sukhatme: Persistent ocean monitoring with underwater gliders: Adapting sampling resolution, J. Field Robotics **28**(5), 714–741 (2011)

19.123 A. Garcia-Olaya, F. Py, J. Das, K. Rajan: An online utility-based approach for sampling dynamic ocean fields, IEEE-JOE **37**(2), 185–203 (2012)

19.124 M. Mesbahi, M. Egerstedt: *Graph theoretic methods for multiagent networks* (Princeton University Press, Princeton 2010)

19.125 T. Keviczky, R. Babuska: On distributed maximization of algebraic connectivity in robotic networks, Proc. IEEE Am. Control Conf. (2011) pp. 2180–2185

19.126 M.M. Zavlanos, G.J. Pappas: Distributed connectivity control of mobile networks, IEEE Trans. Robotics **24**(6), 1416–1428 (2008)

19.127 A.N. Kopeikin, S.S. Ponda, L.B. Johnson, O. Toupet, J.P. How: Real-time dynamic planning to maintain network connectivity in a team of unmanned air vehicles, Proc. IEEE GLOBECOM Workshops (2011) pp. 1303–1307

19.128 A. Rahmani, M. Ji, M. Mesbahi, M. Egerstedt: Controllability of multi-agent systems from a graph theoretic perspective, SIAM J. Control Optim. **48**(1), 162–186 (2009)

19.129 G.F. Young, L. Scardovi, N.E. Leonard: Rearranging trees for robust consensus, Proc. 50th IEEE CDC (2011) pp. 1000–1005

19.130 P. Dlotko, R. Ghrist, M. Juda, M. Mrozek: Distributed computation of coverage in sensor networks by homological methods, Appl. Algebra Eng. Commun. Comput. **23**, 29–58 (2012)

19.131 M. Pavone, E. Frazzoli, F. Bullo: Adaptive and distributed algorithms for vehicle routing in a stochastic and dynamic environment, IEEE Trans. Autom. Control **56**(6), 1259–1274 (2011)

19.132 A.J. Häusler, A. Saccon, A.P. Aguiar, J. Hauser, A.M. Pascoal: Cooperative motion planning for multiple autonomous marine vehicles, Proc. IFAC Conf. Manoeuvring Control Mar. Craft (2012)

19.133 J.K. Parrish, L. Edelstein-Keshet: Complexity, pattern, and evolutionary trade-offs in animal aggregation, Science **284**(5411), 99–101 (1999)

19.134 J. Krause, G.D. Ruxton: *Living in Groups* (Oxford Univ. Press, Oxford 2002)

19.135 C. Brown, K. Laland, J. Krause (Eds.): *Fish Cognition and Behavior* (Blackwell Publishing, Oxford 2006)

19.136 I.D. Couzin: Collective cognition in animal groups, Trends Cogn. Sci. **13**, 36–43 (2009)

19.137 I.D. Couzin, J. Krause, N.R. Franks, S.A. Levin: Effective leadership and decision making in animal groups on the move, Nature **434**, 513–516 (2005)

19.138 T.D. Seeley, P.K. Visscher, T. Schlegel, P.M. Hogan, N.R. Franks, J.A.R. Marshall: Stop signals provide cross inhibition in collective decision-making by honey bee swarms, Science **335**, 108–111 (2012)

19.139 D.J.T. Sumpter: *Collective Animal Behavior* (Princeton University Press, Princeton 2010)

19.140 N.E. Leonard, T. Shen, B. Nabet, L. Scardovi, I.D. Couzin, S.A. Levin: Decision versus compromise for animal groups in motion, Proc. Natl. Acad. Sci. USA **109**, 227–232 (2012)

19.141 D.A. Paley, N.E. Leonard, R. Sepulchre, D. Grünbaum, J.K. Parrish: Oscillator models and collective motion: Spatial patterns in the dynamics of engineered and biological networks, IEEE Control Syst. Mag. **27**(4), 89–105 (2007)

19.142 C. Torney, Z. Neufeld, I.D. Couzin: Context-dependent interaction leads to emergent search behavior in social aggregates, Proc. Natl. Acad. Sci. USA **106**, 5464–5469 (2009)

19.143 W. Wu, F. Zhang: Robust cooperative exploration with a switching strategy, IEEE Trans. Robotics **28**(4), 828–839 (2012)

19.144 G.F. Young, L. Scardovi, A. Cavagna, I. Giardina, N.E. Leonard: Starling flock networks manage uncertainty in consensus at low cost, PLoS Comput. Biol. **9**(1), e1002894 (2013)

19.145 Y. Boussemart, M.L. Cummings: Predictive models of human supervisory control behavioral patterns using hidden semi-markov models, Eng. Appl. Artif. Intell. **24**(7), 1252–1262 (2011)

19.146 D. Baranov, J. Baillieul: Decision making for rapid information acquisition in the reconnaissance of random fields, Proc. IEEE **100**, 776–801 (2012)

19.147 V. Srivastava: Stochastic Search and Surveillance Strategies in Mixed Human–Robot Teams, Ph.D. Thesis (University of California, Santa Barbara 2012)

19.148 J. Casper, R.R. Murphy: Human-robot interaction during the robot-assisted urban search and rescue response at the world trade center, IEEE Trans. Syst. Man Cybern, Part B **33**(3), 367–385 (2003)

19.149 D.J. Bruemmer, D.A. Few, R.L. Boring, J.L. Marble, M.C. Walton, C.W. Nielsen: Shared understanding for collaborative control, IEEE Trans. Syst. Man Cybern, Part A **35**(4), 494–504 (2005)

19.150 R.L. Goldstone, B.C. Ashpole: Human foraging behavior in a virtual environment, Psychon. Bull. Rev. **11**(3), 508–514 (2004)

19.151 W. Mason, D.J. Watts: Collaborative learning in networks, Proc. Natl. Acad. Sci. USA **109**(3), 764–769 (2011)

19.152 A. Nedic, D. Tomlin, P. Holmes, D.A. Prentice, J.D. Cohen: A decision task in a social context: Human experiments, models, and analyses of behavioral data, Proc. IEEE **100**(3), 713–733 (2012)

19.153 A.R. Stewart: Analysis and Prediction of Decision Making with Social Feedback, Ph.D. Thesis (Princeton University, Princeton 2012)

19.154 P. Reverdy, V. Srivastava, N.E. Leonard: Modeling human decision-making in generalized gaussian multi-armed bandits, Proc. IEEE **102**, 544–571 (2014)

19.155 D.E. Acuña, P. Schrater: Structure learning in human sequential decision-making, PLoS Comput. Biol. **6**(12), e1001003 (2010)

20. Nested Autonomy for Distributed Ocean Sensing

Henrik Schmidt, Michael R. Benjamin, Stephanie M. Petillo, Raymond Lum

Nested autonomy is a new control paradigm for distributed undersea surveillance, specifically designed for the layered communication infrastructure provided by the low-bandwidth underwater acoustic communication and the intermittent RF connectivity. Implemented using the open-source MOOS-IvP behavior-based, autonomous command and control architecture, it provides the fully integrated sensing, modeling, and controlling that allows each platform to autonomously detect, classify, localize, and track an episodic event in the ocean, without depending on any operator command and control. The prosecution of an event, such as the detection and tracking of a subsea volcanic plume or an oceanographic feature, may be initiated by the operators or entirely autonomously by an onboard detection capability. The event information collected by each node in the network is reported back to the operators by transmitting an event report, using a dedicated command and control language. Collaborative processing and control is exploited when the communication channel allows, for example, collaborative tracking of a coastal front, or the tracking of manmade sources or marine mammals.

20.1 **Nested Autonomy** 460

20.2 **Concept of Operations (CONOPS)** 461
- 20.2.1 Field Level 461
- 20.2.2 Cluster Level 462
- 20.2.3 Node Level 462

20.3 **Autonomy** 463
- 20.3.1 MOOS-IvP Autonomy Architecture and System 463
- 20.3.2 The Payload Autonomy Paradigm 463
- 20.3.3 The MOOS-IvP Autonomy Architecture 464

20.4 **Acoustic Communication Infrastructure** . 466

20.5 **On-Board, Real-Time Signal Processing** 467

20.6 **Application Examples** 467
- 20.6.1 Unified Command, Communication, and Control Infrastructure 467
- 20.6.2 Adaptive Thermocline and Acousticline Tracking 469
- 20.6.3 Bistatic Target Tracking (GLINT'10) 474

20.7 **Conclusion** 478

References 479

The primary motivation for designing a distributed command and control architecture for undersea monitoring and observation is to achieve the ability to deploy a fleet of autonomous mobile marine platforms over a wide area of the ocean environment and over a long period of time with little or no human supervision. Concerns over effective coverage, communication range, and safe operation of the platforms are all primary motivations of an effective form of autonomous control. The long duration of missions and unpredictable nature of the environment require the vehicles to adapt their missions and behave autonomously as events unfold. Conversely, practical concerns of marine operations over large areas require an element of operator intervention over the course of time. These two characteristics can be at odds with each other in practice, but can be tempered by effective periodic communication through a network of fixed and mobile nodes co-deployed in a coordinated manner designed to balance individual platform and network objectives. The connectivity with and between the submerged assets of such networks is almost entirely dependent on underwater acoustic communication, except for rare and time-limited surfacings. Consequently, the undersea network nodes must operate with a communication infrastructure with severely limited bandwidth. Current underwater communication technology can robustly provide a point-to-point channel capacity in shallow water of less than a few hundred byte km/min, close to ten orders of magnitude smaller than modern electromagnetic communication

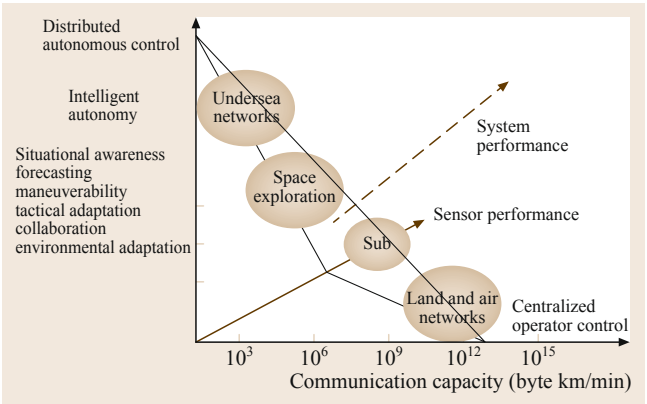


Fig. 20.1 Performance trade-off between sensor performance, communication channel capacity, and autonomy for net-centric sensing and observation systems

protocols used for land- and air-based distributed, net-centric systems. Equally critical is the high latency and short communication windows inherently associated with communication between the human operator and the submerged assets, more severe than that experienced in interplanetary space exploration. Operational constraints for some applications prohibit the existence of permanent surface assets which can provide a high-speed communication link with the operators. The connection of the operator to such systems is instead restricted to gateway vehicles, such as underwater gliders, which occasionally surface for a limited time and quickly relay short messages received acoustically from the submerged network nodes, and receive command and control commands which will subsequently be transmitted via the acoustic channel to the other nodes. The latencies using such a gateway vehicle on the continental shelf will typically be on the order of 10–60 min.

A typical acoustic or optical sensing system will generate data at a rate of the order of megabytes per second, for which the acoustic communication capacity of the undersea environment is totally inadequate for the transmission of raw data back to the operators. Therefore, in contrast to the air- and land-based equivalents, the data processing cannot be performed centrally but must be largely distributed to the individual nodes. Similarly, real-time *tethered* control of the underwater assets is made impossible by the latencies imposed by the use of occasionally surfacing gateway nodes. Consequently, real-time command and control decisions must be made locally on the nodes, in turn requiring that not only the data processing, but also the analysis and interpretation, traditionally performed by human operators, must be performed locally on the nodes. This requires fully integrated *sensing, modeling*, and *control*, a significantly higher level of autonomy than required in most current applications of autonomous underwater vehicles (AUVs) – where the data collection and the control have been handled independently.

In addition to allowing for autonomous reaction to sensor input, the higher degree of autonomy enables the adaptive control of the mobile nodes to take optimal advantage of the environmental and tactical situation through modeling and forecasting. As illustrated in Fig. 20.1, such onboard intelligent autonomy may compensate for the reduction in performance associated with the limited sensing capabilities of small underwater vehicles and the limited undersea communication channel capacity and latency.

For ocean monitoring and observation systems, an important mission objective for the network is the detection, classification, and tracking of episodic – usually unpredictable – events. Such events include chemical plumes from undersea volcanoes or man-made systems, and biological phenomena such as algal blooms. Another important application of undersea sensing systems is the detection and tracking of marine mammals and man-made sources of sound in the presence of ambient noise. Without the possibility of transmitting large amounts of data back to the operators, the on-board autonomy must be capable of fully completing the mission objective of sampling and characterizing the event entirely autonomously, without any human intervention or assistance.

In addition to autonomously *adapting* to such episodic events, the individual nodes may take advantage of *collaboration* with other nodes, again without requiring the human operator in the loop [20.1]. Thus, a cluster of network nodes within – at least occasional – acoustic communication range with each other may fuse its own data collected for the event with those obtained by and broadcast by other network nodes in the vicinity. For example, two AUVs with acoustic arrays may each track a marine mammal and collaboratively create an accurate localization solution by triangulation.

20.1 Nested Autonomy

To enable effective and fully autonomous adaptation and collaboration for an undersea network with its inherently severe communication constraints, MIT has developed an operational *nested autonomy* architec-

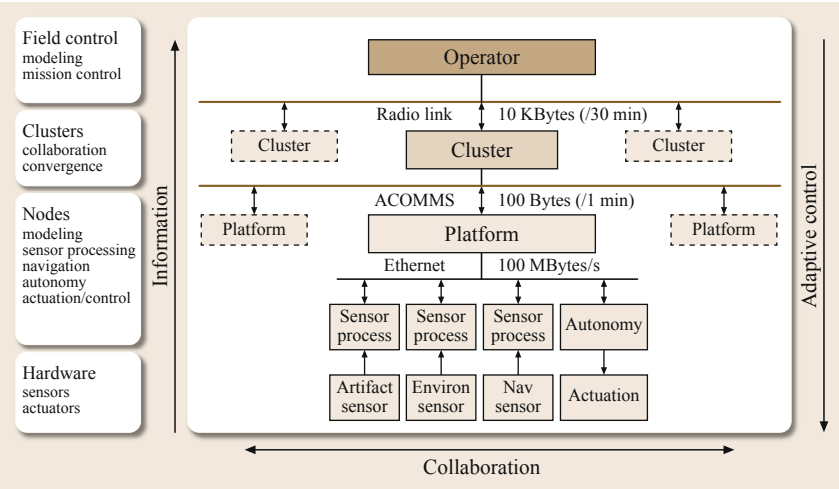


Fig. 20.2 Nested Autonomy. The field operator is communicating with clusters of autonomous nodes through gateway assets occasionally surfacing for transmitting target reports and receiving network commands. The node and cluster autonomy are designed accordingly

ture with fully integrated sensing, modeling and control within each AUV, clusters of assets, and the entire network [20.2].

The *nested autonomy* paradigm is exploiting the inherent layering of the communication infrastructure, illustrated in Fig. 20.2. The underwater network connectivity is being provided by low-bandwidth acoustic communication (ACOMMS), while the above-surface networking is handled by high-bandwidth, but latent, radio frequency (RF) communication through a regularly surfacing gateway node. On-board each node, the computer bus and Ethernet networking provides very high-bandwidth communication between the sensing, modeling and control processes. The three layers of horizontal communication have vastly different bandwidths, ranging from 100 byte/min for the inter-node ACOMMS to 100 Mb/s for the on-board systems. Equally important, the layers of the vertical connectivity differ significantly in latency and intermittency, ranging from virtually instantaneous connectivity of the on-board sensors and control processes to latencies of 10–60 min for information flowing to and from the field control (human) operators. As a result, adaptive control of the network assets with the operator in-the-loop is at best possible on an hourly basis, allowing the field operator to make tactical deployment decisions for the network assets based on, for example, environmental forecasts and reports of interfering shipping lane distributions, etc. Shorter time scale adaptation, such as autonomously reacting to episodic environmental events or a node tracking a marine mammal acoustically must clearly be performed either at the *node* level, or, if collaborative sensing is feasible, at the *cluster* level.

The *nested autonomy* concept of operations (CONOPS) does not entirely eliminate the operator from the decision process. Thus, whenever a communication opportunity arises, the operational paradigm will take advantage of any information that can be received from the operator or collaborators in the cluster. On the other hand, the intermittency of the underwater acoustic communication channel makes it imperative that each node is capable of completing the mission objectives in the total absence of communication connectivity.

20.2 Concept of Operations (CONOPS)

20.2.1 Field Level

The layered and clustered communication infrastructure illustrated in Fig. 20.2 naturally leads to a nested or layered concept of operations, which, as mentioned earlier, provides some optimal mixture of distributed autonomy and centralized control. Figure 20.3 shows a possible field-level concept of operations for an oceanographic observation system for capturing an episodic event, such as a chemical plume released by an undersea volcanic event. The target area is populated by a number of clusters, each with a number of mobile assets such as AUVs and gliders.

One of the adaptive responsibilities of the operators is to deploy the finite number of clusters in a pattern which is optimal for the current environmental situation and with the highest probability for capturing the episodic event of interest. The time scales for deploy-

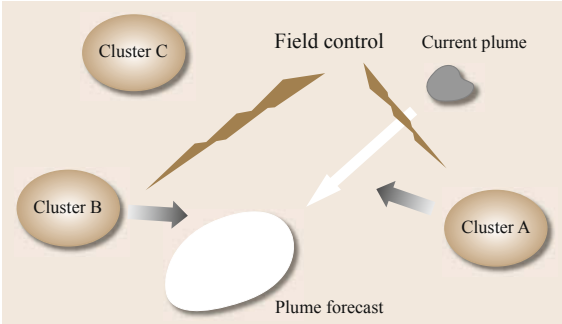


Fig. 20.3 Field-level CONOPS. The field control is dispatching clusters to autonomously prosecute a chemical plume with a forecast path and expansion. Cluster A is instructed to initiate prosecution immediately since it is closest to the projected path. The *downstream* cluster B is alerted to be ready for action, while field control decides not to activate cluster C, which is not in the path of the plume

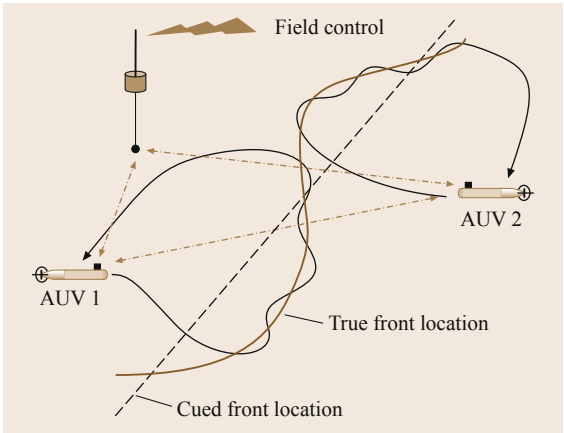


Fig. 20.4 CONOPS for a cluster of AUVs with oceanographic sensors, and a gateway buoy for communication with field control

ment and re-deployment are inherently long – on the order of hours to days – and is therefore highly dependent on reliable environmental and situational forecasts, often requiring a significant modeling and data assimilation infrastructure. Once deployed, it is assumed that each cluster is capable of autonomously detecting, classifying, localizing, and tracking (DCLT) the episodic event of interest. This event *prosecution* may be either cued by the operators through a surface communication gateway or performed fully autonomously. Once a tracking solution and the nature of the event are determined, the result of the prosecution will be reported back to the operators in the form of an *event report*. The human operators may then cue other clusters in the projected path of the event with whatever information is available, packaged into the format suitable for transmission through the network, for example, using the dynamic message coding scheme DCCL (dynamic compact control language) [20.3]. The final crucial role of field control is the fusing of the *event reports* from the various clusters in the path of the event, gradually building up a more and more complete event track and description.

20.2.2 Cluster Level

Depending on the available assets, a wide spectrum of cluster compositions is conceivable, including gliders and propelled AUVs with chemical, biological, and acoustic sensors. Figure 20.4 schematically shows how such cluster assets may be applied in response to an event cueing message from field control. The message identifies a front with a location and heading indicated by the dashed line. After the message is received by a surface gateway buoy, it is broadcast using the acoustic modems. Nearby nodes, such as dormant, drifting, or bottomed AUVs which pick up the message, will initiate a prosecute behavior sequence – in this case the detection and subsequent mapping and tracking of a frontal boundary. Depending on the level of autonomy authorized by field control, the AUV may decide not to pursue the target event if there is little probability that it will come within detection range.

If two or more nodes are prosecuting the event, each node may fuse the event information from the other nodes to produce a more accurate event characterization, and to optimize the coverage or resolution. Thus, in Fig. 20.4 the two vehicles coordinate their survey in order to not overlap and to increase coverage. Another example of collaborative control is a node which did not receive the original *prosecute* command, but which, following the receipt of an event report from a prosecuting node, will determine whether the target event is likely to come within the range, and then autonomously initiate a Prosecute sequence. All event reports generated by the prosecuting nodes are then collected by the communication gateway and transmitted back to field control via RF communication.

20.2.3 Node Level

A suite of node-level CONOPS has been developed for both single node and collaborative detection and tracking of a variety of episodic events, such as the adaptive mapping of a front or a thermocline, and for tracking an acoustic source, such as a marine mammal or a man-made source of sound.

As an example, Fig. 20.5 shows the core adaptive prosecute sequence developed for a propelled AUV for

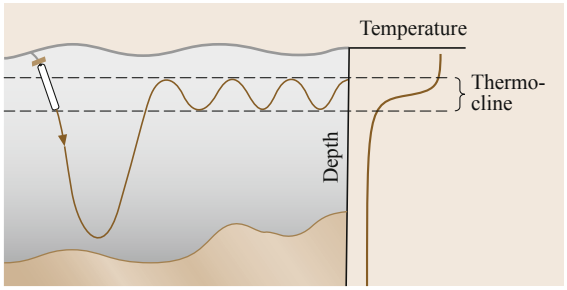


Fig. 20.5 Concept of operations for an AUV detecting, classifying, and tracking a coastal thermocline (after [20.4], courtesy of *Petillo* et al.)

DCLT a shallow water thermocline. The node level, CONOPS are described here for the tracking of a thermocline, but they can be directly mapped onto any other episodic event in the ocean environment, for example, the tracking of a plume, where the collaborative, adaptive cluster autonomy is even more important by providing simultaneously the resolution and coverage required for accurately localizing, classifying, and tracking the event. Thus, the network must first detect and localize the plume, and then adaptively track its boundaries, a mission which obviously requires the vehicles to collaborate to cover the expanding spatial extent of the plume.

20.3 Autonomy

20.3.1 MOOS-IvP Autonomy Architecture and System

The core of the nested autonomy paradigm is MOOS-IvP (the mission-oriented operating suite, with interval programming (IvP)): the autonomous, integrated sensing, modeling, and command and control framework on each individual platform. In combination with the collaborative cluster autonomy, the integrated node autonomy enables the adaptation which may compensate for the reduced physical sensor apertures of the unmanned underwater vehicles. The design of the autonomy system is based on three basic architectural components, extending the nesting into each individual sensor node:

- *Payload/Backseat vehicle architecture*: Low-level vehicle control is separated from the platform autonomy software, with the latter operating on a separate payload computer. This allows the same payload software and payload hardware to be integrated in vehicles of different size and different vehicle manufacturers.
- *Publish and subscribe software application architecture*: The payload software system is comprised of several distinct applications. The decision-making, sensor processing, communications handling, data logging, and many other applications are coordinated by the MOOS publish–subscribe middleware. The core MOOS middleware is lightweight, having no external dependencies and less than 1 Mb compiled size. The 2012 release, MOOS V10, allows for improved high-bandwidth and low-latency communications, suitable for use on the Oxford autonomous car project. Hundreds of MOOS applications have been written for use on at least 20 different marine vehicle platform types. Applications are largely independent, defined only by their interface. Any application is easily replaceable with an improved version with a matching interface. Core MOOS and many common applications are publicly available along with source code under an open source LGPL license (GNU Lesser General Public License).
- *Behavior-based decision-making architecture*: The IvP-Helm is a single MOOS application containing its own architecture of modular components – behaviors. The mission mode determines which behaviors are active, and competing behaviors are coordinated using multiobjective optimization using interval programming [20.5]. Several common and powerful behaviors are available at www.moos-ivp.org, but users may augment this core capability with their own public or proprietary behaviors to suit new mission objectives.

20.3.2 The Payload Autonomy Paradigm

To allow the MOOS-IvP network control to be applied on a variety of fixed and moving nodes with different control software, a *payload autonomy* paradigm was adopted and integrated with the MOOS-IvP control software infrastructure, and has subsequently been integrated into a wide variety of autonomous underwater vehicles and surface craft. This is achieved by adopting a hardware and software architecture that physically separates the sensing, communication, data processing, and associated adaptive autonomy from the basic platform control, illustrated schematically in Fig. 20.6. The idea is that all high-level control including the adaptation to measured and estimated parameters, is performed on a payload computer (PLC) running MOOS middleware, and including the IvP-Helm autonomous decision-making engine. The payload will also handle all communication with the network, ei-

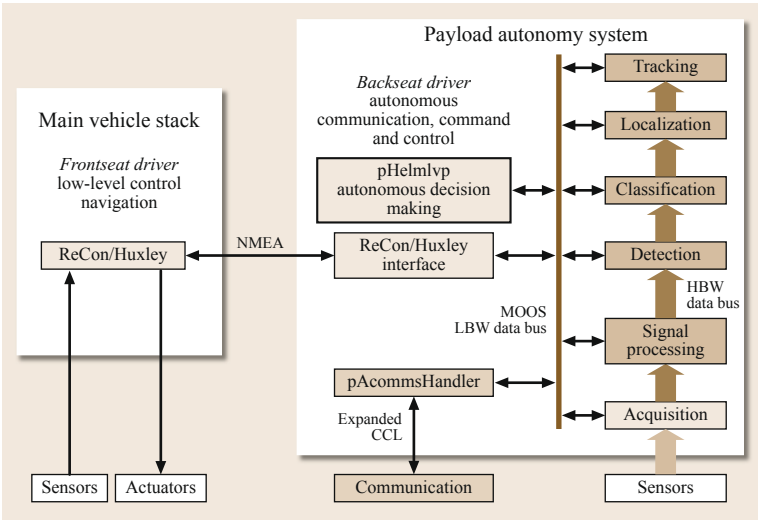


Fig. 20.6 Schematic representation of the payload autonomy paradigm, where higher level adaptive control and network communication are handled by MOOS-IvP on the payload computer, while lower level control, navigation, and vehicle safety are handled by the main vehicle computer using the native control software

ther through a radio link while surfaced, or an acoustic modem when submerged. All lower level control, and basic navigation and platform safety tasks are handled by the native vehicle control software running on the main vehicle computer (MVC), for example, *Huxley* on Bluefin vehicles and *Recon* on Remus vehicles. The communication between the PLC and the MVC is performed over a manufacturer-specific NMEA-type interface (National Marine Electronics Association), operated by a dedicated MOOS process. The commands passed from the PLC to the MVC are simply continuous updates of desired heading, speed, and depth, which the MVC then translates to desired rudder, thrust and elevator signals to the tail cone. The MVC will provide the PLC with a data stream containing all relevant navigation data. Thus, in a traditional ship analogy, the PLC represents the bridge, the radio room and the sensing infrastructure, while the MVC represents the engine room and the navigation resources of the ship. In the same analogy, the Helm represents the *Captain*, while the interface MOOS module represents the *Helmsman*.

The MVC will also perform a series of basic safety tasks, including mission aborts due to bottom altitude limit violations, lack of commands from the PLC within a specified time, or an overall mission timeout. Higher level safety tasks such as exceeding the specified operational area, and individual behavior timeouts or failures, are handled by the PLC.

20.3.3 The MOOS-IvP Autonomy Architecture

The *nested autonomy* paradigm for distributed undersea sensing inherently involves reaction to situations and events that are deterministically unpredictable. Thus, the autonomy architecture cannot be based on the availability of a world model that can form the basis for the autonomous decision making. Instead, it requires a capability of fully autonomously adapting to the environmental and tactical situation associated with the phenomenon it is intended to measure. As such, it forms a clear example of the type of robotic system for which the IvP model for multiobjective behavior coordination was intended and developed. Thus, for example, an underwater vehicle tasked with detecting and tracking an acoustic source is faced with several, often conflicting objectives. It will likely have been assigned a station point, from which it should not move too far, while at the same time having to get close to the source to develop a reliable tracking solution. Also, depending on its sensing capability it may have a preferred heading for achieving tracking resolution. Also, if other vehicles in the vicinity are already tracking the target event, it may not be desirable for it to pursue the same source aggressively, but instead preserve power for future sensing tasks. MOOS-IvP provides exactly the flexibility and inherent multiobjective capability for implementing such high-level autonomy with adaptive and collaborative capabilities.

The MIT-LAMSS group's implementation of the *nested autonomy* concept of operations is using the MOOS middleware, supporting the modular *star* software architecture illustrated in Fig. 20.7. As in all MOOS communities, the MOOS Database (MOOSDB) process is the core of the MOOS architecture and handles all communication between the processes (applications) using a publish-and-subscribe architecture. The various MOOS processes include all necessary control functions as well as sensing and processing modules, with the MOOSDB providing the unified interface standard that enables the fully autonomous integration of

sensing, modeling, processing, and control. MOOS ensures a process executes its *iterate* method at a specified frequency and handles new mail on each iteration in a publish-and-subscribe manner. The autonomy (IvP) helm runs as the MOOS process pHelmIvP.

The fundamental architecture of the IvP-Helm autonomy configuration at the core of the nested autonomy paradigm is illustrated in Fig. 20.8. Because of the inherent latency and intermittency of the underwater communication environment, the mission-dependent autonomy configuration defines a finite set of autonomy *modes* in which the autonomy will remain perpetually until specifically reassigned through a high-level transition command, either from the topside command and control or an onboard mission control process. The autonomy mode structure is configured in a hierarchical tree structure, with the mode transitions achieved by a simple change in a MOOS control variable, thus requiring very limited communication from the operators. Note that this paradigm is in distinct contrast to traditional *scripted* autonomy, where mode transitions are in general predefined.

Another fundamental architectural principle illustrated in Fig. 20.8 is that each mode has a preconfigured set of *behaviors*, each of which defines a set of objective functions for *speed*, *heading*, and *depth*, representing the utility of all allowed values for these variables. As in the case of the modes, the behaviors are perpetual, which means that they are running throughout the mission whenever the mode is active, although they may not contribute an objective function unless certain conditions are met. For example, the vehicles will continuously run a collision avoidance behavior, but it will not be active unless the node to avoid is within a configurable spatial envelope. The central IvP algorithm (the *captain*) will then continuously calculate and publish an objective compromise, which will be passed on to the MVC via the MOOSDB and the interface process (the *Helmsman*).

The *nested autonomy* paradigm in general allows the mode transitions to be entirely arbitrary, which makes it inherently suited to sensor-adaptive mission execution. Thus, for example, an underwater vehicle operating in a *deploy* mode such as a hexagonal loiter will use its onboard sensor processing to *detect* and subsequently *classify* an episodic event to which the node must respond. Once the processing is confident in assessing the desired nature of the event, it will simply change a MOOS variable which will trigger a mode transition in the Helm to a *prosecute* mode, which will activate a set of behaviors that allows the vehicle to map and track the event.

Mode transitions may alternatively be triggered by a simple command from the operators, received via the

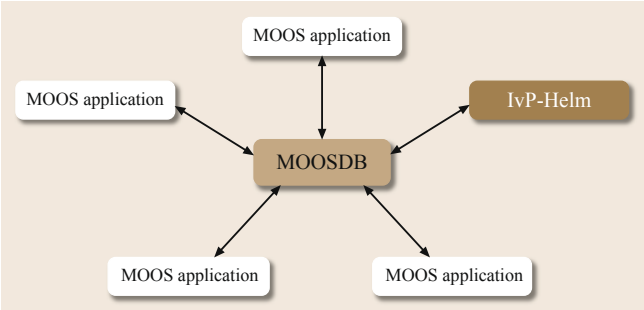


Fig. 20.7 Modular software architecture based on the MOOS middleware

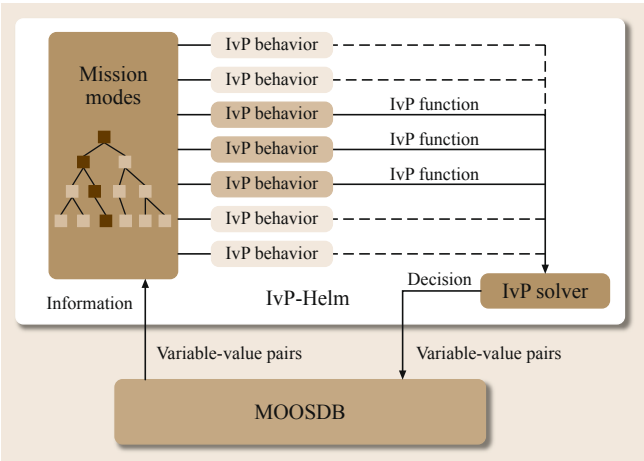
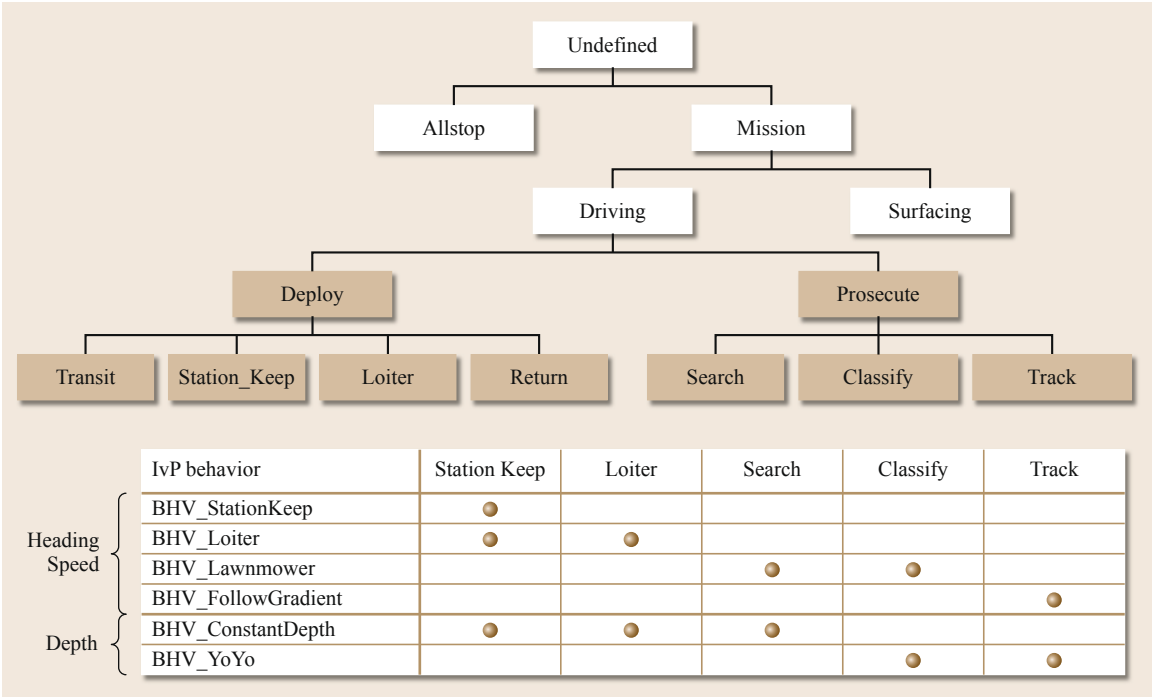


Fig. 20.8 IvP-Helm autonomy architecture

communication infrastructure. Also, depending on the configuration, transitions may be initiated by an *event report* issued by a collaborating fixed or mobile node or the operators. The fact that mode transitions can be initiated through various channels is a key feature ensuring robustness. Thus, for example, a node which has not itself been able to detect an event can be alerted through one of the other channels and consequently participate fully in the event prosecution.

An example autonomy mode hierarchy for vehicles in a network deployed for capturing episodic oceanographic events is shown in Fig. 20.9. During a mission, a vehicle will always reside in one of the modes at the end of a branch, and it will as a fundamental principle remain in that mode until it is commanded, internally or externally, to transition to another mode. Each mode defines a set of behaviors, most of which are generally available off-the-shelf. A typical behavior set is shown in the table in Fig. 20.9.

The perpetual mode/behavior architecture provides an extremely effective basis for executing field missions. Thus, except for configuration variables such



	IvP behavior	Station Keep	Loiter	Search	Classify	Track
Heading Speed	BHV_StationKeep	●				
	BHV_Loiter	●	●			
	BHV_Lawnmower			●	●	
	BHV_FollowGradient					●
Depth	BHV_ConstantDepth	●	●	●		
	BHV_YoYo				●	●

Fig. 20.9 Autonomy mode hierarchy for autonomous detection, classification and tracking of an episodic oceanographic event such as a coastal front

as the operational area and initial deploy location, the autonomy software, in general, does not require modifications before each individual launch, and even the sensing mission itself may be modified after deployment because of the capability of issuing the vehicle a high level, simple command which switches the mode, modifies variables such as waypoint locations, or activates and de-activates onboard sensing resources. Thus, for example, a vehicle executing a sonar survey can with a single command be switched to a mission mapping the temperature and salinity over depth, as long as that particular survey mode and the associated behaviors (race track, depth-yoyo, etc.) have been defined in the autonomy configuration.

20.4 Acoustic Communication Infrastructure

The nested autonomy paradigm is inherently autonomy centric, with the objective of making it robust to the severely constrained undersea communication environment with low bandwidth, high latency, and most importantly, severe intermittence imposed by the underwater acoustic environment. However, the operation of the observation network is still dependent on occasional communication to the vehicle in the form of commands changing the platform mode and configuration variables. Also, there is a need to send *status*, *contact*, and *track* reports to the operators to allow them to make informed decisions regarding progress of the mission. Finally, for many ocean sensing missions, the environmental assessment performance can benefit significantly from collaborative platform behaviors because it has the potential of breaking the space-time ambiguity inherent to measurements made by individual moving platforms.

The MIT nested autonomy operational paradigm is using the Goby [20.6] communication infrastructure and a highly flexible dynamic compact control language (DCCL) (*Schneider* et al. [20.3]), which together with a new Goby-Acomms communication marshalling, queuing and link layer [20.6, 7] provides a highly portable and efficient, unified command and control architecture. This allows field deployments of undersea networks of modem-equipped AUVs with MOOS-IvP autonomy to become routine exercises [20.2].

With each network node being directed by the MOOS-IvP platform autonomy system, the operational

paradigm enables fully autonomous adaptation of the mobile network nodes to the environmental and tactical picture, collaborative target event tracking by multiple platforms, and safe and efficient operation in uncharted environments without the need for re-programming. Once deployed, the entire network is operated using only the DCCL messages for communication between nodes and human operators for changing mission objectives and platform states. The Goby-DCCL is interfaced to the MOOS-IvP platform autonomy by the process *pAcommsHandler*, as shown in Fig. 20.6, providing the following capabilities and properties to the autonomy system:

- Highly portable with most of the software being hardware independent, with generic message handling all the way down to the physical modem driver.
- The DCCL encoding/decoding provides highly efficient data compression through a user-defined message composition with arbitrary value intervals and resolution.
- Dynamic queuing allows for high-priority messages to move to the head of the queue, with the priority of less time critical low-priority messages such as *status* reports increasing with time. This ensures that the message queue not be saturated by high-priority, short-time validity messages such as *track* reports.
- User defined time division multiple access (TDMA) communication scheduling, either using a centralized polling scheme, a fixed slotted scheme, or a dynamic self-discovering slotted scheme.

20.5 On-Board, Real-Time Signal Processing

A key to the autonomous, adaptive sampling of chemical, biological, physical, or acoustic fields in the ocean is an efficient on-board implementation of a data analysis package that allows for real-time feedback to the platform control, allowing for the sensor-adaptive autonomous control of the platforms. MOOS-IvP provides a very effective infrastructure for achieving this due to its modular structure and well-defined communication infrastructure. Thus, as illustrated in Fig. 20.6, separating the signal processing chain into a sequence of MOOS processes will allow each step in the processing to not only take advantage of the navigation information available in the MOOSDB, but also have the possibility of providing feedback to the Helm for optimizing the processing performance. This real-time, closed-loop feedback is the key enabler of environmentally *adaptive* sampling by making processed event data immediately available to the autonomous control. Also, it allows the data processing to take advantage of information arriving from other network nodes and published in the MOOSDB, enabling more effective *collaborative* sampling when the communication environment allows.

20.6 Application Examples

In a collaborative effort between the NATO Undersea Research Centre (NURC), MIT, Woods Hole Oceanographic Institution (WHOI), and the Naval Undersea Warfare Center (NUWC), a series of experiments were carried out between 2008 and 2010, with the primary objectives being the demonstration of the performance of a network of underwater vehicles as receiver platforms for multistatic active sonar tracking, and their communication and control networking. In these experiments, up to seven AUVs were deployed, six of which were equipped with towed hydrophone arrays and operating in a common underwater acoustic communication network [20.2]. These experiments provided a unique opportunity for testing and demonstrating the performance of the nested autonomy control paradigm under realistic underwater communication constraints. The experiments were designed, in addition to the multistatic acoustic data collection, to allow the demonstration of fully autonomous oceanographic mapping and adaptive autonomous behaviors for optimal acoustic sensing and communication. They also provided a comprehensive testbed for the MOOS-IvP platform autonomy in general.

20.6.1 Unified Command, Communication, and Control Infrastructure

The GLINT'08, '09, and '10 experiments were carried out as part of a joint research project (JRP) on undersea sensing network technology (NURC project 4G4), involving NURC, MIT, WHOI, NUWC, and several Italian organizations. The experiments had several scientific objectives, relating both to the sensing

concepts, communication networking, and distributed, autonomous control.

The principal objective was to demonstrate the communication, command, and control of a hybrid platform suite, using a common communication infrastructure based on the WHOI Micromodem, and a common autonomy system for operating all mobile and fixed assets, based on the *payload autonomy* paradigm and the MOOS-IvP behavior-based autonomy software suite. The architecture had previously been integrated and demonstrated on the SCOUT kayaks, the Bluefin BF21 AUVs (Fig. 20.10), and several land robots at MIT. In preparation for and during GLINT'08, it was successfully integrated into the NURC OEX AUV and the NUWC Iver2 AUVs, both deployed in the experiment towing hydrophone arrays for multistatic acoustics. In addition, the architecture was partially integrated into the NUWC Folaga environmental sampler and two bottom moorings equipped with micromodems for undersea networking. The hybrid network with these assets is schematically shown in Fig. 20.11.

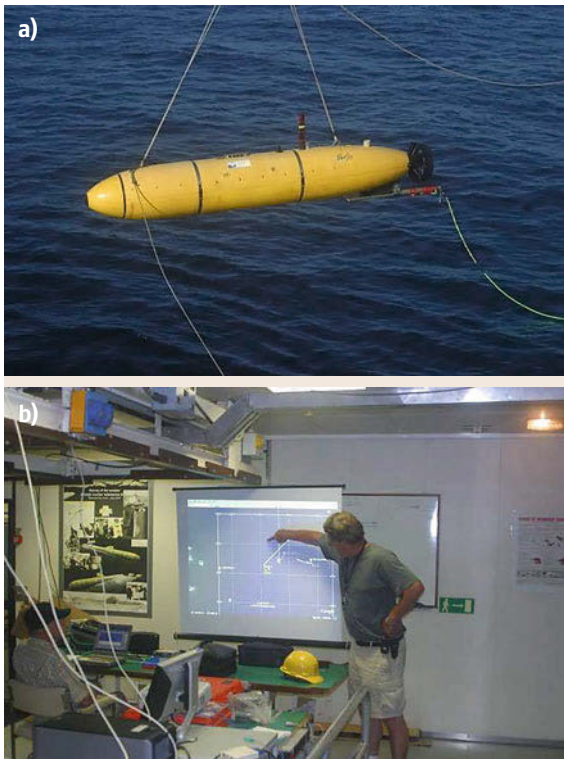


Fig. 20.10a,b GLINT'08 experiment, Pianosa, Italy, July–August 2008. **(a)** The Unicorn BF21 AUV with towed DURIP array (out of photo) being deployed from the NRV (NATO Research Vessel) Alliance. **(b)** The command and control center on the NRV Alliance with situational display

The principal scientific objective of GLINT'08 was to collect a comprehensive multistatic active dataset using three AUVs with towed hydrophone arrays, which will support the development of robust multistatic active processing approaches suited for operation in the limited computational environment of AUVs. The three vehicles were the NURC OEX with the 48-element SLITA array, the MIT Unicorn BF21 with the 32-element DURIP array, and the NUWC Iver2 vehicle towing a 16-element hydrophone array. The two large vehicles – the OEX and Unicorn – had fully integrated MOOS-IvP autonomy systems early in the experiment and were routinely used in coordinated data collection missions. On the last day of the experiment, all three array-towing vehicles were operated together. Also, the MOOS-IvP-DCCL communication infrastructure [20.2] allowed several demonstrations of fully autonomous obstacle and collision avoidance to be performed by Unicorn and OEX, as illustrated in Fig. 20.12, which shows the topside real-time situational display, which graphically displays all status and contact information transmitted from the vehicles via the undersea communication network.

A major accomplishment in GLINT'08 was the development of an enhanced report and command structure which allows for dynamic, optimally compressed, encoding and decoding of messages [20.2]. This new dynamic compact control language (DCCL, [20.3]) communication handler was implemented in MOOS-IvP and demonstrated for the real-time interleaved transmission of regular low-bandwidth FSK (frequency shift keying) messages with high-rate PSK (phase shift keying) coded messages, for up to 2 kb messages at 5.4 kb/s, allowing for the real-time transmission of CTD (conductivity, temperature and depth) (measurements and array signal processing products such as beam-time records (BTR) for real-time display on the topside situational display. The real-time topside display of BTR data from an AUV had not previously been achieved in the field. Acoustic communication messages from Unicorn and the other AUVs were assimilated with a heterogeneous mixture of other data sources (AIS, ship's NMEA, etc.) to give a unified situational display available to both the science crew and the ship's captain, as illustrated in Fig. 20.10b and Fig. 20.12a. Figure 20.12a shows an example of the usefulness of the situational display in a case of a run-away of one of the NUWC Iver2 AUVs. The last reported navigation for the vehicle was extrapolated in the topside command center to determine a possible grounding site on the island of Pianosa. The workboat was subsequently sent to the predicted site at the northern tip, and the vehicle was recovered from the rocks within 10 m of the predicted location.

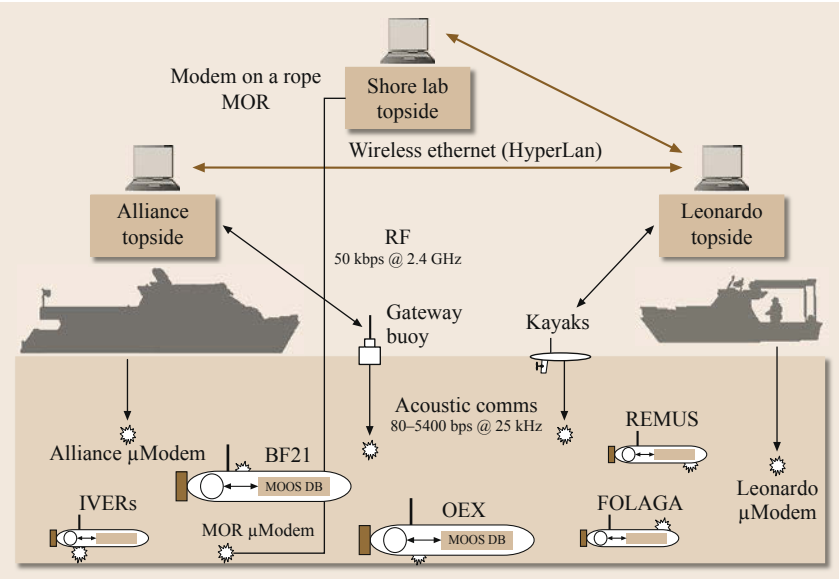


Fig. 20.11 GLINT'08 hybrid vehicle and communication networking

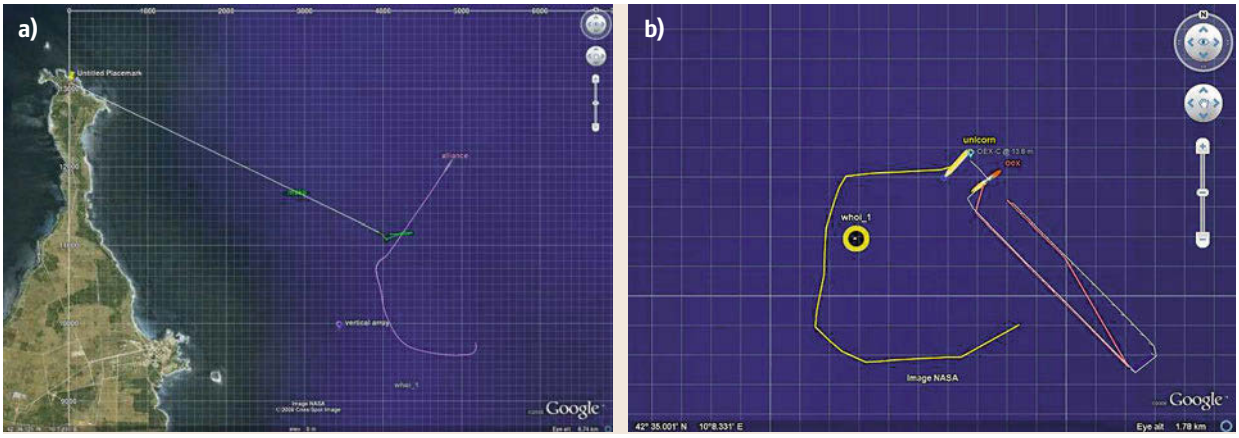


Fig. 20.12a,b Real-time topside situational display in GLINT'08 command center on-board the NRV Alliance. **(a)** Extrapolation of navigation data for a runaway Iver2 vehicle. **(b)** The topside rendering of Unicorn performing its obstacle avoidance and collision avoidance behaviors, with the WHOI Gateway buoy and the OEX AUV, respectively

20.6.2 Adaptive Thermocline and Acousticline Tracking

One of the primary applications of this nested autonomy system is the autonomous and adaptive detection and tracking of oceanographic features with AUVs. To this end, algorithms for autonomous and adaptive thermocline tracking have been developed and extensively tested in field experiments, demonstrating the feedback loop between AUV sensor measurements and adaptive motion using the nested autonomy system.

To implement adaptive thermocline tracking onboard AUVs, a MOOS application named pEnvtGrad was written to interface thermocline detection algorithms with the MOOS and IvP-Helm autonomy system. pEnvtGrad's final output simply consists of the upper and lower depth bounds of the thermocline region, as well as the depth at which the thermocline's temperature gradient ($|\Delta T/\Delta z|$) is largest. The upper and lower bounds of the thermocline are then used by an IvP-Helm behavior (BHV_ToggleDepth) to bound the vertical motion of the AUV, producing a depth-adaptive yo–yo pattern in depth and effectively tracking the thermocline depth (Fig. 20.5). In fact, pEnvtGrad also allows for similarly tracking the acousticline and pycnocline with the same algorithms, where sound

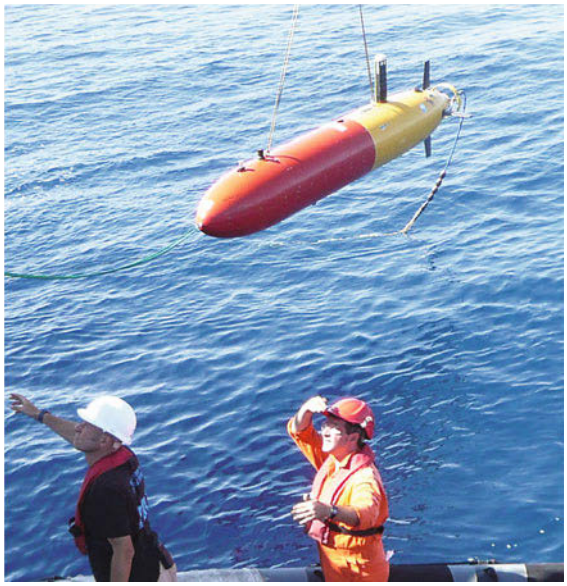


Fig. 20.13 The NURC OEX AUV during GLINT'09. This AUV uses acoustics to communicate with the ship while underwater and gets position updates via global positioning system (GPS) when surfacing (after [20.4], courtesy of *Petillo* et al.)

speed and density values are based on temperature and salinity measurements from the AUV's on-board CTD. The details of the thermocline detection algorithms and pEnvtGrad are given in [20.4].

Adaptive thermocline and acousticline tracking were demonstrated during the GLINT'09, Champlain'09, and GLINT'10 field trials, which are described below. The GLINT'10 experiment in particular used adaptive thermocline tracking missions in the broader context of collecting a synoptic multi-AUV data set displaying evidence of internal waves.

Acousticline Tracking (GLINT'09)

The GLINT'09 experiment was a collaborative effort between MIT and the NATO Undersea Research Centre (NURC, based in La Spezia, Italy) that took place in the Tyrrhenian Sea near Porto Santo Stefano, Italy. Adaptive acousticline tracking missions took place on 13–14 July, 2009, using the NURC OEX AUV (Fig. 20.13) running MOOS and IvP-Helm autonomy. The AUV was deployed from the NRV Alliance, where the topside AUV operators monitored the AUV's status via acoustic communication systems.

In preparation for at-sea testing, pEnvtGrad underwent development and testing in a simulation environment constructed from CTD data collected by the AUV in the same region earlier in the cruise. In developing pEnvtGrad, the acousticline was defined as the depth range over which the sound speed changes most rapidly per unit depth. For the associated in-water acousticline tracking missions that took place, the AUV was commanded into a north–south 1 km × 200 m racetrack pattern and performed the acousticline tracking as an adaptive-depth yo–yo pattern determined and autonomously updated by pEnvtGrad.

The actual depth history of the OEX AUV over its ≈ 2-hour acousticline tracking mission is shown in Fig. 20.14. The initial yo–yo is apparent as the deep dive from 7 to 70 m (*B*), which is followed by the adaptive acousticline tracking between 9 and 28 m depth (*C*). The vertical resolution of the acousticline tracking is based on sound speed values averaged over depth bins to smooth out any higher frequency variations in sound speed. In this case, the depth bins were chosen to be 1 m deep (given a water depth of about 105 m). As the AUV collected more sound speed measurements, these got averaged into the acousticline depth determination algorithms to update the acousticline bounds autonomously and adaptively. To avoid smoothing out all sound speed variations over time, a 30 min periodic reset was implemented to essentially restart the algorithm with a new initial yo–yo (*D*).

The postprocessed vertical sound speed and temperature profiles from the full 2+ hours of data collected during the acousticline tracking mission are plotted in Fig. 20.15a,b, respectively. The *solid vertical blue lines*

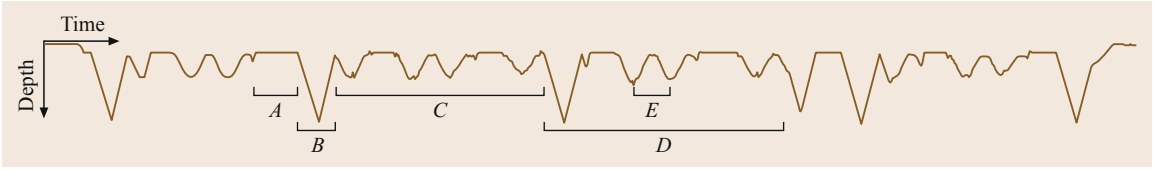


Fig. 20.14 Depth history of the OEX AUV during an adaptive acousticline tracking mission. (*A*) Default shallow turning and transiting depth (7 m). (*B*) Initial yoyo (7–70 m) performed by the AUV to ensure sampling of the entire water column down to the vehicle's maximum dive depth. (*C*) Adapted yo–yo tracking the acousticline between 9 and 28 m depth. (*D*) Is a 30 min tracking period after which the AUV re-initializes the yo–yo through the full water column to account for acousticline depth variation over space and time. (*E*) Is the 400 m period (length) of a single yo–yo (after [20.4], courtesy of *Petillo* et al.)

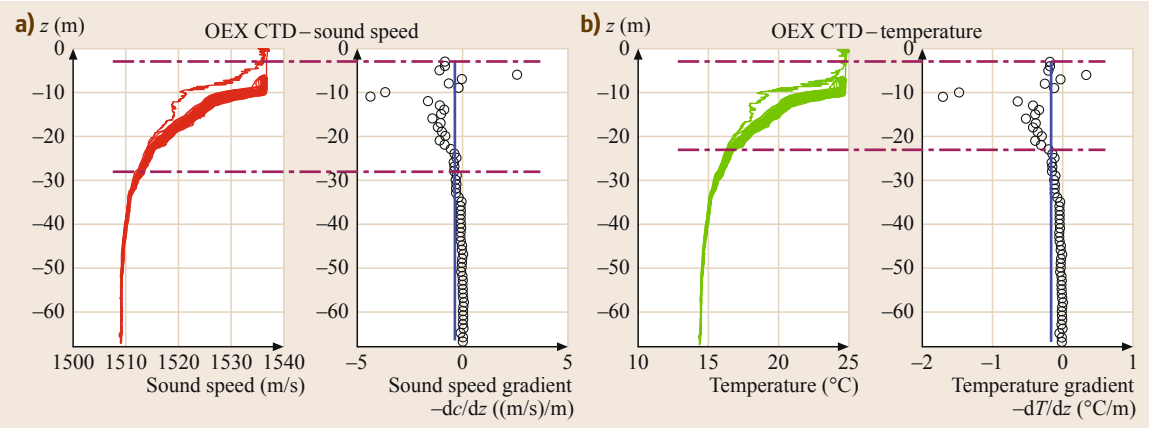


Fig. 20.15a,b Postprocessed sound speed **(a)** and temperature **(b)** profiles with acousticline and thermocline regions delinieated by the dashed lines in **(a)** and **(b)**, respectively (after [20.4], courtesy of *Petillo* et al.)

(on the gradient plots) represent the threshold values (average gradient over all sampled depths). A gradient greater in magnitude than the threshold magnitude is determined to be within the depth range of the acousticline or thermocline, respectively. The acousticline and thermocline regions are bounded by the *dashed lines* shown. When calculating the sound speed with the *MacKenzie* sound speed equation [20.8], the sound speed is dominated by temperature in shallow water (as it is here) and by pressure deep in the ocean. This results in similarities in the shapes of the temperature and sound speed profiles in Fig. 20.15. A thresholding method was used to delineate the acousticline and thermocline depth range, where the threshold was defined as the average of the sound speed and temperature gradients, respectively, over all depth bins. The threshold total average sound speed gradient from postprocessing $(\Delta c/\Delta z)_{\text{tot_avg}}$ was 0.427 (m/s)/m and the average acousticline depth range was calculated to be 3–28 m, where c is the sound speed through the water in m/s and z is the negative of depth in meters. Similarly, the threshold total average temperature gradient from postprocessing $(\Delta T/\Delta z)_{\text{tot_avg}}$ was 0.162 °C/m and the average thermocline depth range was calculated to be 3–23 m, where T is the temperature in °C.

The discrepancy between the minimum depth boundary from postprocessing and that calculated on board the AUV during acousticline tracking (3 m versus 9 m, respectively) is due to the depth range over which the calculations are being bounded (3–70 m versus 7–70 m, respectively), where the postprocessed data additionally include measurements taken during the AUV deployment and surfacing for GPS that skew the upper acousticline depth shallower by slightly decreasing the threshold value.

Fig. 20.16 The NUWC *Hammerhead* Iver AUV used during Champlain'09. This AUV carries a complete environmental package in its nose and communicates with the ship via RF (on the surface) and acoustics (underwater). It also carries a GPS and Doppler velocity logger (DVL) for positioning (after [20.4], courtesy of *Petillo* et al.)

Thermocline Tracking (Champlain'09)

The Champlain'09 experiment took place in Lake Champlain, VT, USA from 3rd to 5th October, 2009. A combined group from MIT and the Naval Undersea Warfare Center (NUWC, based in Newport, RI, USA) deployed an Iver AUV (Fig. 20.16) running MOOS and IvP-Helm autonomy software into this freshwater lake to test adaptive thermocline tracking missions. Since the Iver is a human-portable AUV, it was deployed off the side of a small motorboat and communicated with the operators on the boat via a 25 kHz WHOI Towfish acoustic transducer and Micromodem system. Lake Champlain was chosen due to its proximity to MIT and NUWC and for its deep center channel (> 100 m), which helps it support a stratified thermal structure that allows a thermocline to develop over the warmer months.

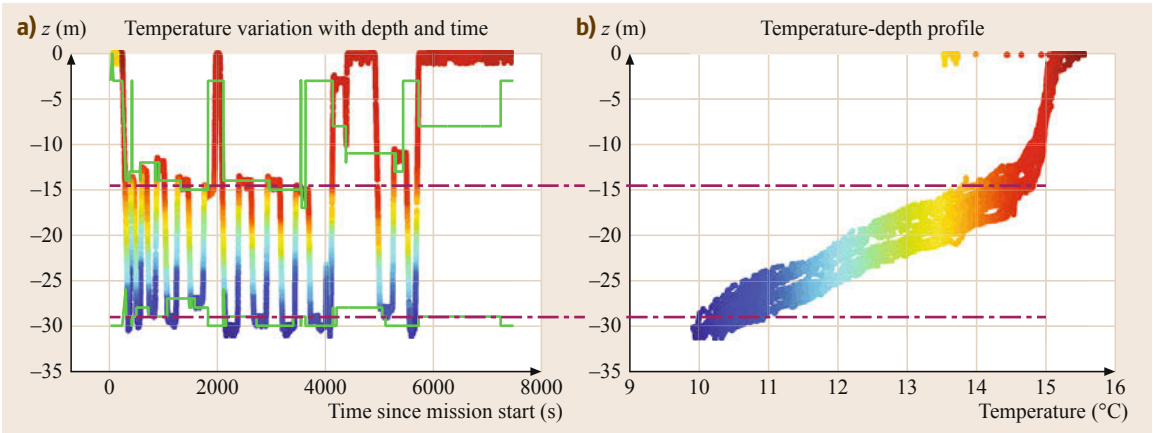


Fig. 20.17a,b Data taken from the real-time topside CTD display showing temperature variations over depth and time. (after [20.4], courtesy of *Petillo* et al.)

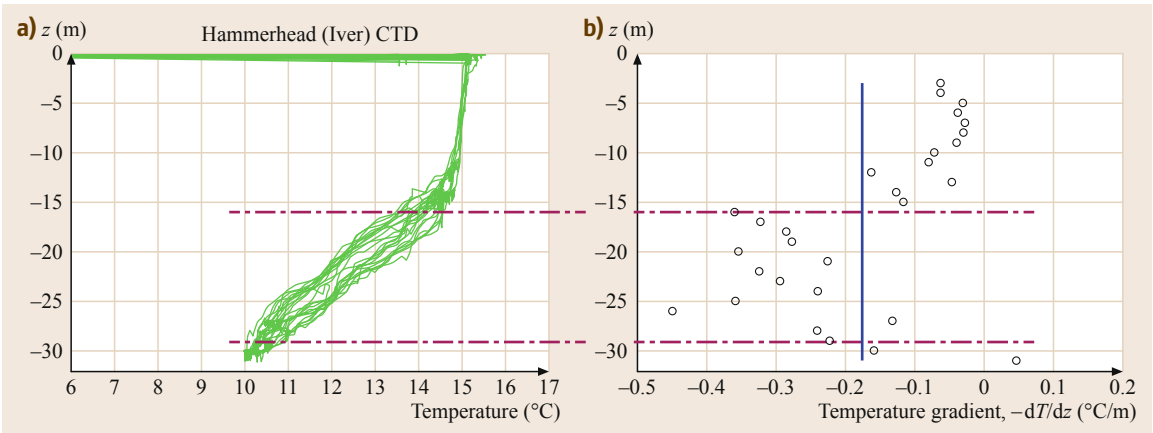


Fig. 20.18 **(a)** Temperature-depth profile over the entire mission (multiple dives), and **(b)** vertical temperature gradients averaged over 1-meter depth bins. (after [20.4], courtesy of *Petillo* et al.)

Champlain'09 was the second field trial of pEnvtGrad, where it underwent further testing and improvement while conducting adaptive thermocline tracking missions. In the horizontal plane, the AUV was deployed into a northwest–southeast straight line transect 1 km long. In the vertical, the AUV performed a depth-adaptive yo–yo pattern across the thermocline depth, which was determined autonomously by pEnvtGrad.

The results of one of these thermocline tracking missions are shown in Fig. 20.17, where the AUV was deployed for about 2 h total and was deployed into a thermocline tracking mission for the first 1.5 hours. Figure 20.17a shows the actual depth of the Iver AUV (multicolor points), with the colors corresponding to the temperature at the given depth, time, and location along the horizontal transect (not shown). The associated temperature values are plotted in the temperature-depth profile in Fig. 20.17b with the same color scale. In Fig. 20.17a, the AUV performs an initial yo–yo from 3 to 30 m as the first dive, gathering temperature data, and then determines autonomously that the thermocline is between about 14 and 29 m depth (smaller amplitude undulations) and starts tracking the thermocline. The thermocline depth bounds actively calculated on the AUV by pEnvtGrad are plotted as the green lines in Fig. 20.17a, which demonstrate the ability of the AUV to actively and autonomously adapt to changes in the thermocline depth boundaries (as small as 1 m) in real time. We chose 1 m depth bins because the water depth at the deployment location was of the order of 100 m, and the periodic reset was set at 30 min. The *dashed red lines* in Fig. 20.17 approximate (by inspection) the average thermocline bounds as determined by pEnvtGrad.

The postprocessed temperature data from the entire mission in Fig. 20.17 is shown in Fig. 20.18.

The average thermocline depth range was calculated as about 16–29 m (dashed lines, Fig. 20.18a,b) in postprocessing, with the total average temperature gradient $(\Delta T/\Delta z)_{\text{tot_avg}}$ of 0.168 °C/m set as the threshold value for bounding the thermocline range (solid vertical line, Fig. 20.18b). A gradient greater in magnitude than this average value's magnitude is determined to be within the depth range of the thermocline. When comparing the thermocline ranges determined by inspection of the AUV's actions in real time versus those calculated in postprocessing (dashed lines, Figs. 20.17a,b and 20.18a,b, respectively), it is apparent that they are very similar (within a couple of meters), verifying the effectiveness of the real time calculations and adaptation.

In this experiment, the muddy lake bottom at an imprecisely known depth posed a risk to the recovery of the AUV, thus we could not let it dive deeper than about 35 m. This directly bounded the range of depths over which we could collect temperature data, which affected the threshold value used to bound the thermocline. Thus, it is likely that we did not capture the full range of the thermocline during the mission. However, since the thermocline range was determined by pEnvtGrad to extend down to 30 m in the real-time data (for safety the maximum thermocline depth was bounded at 30 m from the initial yo–yo settings), this demonstrates the ability of the AUV to detect the majority of the sampled thermocline range with pEnvtGrad algorithms even in cases where data are unavailable over part of water column.

Thermocline Tracking for Internal Wave Detection (GLINT'10)

On August 13, 2010, MIT and NURC conducted the internal wave detection experiment (a single-day experiment in the larger GLINT'10 experiment) in the northern coastal basin of the Tyrrhenian Sea bordered by the Tuscan Archipelago and the western coast of Italy (Fig. 20.19). This experiment aimed to use nested autonomy and multiple AUVs to detect the presence of internal waves (or lack thereof) in this region of the Tyrrhenian Sea.

This experiment took a novel approach to internal wave detection by tasking two collaborating AUVs to autonomously adapt their motion in relation to each other and to their dynamic environment, resulting in a greater efficiency of sampling given a restrictive mission duration and in collection of fully synoptic data sets capturing internal waves.

The internal wave detection experiment involved two AUVs running the MOOS autonomy system guided by the IvP-Helm. These AUVs used acoustic communication during the experiment to send and receive real-time data and status updates, which they used to autonomously coordinate their motions in the horizontal plane through a track-and-trail behavior, as seen in the topside display in Fig. 20.20. In the vertical axis, the Unicorn AUV autonomously adapted to changes in the environment using the adaptive thermocline tracking behavior while the Harpo AUV (which would have also adapted, if the thermocline depth allowed for more reliable acoustic communication) swam just below the thermocline. A thermistor

Fig. 20.19 The region of the Tyrrhenian Sea bounded by the western coast of Italy and the islands of the Tuscan Archipelago. The Tuscan Archipelago basin is outlined by the dashed line. The GLINT'10 AUV operation region is delineated by the box. The numbering shows the five inlets of the basin (after [20.9], courtesy of *Petillo* and *Schmidt*)

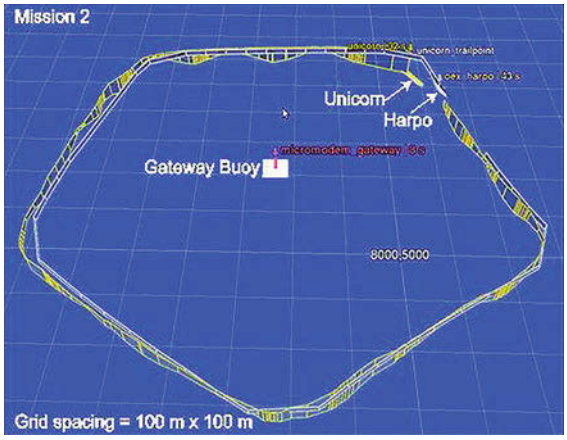


Fig. 20.20 Mission 2 loiter pattern around the gateway buoy, as seen from above, at an angle to the horizontal. Harpo performs a horizontal loiter pattern at constant depth (12 m) just below the thermocline. Unicorn trails directly behind Harpo while performing an adaptive yo–yo pattern through the thermocline depth range. *Vertical bars* along the loiter indicate the AUVs' depths (*yellow* is Unicorn's track, *white* is Harpo's track), and their current positions are shown by the *arrows* (after [20.9], courtesy of *Petillo* and *Schmidt*)

chain was also deployed for the duration of the experiment.

In examining the resulting AUV and thermistor data sets from this experiment, there is strong evidence of internal wave propagation along the thermocline near the buoyancy frequency of the thermocline interface ($N_{max} = 0.05747\,\mathrm{rad/s}$). Internal waves with nearly identical and lower frequencies were seen in the Unicorn, Harpo, and thermistor data collected near the thermocline depth. The 12 −m depth AUV and 11 −m depth thermistor results suggest the presence of buoyancy-supported internal waves along the thermocline (about 11 m depth) in the AUV operation region throughout the day on August 13, 2010. This conclusion may also be extrapolated to say that internal waves are likely detectable along the thermocline throughout the rest of the Tuscan Archipelago basin during the summer, when the thermocline is fairly well defined. Given the lack of previous literature regarding internal waves in the Tuscan Archipelago basin, this finding is rather significant to the scientific groups that conduct acoustic (and other) experiments in this region. See [20.9] for more detailed results and analysis from the internal wave detection experiment.

Overall, this experiment was novel in its use of multiple AUVs collaborating autonomously with each other and autonomously collecting environmentally adaptive data sets for more synoptic spatiotemporal data coverage. Not only does this increase the efficiency of data collection (environmentally-adaptive autonomy behaviors allow us to collect the exact data set we need without a human in the loop), but it also gives us the ability to collect the specific data set a scientist is interested in by using AUVs running autonomy. The use of intelligent acoustic communication networking also allows the AUV operators and scientists to monitor (from the topside on a ship or shore) the data collected in near real time. These abilities are invaluable when ship time for data collection is so expensive, and we hope that such improvements in AUV autonomy, adaptive environmental sampling techniques, and acoustic communications will allow us to further reduce necessary ship time for scientists and engineers to collect the specific data sets they need in the future.

20.6.3 Bistatic Target Tracking (GLINT'10)

As part of the GLINT experiment program and associated research, the nested autonomy paradigm was extended to the bistatic tracking of underwater targets. The concept of this application can be illustrated by the two-dimensional bistatic geometry in Fig. 20.21. The passive acoustic receiver (denoted as R), constructed as a hydrophone array, is towed by an AUV in this application. The active acoustic source (denoted as S) is deployed to transmit known sonar pulse signals at time t_k, where $k = 1, \ldots, N$ denotes the sonar ping number. The sonar pulse signals reach the receiver as two distinct dominant signals – the direct blast is the signal traveling straight from the source to receiver, and the indirect blast is the signal traveling from the source to receiver via acoustic scattering at the target (denoted as T). These signals are then recorded as hydrophone array data on the receiver. By processing the data, the target state – consisting of the Cartesian position and velocity – can be accurately estimated.

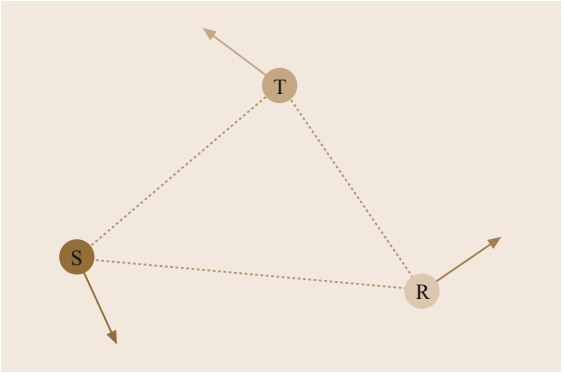


Fig. 20.21 Two-dimensional (2-D) bistatic geometry involving source (S), target (T), and receiver (R). The directional arrows indicate the velocities of the corresponding bistatic assets

Integrated Perception, Modeling, and Control

A new *integrated perception, modeling, and control paradigm*, as shown in Fig. 20.22, has been introduced for the AUV/receiver to solve the bistatic tracking problem [20.10, 11]. The *automated perception*, on the

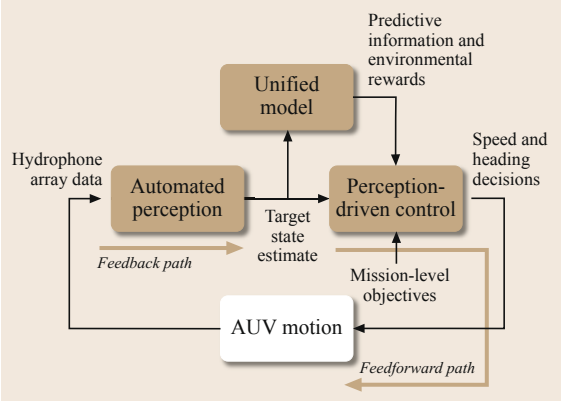


Fig. 20.22 Integrated perception, modeling, and control paradigm for the AUV/receiver to solve the bistatic tracking problem

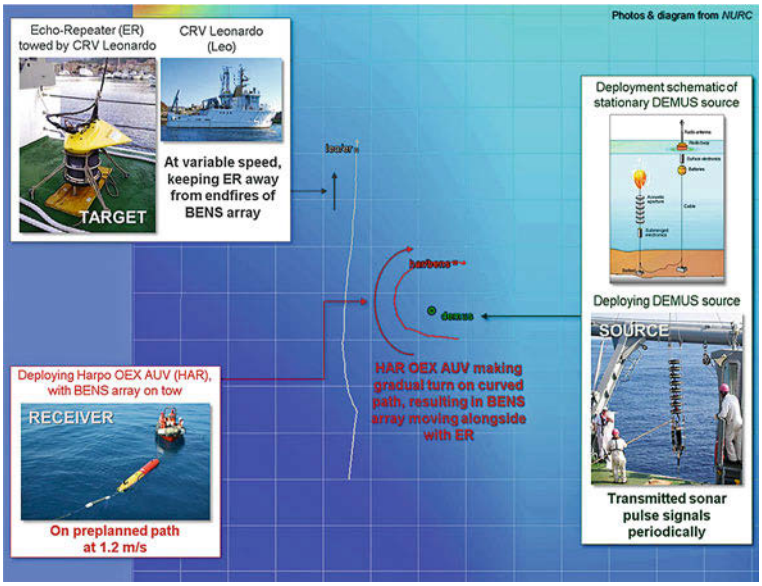


Fig. 20.23 Ground truths of bistatic assets for a GLINT'10 experimental run with photos and schematic diagram. The DEMUS source is plotted in *green*, the Harpo OEX AUV towing BENS array is plotted in *red*, and the CRV Leonardo towing the echo repeater (ER) is plotted in *gray*. The historical Cartesian positions of bistatic assets are depicted, with their current speed and heading indicated by the length and direction of the *arrow* at their current positions. The AUV was commanded to move along a preplanned path at 1.2 m/s – heading west first, before making a gradual turn on the curved path, and then heading east. CRV Leonardo was cruising northward at variable speed, trying to keep the target away from the endfires of the receiver

feedback path, is used to process the hydrophone array data from the receiver, and provide the target state estimate. This perception is performed using sonar signal processing and target tracking algorithms. The estimate is the perceptive feedback in the closed-loop control system for driving both the *perception-driven control* and *unified model*. The *perception-driven control* then deliberates the perceptive feedback against the mission-level objectives to make unsupervised decisions on the speed and heading for the *AUV motion* in the feed-forward path. No depth decision is considered here since Fig. 20.21 is dealing with two-dimensional geometry. By controlling the *AUV motion*, the vehicular path is adjusted adaptively. A new *bistatic behavior* has been developed in the *perception-driven control* to execute, in real time, a new nonmyopic and adaptive control for the vehicle. The predictive information and environmental rewards from the *unified model* are used to provide the IvP function for this new behavior. The *unified model* is amalgamated from both information theoretic and environmental acoustic models. These models are then used to predict the pertinent information and acoustic characteristics of the target state estimate for the particular bistatic geometry. The pertinent characteristics predicted for different future bistatic geometries, resulting from different discrete speed and heading decisions at the current target state estimate, are then encapsulated in the predictive rewards. In essence, the control formulation presents a new vehicular control that applies both information-theoretic and environmental-based controls concurrently.

AUV on Preplanned Path

To demonstrate the advantage of this new *integrated perception, modeling, and control paradigm*, the bistatic tracking results of the underwater target are compared with the AUV operating with and without this new paradigm. We will first examine the tracking results without this paradigm. The data we use are from the GLINT'10 experimental run, conducted jointly between MIT and NURC, in the Tyrrhenian Sea off the coast of Italy. The ground truths of the bistatic assets utilized for this experimental run, together with their respective photos and schematic diagram, are depicted in Fig. 20.23. The deployable experimental multistatic undersea surveillance (DEMUS) source was deployed as stationary asset, and transmitting sonar pulse signals. The echo-repeater (ER) was used to simulate a target by retransmitting the recorded sonar pulse signals from the source. It was towed by the CRV Leonardo (Leo). The receiver was the BENS (Broadband Environmental Networked Sensor) array, towed by the

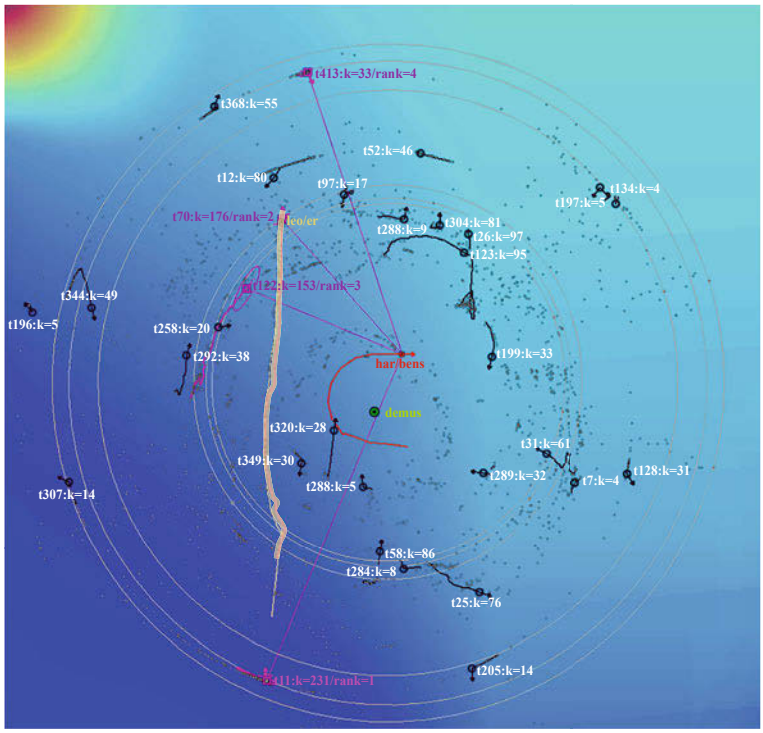


Fig. 20.24 Bistatic tracking results from *automated perception* for the GLINT'10 experimental run at ping $k = 234$. The ground truth of the DEMUS source is plotted in *green*, the Harpo OEX AUV towing the BENS array is plotted in *red*, and the CRV Leonardo towing the ER is plotted in *gray*. Confirmed tracks are plotted in *magenta*, terminated tracks are plotted in *black*, and information pertaining to these tracks is depicted

Ocean Explorer (OEX) AUV. The AUV was commanded to move along a preplanned path at 1.2 m/s, for the entire run from ping $k = 1$ to 235, without adopting the new paradigm. The AUV headed west first, before making a gradual turn on a curved path, and then headed east. The *automated perception* was used to process the hydrophone array data from the receiver and provided the target state estimate. However, this estimate was not used in the control. CRV Leonardo was cruising northward at variable speed, trying to keep the target away from the forward and aft endfires of the receiver array. This was done to ensure good tracking results on the target.

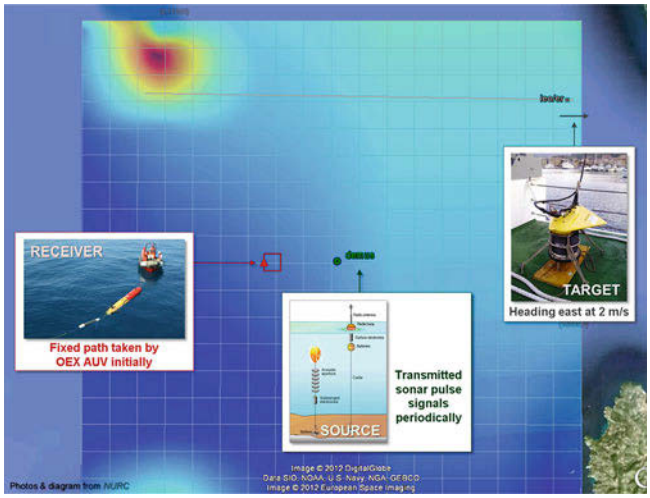


Fig. 20.25 A simulated tactical scenario with the stationary DEMUS source plotted in *green*, the initial fixed path taken by the OEX AUV plotted in *red*, and the CRV Leonardo towing the ER plotted in *gray*. CRV Leonardo was simulated to head east at 2 m/s

The bistatic tracking results from the *automated perception* for this experimental run at ping $k = 234$ are shown in Fig. 20.24. Both the confirmed and terminated tracks are clearly plotted, and the information pertaining to these tracks is depicted. For the confirmed track due to the ER, the information $t70 : k = 176/\text{rank} = 2$ implies that track #70 has been active for 176 pings since initiation and is ranked #2 in the active track priority. This track depicts the state estimated for the target over the past 176 pings. Track #122 is due to the ghost of the ER. This is caused by the starboard/port side ambiguity of the BENS array. However, this ghost is not well tracked because the constant maneuvering by the AUV, particularly during the gradual turn on the curved path, causes the track to violate the nearly constant velocity (NCV) target dynamic model assumed in the *automated perception*. Tracks #11 and #413 are due to the prominent fixed but unknown underwater objects that were persistently present since the start of the experimental run. With the target speed, good target tracking results are obtained for track #70.

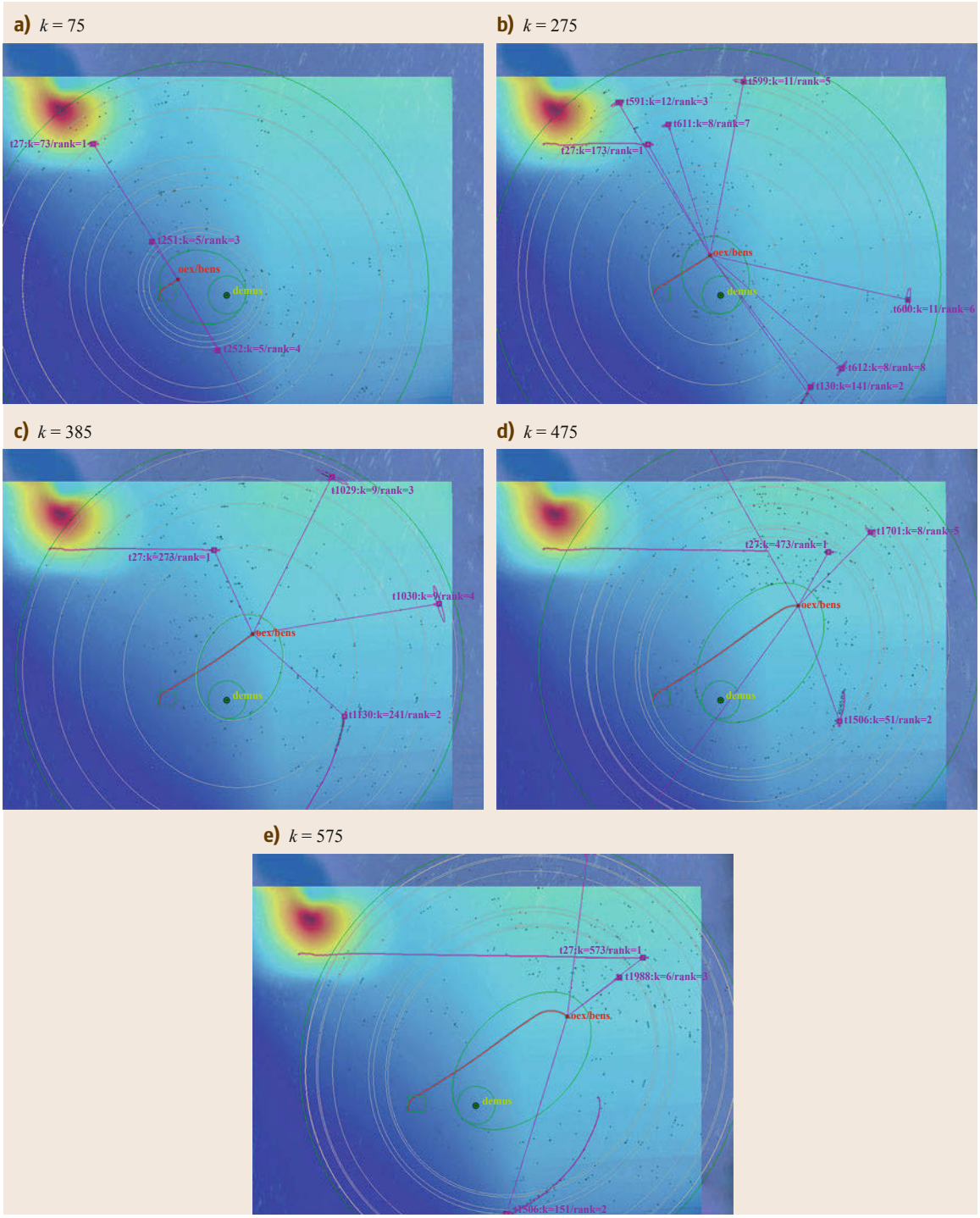


Fig. 20.26a–e Simulated results of a tactical scenario at pings $k = 75$ **(a)**, $k = 275$ **(b)**, $k = 385$ **(c)**, $k = 475$ **(d)**, and $k = 575$ **(e)** with the *bistatic behavior* applied on the OEX AUV and driven by target state estimates of the confirmed track #27 from the ER. The stationary DEMUS source is plotted in *green*. The OEX AUV towing the BENS array is plotted in *red*. The OEX AUV was initially moving on a fixed path, but switched to an adaptive path upon tracking the entry of the ER. The moving ER, towed by the CRV Leonardo, is plotted in *gray* (not visible) and was heading east. Confirmed tracks are plotted in *magenta*

AUV on Adaptive Path

In the previous section, the bistatic tracking results were obtained with the AUV on a preplanned path. The AUV was not reacting to the target state estimate obtained from the *automated perception*. The target speed was adjusted to keep the target away from the endfires of the receiver. This was done to ensure good tracking results on the target. Operationally, this is not realistic since the target is never cooperative. In this section, we will implement the *integrated perception, modeling, and control paradigm* on the AUV. The synthetic data, generated from a high-fidelity simulator, is used here with the same bistatic assets as the experimental data in Fig. 20.23. The simulated tactical scenario is shown in Fig. 20.25. For this scenario, the stationary DEMUS source was simulated to transmit sonar pulse signals. The ER, towed by the CRV Leonardo, was used to simulate a noncooperative target cruising eastward at a fixed speed of 2 m/s from ping $k = 1$ to 650. The OEX AUV, towing the BENS array, was commanded to move along the fixed path initially. With the new paradigm implemented on the AUV, the target state estimate of the ER from the *automated perception* is now used to drive both the *perception-driven control* and *unified model*.

The simulated results of the tactical scenario at different pings k are obtained in Fig. 20.26. The target state estimates of the confirmed track #27 due to ER from the *automated perception* were used to drive the *perception-driven control* and *unified model*. The AUV was initially moving on a fixed path, but switched to an adaptive path upon tracking the entry of the ER. This adaptive path was planned with the *bistatic behavior* in the *perception-driven control*, with the objective of maximizing the information and acoustic performance of the target state estimate. With the adaptive path taken by the AUV, good target tracking results are obtained for track #27.

From the results obtained in Figs. 20.24 and 20.26, it is clear that the *integrated perception, modeling, and control paradigm* allows the path of the AUV to be adaptively adjusted in reaction to the target state estimated from the tactical situation. The path has been planned with the objective of optimizing the target tracking performance. Such methodology can be easily extended for distributed nodes in an AUV network.

20.7 Conclusion

Being dependent on acoustic communication with a channel capacity many orders of magnitude smaller than the air- and land-based equivalents, the operation of distributed ocean sensing networks requires a much higher level of autonomous, distributed data processing, and control than land- and air-based equivalents. The *nested autonomy* paradigm and its underlying principle of *integrated sensing, modeling and control* described here are inherently suited for the layered communication infrastructure provided by the low-bandwidth underwater acoustic communication and the intermittent RF connectivity. Implemented using the open-source MOOS-IvP behavior-based, autonomous command, and control architecture, nested autonomy allows each platform to autonomously detect, classify, localize and track (DCLT) an episodic ocean event without depending on operator command and control. The prosecution of a particular event may be initiated by the operators through an acoustically transmitted compact command message, cued by other sensing systems, such as satellite remote sensing, for example, or fully autonomously by an on-board detection algorithm, but the fundamental principle of the paradigm is that each network node shall be capable of completing its missions objectives, irrespective of communication connectivity. The inferred properties of the event are reported back to the operators by transmitting a compact *event report*, using a dedicated command and control language. Collaborative processing and control may be exploited when the communication channel allows, for example, for collaborative tracking of extended features such as plumes and fronts. The nested command and control paradigm has here been demonstrated for the fully autonomous, adaptive tracking of the space- and time-varying coastal thermocline and acousticline, using relatively simple onboard signal processing, but the highly modular MOOS-IvP architecture allows the event characterization and tracking algorithms to be exchanged with customized and advanced processing software without altering the operational functionality of the autonomy system.

References

20.1 M. Benjamin: Multi-objective autonomous vehicle navigation in the presence of cooperative and adversarial moving contacts, Proc. IEEE OCEANS'02 (2002)

20.2 T. Schneider, H. Schmidt: Unified command and control for heterogeneous marine sensing networks, J. Field Robotics **27**(6), 876–889 (2010)

20.3 T. Schneider, S. Petillo, H. Schmidt, C. Murphy: The dynamic compact control language version 3, Proc. IEEE OCEANS'15 (2015)

20.4 S. Petillo, A. Balasuriya, H. Schmidt: Autonomous adaptive environmental assessment and feature tracking via autonomous underwater vehicles, Proc. IEEE OCEANS'10 (2010)

20.5 M. Benjamin, H. Schmidt, P. Newman, J. Leonard: Nested autonomy for unmanned marine vehicles with MOOS-IvP, J. Field Robotics **27**(6), 834–875 (2010)

20.6 T. Schneider, H. Schmidt: Goby-Acomms version 2: extensible marshalling, queuing, and link layer interfacing for acoustic telemetry, Proc. IFAC 9th Conf. Maneuvering Control Mar. Craft (2012)

20.7 T. Schneider, H. Schmidt: Model-based adaptive behavior framework for optimal acoustic communication and sensing by marine robots, IEEE-JOE **38**(3), 522–533 (2013)

20.8 K.V. MacKenzie: Nine-term equation for the sound speed in the oceans, J. Acoust. Soc. Am. **70**(3), 807–812 (1981)

20.9 S. Petillo, H. Schmidt: Exploiting adaptive and collaborative AUV autonomy for detection and characterization of internal waves, IEEE-JOE **39**(1), 150–164 (2014)

20.10 R. Lum, H. Schmidt: Exploiting adaptive processing and mobility for multistatic tracking by AUV networks, Proc. 4th Int. Conf. Underw. Acoust. Meas. Technol. Results (2011) pp. 1487–1495

20.11 R. Lum: Integrated Perception, Modeling, and Control Paradigm for Bistatic Sonar Tracking by Autonomous Underwater Vehicles, Ph.D. Thesis (MIT, Cambridge 2012)

21. Science of Autonomy: Time-Optimal Path Planning and Adaptive Sampling for Swarms of Ocean Vehicles

Pierre F.J. Lermusiaux, Tapovan Lolla, Patrick J. Haley Jr., Konuralp Yigit, Mattheus P. Ueckermann, Thomas Sondergaard, Wayne G. Leslie

The science of autonomy is the systematic development of fundamental knowledge about autonomous decision making and task completing in the form of testable autonomous methods, models and systems. In ocean applications, it involves varied disciplines that are not often connected. However, marine autonomy applications are rapidly growing, both in numbers and in complexity. This new paradigm in ocean science and operations motivates the need to carry out interdisciplinary research in the science of autonomy. This chapter reviews some recent results and research directions in time-optimal path planning and optimal adaptive sampling. The aim is to set a basis for a large number of vehicles forming heterogeneous and collaborative underwater swarms that are smart, i. e., knowledgeable about the predicted environment and their uncertainties, and about the predicted effects of autonomous sensing on future operations. The methodologies are generic and applicable to any swarm that moves and senses dynamic environmental fields. However, our focus is underwater path planning and adaptive sampling with a range of vehicles such as autonomous underwater vehicles (AUVs), gliders, ships or remote sensing platforms.

21.1 **Time-Optimal Path Planning for Swarms of Ocean Vehicles** 482
21.1.1 Canonical Steady Flows 483
21.1.2 Time-Dependent 2-D Flows 483
21.1.3 Maintain Swarms Formations 485
21.1.4 Forbidden Regions 485
21.1.5 Uncertain Flow Fields 486
21.1.6 Realistic Ocean Conditions: Three-Dimensional Multi-Scale Flows 487
21.1.7 Realistic Ocean Conditions: Swarms in Multi-Scale Flows with Complex Geometries 489
21.2 **Adaptive Sampling for Swarms of Ocean Vehicles** 489
21.2.1 Early Results: Approximate Schemes for Realistic Applications 490
21.2.2 Recent Progress: Towards Rigorous Schemes for Realistic Applications 492
21.3 **Conclusions and Outlook** 494
References 496

Ocean observing is the process of collecting ocean measurements to provide the most useful information about the ocean systems and dynamics of interest. Such marine sensing is expensive and challenging. In the ocean, not everything can be measured on a sustained basis. Ocean modeling is the process of developing and utilizing theoretical and computational models for the understanding and prediction of ocean dynamics. Similar to ocean observing, ocean modeling is challenging, and not every naval and environmental process can be modeled exactly. Stochastic ocean modeling explicitly acknowledges these uncertainties in modeling and thus augments deterministic predictions with probabilistic information. Data assimilation is the process of quantitatively estimating dynamically evolving fields by combining information from observations with those predicted by models, hence limiting uncertainties. With stochastic predictive capabilities and data assimilation, one can estimate future conditions and their uncertainties, and make informed decisions and complete desired tasks. Systems that integrate some of these activities without continuous human guidance are often referred to as autonomous systems.

The science of autonomy is the systematic development of fundamental knowledge about autonomous decision making and task completing in the form of testable autonomous methods, models, and systems. In ocean applications, it involves varied disciplines that are not often connected. For example, ocean dynamics, sensing, robotics, communications, predictions, and uncertainty, as well as varied types of human–computer–vehicle interactions are inputs to ocean autonomy. There are two important ocean modeling feedbacks to autonomy. The first one is the use of models to predict optimal

paths for swarms of ocean vehicles and the second is the use of models to predict the sensing expected to be most useful. Prediction of optimal paths for ocean vehicles is referred to here as *path planning*. This search for optimality often focuses on time-optimal or energy-optimal paths. The prediction of optimal sensing strategies is referred to as *adaptive sampling*. We note that in both cases, the prediction of paths can itself be improved adaptively in the sense that observations collected along the path can be utilized to update the predictions. Such updates can be completed onboard vehicles in the form of *onboard routing* or by central/hierarchical computational entities for *high-level routing*.

Recent autonomous marine activities have started a revolution that was possibly first imagined by *Henry Stommel* in his now famous article, *The Slocum mission* [21.1]. Today, the number of autonomous platforms used in semi-coordinated sea operations can be larger than 10, and this number is increasing. This new paradigm in ocean science and operations motivates the need to carry out interdisciplinary research in the science of autonomy. This chapter reviews some recent fundamental results and research directions in time-optimal path planning and optimal adaptive sampling, setting a basis for a large number of vehicles forming heterogeneous and collaborative underwater swarms that are smart, i. e., knowledgeable about the predicted environment and their uncertainties, and about the predicted effects of autonomous sensing on future operations. The methodologies are generic and applicable to any swarm that moves and senses dynamic environmental fields. However, the focus of this chapter is underwater path planning and adaptive sampling with a range of vehicles such as AUVs, gliders, ships, or remote sensing platforms.

When compared to related control and robotics problems of large dimensions, the differences with the present focus on ocean applications are that:

- Naval platforms are heterogeneous and their data are gappy but multivariate.
- Marine fields can strongly affect vehicles' operations, are dynamic on multiple-scales, and have very large dimensions, but are predictable to some degree.
- The measurements to be collected affect these future predictions.

Therefore, there are feedbacks between optimal planning, sampling and predicting, in time and space, and across variables. The field of ocean autonomy thus uses guidance from ocean flow and acoustic propagation modeling, information theory, dynamical system theory, uncertainty prediction, decision-making under uncertainty, machine learning, bio-inspired algorithms, and distributed computing. In all cases, our focus here is to predict global dynamic swarm patterns and perform high-level optimization. This is not the detailed fully-resolved control of a single robot. In our focus, a global objective function defines the optimal dynamic and collaborative autonomy. In general, objective functions depend on the predicted environment, on the predicted values and positions of the expected measurements, and on the feedbacks between measurements and predicted dynamics.

In what follows, we focus on the science of autonomy results obtained by our multidisciplinary simulation, estimation, and assimilation systems (MSEAS) group and recent collaborators. We first outline and present a subset of our results on time-optimal path planning for swarms of ocean vehicles (Sect. 21.1). This is followed by an overview of some of our results in adaptive sampling (Sect. 21.2). Our objective is to outline overall schemes and approaches, and to illustrate their applications in idealized and realistic ocean conditions. For detailed descriptions of the algorithms and methodologies we developed and utilized, we refer to references cited in each of these sections. For general reviews on oceanic path planning, we refer to [21.2] and [21.3], and for general reviews on oceanic adaptive sampling, to [21.4–8], and [21.9]. Brief conclusions and outlook are given in Sect. 21.3.

21.1 Time-Optimal Path Planning for Swarms of Ocean Vehicles

Our new level-set based method for time-optimal path planning of vehicle swarms in time-varying uncertain velocity fields (ocean currents) was developed and described in [21.2, 3, 10–14]. To forecast the time-optimal trajectories of large numbers of robotic vehicles navigating in strong and dynamic flows, our approach first evolves the reachable sets from the starting points. A reachable set is the set of points that can be visited by a vehicle at a given time. The boundary of such a set is called the reachability front. Our approach keeps track of the reachability front at all times, so that we can determine when the front first reaches the end point. A path traced by a point on the reachability front that first reaches the end point is an optimal path that we wish to compute.

We obtained a modified level set (21.1) for ϕ^o, the scalar field that governs the evolution of reachability fronts [21.3, 10, 11, 14], linking it to the Hamilton–

Jacobi equation governing the minimum-time navigation problem in dynamic flows. The reachability fronts are evolved from the vehicle start points y_s until they reach the desired end points y_f, combining nominal vehicle motions due to steering and advection by the dynamic flow environment, as shown by (21.1). An optimal trajectory $X_P^*(y_s, t)$ and the corresponding vehicle heading directions are then extracted from the evolution of the reachability fronts. This is done by solving the backtracking (21.2) where the solution ϕ^o is differentiable, starting from an end point y_f and integrating backward in time to the start point y_s

$$\frac{\partial \phi^o}{\partial t} + F|\nabla \phi^o| + v \cdot \nabla \phi^o = 0\,, \tag{21.1}$$

$$\frac{\mathrm{d}\, X_P^*(y_s, t)}{\mathrm{d}t} = -v(X_P^*, t) - F\frac{\nabla \phi^o(X_P^*, t)}{|\nabla \phi^o(X_P^*, t)|} \quad \text{with } X_P^*(y_s, T^*(y_f)) = y_f\,. \tag{21.2}$$

We also extended the methodology to the case of coordinated motions of swarms of vehicles and to the case of uncertain stochastic flows. Coordinated motion in this context refers to maintaining swarm formations of specific geometrical shapes by the vehicles, regardless of the ocean currents. Accounting for uncertainties in the predicted ocean fields allows us to yield the paths that best compromise between least uncertainty and least total travel time.

Our approach is interdisciplinary: It is inspired by ideas in fluid and ocean sciences, computational sciences, numerical methods, and uncertainty quantification, and applies them to path planning, which has roots in robotics and optimal control. The methodology is based on solving (stochastic) partial differential equations. In this sense, it is rigorous and obviates the need for heuristics. To compute the level set evolution, i. e., the viscosity solution to the Hamilton–Jacobi equation, we either use our finite-volume code for partial differential equations (PDE) (with varied orders, limiters, etc.) or narrow-band level set and fast marching methods for efficient computations. For the uncertain flow case, we employ our stochastic PDEs solvers [21.15]. In all cases, the computational cost increases only geometrically with the number of grid points (and not with the path length). We refer to the above-mentioned references for descriptions of numerical algorithms. Next, we illustrate results under diverse fluid and ocean conditions, as well as diverse vehicle behaviors.

21.1.1 Canonical Steady Flows

We first evaluated our level-set-based methodology by computing time-optimal paths in canonical steady ocean flow features. Such problems include, for example, determining how to time-optimally enter or leave an ocean eddy, cross an ocean jet, or maneuver in a meander/wave field [21.13]. The results can be very useful as rules-of-thumb for time-optimal operations.

To exemplify such canonical steady flow problems, we consider idealized 2-D-in-space (x, y) non-dimensional *highway* scenarios (2-D: two-dimensional), as illustrated in Fig. 21.1 [21.13]. In the first case (Fig. 21.1a), the vehicle is deployed from $(0, -0.4)$, with the goal of reaching the point at $(0, 0.8)$. The speed of vehicle is 1. The flow consists of two different jets, both flowing in the positive y direction. The first jet (narrow band with black arrows on the left) has a higher speed, $v_1 = 2$, than the vehicle. The second jet (narrow band with red arrows on the right) has a constant speed equal to that of the vehicle, $v_2 = 1$. From (Fig. 21.1a), the optimal path takes advantage of the higher jet flow on the left to reach the goal point. In the second case (Fig. 21.1b), the vehicle is deployed from $(-0.6, -0.6)$ and the goal point is at $(0.6, 0.6)$. The speed of the vehicle is 1. In this case, we have two different jets flowing in opposite directions, both with a speed of 1.5 (faster than the vehicle). The first jet (upper narrow band with black arrows) flows in the negative x direction. The second jet (lower narrow band with red arrows) flows in the positive x direction. As can be seen from (Fig. 21.1b), the vehicle rides the first jet to an upstream position of the second jet in order to compensate for the unfavorable flow of the lower jet. Then, the vehicle crosses the second jet as quickly as possible to reach its goal. In another *highway* test (Fig. 21.1c), the vehicle is deployed from $(-0.6, 0.6)$ and the goal point is at $(-0.6, -0.6)$. The speed of the vehicle is 1. In this case, we have two different jets flowing with a speed of 1.5 (faster than the vehicle). The first jet (upper narrow band with red arrows) flows in the positive y direction. Therefore, it is infeasible to use any trajectory crossing this jet flow. The second jet (lower narrow band with black arrows) flows in the positive x direction. As can be seen from (Fig. 21.1c), the vehicle first skirts the upper flow on the right-hand side. Then it moves to an upstream position of the second jet and crosses it as quickly as possible to reach the goal.

21.1.2 Time-Dependent 2-D Flows

Time-dependent flows are now considered. To evaluate robustness, a large number of 2-D flows were utilized, including flows at the exit of a strait, flows behind islands, and wind-driven flows. Such flows have also been utilized to evaluate our approach for the path planning of swarms of vehicles [21.3, 10–12, 14].

Next, we present three examples of time-optimal path planning for swarms of vehicles in the context of

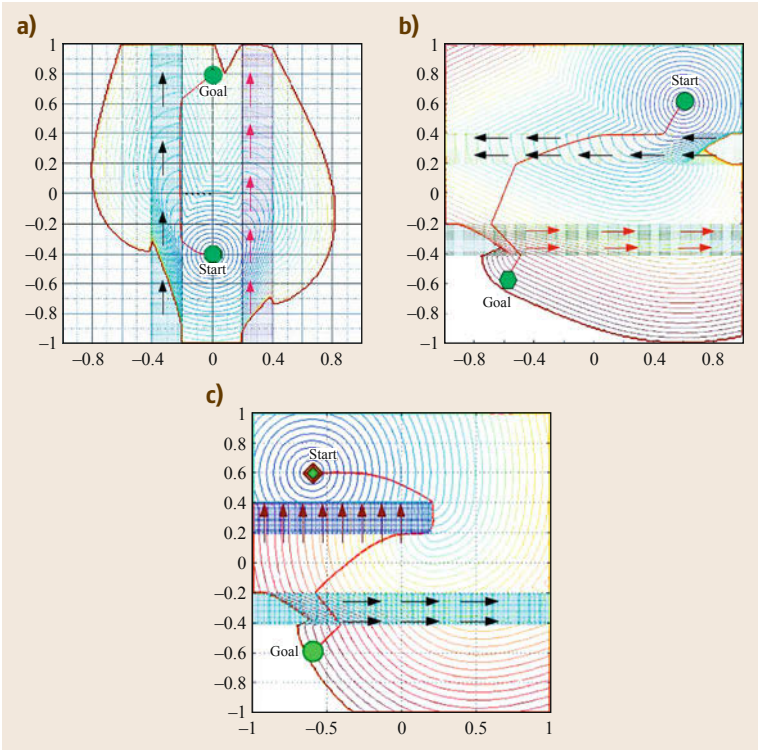


Fig. 21.1a–c Highway test cases. **(a)** 2 ideal jets. Faster jet on left (*black arrows*). Level-set optimal path rides the faster jet. **(b)** Two ideal jets in opposite directions. Level-set optimal path first rides top jet to get to an upstream position of opposing lower jet. It then crosses the opposing jet as quickly as possible to reach goal. **(c)** This case includes one truncated jet (at twice the vehicle velocity) that is effectively an impenetrable obstacle. The second jet is a transverse jet, similar to the bottom jet in **(b)**. Here the level-set optimal path moves around the truncated jet to reach an upstream position of lower jet. It then crosses the lower jet as quickly as possible to reach the goal (after [21.13])

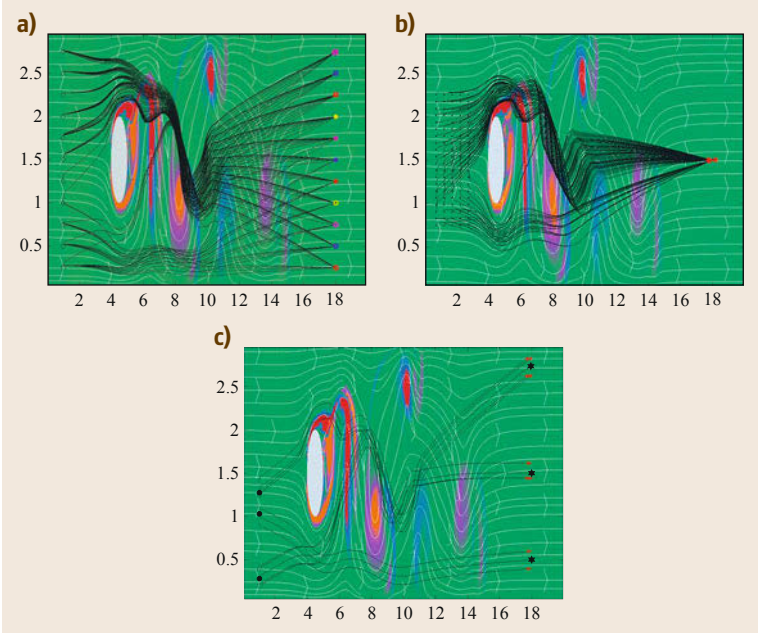


Fig. 21.2a–c Vorticity of the flow past an island (*gray oval*) overlaid with white streamlines. **(a)** *All-to-all broadcast* with swarms while crossing strong eddy field in optimal time – 11 swarms of 11 members each, avoid island, utilize eddies/wakes, reform swarms, and communicate. **(b)** *All back to ship* optimal recovery of large swarm formation of vehicles – formation of 100 vehicles to return to ship for pick up, avoid island, utilize eddies/wakes to return in shortest time. **(c)** *Maintain swarm formation* with coordination based on local gradients or dynamics – 3 swarms of 2–4 members (*line*, *triangle*, *square*), avoid island, shortest time (after [21.10, 12, 14])

flows past an island. The circular island is an obstacle to both the flow and to the vehicles. The flow is fast enough to cause the shedding of vortices off the leeward side of the island.

The first swarm example is an *all-to-all broadcast* to swarm leaders, in which 11 swarms of 11 members each optimally cross the strong eddy field in the wake of the island while exchanging information with each other:

Specifically, one member of each swarm switches to another swarm and all subsequently reform the swarms in fastest time at 11 target locations (Fig. 21.2a). In other words, we start with 11 swarms of 11 vehicles each at the 11 starting points upstream of the island (shown by open circles). Each of the initial swarms is to break up and distribute one member to each of the 11 target points downstream of the island (shown by colored markers), resulting in 11 new swarms, each made up from 1 member of each of the original 11 swarms. The goal is to achieve the new configuration in minimum time, communicating data to each swarm leader, while avoiding the island. The optimal paths computed by our level-set-based method are the black curves in (Fig. 21.2a). We see three classes of paths. Most paths go above the island and get a boost from the first shed eddy. A second group goes below the island and intercepts the eddy at a later point. The final grouping remains below the island and avoids the eddy.

The second example is an *all-back-to-ship* fastest-time recovery scenario (Fig. 21.2b). A formation of 100 vehicles starts upstream of the island (black dots). The goal is to find the quickest paths to bring all the vehicles to a single pick-up point (18, 1.5) downstream of the island, utilizing wakes and eddies while avoiding the island (simulating the recovery of a swarm of AUVs by a single ship). The optimal set of paths are found by the level-set-based method and are shown as black curves in Fig. 21.2b. Roughly three quarters of the vehicles are sent over the island to take full advantage of the first eddy. Most of the remaining vehicles avoid the eddy by passing under the island, although some go under the island and catch the eddy later.

21.1.3 Maintain Swarms Formations

Building on the above, we created algorithms for *maintaining swarm formations*. The first example consists of three swarms of two to four members each, organized in the shape of a line, triangle or square, coordinating their paths based on gradients or dynamics (Fig. 21.2c). For the algorithm illustrated, the coordination is based on local gradients (Fig. 21.2c), using a local control algorithm.

Here, the goal is to bring each of the three swarms from their start locations to their target locations in minimum time while maintaining their formation and avoiding the island. We utilize our level-set-based method to compute the optimal paths for the center of mass of each swarm (black curves starting from black dot and finishing at black star). Around each center-of-mass-path we construct the swarm paths by using the headings for the optimal center-of-mass-path and adding a velocity component at each step to maintain each vehicle position relative to the center of mass. To prevent the vehicle paths from diverging, we allow the center of mass to move at a speed lower than the maximum vehicle speed (i. e., the center of mass moves with a velocity that is a function of the instantaneous flow).

A drawback of this local control approach is that the allowable separation between vehicles depends on the flow dynamics. The distance between vehicles has to be small if the flow speed is highly variable in space (large flow gradients). The smoother the flow in space, the larger the spacing between vehicles can be. In [21.12], we develop a more powerful methodology based on local level-sets which allows maintaining complex shapes in large flow gradients; a time optimal path is computed for the group leader and finite-time local level sets are computed continuously in time for the other vehicles in the swarm. The latter provide reachable sets for each of these vehicles for each finite-time horizon (which could be a time step) and allow a subsequent shape optimization algorithm to find and sustain an optimal swarm formation.

21.1.4 Forbidden Regions

Forbidden regions are areas which are denied to the vehicles (for safety reasons or security restrictions) but do not affect the flow field [21.3]. In the example that follows, this constraint information is provided to the vehicles ahead of time, just as restricted regions or forecasts of hazardous weather can be provided to an aircraft. In these conditions, the goal is still to compute time-optimal paths, but under the additional constraint of avoiding the provided or forecast forbidden regions.

For example, consider a uniform barotropic jet exiting a strait or estuary (sudden expansion or 2-D coastal flow), as illustrated in Fig. 21.3. The non-dimensional speed of the vehicle is 0.5, while the largest current speed is 2.5 (Fig. 21.3a,b). The width of the inlet is one third of the total width of the channel. In the simulation (Fig. 21.3c), nine vehicles are released from the lower edge of the inlet (black dot) and are required to achieve a triangular formation (colored dots). The unconstrained time-optimal paths constructed from the level-set-based algorithm are shown by colored lines. The vehicle heading for the tip of the triangle rides along the central jet, while the vehicles targeting the four endpoints closest to the inlet use the upper and lower recirculation eddies. Overall, we can see from Fig. 21.3c that the algorithm correctly predicts the shapes of the optimal paths.

In the companion case (Fig. 21.3d), two forbidden regions are added (gray shaded area), which block seven of the nine paths in the free run (Fig. 21.3c). The

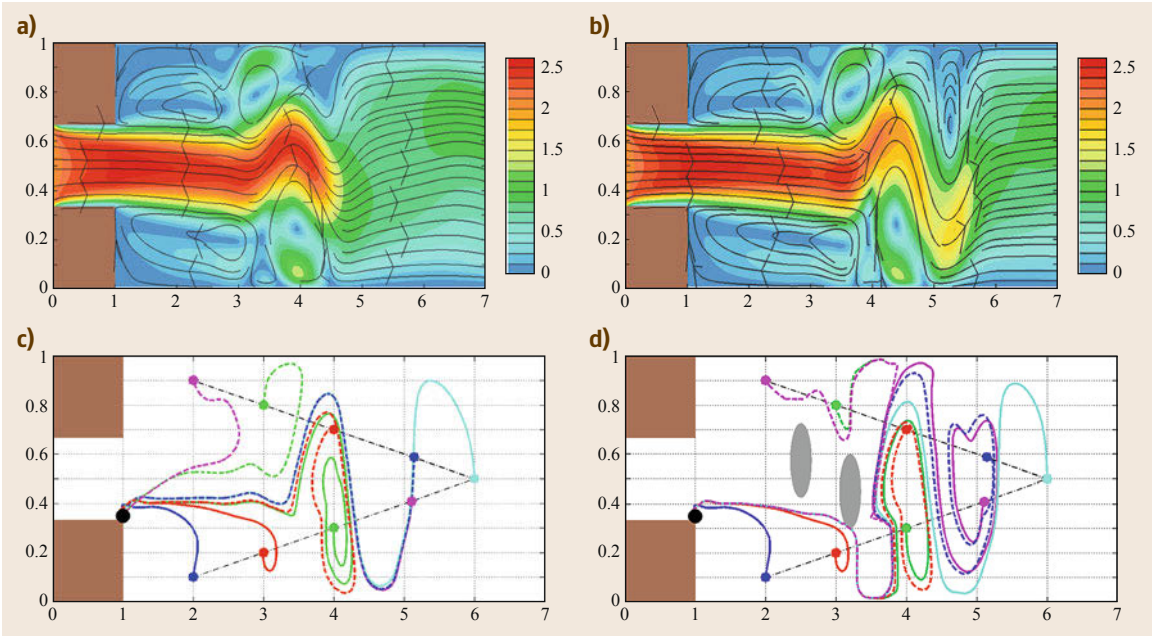


Fig. 21.3a–d Time-optimal paths without **(c)** and with **(d)** forbidden regions. **(a,b)** Snapshots of flow field for a jet exiting a strait or estuary (sudden expansion flow) **(a)** At time of initial vehicle deployment and **(b)** at final time of vehicle maneuvers in **(c)**. **(c,d)** Optimal vehicle paths for nine vehicles deployed from a single point (*black dot*) in the flow field of **(a,b)**. Results for two situations: **(c)** no constraints or forbidden regions. Vehicle paths take full advantage of the evolving jets and eddies to reach their final positions (*colored dots*) in the shortest times. **(d)** Two forbidden regions. Vehicles are denied access to the *gray shaded* regions. The level set algorithm provides seven new time optimal paths for the paths computed in **(c)** that are blocked while correctly leaving unchanged the two paths that are not blocked (after [21.3])

new optimal paths for these seven vehicles all ride the lower edge of the main jet, just skirting the bottom of the second forbidden region. They then ride down one eddy and up an adjoining eddy (Fig. 21.3a), to rejoin the main jet behind the forbidden regions. The two paths from Fig. 21.3c that did not pass through the forbidden areas remain unaffected.

We note that forbidden regions can be learned by vehicles as they progress, using onboard routing (Sect. 21.2.1). In that case, data about forbidden regions are collected by the vehicle along its optimal forecast route. As these new data are utilized, optimal paths are re-computed, leading to a new optimal forecast route for the remaining travel plan.

21.1.5 Uncertain Flow Fields

It is well known that ocean predictions are limited in accuracy because of the predictive capabilities of present modeling systems [21.16, 17] and also due to fundamental limits of predictability for nonlinear dynamics. As a result, the practical horizon for skillful forecasts of time-optimal paths is also limited. In addition, some paths will be more uncertain than others, for example, some fast paths may be very uncertain, while some somewhat slower paths may be very certain. The path planner can thus aim to find a compromise between time optimality and certainty of arrival.

To illustrate this compromise, we consider the case of time-optimal path planning in a 2-D stochastic wind-driven double-gyre flow with large uncertainties (Fig. 21.4), [21.10]. This is a strong flow field in which the uncertainties are predicted using dynamically orthogonal (DO) stochastic PDEs [21.15, 18, 19]. Questions that we consider include:

- What are the effects of flow uncertainties on the time optimal path?
- Which path should be selected?
- Is there a path which is both fastest and least uncertain (most robust or most likely)?

For the chosen start and end points, the shortest-distance path is against the jet flow that is stronger than the vehicle speed. We thus expect at least a bi-modal behavior for the probability density functions (PDFs) and

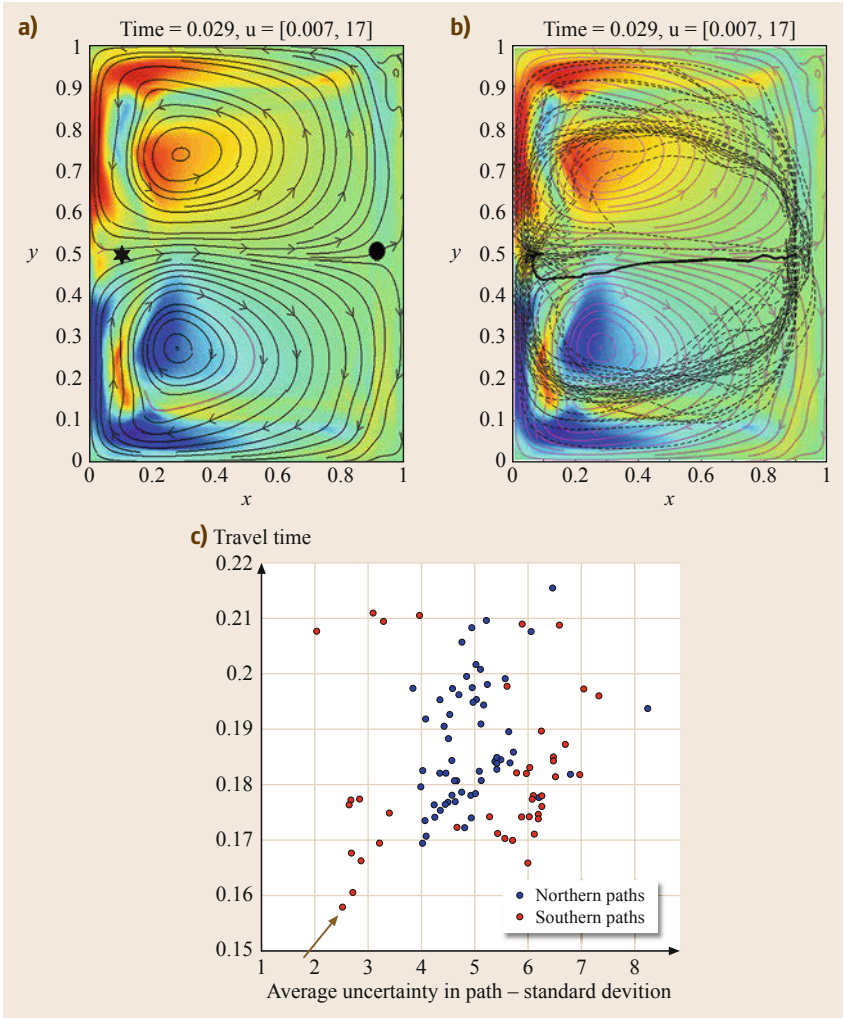


Fig. 21.4a–c Path planning in uncertain flow. **(a)** One realization of strong wind-driven, double-gyre flow. Vehicle starting point indicated by *filled circle*. End point indicated by *black star*. Wind forcing is stochastic. **(b)** Ensemble of optimal paths for different winds shown by *dashed lines*. Mean path shown by *solid line*. The mean path is unrealizable; the ensemble of paths is close to a bi-modal distribution (either following northern or southern gyre). **(c)** Travel time vs. uncertainty of ensemble of gyre paths. Most probably path follows southern gyre (*green arrow*)

sensitivity to initial conditions. We use the DO equations to quantify the uncertainty and the level-set-based path-planning (21.1)–(21.2) to solve for optimal paths for a number of realizations of the DO simulation. As is shown in Fig. 21.4b, as expected, we find that the mean path is not realizable as the paths generated are close to a bi-modal distribution (the flow field PDF leads to a PDF for the paths). For each path, we integrate the standard deviation of the flow velocity along the path, which provides a measure of the extent of *uncertainty* along the path. We then plot the travel time versus this uncertainty (Fig. 21.4c), providing a Pareto-like quantity. This plot indicates that a fast and least uncertain path is a path to the south. We note that this preferred south path arises mainly because the initial mean and PDF conditions are not symmetric with respect to the shortest-distance (horizontal) path between the start and end points.

21.1.6 Realistic Ocean Conditions: Three-Dimensional Multi-Scale Flows

Our level-set-based method has also been employed for path planning in realistic (data-driven) three-dimensional (3-D) ocean flow fields [21.13], as illustrated in Fig. 21.5. In this example, the MSEAS ocean prediction system provides the three-dimensional currents. The simulations are realistic (data-driven) ocean reanalyses based on the real-time exercises that occurred during August–September 2006 in the New Jersey Shelf/Hudson Canyon region [21.20, 21], specifically the autonomous wide aperture cluster for surveillance (AWACS) and Shallow Water-06 (SW06) exercises. The simulations employ nonlinear free surface primitive equations with fully implicit two-way nesting (Fig. 21.5a,b) and tides. The larger domain (Fig. 21.5a)

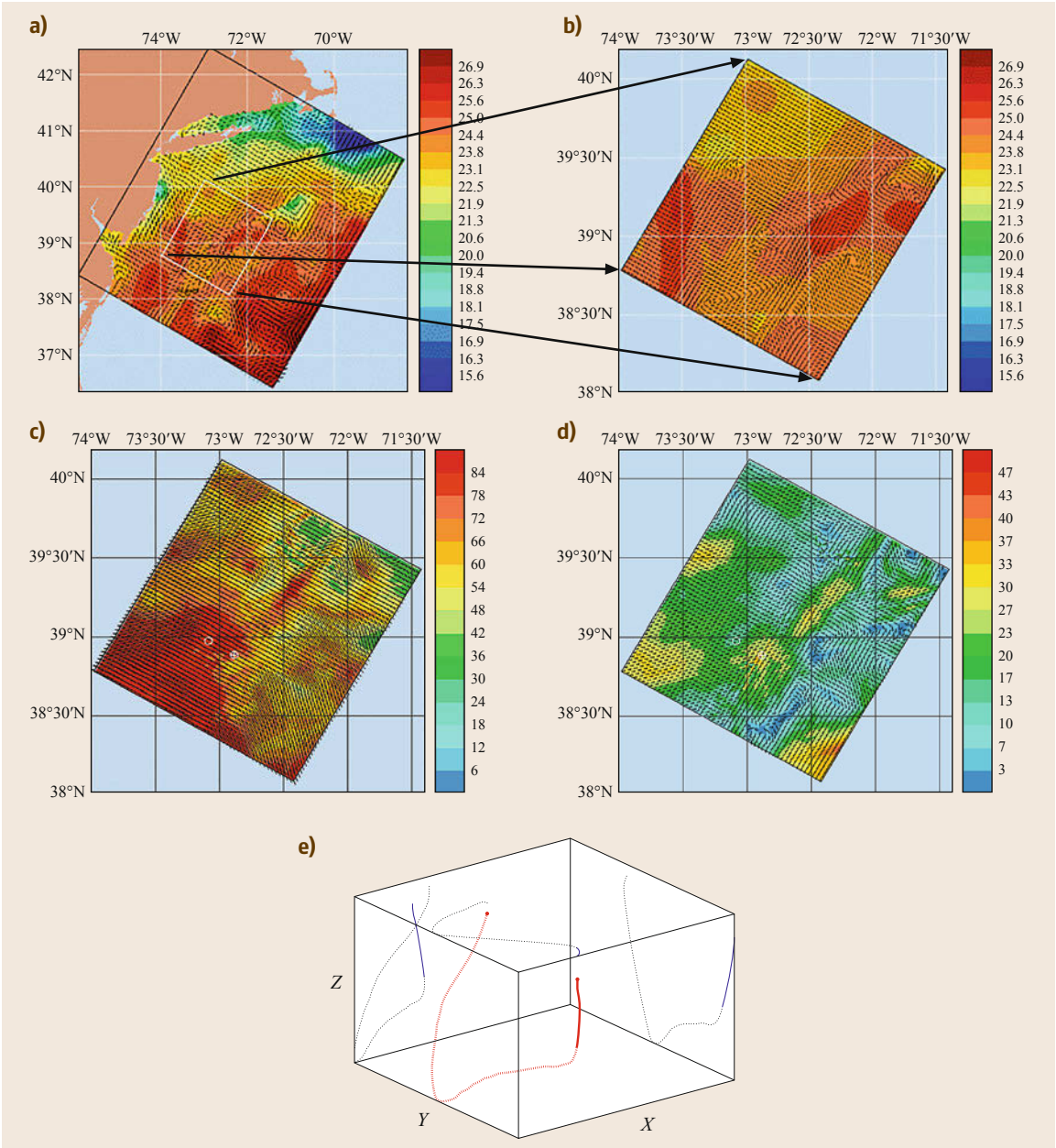


Fig. 21.5a–e Optimal path planning in 3-D with time for *fully four-dimensional (4-D) realistic ocean flows*. **(a–b)** Surface temperature and velocity from fully implicit nested SW06 simulation. **(c)** Surface flow during path planning from open circle to circle+cross. Note the strong flow opposing path. **(d)** 80 m flow during path planning. Weaker flow with some flow into goal from NE. **(e)** 3-D optimal path initially dives deep to weaker flows. The final approach to goal is from NE along the favorable flow (after [21.13])

is a 522×447 domain with 3 km resolution and 30 vertical levels, while the smaller domain (Fig. 21.5b) is a 172×155 km domain, with 1 km resolution and 30 vertical levels. The ocean initial conditions utilize real data gathered by gliders, AUVs, ship deployed conductivity temperature depth (CTD) and historical data taken from the World Ocean Database, Gulf Stream feature analyses, National Marine Fisheries Service, etc. The duration for this real-time simulation was 43.5 days. Here, we focus on the period during which Tropical

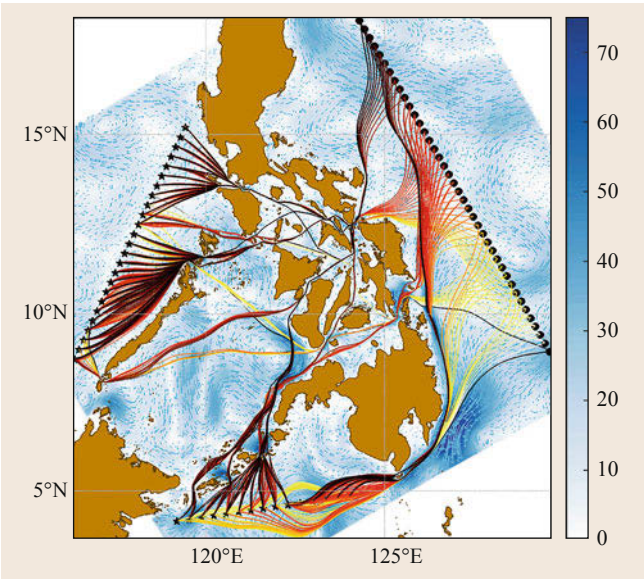


Fig. 21.6 *Time-optimal path planning in the Philippines Archipelago for vehicle swarms*: time-optimal paths of 1600 vehicles deployed from 40 different release points in the Pacific (marked by *circles*, in the east of the picture). The vehicles cross the Philippines Archipelago from east to west in the fastest time. Most of the paths are indirect as they take advantage of currents and other multi-scale ocean features ►

Storm Ernesto passed by the region (late August to early September 2006).

In order to simulate realistic and most practical conditions, we set our vehicle speed to $0.25\,\mathrm{m\,s^{-1}}$, which is often a maximum speed for most currently operated underwater gliders. The vehicle is deployed at the ocean surface (white open circle in Fig. 21.5c) and needs to travel to a second, offshore surface location (white circle with cross, Fig. 21.5c). During the path planning exercise, Tropical Storm Ernesto generates a strong, opposing, onshore current of around $0.8\,\mathrm{m\,s^{-1}}$ (Fig. 21.5c). Therefore, the glider cannot use the straightforward surface trajectory from the starting location to its offshore goal. However, the level-set algorithm finds a time-optimal 3-D path (Fig. 21.5e) in which the glider initially dives and *meanders* in 3-D to take advantage of the weaker flow field in the deeper ocean region (Fig. 21.5d).

21.1.7 Realistic Ocean Conditions: Swarms in Multi-Scale Flows with Complex Geometries

Our level-set-based methodology has been parallelized to efficiently plan optimal paths for swarms of vehicles in realistic multi-scale ocean flows with complex geometries. To illustrate this, we consider tidal-to-mesoscale ocean simulations for the Philippines Archipelago [21.22]. In this example (Fig. 21.6), there are 40 different swarms of vehicles, each released at a different location in the Pacific (east of the Archipelago). Each swarm and location consists of 40 different vehicles, one vehicle per swarm is the leader and aims to reach one of the end points in the west while the other vehicles aim to reach another of the 39 end points. When all vehicles reach their final destination, vehicles in the reformed-swarms are close to each other and can communicate underwater to exchange all needed information from their previous swarms. In other words, at the final time, an information exchange among all swarms can occur, in the form of an *all-to-all broadcast* (as in Fig. 21.2, but for realistic complex flows). The key is that each of the vehicles is navigated in fastest time to its specific end point, to the west (South China Sea) and the south (Sulawesi Sea) of the Archipelago. To do so, each vehicle optimally combines the following behaviors for time-optimality:

- Accounting for the wind events and the corresponding strong currents
- Waiting for favorable strong tidal flows in the many straits
- Taking the major currents in the region, following their meanders and eddies
- Looping in rings as needed or avoiding them all together, depending on their final destinations and variability of the 3-D ocean flows.

Figure 21.6 shows the fastest paths for all the vehicles overlaid on the map of the Philippines [21.14]. In this example, paths for 1600 vehicles have been planned by efficiently parallelizing the level-set-based algorithm. As observed, none of the vehicle trajectories is a direct path to its target. Vehicles clearly utilize ocean currents and multi-scale features to reduce their travel time.

21.2 Adaptive Sampling for Swarms of Ocean Vehicles

With our collaborators, we have developed a range of adaptive sampling schemes and applied them in varied coastal regions. In what follows, we start by outlining some of the schemes that were first applied in real

ocean situations, as well as the results of these schemes (Sect. 21.2.1). The constraints of the realistic ocean forecasts and the computational capabilities of the last decade led to a range of approximations necessary to issue optimal sampling paths forecasts in real time. After these results, we have started new theoretical research towards schemes and methods that are both rigorous and applicable to real ocean systems. This progress is summarized in Sect. 21.2.2.

21.2.1 Early Results: Approximate Schemes for Realistic Applications

During the last decade, we have developed and applied a range of adaptive sampling and onboard routing schemes for optimal sensing in varied regions of the world's oceans. They include:

- Adaptive sampling via error subspace statistical estimation (ESSE) with nonlinear predictions of error reductions [21.5, 23]
- Mixed integer linear programming (MILP) for optimal-sampling path planning [21.24]
- Nonlinear optimal-sampling path planning using genetic algorithms [21.25]
- Dynamic programming and onboard routing for optimal-sampling path planning [21.26]
- Command and control of surface kayaks over the Web, directly read from model instructions [21.27].

These schemes and results are outlined next.

Real-Time Adaptive Sampling via ESSE

The results of the ESSE adaptive sampling scheme [21.20, 28] utilized in real time are illustrated in Fig. 21.7. For data assimilation, the ESSE scheme integrates an ensemble of nonlinear simulations and assimilates data by Kalman updates in the adaptive error subspace obtained from this ensemble. For adaptive sampling, the same is carried out, but in a forecast sense and for many candidate sampling plans, since the goal is the nonlinear prediction of the ideal future sampling. Specifically, a set of candidate sampling plans (data locations and sensor types) was chosen a priori to reduce computational costs and directly satisfy operational constraints. For each plan, a nonlinear ESSE prediction and data assimilation was then carried out. This estimated the error reduction for each plan and so determined the candidate plan with the best impact.

Adaptive Sampling with Routing Schemes

Predictive adaptive sampling and onboard adaptive routing have been combined for thermocline tracking and adaptive sampling for acoustic fields with AUVs [21.26], as illustrated in Fig. 21.8. The method consists of two complementary steps. First, data assimilative environmental and acoustic propagation ESSE ensemble modeling provides input to a dynamic scheme that computes parameter values for autonomous sensing behaviors, with the goal of optimally reducing the forecast acoustic uncertainties. Behavioral parameters include, for example, the dive angle or the times to switch depths. Second, these parameters for the autonomous sensing behavior are refined onboard the vehicle in realtime, in response to the data sampled.

A related approach consists of computing the high-level routing remotely (in a centralized computer). A vehicle is then guided by downloading remotely-computed optimal waypoints from a website. During

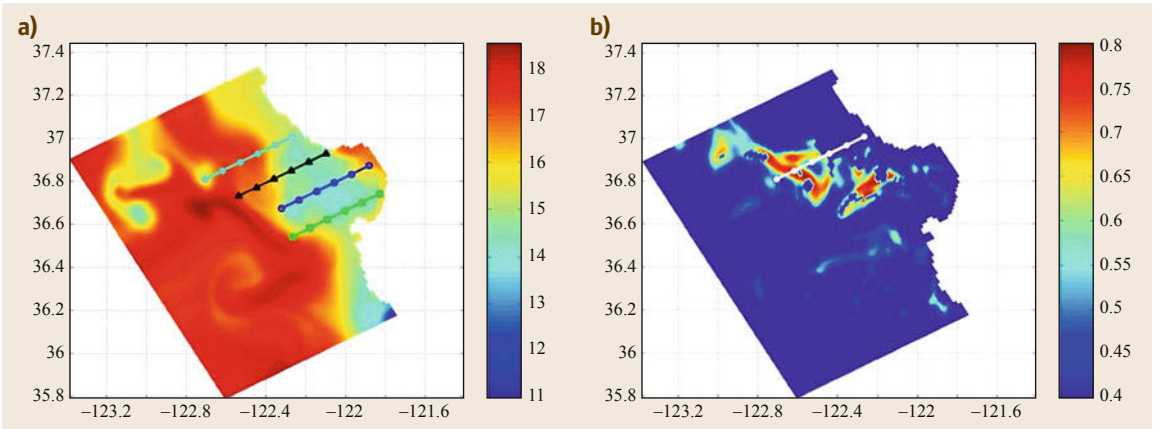


Fig. 21.7a,b Adaptive sampling via ESSE, for an 24–27 August sample forecast. **(a)** Central forecast of the surface T for August 26, overlaid with the four candidate tracks. **(b)** Predicted optimal track for August 26 and its relative error reduction for August 27 (after [21.5]). The question answered is *which of the four sampling track tomorrow will reduce uncertainties the most the day after tomorrow?*

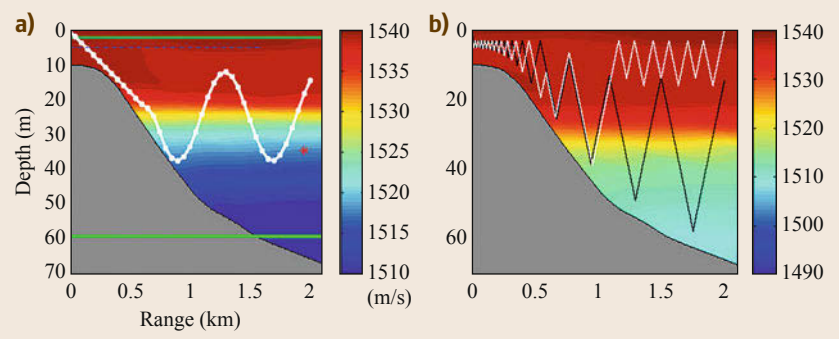


Fig. 21.8a,b Simulated adaptive thermocline tracking using yo-yo adaptive sampling and onboard routing, overlaid on sound–speed ($m\,s^{-1}$). **(a)** AUV path; *green lines* are the minimum and maximum depths allowed. **(b)** The optimized AUV automatically captures the *afternoon effect* on the surface thermocline, i. e., the warming of the upper ocean layers due to the strong daylight sun (after [21.26])

the Persistent Littoral Undersea Surveillance Network 2007 exercise (PN07) in Dabob Bay [21.27, 29]; such a website was used to optimally guide kayaks at sea, directly based on data-assimilative ocean model predictions of the optimal sampling plans, without humans in the loop. Computers ran the ocean model forecasts and adaptive sampling forecasts and then wrote the waypoints file on a website that was read by a kayak, all without human input. This was a technical achievement compared to other larger and more extensive exercises, e.g., Monterey Bay 06 [21.8, 9, 30], where humans were intermediaries between ocean models and vehicles, e.g., entering waypoints or interpreting ocean model forecasts and then selecting waypoints.

Path Planning for Fixed-Uncertainty Optimal Sensing

Path planning for fixed-uncertainty optimal sampling aims to answer the following type of questions: Assuming the error field is not dynamic during the sampling, which sampling path is optimal? This approach often assumes simple updates of field estimates due to data but more importantly neglects how dynamics over time evolves future observations, evolution which affects the truly optimal adaptive sampling. The planning then amounts to the optimization of a fixed objective function whose solution often amounts to visiting the locations of highest uncertainty. Here, we outline two schemes for such planning. The first one solves the optimization using MILP [21.23, 24, 31], as shown in Fig. 21.9. The solution is obtained by a *branch and bound algorithm*, which provides exact and globally optimal solutions. The path is segmented by waypoints, and the goal is to solve for their coordinates. The number of waypoints is a parameter fixed as a function of vehicle range and grid dimensions. An extension of this approach for time-dependent fields has been obtained [21.32].

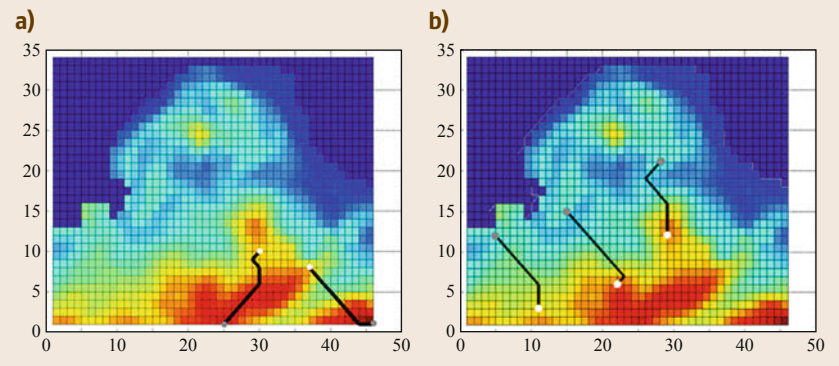


Fig. 21.9a,b Generation of sampling paths for fixed objective fields using mixed-integer programming (MIP). Fields are the ESSE T error standard deviation averaged over 0–50 m. *Gray dots* are starting points for the AUVs and *white dots* are the MIP optimal termination points. **(a)** Optimal path of two vehicles. **(b)** Optimal path of three vehicles (after [21.24]). The question answered is *assuming the error field remains constant for the next day, on which path do I send my AUVs?*

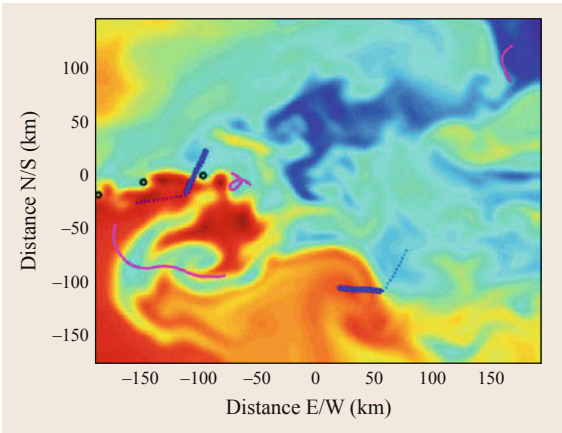


Fig. 21.10 Optimal sampling path solution computed by a genetic algorithm (GA) in the Middle Atlantic Bight and Shelfbreak Front regions (after [21.25]). The plan is computed for three moorings, three gliders, and two REMUS vehicles. The GA places all moorings (*black circles*) directly within the Shelfbreak Front, which passes over them during the simulated 5 days of deployment. The REMUS vehicles (*blue tracks*) pass over Shelfbreak events and fronts, and the gliders (*magenta tracks*) track large gradients with time. A distance-potential constraint maintains the paths of platforms apart by at least a Rossby radius (key horizontal length scale for mesoscale ocean dynamics, 15 km in this example)

The second scheme uses genetic algorithms to solve the optimization problem [21.25]; the results are illustrated in Fig. 21.10. With genetic algorithms, one can carry out nonlinear path planning for any type of cost function, which is a significant practical advantage. The weights of each term in the cost function are chosen by users. However, a disadvantage of the genetic algorithm solution is that one cannot always theoretically guarantee that it is at, or is near, the optimum. In the results illustrated, the cost function includes terms representing ocean variability, acoustic transmission loss sensitivity, and temperature uncertainty. Results illustrate that the temporal dependence of the ocean fields strongly influences the optimal location of the moorings that are fixed in space and the gliders that are slowly drifting. The remote environmental monitoring unit (REMUS) vehicles move quickly and are less sensitive to time-dependent ocean effects.

21.2.2 Recent Progress: Towards Rigorous Schemes for Realistic Applications

Our recent research in adaptive sampling has been motivated by the ESSE examples presented above. The first research theme that we investigate is to increase the number of candidate sampling paths towards functional paths covering the whole physical domain and to increase the durations of look-aheads towards durations up to the end of the experiment or the predictability limit. Ideally, there should be no limitation on the candidate paths other than the operational constraints. A second theme is to continue the utilization of nonlinear uncertainty evolution and estimation of future data impacts, but replace the ESSE ensemble approach by our DO PDEs and replace the ESSE Kalman updates by non-Gaussian updates, including Gaussian Mixture Models updates [21.33, 34]. Mixture models are semiparametric frameworks consisting of a sum (mixture) of parametric distributions whose parameters and number (total number of distributions themselves, referred to as mixture complexity) are commonly fit to an ensemble of realizations. A third theme is to extend the adaptive sampling optimality criterion to information theory (mutual information schemes) and to include nonlinear smoothing effects. In what follows, we outline a subset of results related to these research themes.

Adaptive Sampling Using Partially Observable Markov Decision Processes (POMDPs)

Computational research has begun on the use of POMDPs as a means of adaptively sampling fields of interest in the ocean. A POMDP in this case is a mathematical framework that allows one or more agents to optimize an objective function under uncertainties both in sensing and knowledge of the field dynamics. We have demonstrated its use on two simplified dynamical model testbeds. The first is a stochastic game of life where a robot aims to optimally reach an end goal while sampling minor rewards and avoiding minor losses along the way. The second is the prediction of optimal sampling paths for two vehicles operating in the idealized Lorenz-95 atmospheric model [21.35]:

1. *Stochastic game of life*: We have adapted John Conway's game of life [21.36] by including stochasticity in the governing equations, as described in [21.37]. As such, we consider this to represent a (simplified) discrete flow model. Figure 21.11 shows one time instant in that game, all actors of the game are dynamic: the black circle is the AUV; the green circle is a slowly moving global feature of high reward; blue plus signs are minor rewards; and red crosses are minor cost/loss. Both of the latter fields evolve independently according to the rules of the game. The route of the AUV (not plotted in the snapshot) is being optimally planned. The domain is discretized such that all motions are restricted to be up or down, or left or right. The AUV applies the POMDP-like algorithm, intelligently maneuvering

through the field such as to maximize its reward. Specifically, it seeks to approach the green circle, primarily sampling blue dots along its path. This test bed naturally extends to multiple vehicles. The game of life results (Fig. 21.11) may also directly apply to real ocean fields in the sense that green, blue, and red regions could be specific ocean field properties or discretized regions of instability or of interesting dynamics that one wishes to sample.

2. *Lorenz-95 model*: The system studied is the Lorenz-95 atmospheric model, which is a time-varying one-spatial dimension model, as illustrated in Fig. 21.12. The Lorenz-95 model replicates a number of phenomena observed in geophysical fluid dynamics, and as such lends itself well to this analysis. Each vehicle (i. e., adaptive observation) locally predicts future consequences of proposed sampling paths, independently of other vehicles, given current knowledge of the system. In this test bed, the adaptive DO equations are used to predict the PDF of the Lorenz-95 state and artificial potential-like functions are employed to ensure that adaptive vehicles remain sufficiently apart from each other. We implemented a POMDP-like algorithm for adaptive sampling, using a GMM-DO (GMM: Gaussian Mixture Model) filter for data assimilation [21.33, 34]. With this example, we show that adaptively sampling the field using the POMDP approach outperforms the case in which the adaptive observations remain fixed in location (see the bottom left plot of Fig. 21.12).

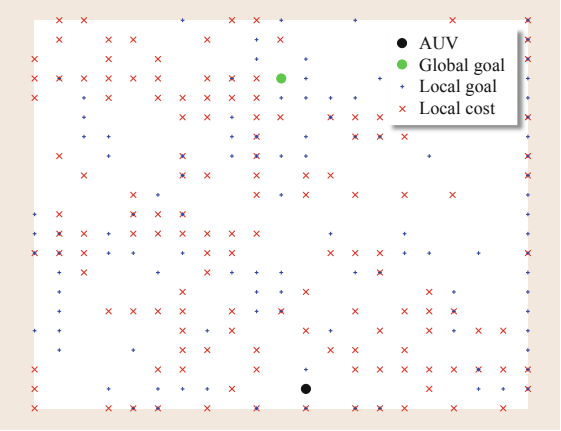


Fig. 21.11 Stochastic game of life with intelligent sampling using POMDPs. A fixed-time snapshot in the time-dependent game is shown. The route of the AUV is being optimally planned. The domain is discretized such that all motions are restricted to be up or down, or left or right. The goal for the AUV, *black circle*, is to track and reach the *green circle* as fast as possible and, if efficient along the way, sample the *blue plus signs* (e.g., food) and avoid the *red crosses* (e.g., predators), so as to increase the total reward

While POMDPs are computationally expensive, techniques are being developed that continually allow their use in domains of increasing complexity. In the future, with a detailed probabilistic description of the currents via the adaptive DO equations, one could derive novel metrics using information theoretic measures (specifically *mutual information*) to represent rewards. Our scheme may effectively be applied among multiple agents, where each agent locally applies a POMDP, communicating across a subset of its information to neighboring vehicles. Knowledge of the ocean fields, derived from the MSEAS ocean prediction model, is naturally of significant advantage. By subdividing the POMDP framework into a set of hierarchies, itself inspired by biological swarming formations, one could optimize the sampling behavior of the team of AUVs. Then, a top layer POMDP controls a number of lower layered POMDPs, themselves controlling a group of AUVs.

Adaptive Sampling with Limited-Time Look-Ahead POMDPs and ESSE in 2-D Fields

We now extend the above approach to two-spatial dimension and time-dependent problems and evaluate it on simulated ocean fields. We still employ DO equations for uncertainty prediction and a limited time look-ahead POMDP approach to select the sampling paths. However, we now use ESSE for data assimilation instead of the GMM update.

As shown in Fig. 21.13, we consider an uncertain flow behind a circular cylinder/island. The uncertainty arises due to the initial conditions and is evolved in time using the DO equations. Assuming that we have access to a single sampling vehicle (a glider), we predict a trajectory for this asset that optimally reduces the uncertainty in the estimated velocity field (horizontal components: u and v). In Fig. 21.13a, we plot the field of variance in u and v along with the vehicle trajectory obtained using a three-time-steps look-ahead approach for adaptive sampling. Figure 21.13b shows the time evolution of the variance of two stochastic DO modes. Initially, the variances of the modes are large (order 1). As the vehicle makes measurements at integer times, the variance of each of the modes is reduced. With time, the variance of the modes eventually decreases to an order of 10^{-1} for the first mode and 10^{-5} for the last mode.

Figure 21.13c,d shows the true solution for the velocity field and the mean of the adaptive-sampling-based estimate of that velocity field at two different times. At time $t = 6$ (Fig. 21.13c), we see that the mean of the estimated solution is still very different from

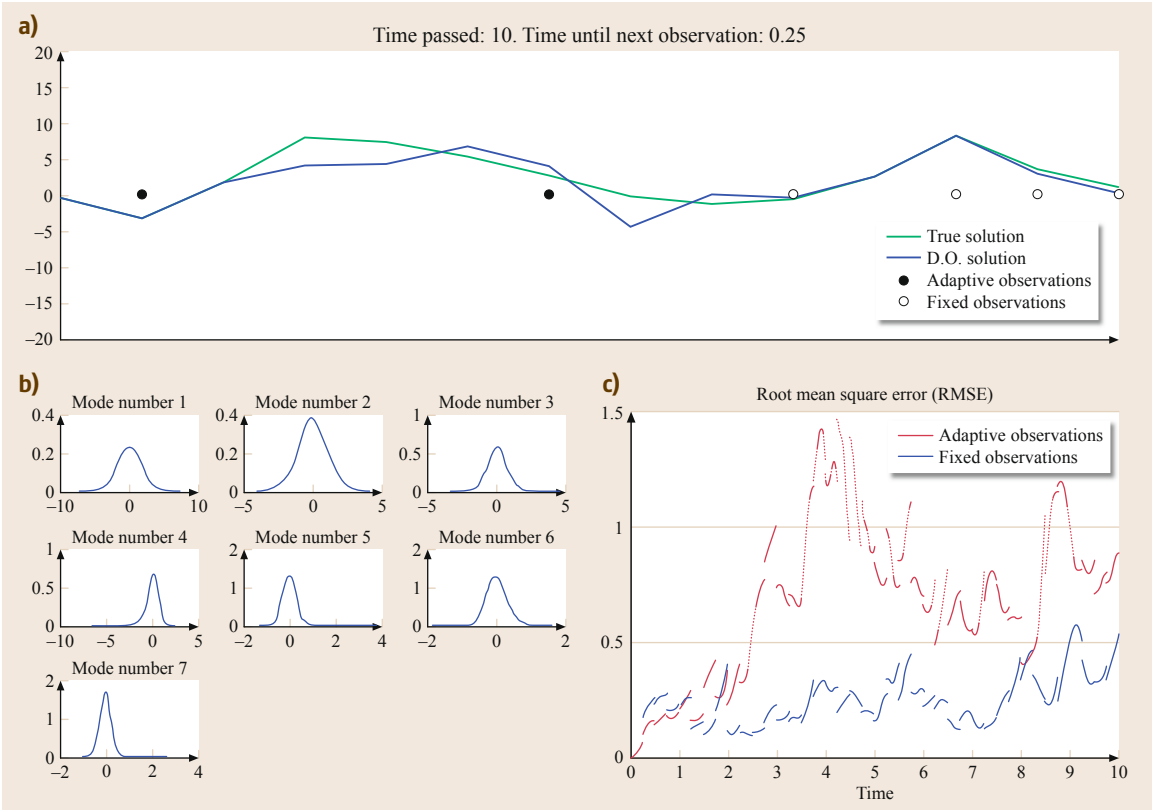


Fig. 21.12a–c Adaptive sampling one-dimensional (1-D) example: the Lorenz-95 application. The adaptive sampling is based on: i) non-Gaussian data assimilation (GMM-DO Filter) combining Gaussian Mixture Models (GMMs) with the DO equations; ii) three-step look-ahead POMDPs to decide which direction (left or right) to sample for each of the two adaptive observations; and iii) an artificial potential-like function to ensure the two adaptive observations remain separated in space. **(a)** True (unknown) and estimated solutions at final non-dimensional time 10. The latter estimated solution is obtained by adaptive sampling by GMM-DO, three look-ahead POMDPs and artificial potential-like functions. Three routine observations are fixed in space (*open circles*) while two observations have paths that are forecast by the adaptive sampling (*black circles*). **(b)** Marginal PDFs for the seven DO modes active at final time 10. **(c)** Comparison between the time-evolution of the averaged true error of the estimates obtained using adaptive sampling or not using it (i. e., assimilating data but keeping the two *black circle* observation locations fixed in place). The true error reduction obtained by predicting the best sampling location is substantial (about 100% better)

the true solution because very few measurements have been collected so far. As time progresses (Fig. 21.13d) the mean of the estimate matches quite well with the true solution, and the algorithm correctly predicts the shape, phase, and frequency of vortex shedding behind the island. Another interesting result in this example is that the vehicle automatically navigates towards the cylinder, indicating that it is the optimal location where measurements must be made to gain maximum information about the field. It also follows the locations where eddies are shed. Clearly, the dynamics of the problem, the limited-time look-ahead POMDP behavior and the assimilation scheme govern the track of the vehicle.

21.3 Conclusions and Outlook

Some of the results obtained by our MSEAS group and recent collaborators have been summarized here. We first presented a subset of our results on time-optimal path planning for swarms of ocean vehicles. We then

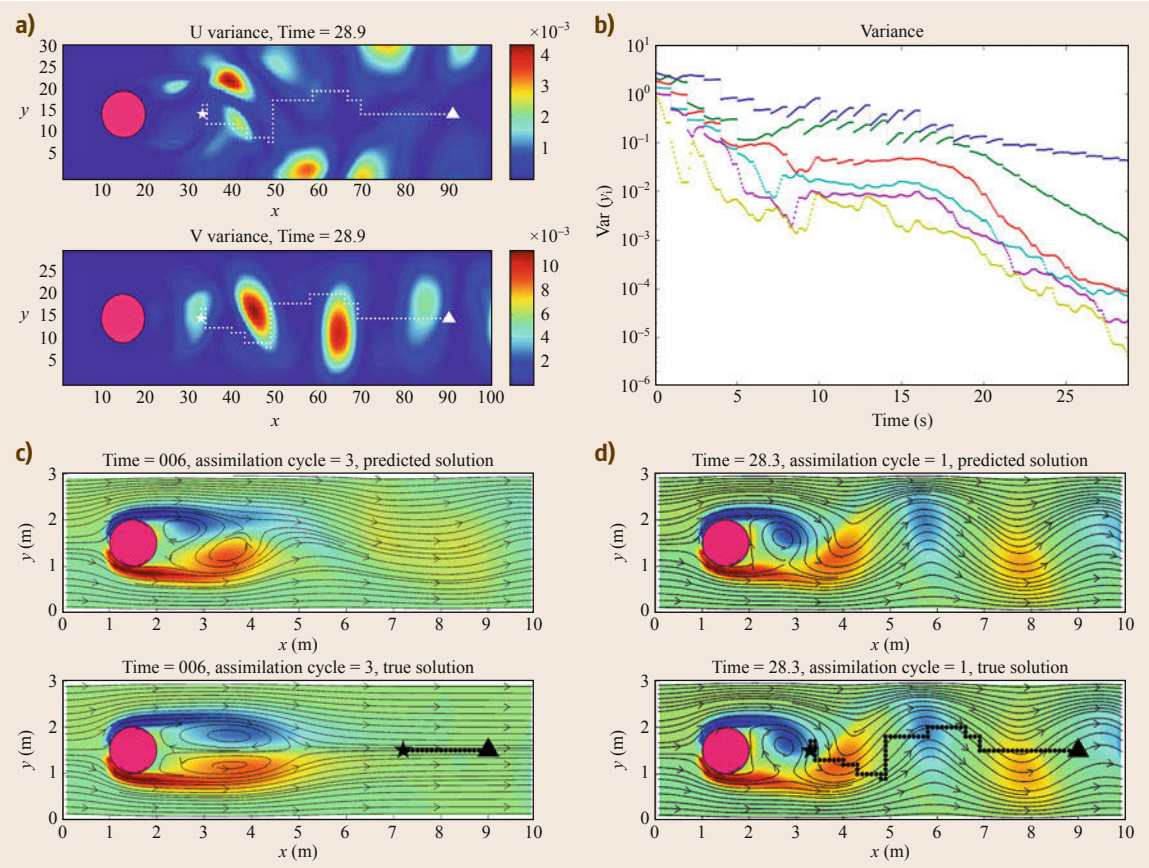


Fig. 21.13a–d *Adaptive sampling for flow past a circular island.* We predict the trajectory of a vehicle that makes measurements in the field so that the uncertainty in the field is optimally reduced. The uncertainty due to initial conditions is propagated forward in time by the DO equations and an ESSE update is used for the assimilation. **(a)** Variance in the u and v velocities of the flow field. **(b)** Reduction of variance of each of the DO modes due to the measurements made by the vehicle. **(c)** The mean of the predicted solution compared with the true velocity field. The predicted solution is significantly different from the true solution at this time. **(d)** As the vehicle measurements become more effective, the predicted solution is driven towards the true solution and eventually, both these fields match up well

reviewed some of our research in adaptive sampling. The path planning and adaptive sampling schemes and approaches presented were illustrated by their applications in idealized and realistic ocean conditions.

Of course, many more results, complementary to ours, have been obtained, both in optimal path planning and adaptive sampling. Even if we were to limit ourselves to oceanic and atmospheric applications, the depth and breadth of these complementary results are significant. For general reviews on oceanic adaptive sampling, we refer the reader to [21.4–9]. For general reviews on oceanic path planning, we refer the reader to [21.2], as well as to Chaps. 14, 19, and 20 of this Handbook.

In coming decades, in light of the growing human population on earth, the ocean, with the water and life it contains, is likely to become increasingly important. Since data collection and operations at sea are expensive, utilizing all of the knowledge we have to plan such collections and operations is logical and important. Useful knowledge here combines science and engineering, including observation, modeling, prediction, estimation, and control. The science of autonomy will necessarily play an increasing role in ocean applications. This is because sustaining healthy interactions between the ocean and humans is vital to life on earth. These interactions will have to be intelligently and autonomously managed.

References

21.1 H. Stommel: The Slocum mission, Oceanography **2**(1), 22–25 (1989)

21.2 T. Lolla: Path Planning in Time Dependent Flows Using Level Set Methods, Master's Thesis (MIT, Cambridge 2012)

21.3 T. Lolla, M.P. Ueckermann, K. Yigit, P.J. Haley Jr., P.F.J. Lermusiaux: Path planning in time dependent flow fields using level set methods, Proc. ICRA (2012) pp. 166–173

21.4 N.E. Leonard, D. Paley, F. Lekien, R. Sepulchre, D. Fratantoni, R. Davis: Collective motion, sensor networks and ocean sampling, Proc. IEEE OCEANS'07, Vol. 95 (2007) pp. 48–74

21.5 P.F.J. Lermusiaux: Adaptive modeling, adaptive data assimilation and adaptive sampling, Physica D **230**, 172–196 (2007)

21.6 N. Roy, H.-L. Choi, D. Gombos, J. Hansen, J. How, S. Park: Adaptive observation strategies for forecast error minimization, Lecture Notes Comput. Sci. **4487**, 1138–1146 (2007)

21.7 D.A. Paley, F. Zhang, N.E. Leonard: Cooperative control for ocean sampling: The glider coordinated control system, IEEE Trans. Control Syst. Technol. **16**(4), 735–744 (2008)

21.8 S.R. Ramp, R.E. Davis, N.E. Leonard, I. Shulman, Y. Chao, A.R. Robinson, J. Marsden, P.F.J. Lermusiaux, D. Fratantoni, J.D. Paduan, F. Chavez, F.L. Bahr, S. Liang, W. Leslie, Z. Li: Preparing to predict: The second autonomous ocean sampling network (AOSN-II) experiment in the Monterey Bay, Deep Sea Res. Part II **56**(3–5), 68–86 (2009)

21.9 T.B. Curtin, J.G. Bellingham: Progress toward autonomous ocean sampling networks, Deep Sea Res. Part II **56**(3–5), 62–67 (2009)

21.10 T. Lolla, P.F.J. Lermusiaux, M.P. Ueckermann, P.J. Haley Jr.: *Modified Level Set Approaches for the Planning of Time-Optimal Paths for Swarms of Ocean Vehicles*, MSEAS Report-14 (MIT, Cambridge 2012)

21.11 T. Lolla, P.F.J. Lermusiaux, M.P. Ueckermann, P.J. Haley Jr.: Time-optimal path planning in dynamic flows using level set equations: Theory and schemes, Ocean Dynamics **64**(10), 1373–1397 (2014)

21.12 T. Lolla, P.J. Haley Jr., P.F.J. Lermusiaux: Time-optimal path planning in dynamic flows using level set equations: Realistic applications, Ocean Dynamics **64**(10), 1399–1417 (2014)

21.13 K. Yigit: Path Planning Methods for Autonomous Underwater Vehicles, Master's Thesis (MIT, Cambridge 2011)

21.14 T. Lolla, P.J. Haley Jr., P.F.J. Lermusiaux: Path planning in multi-scale ocean flows: Coordination and dynamic obstacles, Ocean Dynamics **94**, 46–66 (2015)

21.15 M.P. Ueckermann, P.F.J. Lermusiaux, T.P. Sapsis: Numerical schemes for dynamically orthogonal equations of stochastic fluid and ocean flows, J. Comput. Phys. **233**, 272–294 (2013)

21.16 P.F.J. Lermusiaux, P. Malanotte-Rizzoli, D. Stammer, J. Carton, J. Cummings, A.M. Moore: Progress and prospects of U.S. data assimilation in ocean research, Oceanography **19**(1), 172–183 (2006)

21.17 P.F.J. Lermusiaux, C.-S. Chiu, G.G. Gawarkiewicz, P. Abbot, A.R. Robinson, R.N. Miller, P.J. Haley, W.G. Leslie, S.J. Majumdar, A. Pang, F. Lekien: Quantifying uncertainties in ocean predictions, Oceanography **19**(1), 92–105 (2006)

21.18 T.P. Sapsis, P.F.J. Lermusiaux: Dynamically orthogonal field equations for continuous stochastic dynamical systems, Physica D **238**, 2347–2360 (2009)

21.19 T.P. Sapsis, P.F.J. Lermusiaux: Dynamical criteria for the evolution of the stochastic dimensionality in flows with uncertainty, Physica D **241**(1), 60–76 (2012)

21.20 P.F.J. Lermusiaux: Uncertainty estimation and prediction for interdisciplinary ocean dynamics, J. Comput. Phys. **217**, 176–199 (2006)

21.21 P.J. Haley Jr., P.F.J. Lermusiaux: Multiscale two-way embedding schemes for free-surface primitive-equations in the multidisciplinary simulation, estimation and assimilation system, Ocean Dyn. **60**, 1497–1537 (2010)

21.22 P.F.J. Lermusiaux, P.J. Haley Jr., W.G. Leslie, A. Agarwal, O. Logutov, L.J. Burton: Multiscale physical and biological dynamics in the Philippines archipelago: Predictions and processes, Oceanography PhilEx Issue **24**(1), 70–89 (2011)

21.23 P.F.J. Lermusiaux, P.J. Haley Jr., N.K. Yilmaz: Environmental prediction, path planning and adaptive sampling: Sensing and modeling for efficient ocean monitoring, management and pollution control, Sea Technol. **48**(9), 35–38 (2007)

21.24 N.K. Yilmaz, C. Evangelinos, P.F.J. Lermusiaux, N. Patrikalakis: Path planning of autonomous underwater vehicles for adaptive sampling using mixed integer linear programming, IEEE-JOE **33**(4), 522–537 (2008)

21.25 K.D. Heaney, G. Gawarkiewicz, T.F. Duda, P.F.J. Lermusiaux: Non-linear optimization of autonomous undersea vehicle sampling strategies for oceanographic data-assimilation, J. Field Robotics **24**(6), 437–448 (2007)

21.26 D. Wang, P.F.J. Lermusiaux, P.J. Haley, D. Eickstedt, W.G. Leslie, H. Schmidt: Acoustically focused adaptive sampling and on-board routing for marine rapid environmental assessment, J. Mar. Syst. **78**, S393–S407 (2009)

21.27 J. Xu, P.F.J. Lermusiaux, P.J. Haley Jr., W.G. Leslie, O.G. Logutov: Spatial and temporal variations in acoustic propagation during the PLUSNet'07 exercise in Dabob Bay, Proc. POMA, Vol. 4 (2008) p. 070001

21.28 P.F.J. Lermusiaux: Evolving the subspace of the three-dimensional multiscale ocean variability: Massachusetts Bay, J. Mar. Syst. **29**, 385–422 (2001)

21.29 J. Curcio, T. Schneider, M. Benjamin, A. Patrikalakis: Autonomous surface craft provide flexibility to remote adaptive oceanographic sampling and modeling, Proc. IEEE OCEANS'08 (2008) pp. 1–7

21.30 N.E. Leonard, D.A. Paley, R.E. Davis, D.M. Fratantoni, F. Lekien, F. Zhang: Coordinated control of an underwater glider fleet in an adaptive ocean sampling field experiment in Monterey Bay, J. Field Robotics **27**(6), 718–740 (2010)

21.31 N.K. Yilmaz, C. Evangelinos, N.M. Patrikalakis, P.F.J. Lermusiaux, P.J. Haley, W.G. Leslie, A.R. Robinson, D. Wang, H. Schmidt: Path planning methods for adaptive sampling of environmental and acoustical ocean fields, Proc. OCEANS'06 (2006)

21.32 N.K. Yilmaz, P.F.J. Lermusiaux: Mixed Integer Linear Programming MILP Path Planning of AUVs for Adaptive Sampling: Real-World Simulation Results, MSEAS Tech. Rep., Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge (2012)

21.33 T. Sondergaard, P.F.J. Lermusiaux: Data assimilation with Gaussian Mixture Models using the dynamically orthogonal field equations. Part I: Theory and scheme, Mon. Weather Rev. **141**(6), 1737–1760 (2013)

21.34 T. Sondergaard, P.F.J. Lermusiaux: Data assimilation with Gaussian Mixture Models using the dynamically orthogonal field equations. Part II: Applications, Mon. Weather Rev. **141**(6), 1761–1785 (2013)

21.35 E.N. Lorenz, K.A. Emanuel: Optimal sites for supplementary weather observations: Simulation with a small model, J. Atmos. Sci. **55**(3), 399–414 (1998)

21.36 M. Gardner: The fantastic combinations of John Conway's new solitaire game *life*, Scientific American **223**, 120–123 (1970)

21.37 R.A. Monetti, E.V. Albano: Critical edge between frozen extinction and chaotic life, Physical Review E **52**(6), 5825–5831 (1995)

22. Cooperative Vehicle Target Tracking

Frank Ehlers

As autonomous ocean vehicles become more affordable and reliable, applications of multivehicle teams become more feasible. Cooperative vehicle target tracking is a promising application since in many sport, military, and biological endeavors cooperative strategies have proven themselves to be advantageous over noncooperative strategies. The sophistication of winning human team interactions indicates the challenge inherent in programming a fleet of autonomous vehicles to work cooperatively to effectively and efficiently accomplish a goal. For a multivehicle team operating in the continuous and transient ocean environment, meeting this challenge involves optimizing a high-dimensional parameter space, even more so in cases when the target has the ability to make intelligent choices to avoid being tracked. In this chapter, recent progress to construct a theoretical framework and recent applications for maritime surveillance are presented. A methodology is developed that can help bridge the gap between the top-down view starting with theoretical concepts and the bottom-up view dealing with all details of real experimentation and execution at sea.

22.1 **General Theoretical Framework** 499
22.1.1 Control and Estimation Approach 501
22.1.2 Multiagent Approach 502
22.1.3 Cellular Automata Approach 502
22.1.4 Probabilistic Graphical Models Approach 502
22.2 **Distributed Sensing, Control, and Decisions** 503
22.3 **Multistatic Sonar** 505
22.4 **Maritime Surveillance** 510
22.5 **Effective Coordination Schemes** 511
22.5.1 The *Understanding Phase* 511
22.5.2 The *Trading Methodology* 511
22.6 **Conclusions and Recommendations** 512
References 513

22.1 General Theoretical Framework

Target tracking in surveillance applications is essentially a problem of connecting signature dots (glimpses of a target) in a noisy, uncertain environment. Cooperative vehicle target tracking is a method in which multiple sensing platforms (vehicles) jointly react to a target's behavior, which may or may not be evasive, to maintain contact and extend previously observed tracks. The maritime environment is noisy, uncertain, and spatially extensive. In maritime surveillance applications noise and uncertainty can be reduced through persistence and adaptability. For operations over large areas, the scalability of any target-tracking methodology is an important practical consideration. Autonomous ocean vehicles are useful tools for maritime target tracking since they enable both persistence and adaptive, scalable approaches [22.1, 2].

Teamwork can be efficient and effective in successfully completing difficult tasks. Teamwork is a cooperative effort by a group of individuals or agents to achieve a common goal, wherein individual team members contribute different skills, knowledge, and interests to the effectiveness of the group. In a game setting, targets are modeled to make locally optimal decisions based on their established tactics (or internal *mental state*) independently (from their perspective) of the activities of the surveillance team. Constrained by established tactics, the target's dynamics can be expressed by a control algorithm that formulates the objective of the chosen *mental state*. For example, if the *mental state* is to drift to save energy, the target dynamics can be linked to the oceans dynamics. The surveillance team's goal is to track the target with distributed mobile sensors under centralized and/or distributed control in a coordination scheme that may be called *synchronized covertness*.

In this chapter, the advantages (or even necessity) of teamwork for target tracking using autonomous ocean vehicles are investigated. A *trading methodology* is described which enables the generation of coordination

schemes for cooperative target tracking with measurable effectiveness. This *trading methodology* can be used to evaluate the effectiveness of different schemes, and to predict the effectiveness of a given scheme under changing environmental and operational constraints. The trade-offs are, in general, effectiveness versus efficiency (for a specific scenario, time to find a target versus number of assets). Applying the methodology to a specific scenario avoids the computational limitations of brute-force processing methods. Two example scenarios are discussed stemming from multistatic sonar and maritime surveillance.

Cooperative target tracking using autonomous ocean vehicles is a multidisciplinary endeavor that benefits from recent progress in distributed sensor networking, networked robotics, collaborative signal processing, control and game theory, and interactions of multiagent communities [22.3–11]. However, maritime applications have unique challenges including transient environmental conditions, high noise levels at sensors, sparse coverage over large areas, intermittent and low bandwidth communication channels, navigation uncertainty, and limited energy availability.

Communication is essential for coordinating the behavior of individuals or agents. Communication can be preplanned or adaptive. To minimize communication bandwidth, two questions must be answered: What information should be shared and with whom should this information be shared? More specifically, for the target-tracking task: How much information is sufficient to allow data association and pattern recognition? Given the assumption of intelligent targets with detection and avoidance capability, communication events should be as rare as possible. From a system design point of view, the communication process should minimize interdependency, since interdependency can result in unpredictable and perhaps uncontrollable system behavior.

Here, the target-tracking team consists of an array of mobile sensors distributed throughout an ocean domain of interest. If the spatial domain is large, and the number of mobile sensors relatively small, then part of the solution for robust tracking is determining curves of intercept for sensors along a projected track. Such curves can be mapped covertly with minimal target reaction to the sensing or they can be manipulated through deliberate interaction with the object. The tracking process can be divided into two phases:

- A search phase (I) in which the cooperative sensor interaction is used to fuse measurements in time and space to increase the signal-to-noise ratio of potential target signatures.
- A holding phase (II) in which the highest ranked signals from the first phase are correlated with tracks and the tracking team adapts to best follow the tracks.

In Phase I, the performance of a single sensor can be enhanced to increase range, or multiaspect measurements made to better determine a signature.

In Phase II, a track can be followed with a single-sensing system whose speed and maneuverability matches that of the target (*track and trail*), or with a network of sensors distributed over the domain in a *handoff* mode.

In both phases, if the sensors are stationary and the detection ranges are modest relative to the domain size, the number of sensors must be large. If instead the sensors have mobility and can adapt to the environment and to the projected tracks, then their effectiveness increases and the number of sensors is reduced. Further effectiveness gains can be achieved if the sensors can adapt to one another (cooperative behavior). This chapter focuses on the latter approach to tracking performance enhancement.

When the target is reactive (perceives the tracking and maneuvers evasively), cooperative target tracking becomes a game. Each player in this game has a given system capability. Both must be modeled consistent with the stochastic nature of the control and estimation processes operative during game execution.

Cooperative target tracking over realistic maritime time and space scales is a complex adaptive system. The specific application dictates the level of cooperation needed:

- From situations where the cooperation is not really necessary to achieve sufficient performance, and the built-in cooperation dependencies introduces an overhead which hampers the overall system. *Divide and conquer* is rather the better heuristic then. Stated differently: If the cooperation introduces interdependencies, the risk that these interdependencies produce is that an unstable and unpredictable overall system should be monitored.
- To situations where cooperation is mandatory, especially against stealth targets, in nonstationary environmental conditions. In this case, single sensors or platforms are not capable of detecting targets or holding tracks, and collaboration is the only way to achieve a sufficient overall system effectiveness. The interconnectivity between the collaborating platforms is key to achieving the gain in effectiveness: One system is helping the other or one platform is covering the potential poor performance of another platform. This covering can be done simply (but not very efficiently) by adding more assets, scanning, and acting independently. Or it can be done by implementing synchronization

or coordination with careful attention to interdependencies that may lead to an unstable overall system.

For realistic scenarios, the quantitative evaluation of the effectiveness of cooperative target tracking systems with tractable decision rules for their operation is a subject of current research. Rigorous analytical solutions exist only at the level of methodological ideas, rather far from implementations, and numerical simulations are challenging. To gain insight into the performance advantages (or disadvantages) of such systems and assess trade-offs and sensitivities, a simplified framework must be constructed that models essential features at tractable levels of complexity. Repeatable experimentation within this simplified framework provides sensitivity analysis and uncertainty estimation. When transitioning from single-sensor systems to distributed, mobile systems, an essential feature is the coordination mechanism between spatially separated sensors that may not have the same information due to limited communication between them. The behavior of complex adaptive systems is often nonlinear. A small change in one specific variable, such as the coordination mechanism, may result in large changes in system performance. Modeling the behavior of the overall system as a function of this specific variable is critical for assessing system stability and determining overall measures of effectiveness (MoE) from component measures of performance (MoP).

Four approaches are commonly used for analysis of cooperative target tracking:

- Control and estimation (CE)
- Multiagent systems (MAS)
- Cellular automata (CA)
- Probabilistic graphical models (PGM).

22.1.1 Control and Estimation Approach

In a state space approach, there is a model for the evolution over time of the state of each player. This evolution is corrupted by noise which randomly pushes the player in uncontrolled directions [22.12]. For cooperative target tracking, there is a coupling among all state variables, modeled by a set of stochastic differential equations. In addition to the motion or control equations in the state space mode, there is a set of estimation equations describing the uncertainty of each player about the state of all state variables of all players. Control design and estimation design can be separated, that is, each control equation is conditioned by the result of the estimation process for each state variable [22.13]. For example, the state of a member of the surveillance team will probably change when a target is detected nearby. How each player reacts to the knowledge about the states is defined by the control equations. The performance of the system is modeled by a multiple-objective cost function with multiple constraints. Such a cost function can be expressed in terms of risk. For cooperative tracking of a reactive target, the surveillance team will try to minimize the risk of the target passing undetected, while the target will act such that its own risk of detection is minimized.

Mathematically this risk can be calculated as a function of the MoP of the components of the systems. MoPs can be combined into *attributes* of the system specific to a task. The degree to which overall risk is minimized can be defined as the system's MoE. The art of system design is to create from expert knowledge of the system the attributes which are then used to optimize the setting of free parameters, including parameters that govern coordination. The proof that a chosen combination of MoPs is a good cost function to increase the MoE is possible for mathematically tractable setups [22.14], for example, for linear-quadratic forms [22.12]. However, this simplification is unrealistic since the main features of the motion, control, decisions, and estimation cannot be represented accurately by linear-quadratic equations. A coupled physical-mechanical-fluids control model is a better choice [22.15]. Tools from applied physics can then be used for further analysis including simplification and approximation techniques for the analytical treatment of stochastic differential equations. Other approaches to generating a proper functional representation of the risk are via learning (trial-and-error or more sophisticated reinforced learning [22.16, 17]).

MoE values for specific scenarios can be statistically derived through Monte Carlo simulations. A large number of degrees of freedom in realistic scenarios make such simulations a formidable task, and one that is more likely to yield insight into the sensitivities of specific constraints in specific scenarios rather than the impact of different coordination mechanisms for more generalized application classes.

The impact of coordination mechanisms can be appreciated by noting the advantages of complete data exchange between two sensing platforms. The sensors could exploit the received signals in a coherent or semicoherent manner. Collaborative coherent signal processing allows for an increased signal-to-noise ratio. Semicoherent techniques could, for example, exploit complementary Doppler and target strength features of a target. Joint estimation and classification can lead to higher localization accuracy and a lower false alarm rate. Collaborative signal-processing techniques can be adaptive, or the motion of the sensing platforms preselected [22.18, 19].

An important output of a mobile vehicle controller is path planning [22.20, 21], especially for distributed sensors in a surveillance mission [22.22]. Path planning can be myopic (with receding horizon) or nonmyopic. The planning horizon in an energetic ocean environment is highly variable, and targets are often unpredictable. Executing initially good plans without adapting to new conditions is risky. Thus, extrema in the cost function solution surface, which is a formulation of the risk, will vary with time in the absence of adaptive pathways.

22.1.2 Multiagent Approach

In a surveillance scenario, targets and target trackers can all be viewed as agents. Some agents may communicate and cooperate, some may be passive, some reactive. The agent paradigm is helpful because it explicitly encapsulates features that in the CE representation are implicit. In a cooperative system, each agent seeks to maximize a joint global reward function.

Cooperative target tracking can be formally described as a decentralized partially observable Markov decision process (DEC-POMDP). Since standard procedures for solving POMDPs are computationally intractable for scenarios of large dimensions, approximations play a major role (as in the CE approach), and the art is in adapting the way approximations are made to the structure of the interactions [22.24–28]. The local policy for agent I is a mapping from the local history of observations made by agent I to an action. To program the autonomy in the cooperative target-tracking team members, this local policy has to be found. This is equal to finding the control design in the CE approach.

The tracking of reactive targets is the field of multiagent partially observable stochastic games (POSG). POSG formulations can be found in [22.29]. Formulated for noncooperative multiagent systems, they allow each agent to have a unique reward function. Further analysis of Bayesian games can be found in [22.30].

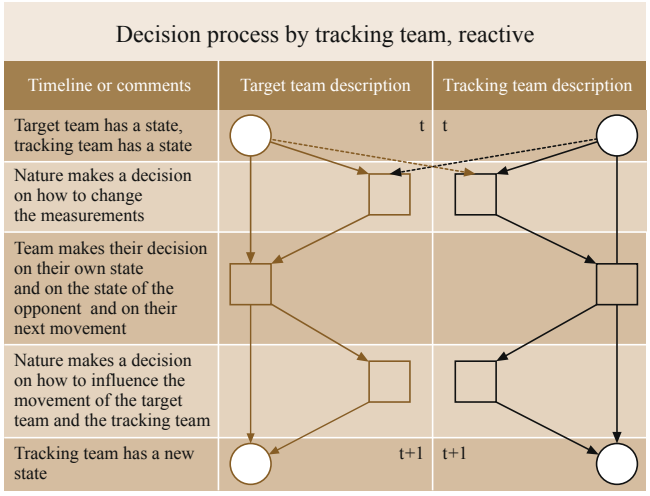


Fig. 22.1 An influence diagram (a special type of PGM) contains nodes that represent chance variables (*circles*), decisions (rectangles), and utility values (the utility values are similar to the attributes discussed in the CE representation) (*diamonds*) (after [22.23])

22.1.3 Cellular Automata Approach

A cellular automaton consists of a regular grid of cells, each in one of a finite number of states. The grid can be in any finite number of dimensions. For each cell, a set of cells called neighborhood is defined relative to the specific cell. An initial state is selected by assigning a state for each cell. A new generation is created by some fixed rule which can include stochastic and asynchronous elements. When analyzing systems for cooperative target tracking by cellular automata (CA), simplifications are made on the internal activities, and the focus is set on the interactions in the neighborhood of the cell. Often these interactions generate group behavior that allows each member to be maximally independent of the target.

Linking the CA approach to the MAS approach, each tracking agent has interactions with the other agents and the target. For many agents, such interactions result in a complex system. Self-organization methods like random graphs and finite state machines are used for discrete event control in distributed surveillance [22.31, 32]. A general description of such complex, nonlinear systems gives a useful insight, in emergent or swarming behaviors of the entire system [22.33]. Starting with many CAs, which are all independent of each other, but depend on the target, they trade their independence among each other, to become more independent of the target. In CA language, this is a formulation of the main idea behind the *trading methodology* discussed in the introduction. In a swarm of fish, for example, the individuals have a strong dependency to their nearest neighbors, but in the swarm the individuals are well protected against enemies.

22.1.4 Probabilistic Graphical Models Approach

The power of this approach is best described by a citation [22.34]:

> *Probabilistic graphical models (PGM) encode our domain knowledge declaratively, using a representation that has its own inherent semantics. Thus the*

conclusions induced by the model are intrinsic to it, and not dependent on a specific implementation or algorithm. This property gives us the flexibility to develop a range of inference algorithms, which may be appropriate in different settings. As long as each algorithm remains faithful to the underlying model semantics, we know it to be correct. The existence of a declarative representation allows us to separate out the two sources of error, modeling error and algorithmic error, and consider each separately.

Influence diagrams (Fig. 22.1) are able to represent agents beliefs and decision-making processes [22.35]. The reason they are well suited is that it is possible to implement factored reasoning (which is exploiting independence) for monitoring team formations [22.36]. They can also be applied to implement decision making in open mixed networks [22.37]. In [22.38], a graph-based model of distributed cooperative navigation for multirobot measurement is presented. Solving the task of cooperative target tracking could start with similar mathematical tools, expanded to include targets acting in a noncooperative manner. The influence diagrams can be used to explicitly visualize the steps of the *trading methodology*. This will be further discussed in Sect. 22.5.

22.2 Distributed Sensing, Control, and Decisions

If the data fusion is performed correctly, multisensor measurements have a better detection performance than the best single sensor among them [22.39]. For example, spatially distributed sensors resolve target ambiguity by providing different aspects of the target. Not all information has to be sent to a fusion center. Depending on the setup, it might be sufficient to send only selected information to specific team members (optimally, this should be the most important information). This links the distributed sensing task to a local decision rule that specifies which data should be sent to whom to increase the overall utility function with the constraint of limited bandwidth.

Based on the incoming data, either from own measurements or via communication channels, each team member has to fulfill its specific control tasks, which again (although individually executed) are to increase the team utility function with the constraint of achieving consensus among team members on sharing the workload [22.41, 42].

Thinking about control as an activity to be done by each team member, distributed decision making is understood as the method to generate agreement on a specific state inside the team (e.g., switching from target search to target trail). It is possible to generate consensus (distributed decision making) within the team without a central controller.

In a distributed architecture, two types of communication are necessary:

1. To establish the data fusion, and
2. To coordinate the sensing activities (Fig. 22.2).

Molinares [22.43] describes a data link layer in support of swarming of autonomous underwater vehicles which would support both types of communication.

In the maritime domain, different communication methods exist: underwater acoustic, underwater laser, underwater cable, underwater electromagnetic, and above water radio (high frequency (HF), very high frequency(VHF), ultra high frequency (UHF), satellite communication (SAT)). An important design task is to find capable and reliable communication links, including knowledge about the probability of errors associated with these links. All channels have different ranges and bandwidths. For maritime applications over large areas, satellite communication offers the highest bandwidth channel to share data over long ranges to enable a centralized network architecture for data fusion. However, underwater vehicles must come to the surface in order to communicate via satellite communication channels. As communicated sensor information by team members is a measurement, the accuracy of this measurement has to be taken into account for fusion.

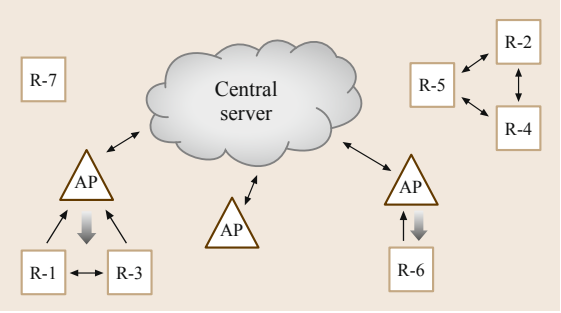


Fig. 22.2 Diagram of the network structure. Robots (*squares*) are able to communicate on a peer-to-peer basis with nearby robots as well as access the central server through access points (*triangles*). The communication links originating from robots are all relatively low bandwidth while the downlink from the server may be higher bandwidth (after [22.40])

To adapt to inherent uncertainties, various learning methods can be employed. Machine learning with reinforcement is a powerful tool; however, its extension to multiagent and large dimensions is not trivial. The mathematics for reinforcement learning are similar to the CE approach. Learning is also equivalent to the setting of specific parameters in a POMDP. Learning for CAs can be executed via genetic algorithms. Learning in Influence Diagrams is formulated as an optimization and/or fixed point problem (i. e., solvable by iterative application of an optimization operation).

In general, learning works better if the parameter space is smaller and if the free parameters can be treated independently in the optimization process. An open research question is whether best preplanned collaboration exists. From a coordination expert system point of view it could be a decision algorithm on what to do under specific situations. This implies that for a large number of specific situations, a large number of coordination settings have to be stored, which becomes impractical in realistic scenarios. A possible solution is the creation of a *team template library* as an algorithmic version of the *trading methodology* mentioned in the Introduction. The algorithmic approach first identifies critical parameters for a specific situation. Optimization with respect to coordination is then performed at this identified *criticality ridge*. Since situations are constantly changing, both a similarity measure for situations (*this situation is similar to the one we had yesterday where we successfully applied the following coordination scheme*) and a fast identification method for critical parameters are needed. For the construction of such a fast identification method, the heuristic of blind separation techniques could be applied: a critical parameter changes the system behavior independently of the value settings of other parameters. Information science offers the tools to design, store and implement situational and process-specific details in ontologies. Ontologies represent all bottom-up (*the devil is in the detail*) information. The *team template library* is where all detailed specific bottom-up information is linked to the criteria of applicable coordination schemes (which can also be described as algorithms representing the top-down approach to cooperative tasks). Sections 22.3–22.5 further elaborate on the *trading methodology* that incorporates both heuristic and expert knowledge to generate and evaluate coordination schemes.

Coordination for cooperative target tracking is a tool to allow cooperation. Coordination methods differ in the amount of data transfer they need, the speed of convergence, and the sensitivity to error.

Coordination schemes applied in cooperative systems (multirobot, human societies, business processes, and information science) are:

- *Divide and conquer*: no interaction necessary after planning phase.
- *Decentralization due to a limited sensor range* [22.44].
- *Decentralization due to limited control access*.
- *Auction*: bidding for a task. An auction is a process of buying and selling services by offering them up for bid, taking bids, and then selling the item to the highest bidder. The auction procedure has to be set up before the execution of the auction.
- *Voting*: recommendations based on ratings or rankings collected from many sources. Voting is a method for a group such as a meeting or an electorate to make a decision or express an opinion. The majority wins in a typical voting procedure.
- *Consensus*: recommendations based on ratings or rankings collected from many sources, Consensus decision making is a group decision-making process that seeks the consent of all participants. Consensus may be defined as an acceptable resolution, one that can be supported by all, even if not the *favorite* of each individual [22.45, 46].
- *Game theoretic*: an equilibrium state in which each player has achieved the maximum that can be achieved (i. e., any further move will decrease the payoff).
- *Bio-inspired*: similar to ants forming teams for transporting heavy pieces of food, bio-inspired cooperative target tracking can be interpreted to be a trial-and-error approach on the information gathering task of finding and holding track of a target.
- *Adaptive hierarchy*: explicitly allowing changes of hierarchical structures at run time leads to coordination schemes adaptive to specific situations the team is facing in cooperative target tracking, for example, the surveillance units can reason about a reorganization of the team structure in the case of a sudden loss of target contact [22.47].
- *Synchronized covertness*: the design idea is to exploit the covertness of receivers, but at the same time allows for coordinated activities among them. The covertness of receivers is important when dealing with stealth targets which can avoid detection only in a limited number of geometric setups. If the stealth targets have no knowledge about the receivers positions, they cannot adjust their path planning to exploit their stealth design in these geometric setups. The mixed architecture which results from the design idea of synchronized covertness

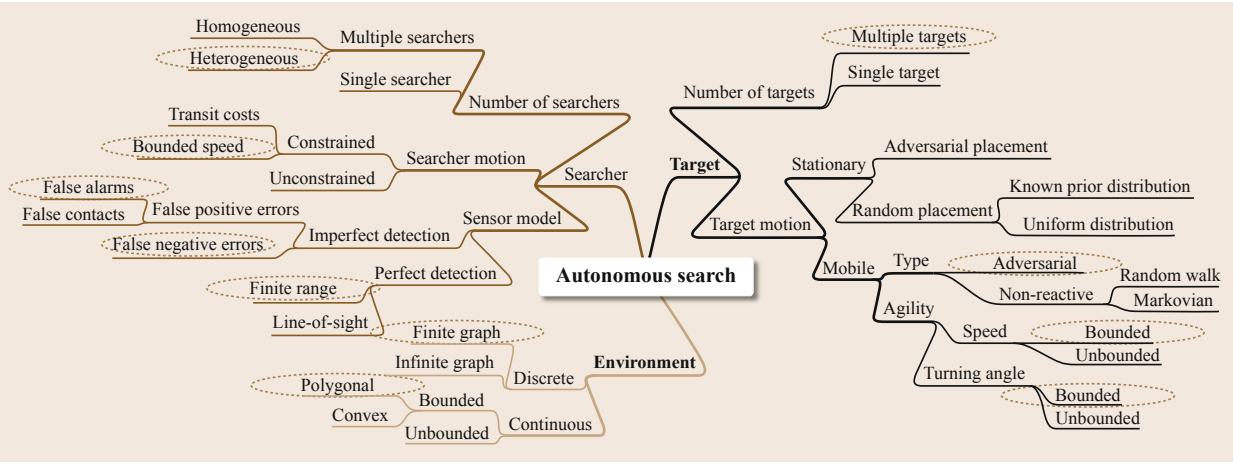


Fig. 22.3 Parameter space for autonomous search models (after [22.48]). Applications discussed in this chapter are indicated by *dotted ellipses*

is developed by the *trading methodology* described in this chapter. The *trading methodology* allows to balance the need for synchronization (or interdependence) with the benefit of covertness (or independence).

Most desirable is a coordination architecture that is a mixture of centralized elements, for example, for synchronization and startup phases, and decentralized elements that are robust, delay, and disruption tolerant. Multilayer architectures developed in the robotics community are able to support cooperative target tracking with operator interaction. Publish and subscribe (e.g., MOOS [22.49], ROS (and, e.g., PRACSYS built on it [22.50])) techniques minimize the interdependencies between the various processes within such a multilayer, multisensor, and multiplatform network. A cooperative architecture for information-theoretic search and track has been examined in [22.51]. Multiple robots are used for cooperative target tracking in [22.52]. Energy management is always an important consideration [22.53].

Few multirobot systems have been tested on the scale necessary to demonstrate pursuit evasion in complex environments such as the ocean (Fig. 22.3). In addition, a reliable communication infrastructure is required to guarantee that a plan will be executed successfully. Decentralized architectures and methods capable of utilizing new information as it becomes available can reduce these requirements, but the development and evaluation of such algorithms remains an active research area [22.48].

The probability of detection of a target depends on the target motion, particularly on the target strength for specific aspects of the target relative to the detector, or on the anomaly of its behavior with respect to many other neutral agents in its vicinity. Systems designed to exploit these two detection methods are discussed below: multistatic sonar target tracking and maritime surveillance.

22.3 Multistatic Sonar

In this section, we describe multistatic sonar from its implementation point of view (the bottom-up description), then we switch to the top-down view of describing the objective for the application of multistatic sonar. Within the top-down view, we review standard concepts for cooperative tracking that could be applied. We specify then a benchmark problem which explicitly includes the main features identified by the expert evaluation of the bottom-up view. We formulate the benchmark problem as a fair game and describe a solution methodology.

Detection and tracking of objects in the ocean by passive acoustic means are ultimately limited by emitted sound source levels. In noisy, multipath environments, passive approaches tend to be inadequate. Active acoustic systems, on the other hand, offer sufficient performance, especially when operating in multistatic mode, using multiple, spatially distributed sources and sensors to activate and receive signals. In shallow water areas (< 100 m), applying data fusion and tracking algorithms to multistatic sonar data are challenging due to noisy and fading channels, reverberation and clut-

ter [22.54, 55], but can be solved [22.56]. A simple and surprisingly well-performing approach in the signal and contact processing for multistatic tracking is to sequentially process the incoming contacts. The motion model inserted in the tracking algorithm (by Kalman filtering) is implicitly concatenating the contacts together with their uncertainties. Sequential handling of contacts has less performance than a full fusion, but in many applications and simulations this procedure is adequate.

For multiple tracking platforms, the data fusion process increases track accuracy and track length, and minimizes track latency. For long-range applications, sequential target tracking can be used as a method for data fusion [22.56]. If the sequential target tracking is not feasible due to too low probability of detection, data fusion before tracking or a multisensor track-before-detect may be possible. [22.57, 58].

Multistatic sonar enables false alarm classification by evaluating hypothesis on the existence or nonexistence of contacts in different source receiver geometries depending on the size and structure of the objects or larger underwater features. These geometric considerations lead to a very important feature of multistatic sonar: It is antistealth, in the sense that the target has less degrees of freedom in deciding what to do next without being detected, which relates directly to a decrease in the number of internal *mental states* available to the target. Multiple acoustic sources can be used to localize autonomous receivers automatically [22.60], resulting in a very high precision of target position estimation. Any mobile array of receivers can be used for spatially adaptive target tracking. Sensor performance is related to the overall aperture and element spacing. System performance is determined by a combination of both sensor performance, networking capabilities, and platform agility.

Multistatics with autonomous systems is advantageous because it allows low-risk sources and receivers in close proximity to the target, resulting in a higher signal excess [22.61]. The fusion of contacts has to account for the quality of each measurement correctly. To improve the measurement of each sensor, two MoPs have to be combined: the detection probability (related to the expected signal-to-noise ratio of target returns) and the accuracy of the localization. How these two MoPs are combined depends on the situation (states of all assets and of the environment). Measurements that are uncorrelated provide maximal information. Information theoretic metrics (e.g., Renyi entropy [22.39]) can be used to calculate the information added by using multiple sources and receivers.

Multistatic sonar can be understood as a semantic sensor network (Fig. 22.4). The realization of such multistatic sensor networks [22.59] reveals the need to specify and implement all elements of ontology. Specific field trials in the ocean have demonstrated that the complete chain from detection (glint detection), communicating, and simultaneous track holding is pos-

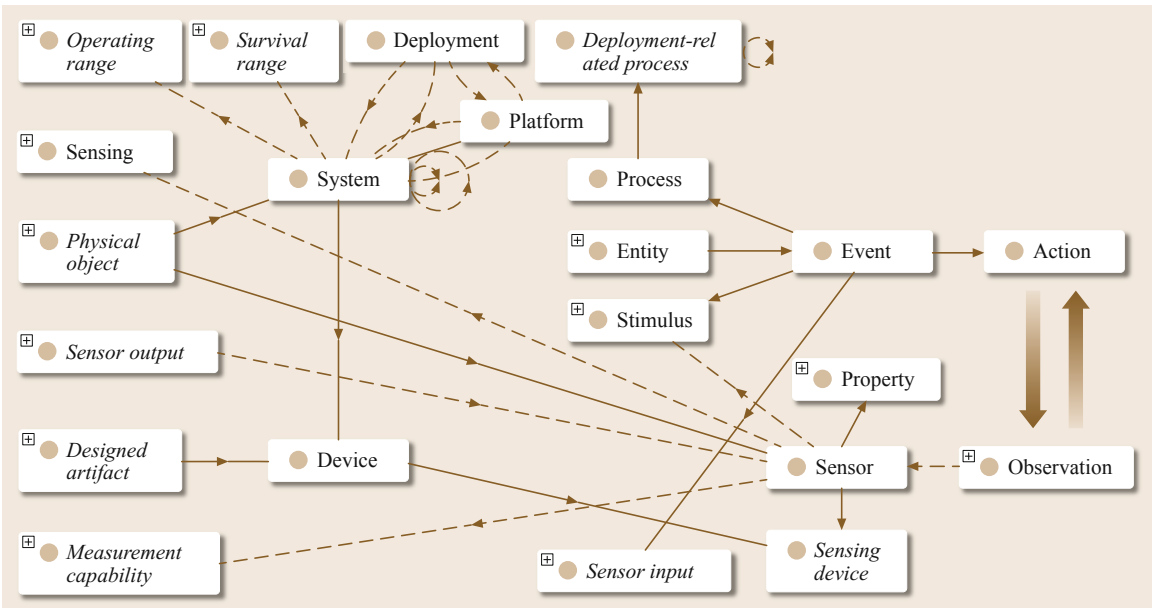


Fig. 22.4 The semantic sensor network (after [22.59]) with the focus on the link between observation and action that has to be statistically separated (become independent) from each other for target's actions and surveillance's observations in order to generate a fair game

sible [22.62–65]. The multistatic approach has also been demonstrated in other domains [22.66, 67], and multirobot patrolling has been applied to other areas, like harbor patrolling [22.68].

Experimental experience provides a valuable estimation of costs to the elements of the ontology. Efficient cost accounting will be needed later (Sect. 22.5) when in the *trading methodology* the ontology design must be changed.

Scenarios for cooperative target tracking using low-cost autonomous vehicles for surveillance in a predefined area covered by several stand-off sources are depicted in Fig. 22.5. We call this scenario *low-frequency cooperative target tracking* (LFCTT). This implementation is a distributed, mobile sensor network with limited inter-platform communication. We assume that the stand-off sources can be used to continuously distribute a small amount of coordination information via the acoustic underwater channel. The receivers should be kept covert as much as possible to limit target adaption, but have the capability of transmitting over the low-frequency acoustic channel information on specific confirmed tracks. This information can be encapsulated into a few hundred bytes every minute.

For these scenarios, we develop a coordination scheme and outline how to evaluate its parametric settings.

A possible approach from control theory is to model the target as a worst-case target, meaning that in every time step of the game, the target is reacting noncooperatively to the maximum extent, trying to avoid detection as much as possible [22.12]. Inserting the resulting reward matrix into the control equations of the surveillance leads to a set of equations dealing separately with the controller and with the observer design of the surveillance team. In the linear-quadratic control case, this leads to closed-form equations for how to handle the observations and how to react after analyzing them.

The worst-case scenario makes a high demand on quality of sensors and number of assets in the surveillance team. The resulting strategy of the surveillance team is probably close to the scenario where all possible target tracks lead to the detection of the target (independent of its strategy). For example, the target might guess that the surveillance team is preparing for its most clever move; hence, it is taking a (from the perspective of the surveillance team) less clever, but (with the surveillance team having expected a different move) at the end successful move. As a result, the surveillance team has to be prepared for everything, which means organizing itself to detect targets under all circumstances, an extreme scenario that is likely to be unrealistic.

To approach cooperative target tracking in a more realistic way, we choose a heuristic generally used for the analysis of complex adaptive systems: avoid interdependencies, strengthen interconnectivity. Following this heuristic, the resulting network becomes more tractable (phase transitions are less complex) [22.69].

Complex phase transitions lead to unpredictable chaotic behavior of the entire system. Interconnectivity is maintained in LFCTT by the use of multiple acoustic sources and receivers. Greater signal diversity (aspects of the target illuminated) results in increased detection probabilities [22.56]. Interdependencies are avoided by the tendencies of both the target and the surveillance team to operate independently under their own control.

We assume that the LFCTT scenario has two different subtasks: subtask 1 where the target's position is unknown to the surveillance team, called search task; subtask 2 in which the surveillance team has made a definitive decision on the target's position and tries to hold the track as long as necessary to initiate an appropriate action [22.39].

In the search task, if the surveillance team is acting such that the target cannot predict its behavior over a longer time frame, an appropriate strategy for the target is to act optimally in the current situation, but to avoid situations in the near future in which it might be trapped.

At this point in the design process, it makes sense to use a benchmark problem to highlight the most im-

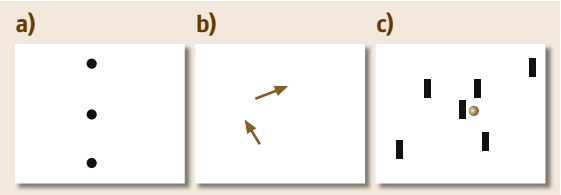


Fig. 22.5a–c Scenario consisting of (c) a surveillance area, and areas for (b) autonomous receivers and (a) stand-off acoustic sources. For low-frequency active sonar, the actual distances between the boxes can be large, especially for the sources (*dots*) which are operating as stand-off illuminators of the surveillance area. The goal of the target (*dot*) is to reach the left side of the surveillance area. It can hide at clutter points (*black rectangles*). The coordination necessary in this LFCTT game is for the receivers (*arrows*) which have to have a simultaneous good view on a specific patch in the surveillance area to make optimal use of the multisource and multireceiver setup. Since the discrimination of clutter and target detections is difficult because of nonstationary fluctuations, we assume that the best data for such discrimination is generated if the sound transmitted from all three source is hitting the same patch to which the receivers are focussing

portant features of the scenario. We further assume that the cellular automaton modeling the target consists of a small number of *mental states* (reckless, offensive, defensive, cautious, conservative) (Fig. 22.6) that influence how the actual path planning of the target is executed. We postulate that each action possible in a *mental state* leads to independence from the surveillance team in the near future. Changing the *mental state* is only possible for the target when:

- The target certainly knows that it is undetected on some time scale.
- The target has reached a point to hide, for example, within known clutter fields in the environment.
- The target has reached a safe point far enough away to be not affected by the surveillance team.
- The target has reached a point in which there are still many *mental states* (and therefore actions) available to be finally successful and the surveillance team is still acting randomly and independently of the target actions.

We explicitly model clutter that can be from databases or from an in-situ clutter analysis. Clutter plays a major role in the LFCTT game, influencing the measurements of the surveillance team and influencing the target movements. We assume that classification is best when three sources converge on one target. This is analogous to the saccades, movements of (human) eyes in surveillance mode, where the two eyes coordinate their focal point quickly within the scene

To simplify the calculations, we model sensor performance crudely, taking only grid cells into account that are under best surveillance conditions. This approach is similar to the *winner-takes-all* (WTA) methodology in artificial neural networks. We call these grid cells the WTAs of the surveillance team.

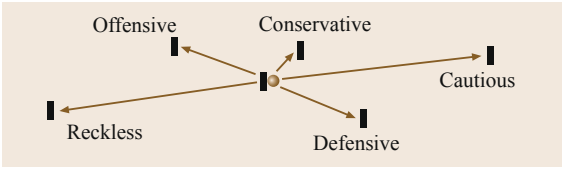


Fig. 22.6 An example for the decision-making process, modeled as a celluar automaton (CA), inside the target (*dot*) on which clutter point (*black rectangles*) to reach next. It is deciding depending on its *mental state* on how the path for the near future actions is calculated. Since clutter points are genarally a place to hide, in the example of Fig. 22.6 we depict a one-to-one mapping of *mental state* (reckless, offensive conservative, defensive, cautions) and actual path. The reality in LFCTT is much more *stochastic* in the sense that uncertainty about positions, resolutions and performance of tracking processes by the surveillance team have to be taken into account. But these effects are understood in the sense that they can be adequately modeled to extend this simplified CA model

The simplification for sensor performances by WTAs helps to infer that the best performance is generated if the crossing beams of the receivers are located in the area defined by the sound sources. This leads to a coordination scheme in which the sources also coordinate the movement of the receivers which might be autonomous underwater vehicles (AUVs) equipped with towed hydrophone arrays (Fig. 22.7).

In the game, the target does not know which actions the receivers are taking. However, it is reasonable for the target to assume their existence (otherwise the game-theoretic setting would not apply). The receiver positions are only unknown until they participate actively in long range communications. Hence, the target's planning horizon should include an evasive reaction on this detected communication event. Maximizing independence means for a receiving platform to keep its covertness as long as possible.

Subtask 1 of the game is finished if either the target transits through to its goal undetected or the surveillance team passes a persistent track to subtask 2.

The benchmark problem just described leads to a game that can be further analyzed. Percolation theory as an instance of renormalization methods can be used to analytically predict the game outcome when pa-

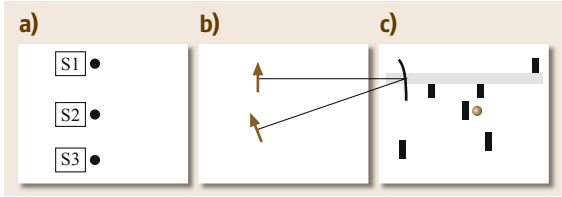


Fig. 22.7a–c Scenario from 22.5 with coordination scheme: The sources are located in **(a)**, the receivers are positioned in **(b)**, and **(c)** is the surveillance area. The three sources coordinate their ping transmission scheme in such a way that a patch in the surveillance area is hit by the sound simultaneously. For example, if the two sources S1 and S2 are pinging at the same time, the area filled with *dots* is specified. The source S3 can arrange a further focus with its ping timing (e.g., the *curve* crossing the the *shaded gray area*). Regardless on their own position, the receivers (*arrows*) can quickly change heading to point with their broadside beam to this patch. By the coordination between the sources, the receiver movement is also specified without any further communication between the sources and the receivers being necessary. If the receivers are AUVs equipped with towed arrays, the changes in heading can be performed within a short time frame

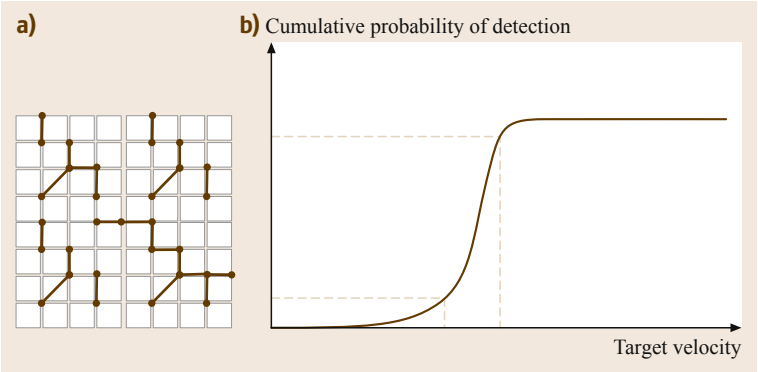


Fig. 22.8a,b **(a)** Percolation theory to describe whether a target breaks through a barrier. Each leg corresponds to a possible path of the target with associated probability of detection. **(b)** Depending on the velocity of the target the probability of the barrier to be open or close goes through a critical phase. The sigmoid shape of the cumulative probability of detection is a typical example for the behavior of complex adaptive systems. The system is transitioning from one phase (barrier open) to another phase (barrier close). The *dashed lines* mark a critical parameter interval around the critical value of the cumulative probability of detection of 0.5

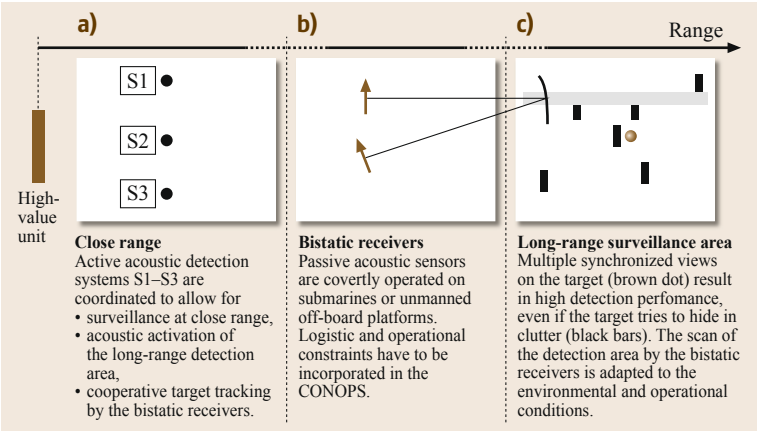


Fig. 22.9a–c Layered defense setup for the scenario from Fig. 22.5. The long-range surveillance in area **(c)** is described by the synchronized covertness process in Fig. 22.7. In area **(b)** the covert receivers can simultaneously act as passive receivers. In area **(a)** the additional installation of standard active sonar protection of the High Value Unit is possible by just adding receivers to the stand-off sources

rameters are chosen to be outside the fair game setting (Fig. 22.8).

Furthermore, once understood at this analytical level, benchmark problems can be combined, in a layered defense setup (Fig. 22.9). The more the layers are designed to be independent from each other, the better can the analysis of the resulting multilayer system inherit elements of the analytical evaluation of the benchmark problem for the long-range detection area (Fig. 22.9c).

Summarizing the approach outlined in this section, we can identify two major steps:

1. An *understanding phase* in which a fair game is constructed as a benchmark problem. The benchmark problem deals only with the important features of the cooperative target-tracking application in multistatic sonar. It separates critical functionalities or agents in the description. These simplifications and compact descriptions allow a further analytical treatment and an extension of the findings by this treatment to larger, multilayer structures.
2. A design phase for multilayer systems where coordination plays a major role for the construction of both effective and efficient systems. Effectiveness and efficiency can be predicted by re-using the results from the analytical treatment of the benchmark problem in the *understanding phase*. This design phase can follow a *trading methodology* as further explained in Sect. 22.5.

22.4 Maritime Surveillance

Marine traffic is one of the backbones of the world's economy. Together with fishing and industrial exploration and recreation activities, an enormous amount of uncorrelated movements take place. The amount of data is tremendous; sensors are erroneous and cannot always been trusted. For example, AIS (automatic identification system) is a locating signal which all large commercial ships must carry, though many vessels transporting illicit cargo turn off their AIS transmitters in order to avoid cooperatively reporting their status. Turning off the AIS transmitter however is an anomalous behavior which can be detected.

For information gathering tasks, the evaluation of sampling strategies depends on the ability to produce quantitative metrics for system performance, which in turn depends on having a clear sense of what aspects of the observed field are important. Much recent effort has been directed toward *anomaly* detection. The concept of an *anomaly* or *anomalous behavior* is dependent on baseline statistics that are problematic when regimes can only be sampled sparsely. In such cases, the role of feedback will be even more critical with perhaps active probing (sending an UAV (unmanned aerial vehicle)) necessary to elicit a response. False positives (the false alarm problem) and their associated costs are typically the most serious limitation of a highly sensitive system. When risk is included as a metric, state variables may have different associated utility and expected values. In the case of systems for anomaly detection, consider a large sensor network including UAVs, AUVs, ships, radars, etc., for the detection of smuggling, or otherwise anomalous behavior in coastal waters. The system would not only detect anomalous behavior, but it could also access databases about vessels, their cargo and crew in order to estimate risk for the particular observation. This imposes high levels of uncertainty in the system as public databases might be inaccessible, or UAV assets not available at the time. The system has to constantly reconfigure in order to deduce the information relevant to the operators. An adaptive architecture is necessary for the coordination scheme to cope with this variability. The implementation of autonomous ocean vehicles could potentially be a game changer.

In maritime surveillance, there are many innocent players who are following the rules but run into strange situations (by equipment failure or environmental conditions). Surveillance of all these players is not possible. Practically, the target always risks being under surveillance (the covert receiver is the general case). Instead of trying to be invisible (which is not possible), the objective of the target is to make only its illegal activities invisible.

The aim of maritime surveillance is to check the sequence of decisions made by the target to see whether this sequence is designed to be maximally independent. That is an anomaly because decisions of the target should be normally following another strategy (the one written in the AIS log or inherently given like bringing goods from A to B, finding fish, etc.). The strategy for the surveillance team is to gather as much data as possible for this task. Coverage is an attribute for the search phase. A quick reaction capability of requesting more data if an anomaly has been detected is an attribute for Phase 2 (verification and hold). More than in the multistatic scenario, the maritime surveillance scenario will be demand based in order to avoid the transmission of all collected data. Demand-based data sharing might look like: *Something is strange here. Do you have more information?* An implementation of this network should build upon the features of a publish and subscribe architecture.

In this case, we are not only gathering information about positions, but about processes. Position information is still important because it can be, for example, the start position of a process or a constraint for a process, but in general we are looking for a measured sequence of positions and activities which is then used to estimate the target's intention. A decision on whether the target's behavior is inside or outside normal bounds is called an anomaly detector.

The target's goal is to disguise the illegal intention as much as possible for a sufficiently long time. The target is modeled to have a small number of mental states, as before, and determines how much probability of letting the surveillance detect its behavior is acceptable. The coordination architecture is again a mix between centralized synchronization and distributed decision making and the coordination method is synchronized covertness (letting the target know that you might be watching, but not whether you are really there or not). This can be visualized as a separation in the related probabilistic graphical model.

22.5 Effective Coordination Schemes

An important aspect of a cooperative target tracking system is measuring its overall performance [22.70]. In particular, connecting MoPs (and their probability density functions (PDFs)) with MoEs requires integration of specific performances into overall effectiveness in achieving objectives and goals (Fig. 22.10).

Simulations or standard reinforcement learning methods [22.17] are not feasible for large parameter spaces. The expert choice of attributes linking MoPs and MoEs is a viable method, but is not rigorous enough to test the effects of different coordination methods.

Complex nonlinear systems often show critical behavior: small changes in parameter settings lead to completely different behavior of the system. Renormalization methods have been used to analyze the criticality of systems with many degrees of freedom [22.69, 71]. Renormalization can mathematically bridge between bottom-up and top-down perspectives. Essentially, gradients in the system are analyzed from the critical point in parameter space. The parameter space is, in fact, of infinite dimension if beliefs have to be handled. This means that the *parameters* are functionals, and it is necessary to find independent subprocesses in order to handle the overall complexity. Equality of processes has been investigated and found to be challenging [22.72]. A sensitivity analysis to initial conditions is the key to allow a qualitative evaluation of the effectiveness of a coordination scheme. A quantitative evaluation becomes possible by high-fidelity simulation for a limited number of parameter settings in small domains previously identified by the qualitative evaluation.

We divide the design process into two parts:

- Creation of benchmark problems to calibrate and standardize the mapping between MoEs and MoPs is the *understanding phase*.
- Constructing an efficient cooperative team by inserting coordination into layers of originally independently operating surveillance assets (and thereby creating efficient and suffciently effective functionalities) is the *trading methodology*.

22.5.1 The *Understanding Phase*

Benchmark problems are simulations characterized by an unbiased initial condition for each team and positioned at a critical point in the parameter space. The likelihood of one team prevailing over the other is initially equal. Benchmark problems are simplifications of the real world that distil key aspects of a system to clarify and calibrate the underlying dynamics. Often the simplifications allow analytic solutions against which more comprehensive (but opaque) numerical models can be tested for fidelity.

In benchmark problems for the evaluation of the effectiveness of coordination methods, the critical design aspect is the transition from one to many agents as a means to achieving the objectives. For example, in multistatic surveillance, cooperative target tracking enables the detection and tracking of targets hidden in clutter by increasing the effective antenna aperture and thus narrowing the focus and increasing the signal-to-noise ratio. In the detection of anomalous behavior, cooperative sensing synchronized with a global observation network enables the construction of a time dependent, multidimensional vector whose magnitude and direction can reveal a target's hidden process. These two scenarios share the common features of multisensor data fusion and a smart adversary.

Approaches to generating benchmark problems vary:

- The heuristic of looking for independent subprocesses or functionalities. Numerical simulations governed by analytical considerations on the importance of these functionalities gives as a by-product the formerly unknown functional connections between components in the entire process. This functional relationship represents the link between the a-priori MoPs and the posterior MoEs.
- A game scenario constructed by experts. Idealistically, the game is set up such that each side has an equal chance of winning at the start. Alternatively, the initial conditions can be deliberately biased to test specific hypotheses. Fundamental to target detection and tracking problems is the process of minimizing uncertainty. Teams that best succeed in such minimization (through cooperative behavior) will have a better chance to win the game.

22.5.2 The *Trading Methodology*

In the *understanding phase* explicitly separated (independent) agents or functionalities are generated. The *trading methodology* starts with uncoordinated independent layers of these agents. From this starting point with a specific position in the effectiveness-efficiency plane and a guaranteed minimal effectiveness the risk harming the effectiveness of the system by inserting dependencies is traded with the predicted gain of efficiency by inserting these dependencies. For example, using the same acoustic sources in the layered defence

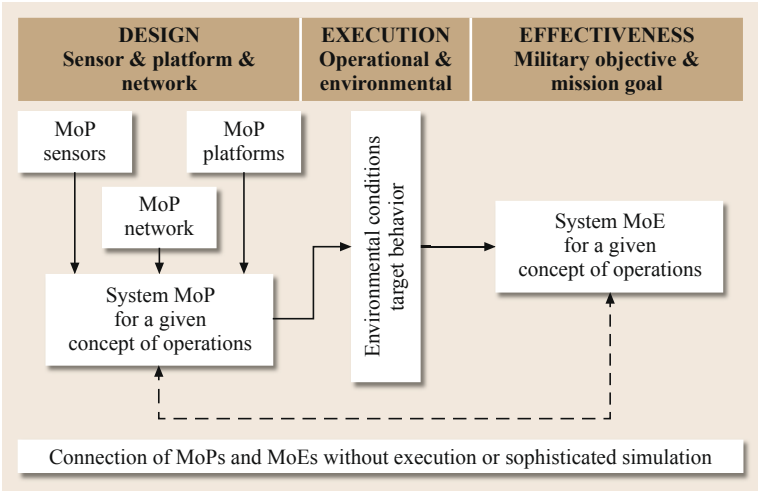


Fig. 22.10 Schematic showing pathways connecting measures of performance with measures of effectiveness

system in Fig. 22.9 for the execution of operations in all three layers is more efficient than having extra sources for each layer, but the risk is introduced of losing all three layers if the single set of sources breaks.

In the context of a probabilistic graphical model, the *trading methodology* is connecting the agents of the cooperative team. By applying the *trading methodology*, the system design takes a path on the effectiveness-efficiency surface. Connections made can be similar to already known coordination schemes (Sect. 22.3). For example in an auction scheme there will be many connections representing the process of bidding and bid evaluation. Inserting this in the multistatic framework would result in a loss of covertness and in a loss of the ability to counter stealth targets with a low number of surveillance assets. In the synchronized covertness applied to a layered defence scheme, there will be less connections due to the preplanned spatial separation of the layers and the covertness of the receivers. Both coordination schemes lead therefore to two different positions in the effectiveness-efficiency plane, relative to the started point, hence, a comparison of both becomes possible.

Instead of comparing complete packages of coordination schemes, as discussed in the paragraph above, the *trading methodology* can also be understood as an iterative process, inserting, or deleting single or small sets of connections between agents. Since inserting or deleting connections results in principle in the necessity to completely re-analyze the entire model, methods are needed to quickly recalculate the effectiveness and efficiency predicted for the updated version of the model. The analytical knowledge gathered in the understanding phase can govern finding such methods by checking the importance of the update made relative to the critical behavior of the entire system. Similarly, maintaining substructures (hierarchies) invariant to chances, which could be, for example, clustering as proposed in [22.73], leads to fast re-evaluation of systems, too. The identification of critical (important) and independent (invariant to changes elsewhere) agents or functionalities is prepared in the *understanding phase*.

22.6 Conclusions and Recommendations

Many practical applications in the maritime domain deal with stealthy targets and large scale, three-dimensional transient environments. Cooperative vehicle target tracking offers the best solution to many surveillance and search/pursuit tasks in this domain. Successful implementations of cooperative vehicle target tracking can be found in multistatic sonar operation and in maritime surveillance systems. The cooperation is important because single sensor, single-platform systems do not provide the necessary persistence, coverage and environmental adaptability. In maritime surveillance, tracking is not plotting positions on a geographic grid, but following a trajectory in a space of behavioral parameters particular to the target.

A distributed coordination architecture is desirable in the implementation of cooperative target tracking with autonomous ocean vehicles. Coordination is advantageous in both the search and tracking phases.

Synchronization of the team in a centralized manner is necessary in large areas with sparse communication,

leading to a hybrid coordination scheme called synchronized covertness.

A methodology to generate coordination schemes has been developed, stemming from a heuristic bias toward independence. Coordination schemes generated with this methodology allow evaluation of the effectiveness of coordination.

Benchmark problems provide a useful way to quantify and compare the effectiveness of coordination schemes on a common basis (Fig. 22.11).

A recommended way forward would be the creation of a library of benchmark problems in realistic maritime environments together with the results of the *trading methodology*, that is, the produced coordinated behavior. Such a library could serve to gauge advances in coordination approaches on a common and understood basis. The scenarios could also serve as training tools for interactions between humans and multirobot systems.

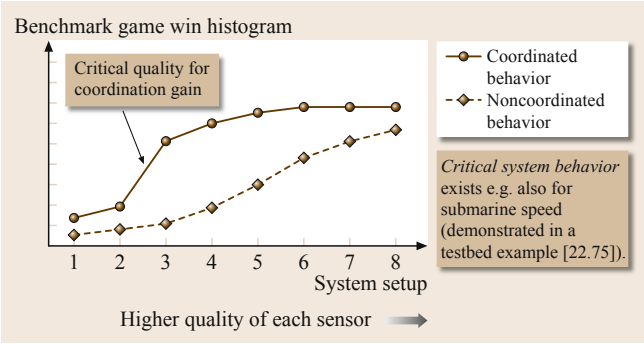


Fig. 22.11 The advantage of coordinated behavior (after [22.74, 75])

References

22.1 T.B. Curtin, D.M. Crimmins, J. Curcio, M. Benjamin, C. Roper: Autonomous underwater vehicles: Trends and transformations, Mar. Technol. Soc. J. **39**(3), 65–75 (2005)

22.2 Department of Defense: Unmanned systems integrated roadmap FY2013–2038 (2013) http://www.defense.gov/pubs/DOD-USRM-2013.pdf

22.3 Y.U. Cao, A.S. Fukunaga, A.B. Kahng: Cooperative mobile robotics: Antecedents and directions, Auton. Robots **4**, 1–23 (1997)

22.4 M.A. Hsieh, A. Cowley, J.F. Keller, L. Chaimowicz, B. Grocholsky, V. Kumar, C.J. Taylor, Y. Endo, R.C. Arkin, B. Jung, D.F. Wolf, G.S. Sukhatme, D.C. MacKenzie: Adaptive teams of autonomous aerial and ground robots for situational awareness, J. Field Robotics **24**(11/12), 991–1014 (2007)

22.5 J.K. Hedrick, B. Basso, J. Love, B.M. Lavis: Tools and techniques for mobile sensor network control, J. Dyn. Syst. Meas. Control **133**, 24001–24005 (2011)

22.6 N. Carlesi, F. Michel, B. Jouvencel, J. Ferber: Generic architecture for multi-AUV cooperation based on a multi-agent reactive organizational approach, Proc. IROS'11 (2011) pp. 5041–5047

22.7 J. Liu, M. Chu, J.E. Reich: Multitarget tracking in distributed sensor network, IEEE Signal Process. Mag. **37**, 36–46 (2007)

22.8 X. Zhao, S.Y. Tu, A.H. Sayed: Diffusion adaptation over networks under imperfect information exchange and non-stationary data, IEEE Trans. Signal Process. **60**(7), 3460–3475 (2012)

22.9 L.E. Parker: Distributed intelligence: Overview of the field and its application in multi-robot systems, J. Phys. Agents **2**(1), 5–14 (2008)

22.10 D.P. Eickstedt, M.R. Benjamin, H. Schmidt, J.J. Leonard: Adaptive tracking of underwater targets with autonomous sensor networks, J. Underw. Acoust. **56**, 465–495 (2006)

22.11 P. Ray, P.K. Varshney: Distributed detection in wireless sensor networks using dynamic sensor thresholds, Int. J. Distrib. Sens. Netw. **4**, 4–11 (2008)

22.12 C.D. Charalambous: Stochastic nonlinear minimax dynamic games with noisy measurements, IEEE Trans. Autom. Control **48**(2), 261–266 (2003)

22.13 H.S. Witsenhausen: Separation of estimation and control for discrete time systems, Proceedings IEEE **59**(11), 1557–1566 (1971)

22.14 F. Morbidi, G.L. Mariottini: Active target tracking and cooperative localization for teams of aerial vehicles, IEEE Trans. Control Syst. Technol. **21**(5), 1694–1707 (2013)

22.15 L.C.A. Pimenta, V. Kumar, R.C. Mesquita, G.A.S. Pereira: Sensing and coverage for a network of heterogeneous robots, Proc. 47th CDC (2008) pp. 3947–3952

22.16 Z. Xu, R. Fitch, S. Sukkarieh: Learning utility models for decentralised coordinated target tracking, Proc. IEEE ICRA'12 (2012) pp. 1753–1759

22.17 L.E. Parker, C. Touzet, F. Fernandez: Techniques for learning in multi-robot teams. In: *Robot Teams: From Diversity to Polymorphism*, ed. by T. Balch, L.E. Parker (A K Peters, Natick 2002)

22.18 T. Kaiser, A. Bourdoux, H. Boche, J.R. Fonollosa, J.B. Andersen, W. Utschick (Eds.): *Smart Antennas – State of the Art*, EURASIP Book Series on Signal Processing and Communications (Hindawi, New York 2005)

22.19 P. Aarabi: The fusion of distributed microphone arrays for sound localization, EURASIP J. Appl. Signal Process. **4**, 338–347 (2003)

22.20 V.A. Ziparo, L. Iocchi, P.U. Lima, D. Nardi, P.F. Palamara: Petri net plans – A framework for collaboration and coordination in multi-robot systems, Auton. Agent Multi-Agent Syst. **23**, 344–383 (2011)

22.21 R. He, A. Bachrach, N. Roy: Efficient planning under uncertainty for a target-tracking micro-aerial vehicle, Proc. IEEE ICRA'10 (2010) pp. 1–8

22.22 A. Ghafforkha, Y. Mostifi: Path planning for networked robotic surveillance, IEEE Trans. Signal Process. **60**(7), 3560–3575 (2012)

22.23 D. Koller, B. Milch: Multi-agent influence diagrams for representing and solving games, Proc. 17th Int. Jt. Conf. Artif. Intell. (IJCAI) (2001) pp. 1027–1034

22.24 W. Yeoh, A. Kumar, S. Zilberstein: Automated generation of interaction graphs for value-factored decentralized POMDPs, Proc. Int. Jt. Conf. Artif. Intell. (IJCAI) (2013) pp. 411–417

22.25 S.S. Nestinger, M.A. Demetriou: Adaptive collaborative estimation of multi-agent mobile robotic systems, Proc. IEEE ICRA'12 (2012) pp. 1856–1861

22.26 S.A. Miller, Z.A. Harris, E.K.P. Chong: A POMDP framework for coordinated guidance of autonomous UAVs for multitarget tracking, EURASIP J. Adv. Signal Process. **2009**, 724597 (2009)

22.27 S. Seuken, S. Zilberstein: Formal models and algorithms for decentralized decision making under uncertaintly, Auton. Agent Multi-Agent Syst. **17**(2), 190–250 (2008)

22.28 S.J. Witwicki, E.H. Durfee: Flexible approximation of structured interactions in decentralized markov decision processes, Proc. AAMAS **2**, 1251–1252 (2009)

22.29 F. Oliehoek: The multi agent decision process toolbox (2015) http://www.fransoliehoek.net/index.php?fuseaction=software.madp

22.30 G. Gottlob, G. Greco, T. Mancini: Complexity of pure equilibria in Bayesian games, Proc. Int. Jt. Conf. Artif. Intell. (IJCAI) (2007) pp. 1294–1299

22.31 R.R. Brooks: Sensor network self-organization using random graphs, Int. J. Distrib. Sens. Netw. **5**, 201–208 (2009)

22.32 M. Zhu, R.R. Brooks: Comparison of Petri net and finite state machine discrete event control of distributed surveillance networks, Int. J. Distrib. Sens. Netw. **5**, 480–501 (2009)

22.33 B. Lindley, L. Mier-y-Teran-Romero, I.B. Schwartz: Randomly distributed delayed communication and coherent swarm patterns, Proc. IEEE ICRA'12 (2012) pp. 4260–4265

22.34 D. Koller, N. Friedman: *Probabilistic Graphical Models: Principles and Techniques* (MIT, Cambridge 2009)

22.35 Y. Gal, A. Pfeffer: Networks of influence diagrams: A formalism for representing agents beliefs and decision-making processes, J. Artif. Intell. Res. **33**, 109–147 (2008)

22.36 A. Pfeffer, S. Das, D. Lawless, B. Ng: Factored reasoning for monitoring team and goal formation, Inform. Fus. **10**, 99–106 (2009)

22.37 Y. Gal, B. Grosz, S. Kraus, A. Pfeffer, S. Shieber: Agent decision-making in open mixed networks, Artif. Intell. **174**(18), 1460–1480 (2010)

22.38 V. Indelman, P. Gurfil, E. Rivlin, H. Rotstein: Graph-based distributed cooperative navigation for a general multi-robot measurement model, Int. J. Robotics Res. **31**(9), 1057–1080 (2012)

22.39 F. Ehlers: Particle filter for information management in an autonomously operating activated distributed sensor network, OCEANS 2007-Europe (2007) pp. 1–6

22.40 P. Dames, V. Kumar: Technical report: Cooperative multi-target localization with noisy sensors, Proc. IEEE ICRA'13 (2013) pp. 1877–1883

22.41 N. Kantas, S.S. Singh, A. Doucet: Distributed maximum likelihood for simultaneous self-localization and tracking in sensor networks, IEEE Trans. Signal Process. **60**, 5038–5047 (2012)

22.42 C. Reinl: Trajektorien- und Aufgabenplanung kooperierender Fahrzeuge: Diskret-kontinuierliche Modellierung und Optimierung, Ph.D. Thesis (TU Darmstadt, Darmstadt 2010)

22.43 D.J. Molinares: A Data Link Layer in Support of Swarming of Autonomous Underwater Vehicles, Ph.D. Thesis (University of South Florida, Tampa 2009)

22.44 P. Dames, M. Schwager, V. Kumar, D. Rus: A decentralized control policy for adaptive information gathering in hazardous environments, Proc. 51st CDC (2012) pp. 2807–2813

22.45 B.J. Julian, M. Angermann, M. Schwager, D. Rus: A scalable information theoretic approach to distributed robot coordination, Proc. IROS'11 (2011) pp. 5187–5194

22.46 W. Ren, R.W. Beard: *Distributed Consensus in Multi-vehicle Cooperative Control*, Communications and Control Engineering Series (Springer, London 2008)

22.47 L. Kester: Interactive collaborative information systems, Method for designing networking adaptive interactive hybrid systems, Stud. Comput. Intell. **281**, 401–421 (2010)

22.48 T.H. Chung, G.A. Hollinger, V. Isler: Search and pursuit-evasion in mobile robotics. A survey, Auton. Robots **31**(4), 299–316 (2011)

22.49 B. Schmidt, L. Newman: MOOS nested autonomy for unmanned marine vehicles with MOOS-IvP, J. Field Robotics, J. Field Robotics **27**(6), 834–875 (2010)

22.50 A. Kimmel, A. Dobson, Z. Littlefield, A. Krontiris, J.D. Marble, K.E. Bekris: Pracsys: An extensible architecture for composing motion controllers and planners, Lect. Notes. Comput. Sci. **7628**, 137–148 (2012)

22.51 D.T. Cole: A Cooperative UAS Architecture for Information-Theoretic Search and Track, Ph.D. Thesis (University of Sydney, Sydney 2009)

22.52 B. Jung: Cooperative Target Tracking Using Mobile Robots, Ph.D. Thesis (University of Southern California, Los Angeles 2005)

22.53 J. Fuemmeler, V.V. Veeravalli: Smart sleeping policies for energy-efficient tracking in sensor networks. In: *Networked Sensing Information and Control*, ed. by V. Saligrama (Springer, New York 2008) pp. 267–287

22.54 S. Coraluppi: Multistatic Sonar Localization, IEEE JOE **31**(4), 964–974 (2006)

22.55 S. Coraluppi, C. Carthel: Distributed tracking in multistatic sonar, IEEE Trans. Aerosp. Electron. Syst. **41**(3), 1138–1147 (2005)

22.56 M. Daun, F. Ehlers: Tracking algorithms for multistatic sonar systems, EURASIP J. Adv. Signal Process. **2010**, 46153 (2010)

22.57 F. Ehlers, D. Orlando, G. Ricci: Batch tracking algorithm for multistatic sonars, IET Radar Sonar Navig. **6**(8), 746–752 (2012)

22.58 R.L. Streit: Multistatic target and sensor field tracking, Proc. IEEE Aerosp. Conf. (2006) pp. 1–12

22.59 W3C Semantic Sensor Network Incubator Group: Semantic Sensor Network XG Final Report (2011) http://www.w3.org/2005/Incubator/ssn/XGR-ssn-20110628/

22.60 M. Daun, L. Broetje, F. Ehlers: Simultaneous localisation and tracking, Proc. UAM (2011)

22.61 R.P. Hodges: *Underwater Acoustics Analysis, Design And Performance Of Sonar* (Wiley, Hoboken 2010)

22.62 D.P. Eickstedt, H. Schmidt: A low-frequency sonar for sensor-adaptive, multi-static, detection and classification of underwater targets with AUVs, Proc. IEEE OCEANS'03 (2003) pp. 1440–1447

22.63 S. Kemna, M. Hamilton, K. Lepage, D.T. Hughes: Adaptive behaviours for autonomous underwater vehicles in anti-submarine warfare, Intell. Serv. Robotics **4**(4), 245–258 (2011)

22.64 S. Kemna, M.J. Hamilton, D.T. Hughes, R. Been: *Behavior Development for Anti-Submarine Warfare: The GLINT09 and GLINT10 Field Trials* (MOOS-DAWG, MIT, Cambridge 2010)

22.65 M. Hamilton, S. Kemna, D. Hughes: *Information theory based multi-vehicle collaboration for multi-static sonar using MOOS-IvP* (MOOS-DAWG, MIT, Cambridge 2010)

22.66 P. Zhan, D.W. Casbeer, A.L. Swindlehurst: Adaptive mobile sensor positioning for multi-static target tracking, IEEE Trans. Aerosp. Electron. Syst. **46**(1), 120–132 (2010)

22.67 P. Zhan, D.W. Casbeer, A.L. Swindlehurst: A centralized control algorithm for target tracking with UAVs, Proc. 39th Asilomar Conf. Signals Syst. Comput. (2005) pp. 1148–1152

22.68 A. Marino, G. Antonelli, A.P. Aguiar, A. Pascoal: A new approach to multi-robot harbour patrolling: Theory and experiments, Proc. IROS'12 (2012) pp. 1760–1765

22.69 Y. Hu, B. Ksherim, R. Cohen, S. Havlin: Percolation in interdependent and interconnected networks: Abrupt change from second- to first-order transitions, Phys. Rev. E **84**, 066116 (2011)

22.70 F.M. Mirzaei, A.I. Mourikis, S.I. Roumeliotis: On the performance of multi-robot target tracking, Proc. IEEE ICRA'07 (2007) pp. 3482–3489

22.71 W.D. McComb: *Renormalization Methods* (Oxford Univ. Press, Oxford 2004)

22.72 D. Sangiorgi, J. Rutten (Eds.): *Advanced Topics in Bisimulation and Coinduction*, Cambridge Tracts in Theoretical Computer Science, Vol. 52 (Cambridge Univ. Press, Cambridge 2011)

22.73 E. Bovio, H. Schmidt: The GOATS joint research project: Underwater vehicle networks for acoustic and oceanographic measurements in the littoral ocean, Goats 2000, SACLANTCEN CP-46, ed. by E. Bovio, R. Tyce, H. Schmidt (2001) pp. 3–16

22.74 F. Ehlers: Coordinated distributed mobile sensors and how to measure their performance, Proc. Sens. Data Fusion (2011)

22.75 F. Ehlers: Multi-agend system for multistatic sonar, Workshop Mach. Intell. (2009)

23. Rules of the Road for Unmanned Marine Vehicles

Joseph A. Curcio

The growing use of unmanned marine vehicles (UMVs) demands safe and reliable operation in order to assure acceptance and integration into public water space. By looking at the evolution of unmanned aerial vehicles (UAVs) and unmanned ground vehicles (UGVs) and examining the existing legal structure related to safe maritime operations, we explore the path ahead for UMVs. The existing COLREGS (rules of the road) for marine vessel operation may be undergoing modifications in order to accommodate UMVs, while simultaneously, designers of UMVs are working on developing advanced autonomy behaviors that exhibit *human-like* on-water performance by UMVs. This paper presents a snapshot of the rapidly changing field of UMV operation, with particular focus on software technology shaping higher level autonomy behavior and on the legal landscape providing the framework for acceptable operation in public water space.

23.1 COLREGS 517
23.2 Sensing the World 518
23.3 Proper Behaviors Yield Compliance 519
23.4 Integrating UMVs into Public Water Space 523
23.5 Developing Standards 524
23.6 The Road Ahead 525
23.7 Conclusion 525
References 526

Unmanned marine vehicles (UMVs) operating in populated waterways may require compliance with existing rules such as the collision regulations defined by the convention on the international regulations for preventing collisions at sea (COLREGS) [23.1] as published by the International Maritime Organization (IMO) and widely accepted by individual nations through international treaty agreements. Every mobile marine platform carries the risk of becoming involved in a collision with both stationary and mobile objects and vehicles. This risk exists regardless of whether the platforms are manned or unmanned. Owners and operators of mobile robotic marine platforms must take certain responsibility and action by designs intended to mitigate risks associated with the potential for collision and errant conduct by these platforms. This notion suggests that UMVs may need to operate in a very *human-like* manner so that their maneuvering will be recognizably executed and accurately interpreted by other vessels operating in their vicinity.

Rules and regulations pertaining to the operation of UMVs in public water space are evolving topics. The existing marine COLREGS *rules of the road* may themselves need to change in order to accommodate UMV interoperability with other manned and unmanned vessels. We will explore here the current state of technology associated with keeping UMV operations compliant with the existing COLREGS requirements and discuss the natural evolution of both the technology and the legal framework that will be necessary in order to keep pace with this technological development.

23.1 COLREGS

COLREGS establish a set of standard procedures that mariners are expected to follow. Comprised of approximately four dozen rules, they cover fundamental vessel maneuvering, vessel markings, situation awareness requirements, and signal specifications. In nautical terms these rules prescribe proper procedures for maneuvering a vessel and specify additional requirements for light and sound signals that alert other mariners to

a vessel's location, heading, intentions, and maneuvering constraints. Safe operation in crowded public water space is achieved through compliance to this widely accepted standard set of rules by all vessel operators.

At the core of the numerous COLREGS regulations are three fundamental driving rules. These three rules form the basis for most of the behaviors necessary for safe operation in the presence of other vessels. In no particular order, these three rules consist of the *head-on*, *overtaking*, and *crossing* situations. In general, the rules provide for specific actions suitable for avoiding collision when two or more vessels are operating in proximity of each other. Typically, in most crossing, overtaking, and approaching situations, one vessel will be the *stand-on* (privileged) vessel and the other the *give-way* (burdened) vessel, but as the rules clearly state, this is not always the case. In addition, there exists a *catch all* rule (rule 2 – responsibility) that requires a last chance to avoid collision by any and all vessel operators, regardless of rights.

As these rules were originally written to be interpreted and acted upon by humans, there exists room for interpretation and a notion of good helming and common sense is required in operating under these rules. In addition, COLREGS do not afford operators a clear *right of way* over other vessels, but rather dictate the accepted standard for safe maneuvering in order to avoid collision. Rule 2 along with rule 7 (risk of collision) are in place to clearly indicate that safe helming is mandatory, and regardless of perceived *rights of way* as defined by the rules, operators must take every precaution to avoid collision.

Fundamental to the rules is that they were originally written for humans to interpret and act upon, as evidenced in the use of terms such as *proper look-out* and the use of sight and sound signals to alert other mariners. Even today with the availability of AIS (automatic identification system), radar, and other navigation aiding devices, the role of the human lookout remains embedded in the rules. It is unlikely that completely automated navigation and vessel control will be accepted and implemented within the existing set of COLREGS rules any time in the near future. Meanwhile, how these rules apply to autonomous (and semi-autonomous) unmanned marine vehicles operating on and below the surface must be addressed, given current deployment trends. UMV platforms might ultimately be required to operate with certain specific behaviors in order to comply with COLREGS and with basic navigational constraints.

23.2 Sensing the World

In order to properly implement the COLREGS rules, UMV programmers and operators will first need to provide unmanned platforms with the capability to detect other vessels through adequate sensor systems, and react appropriately using robust behaviors.

The introduction of all mobile robots (on land, sea, and air) has required significant development in sensory systems capable of providing a *world view* to the autonomy system, producing the equivalent of a map of the environment surrounding the robot. This perceived knowledge of the world around the robot consists of data and meta-data collected by a suite of sensors integrated into a data fusion system, internally representing to the machine an awareness of the physical world in proximity to its operation. This perception is the basis of all that the machine *knows*. It is vital to understand that in order for a mobile robot to ultimately behave in

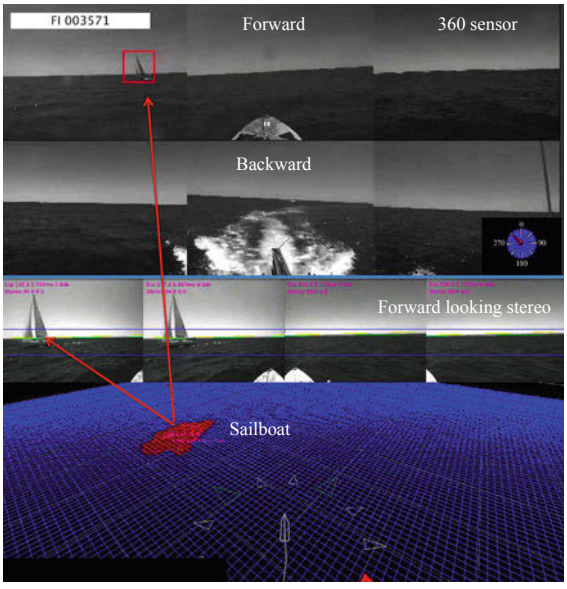


Fig. 23.1 Example of fusion of two sensor modalities (360° sensor, forward looking stereo) for a *proper look-out* of the environment surrounding the UMV. *Top*: the 360° sensor (in this case six cameras). *Bottom*: the forward looking stereo (in this case four cameras). The sailboat is detected in both sensor systems with the explicit range provided from the stereo system and relative bearing fused from both systems (courtesy of Jet Propulsion Laboratory, California Institute of Technology) ►

a *real-world* environment and to do so in a way that appears *human-like*, this perceived world must contain information that maps closely to that perceived by a human observer.

Sensors along with complex computer algorithms that fuse the raw sensor data are used to produce a representation of the surrounding environment much the same way the human eye and brain interact to create a picture of the surroundings. Relative to the evolution of the human eye, robot perception today is primitive. The perceived environment with all of its misperceptions and uncertainties is integrated into the robot's autonomy engine in order to facilitate safe mobilization. *Huntsberger* [23.2, 3] provides a good example of this process, as shown in Fig. 23.1, where a forward looking stereo system and a 360° sensor head are used to give full situational awareness around the UMV [23.2, 4].

Significant progress has been made within the past decade in the area of robot exteroception (perception of the world around the robot), including advances in the technology associated with improved sensors and evolution in the software associated with processing the information obtained by these sensors. Through an array of sensors, each sensitive to a limited spectral window (radar, laser, infrared, etc.), fused into a single *contact report*, the UMV assimilates information about its immediate environment. In general, the higher the fidelity of the fused data, the better the probability that the behavior algorithm will operate the robot with clearly defined (deliberate) movements. This will ultimately instill a high degree of trust of UMVs operating in crowded marine environments. It is not the focus of this paper to discuss the details of sensing the world around the UMV, but rather to discuss the impact that this capability has on advancing the functionality of the UMV and the potential for integration into the human/robot world in which these machines will operate.

As sensory capability and autonomy behaviors continue to evolve, marine robots will continue to become more and more capable of effectively discriminating between stationary and moving objects, sailing and motor driven vessels, channel markers, bridge pilings, and so on. Clutter rejection, false positives, lights on shorelines and other unwanted signal *noise* will be properly rejected, and complex artificial vision systems will produce reliable *contact reports* with a high degree of accuracy and confidence to the UMV *autonomy engine*. Eventually, humans operating in near proximity to these UMVs will have a difficult time distinguishing between human and machine operators.

23.3 Proper Behaviors Yield Compliance

Competent mobile marine vehicles generally operate with deliberate intentions (following a prescribed waypoint path) until a reactive behavior is necessary to avoid collision with a newly detected obstacle or vessel while underway. Various techniques for blending reactive and deliberate actions are practiced by many different groups engaged in programming mobile robots. For example, *Kuwata* et al. [23.5, 6] demonstrate *velocity obstacles* (VO) in conjunction with a *subsumption* approach [23.3] to meet COLREGS requirements with UMVs. *Benjamin* et al. [23.7] employ a mathematical methodology, termed *interval programming* (IvP) coupled with an open source *autonomy middleware* called MOOS [23.8] (mission oriented operating suite – developed at MIT and Oxford University), collectively termed *MOOS-IvP* [23.9]. Using MOOS-IvP, *Benjamin* et al. [23.10] use multi-objective optimization to generate an objective function weighing multiple objectives simultaneously and producing a globally optimized behavior solution. Conflicting objectives are weighed internally using interval programming (IvP) methodology, and a set of optimized vehicle heading and speed commands are issued to the UMV controller. By including COLREGS as one of the objective function sets, a balance is struck between near-term objectives (such as the desired goal to reach a waypoint) and safe operation (avoiding collision while satisfying COLREGS requirements), as shown in Fig. 23.2.

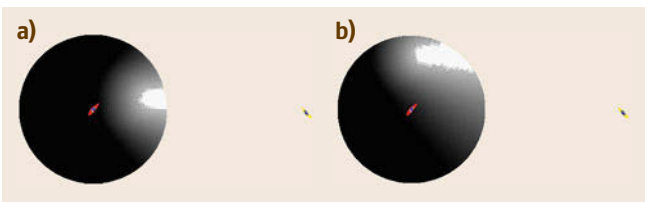


Fig. 23.2a,b Graphical depiction of an objective function as produced in this case by a collision avoidance behavior. The moving contact is shown with a motion vector indicating the contact's trajectory. Ownship has a fixed speed with the contact vehicle in **(a)** moving slightly below ownship speed and in **(b)** the contact is moving slightly faster than ownship. In both cases, *darker colors* represent more favorable trajectories and *white* indicates the closest point of approach between vehicles (courtesy of Massachusetts Institute of Technology, Laboratory for Autonomous Marine Sensing Systems) ►

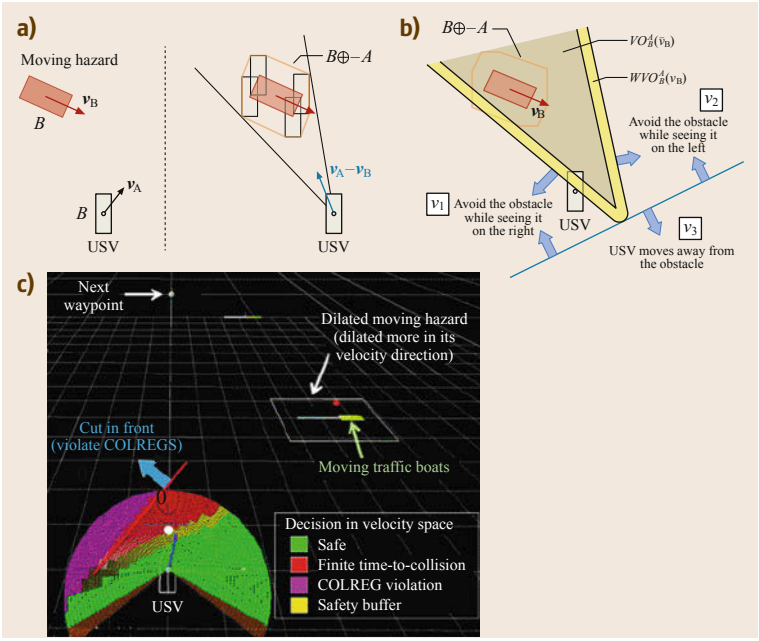


Fig. 23.3a–c A graphical representation of velocity obstacles **(a)** showing how the relative velocities of the unmanned surface vehicle (USV) and a moving hazard predict collision and define heading and speed combinations that allow safe passage. Combining the VO approach with COLREGS further constrains the problem, generating safe maneuvering projections that satisfy the COLREGS requirements **(b)**. Image **(c)** depicts a simulation run where the USV is encountering a potential collision and optimal COLREGS compliant trajectory is determined based on the VO and COLREGS restrictions (courtesy of Jet Propulsion Laboratory, California Institute of Technology)

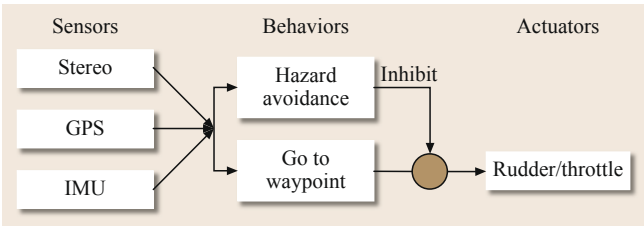


Fig. 23.4 An example of the *subsumption* approach as used in autonomous decision making, where certain behaviors are given higher priority

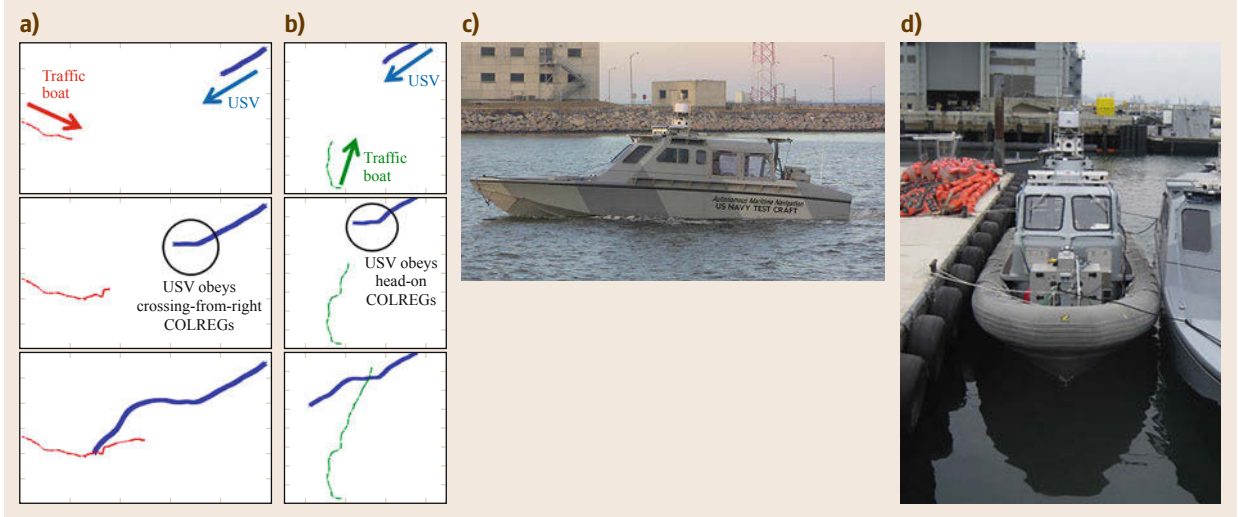


Fig. 23.5a–d Data showing a crossing from a right condition **(a)**, and a head-on situation **(b)** plotted from at-sea experiments using the two boats shown on the *right*. **(c)** The USV – power vent and **(d)** the 11 m RHIB (rigid hull inflatable boat) (courtesy of the Jet Propulsion Laboratory, California Institute of Technology)

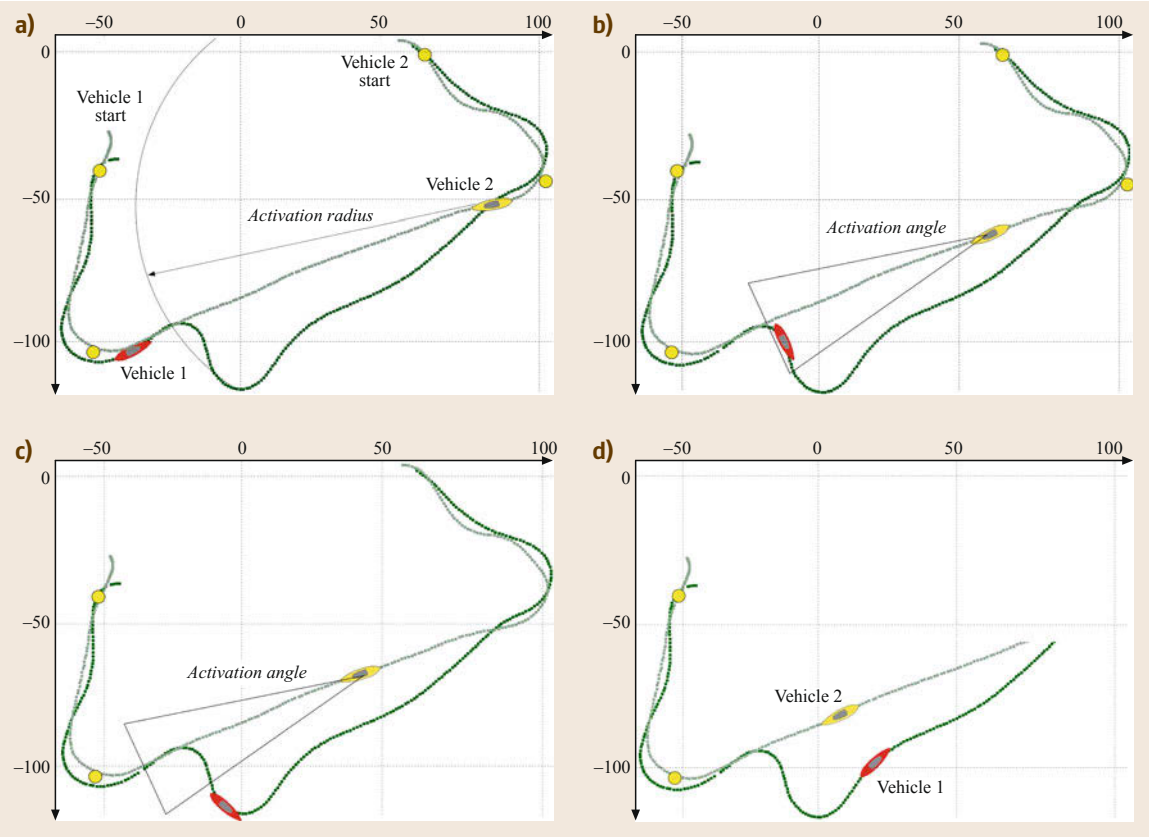


Fig. 23.6a–d In-water demonstration using two SCOUT vehicles in a *head-on* situation. Vehicle 1 is programmed to be COLREGS compliant with a waypoint objective active and vehicle 2 is simply transiting to the next waypoint with no regards to collision avoidance. When Vehicle 2 enters the 'activation radius' that necessitates collision avoidance behavior (from **(a)** to **(b)**), vehicle 1 makes a proper turn to starboard **(c)** to avoid collision and both vehicles continue towards next waypoint **(d)**. (courtesy of Massachusetts Institute of Technology, Laboratory for Autonomous Marine Sensing Systems)

Multi-objective optimization using IvP functions expressing utilities for all maneuvers (not just optimal maneuvers) is fundamental in allowing a common-sense balance between collision avoidance and mission objectives. This technique yields real-world behaviors that come strikingly close to being *human-like* in their implementation.

By comparison to the IvP approach, the velocity obstacle approach generates a cone-shaped area ahead of the UMV, mapped in a *velocity space* (defined as the area that the UMV would be capable of covering in some pre-determined time step, when considering the speed and maneuvering ability of the craft). Collision with other vehicles is prevented by assuring that the UMV's own velocity vector remains outside the region of the VO during transit. Fig. 23.3 graphically depicts the VO concept in conjunction with COLREGS compliant maneuvering restrictions. The subsumption approach to balancing conflicting behaviors uses a pre-established hierarchy in order to rank behavior authority, allowing certain behaviors to take precedence over other conflicting behaviors, as depicted in Fig. 23.4.

Some examples of in-water demonstrations of COLREGS compliant behaviors serve to illustrate the state-of-the-art. For example, as illustrated in Fig. 23.5, data taken from at-sea experiments demonstrate the robust nature of the NASA/JPL (National Aeronautics and Space Administration/Jet Propulsion Laboratory) system using COLREGS algorithms integrated into the JPL autonomy suite, CARACaS (control architecture for robotic agent command and sensing) [23.5]. In this example, it is evident that the USV altered course in proper compliance with COLREGS so as to avoid collision while continuing to pursue a waypoint objective. The NASA/JPL approach has been used successfully

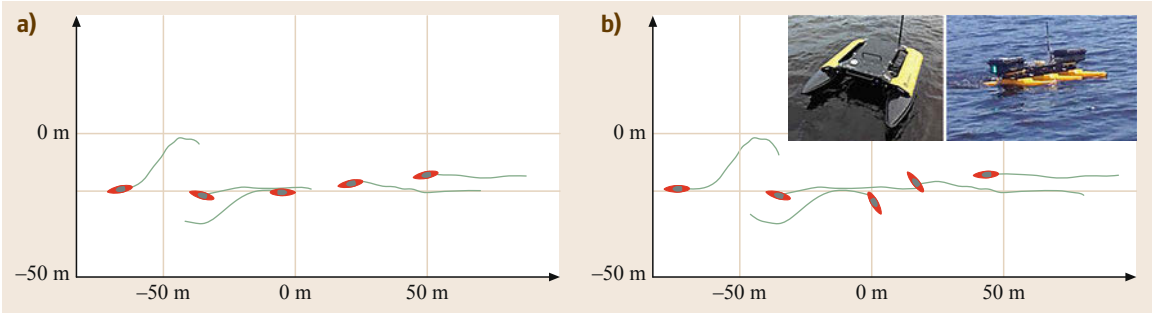


Fig. 23.7a,b Five UMVs demonstrating multiple collision conditions while executing COLREGS compliant MOOS-IvP behaviors (courtesy of Massachusetts Institute of Technology, Laboratory for Autonomous Marine Sensing Systems). Intentional head-on crossing conditions are set up by tasking four vehicles to transit along one trackline and one vehicle to transit in the exact opposite direction. A time series depiction of vessel motion **(a)** to **(b)** shows clearly executed avoidance maneuvers by two vehicles nearest the center of the image, both turning to starboard in accordance with COLREGS requirements

on several 11 meter USVs with on-board sensory capability, with operationally relevant speeds and conditions [23.11].

Benjamin et al. [23.12] have demonstrated the capability of the MOOS-IvP approach to satisfactorily achieve COLREGS compliant vehicle behavior while simultaneously reaching deliberate mission objectives. Using a small fleet of low cost SCOUT (surface craft for oceanographic and unmanned testing) [23.13] autonomous surface craft, *Benjamin* et al. [23.12] at MIT successfully demonstrated safe COLREGS compliant crossing of multiple UMVs in a controlled environment using only on-board sensors or knowledge of other vessels exclusively. A typical in-water demonstration (Fig. 23.6) depicts COLREGS compliant waypoint objective based activity between two vehicles demonstrating safe transit with collision avoidance. The focus of this research was directed at developing a suite of robust and efficient behaviors and as such, the physical hardware for detection of other vessels was substituted with active data sharing among vehicles. All of the SCOUT craft regularly broadcasted their GPS-derived trajectory information on a shared network. Eventually, detected vehicle contact reports would be generated by a fully integrated sensory suite using data fusion engines as developed and demonstrated by *Huntsberger* et al. [23.2–6, 6] and others. Multiple collision situations were generated, providing ample opportunity to focus on improving the robustness of the collision avoidance software.

COLREGS research at MIT has resumed in the last year to include field testing with up to five surface craft ([23.14], Fig. 23.7) with multiple simultaneous inter-vehicle interactions and COLREGS rule instantiations. Ongoing research also includes studying the effects of augmented AIS message structures conveying ownship state and intention, and an on-board acoustic sensing system for localizing other ships using a horn in low-visibility, fog situations. The effect of inter-vehicle voice communication is also being investigated including the automatic recognition of human intentions from manned vessels and the conveyance of intentions of UMVs using computer generated voice over standard radio channels.

As work continues to progress on this front, autonomy will move humans further out of the control loop, thereby requiring a higher degree of reliability in the autonomy system. Each vessel will need to detect all obstacles (moving and stationary) in its surroundings, be acutely aware of any restrictions to maneuvering imposed by physical constraints such as water depth restrictions, channels, obstructions, and so on, and behave in compliance with an established set of governing regulations.

For a glimpse into the likely future of UMV autonomy with superior sensory capability, one need only look at recent demonstrations of unmanned systems operating in real-world air (UAVs such as the Global Hawk and others operating in military theaters) and land (DARPA Grand Challenge and DARPA Urban Challenge) environments.

23.4 Integrating UMVs into Public Water Space

As we follow the evolution of UMV autonomy and the legal landscape associated with acceptance in public water space, observation of the process of the integration of unmanned aerial vehicles (UAV) into public airspace is instructive. The *see and avoid* paradigm in federal aviation regulations is analogous to underlying assumptions and procedures in marine COLREGS. Both are based on perception of the surrounding environment at a human competence level. Establishing autonomous control at acceptable competence levels in airspace will set precedents applicable to operations in waterways. UAV technology is currently more advanced than UMV technology because of the popularity of the R/C industry and because of the early adoption by military users. UAV platforms were able to prove their merit readily by providing operators with an *eye in the sky* and demonstrating the ability to deliver lethal payloads while protecting soldiers from harm. Due to the FAA Modernization and Reform Act [23.15], the FAA is currently working to integrate UAVs into the National Airspace System (NAS) and has developed a roadmap for achieving this objective [23.16]. In addition, the US Department of Defense (DoD) is preparing the Unmanned Systems Integrated Roadmap FY2011-2036 [23.17] so as to lay out a plan for integration of UAVs into the DoD Joint Force Structure and seamless integration of UAVs in the NAS along with other safety, interoperability, and reliability goals.

In the United States, public water space is governed at the local, state, and federal levels, depending on the specific water body and proximity to coastal boundaries. Both domestic and international maritime laws hold varying degrees of governing authority over vessels flagged to a specific country and to commerce carried out in international, territorial, and state waters. There is no single federal agency responsible for setting standards for accepting new types of *vessels* into public waters, as there would be in public air space through the FAA. The US Coast Guard is both an enforcing body and a regulatory agency within the US Department of Homeland Security. As such, certain legal responsibilities are likely to fall on the shoulders of the Coast Guard with regards to managing UMV activity in public waters. It is likely that UMV regulation will evolve in the United States through a comprehensive approach including self-regulation by owners, operators, and designers, along with constraints imposed by COLREGS regulations. Insurance costs and liability exposure will also likely be governing factors in defining legal standards and establishing best practices. No doubt a precedent-setting legal case will eventually be heard and settled, defining the legal liability associated with operating UMVs in public waterways, and the tort liability will most likely fall upon the shoulders of the owners, operators, and programmers of these vehicles.

Presently, there is an active dialog underway regarding the rules regarding operation of UMVs and, in particular, the possible evolution of COLREGS to accommodate unmanned platforms operating in public waters. This discussion involves the US Coast Guard, industry leaders, academia, the AUVSI (Association for Unmanned Vehicle Systems International), and US Navy representatives. All of the parties involved in these discussions are stakeholders to varying degrees, interested in reaching a consensus regarding the best approach to assuring safe and acceptable operation of UMVs in public waters. In particular, at the time of writing, ongoing discussions are being held in open forums with the goal of establishing a complete set of suggested changes to the rules governing UMVs operating in public water space. These changes are being reviewed, discussed, and compiled by the US Department of Homeland Security Navigation Safety Advisory Council (NAVSAC) and then passed on to the Coast Guard representative for final evaluation. From the Coast Guard, these proposed changes are passed on to a set of IMO (International Maritime Organization) member nations for translation, evaluation, and review before final acceptance and incorporation into COLREGS. This process is expected to take between 3 and 3.5 years to complete.

The process involved with revising COLREGS includes a fundamental discussion of the legal status of unmanned systems. In particular, the question of whether or not a UMV will be considered a *vessel* or not, as defined within COLREGS, may bear heavily on the resolution regarding their legal treatment. Most of the COLREGS regulations start out with a description of the term *vessel*, as this remains at the center of the discussion when ascertaining the need for compliance with the rules, as pointed out by *Bederman* [23.18]. A number of articles have been written on the subject of classifying an unmanned (and possibly autonomous) marine vehicle as a *vessel* under maritime law. The very definition of the term *vessel* varies as defined by International Maritime Law and COLREGS and goes undefined in UNCLOS (United Nations Convention on the Law of the Sea) [23.19]. The International Regulations for Avoiding Collisions at Sea (COLREGS) define a vessel as *every description of watercraft, including non-displacement craft and seaplanes, used or capable of being used as a means of transportation on water* [23.20]. Broadly speaking, these definitions do indeed encompass the modern autonomous marine ve-

hicle as a legal *vessel* subject to compliance with the rules of the road. This conclusion is important because it may be the determining factor in establishing admiralty [23.21] jurisdiction in settling any tort liability claims. In American jurisprudence literature [23.22], *vessel* interpretation includes: *For the purpose of determining admiralty jurisdiction, the term vessel has been interpreted liberally and broadly*. (It indicates) *any structure used or capable of being used for transportation on water* [23.23] *and predominantly characterized by movement, rather than fixity or permanence* [23.24]. These interpretations also portend that UMVs will be classified as vessels [23.25–27], demanding compliance to code and equal treatment by other vessels. The legal ramifications for the owners and operators associated with these UMVs will include full liability for damages to other persons and property, as though those same individuals were directly involved in the on-board operation of the UMV. However, allowance is made for vessels that are not fully capable. COLREGS rule 3(g) provides for a vessel which is *restricted in her ability to maneuver* defined as *a vessel which from the nature of her work is restricted in her ability to maneuver as required by these Rules and is therefore unable to keep out of the way of another vessel* [23.28].

Although the initial NAVSAC review held in 2011 reportedly demonstrated some doubt regarding classifying UMVs as *vessels* as defined by COLREGS rule 3(a), they did propose that the US Coast Guard sponsor an amendment to the RAM (restricted in her ability to maneuver) in rule 3(g) to add *a self-propelled vessel while unmanned and operating autonomously*. This proposal suggests that within the NAVSAC review panel there is some level of acceptance that UMVs will be classified as *vessels* within the COLREGS. By logical extension, if accepted as RAM vessels, UMVs will be afforded a *privileged* right over most manned vessels, due to their inherent limitations in keeping out of another vessels' way (in conditions where visibility is not restricted).

From the summary record of the NAVSAC Working Group Report of the May 2011 meeting, the following resolutions provide some insight to the present status of this evolving topic. NAVSAC advised the US Coast Guard to amend COLREGS so as to require vessel automatic identification system (AIS) transponders on all unmanned underwater vehicles, if practicable, suggesting that the overall vessel length be used as the defining characteristic, with 8 m in overall length and 1 m in overall diameter used as the delineation for meeting the AIS and lighting requirements. They also recommended an amendment to rule 23 (both inland and COLREGS), which would require all inconspicuous, partly submerged vessels or objects while unmanned and operating autonomously to exhibit the lights specified in rule 24(g), again suggesting the 8 m length and 1 m diameter minima. For underwater vessels, they recommended amending rule 24(g), both inland and COLREGS, to require that these same vessels exhibit the specified lights at night and by day. Similar suggestions were proposed that affected unmanned surface vessels requiring AIS where practicable, as well as specific lights and day shapes, where practicable. In addition, NAVSAC recommended that rule 3(g), both inland and COLREGS, be modified to include language addressing *self-propelled vessels while unmanned and operating autonomously*. Further, NAVSAC recommended that the lookout requirements in rule 5 be amended to include the term *manned*, thereby relieving unmanned vessels from lookout responsibility. NAVSAC clearly suggested that the US Coast Guard promulgate an interpretive rule under 33 CFR parts 82 and 90 to provide that a vessel being operated remotely is considered to be manned and must comply with the applicable navigation rules and annexes. At present, the NAVSAC Work Group recommendations are being evaluated by industry experts, academia, US Coast Guard representatives, and legal scholars. Should the proposed approach be adopted, a proposal will be carried by the US Coast Guard to IMO for further review and eventual adoption and inclusion into COLREGS. It should be pointed out that while UMV may be relieved of the lookout requirement if the NAVSAC recommendations are accepted by the US Coast Guard and eventually incorporated into the COLREGS, this will not relieve UMV from obligations under COLREGS rule 7 (risk of collision) and rule 2 (responsibility).

23.5 Developing Standards

In parallel to the evolving COLREGS rules as proposed by NAVSAC, the American Society for Testing and Materials (ASTM) established a technical committee (F41) that is responsible for developing standards for UMVS (unmanned maritime vehicle!system) in 2005. This committee includes subcommittees addressing autonomy and control (F41.01), maritime regulations (F41.05), and unmanned surface vehicles (F41.90.01), among others. Presently, the core (F41.01) autonomy and control section is the furthest developed of this set of standards. Through industry, academics, and user group input, the ASTM committees continue to push

the evolution of these guidelines with the objective of establishing common operational and interface standards. These are currently all in a state of ongoing development, but it is likely that the community will eventually come to recognize these as the accepted standards with regards to safe and compliant (COLREGS) operation of UMVs in public water space. The specific subcommittee responsible for the Maritime Regulations (F41.05) originally set out to establish self-regulating COLREGS compliance through these standards. Of late, this subcommittee is reaching out to the NAVSAC Working Group through the AUVSI community and participates in the ongoing dialog with the US Coast Guard, as described previously.

23.6 The Road Ahead

The rules of the road might evolve so as to include a provision for autonomous or unmanned operation in much the same way that there exists a provision for fishing vessels, vessels restricted in their ability to maneuver, draft restrictions, and so on. Provided with a legal designation as such, autonomous marine vehicles could then be afforded a safe operations entitlement, lifting the burden of full COLREGS compliance and designated via a unique AIS code, day shape, etc. This might be provided in the form of a proximity buffer (consider the high-occupancy vehicle (HOV) lane on highways, for example) to be observed by other vessels, or simply as an authorization to relax the vigilant lookout requirement. This would be a step in the direction of integrating UMVs into the world of human-operated vessels in public waterways.

Alternately, UMV developers might produce a *backseat driver* as an interim solution. This embedded autonomy device would employ all of the sensor fusion techniques previously mentioned and deliver driving guidance to a human operator, who would retain manual control of the helm and take full legal responsibility for the safe operation of the vessel. This backseat driver would provide a service to skilled marine operators, acting as a lookout and navigator, sounding alarms when collisions appear imminent and advising maneuvering decisions based on optimal travel paths in compliance with COLREGS. This scenario describes a glorified autopilot capability with enhanced perception capabilities and advanced behavior capability tuned for balancing multiple objectives.

No doubt scientists, engineers, and policy makers will continue to upgrade and refine the capabilities and the requirements placed upon unmanned vehicles that operate in public water space. Using advanced sensors and autonomous control, maritime robots may someday evolve into machines with super-human-like behavior that perform better than humans in noisy stressful environments where flawed perception and emotional irrationality erode the performance of mortals. When that happens, the rules of the road may need to be modified so as to designate manned vessels as those restricted in their ability to maneuver, requiring them to maintain a buffer zone for safe operation in the proximity of the more predictable and reliable unmanned vessels. The road ahead may indeed see the tables turned.

23.7 Conclusion

Improved system performance suggests that UMVs are here to stay and will be providing a wide array of valuable services in the marine environment for years to come. Along with the benefits afforded by these platforms, we should expect increasing interaction between manned and unmanned vessels and a need for clearly defined laws, policies, and procedures for safe and reliable operation of UMVs in public waterways.

In a semi-hybrid fashion, robotic autopilot *assistance* is creeping into how we operate aircraft, automobiles, and boats today. In cars, we now have braking assistance that prevents collision and steering assistance that maintains the vehicle track on the roadway and aids in parking. In aircraft and boats, we have autopilots that can follow prescribed waypoints to safely transit a craft over great distances. Autopilots coupled with obstacle avoidance need only an embedded autonomy engine with COLREGS compliant behaviors and a human helmsman will become a passive observer, albeit the legally responsible observer, but an observer and passenger, nonetheless. The rapid evolution of the technology and the legal framework associated with acceptance of unmanned cars and unmanned aerial vehicles provide precedents for the likely trajectory for unmanned marine vehicles.

It is conceivable that at some point in the near future, two unmanned marine vehicles will encounter one another in a potential collision situation. Or perhaps,

one manned vessel, operating under partial autonomy (autopilot) will encounter an unmanned vessel, both similarly programmed to avoid collision. One can imagine several outcomes to this encounter, including those involving a collision and eventual legal engagement. As with most technological evolutions, the rapidly changing technical and legal landscape will dictate the final regulatory environment that all vessels will operate within. In this world of convergence between manned and unmanned platforms, it is entirely possible that future vessel operators will become passive observers as robotic helms carefully maneuver all manner of water craft in safe and predictable fashion while obeying all of the established and accepted *rules of the road*.

References

23.1 International Maritime Organization: *Convention on the International Regulations for Preventing Collisions at Sea (COLREGS)* (IMO, London 1972)

23.2 T. Huntsberger, H. Aghazarian, A. Howard, D. Trotz: Stereo vision based navigation for autonomous surface vessels, J. Field Robotics **28**(1), 3–18 (2011)

23.3 T. Huntsberger, G. Woodward: Intelligent autonomy for unmanned surface and underwater vehicles, Proc. OCEANS (2011) pp. 1–10

23.4 M.T. Wolf, C. Assad, Y. Kuwata, A. Howard, H. Aghazarian, D. Zhu, T. Lu, A. Trebi-Ollennu, T. Huntsberger: 360-degree visual detection and target tracking on an autonomous maritime patrol vessel, J. Field Robotics **27**(6), 818–838 (2010)

23.5 Y. Kuwata, M. Wolf, D. Zarzhitsky, T. Huntsberger: Safe maritime navigation with COLREGS, using velocity obstacles, IEEE J. Ocean Eng. **39**(1), 110–119 (2014)

23.6 Y. Kuwata, M. Wolf, D. Zarzhitsky, T. Huntsberger: Safe maritime navigation with COLREGS using velocity obstacles, Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS) (2011) pp. 3728–4734

23.7 M. Benjamin, J. Curcio, J. Leonard, P. Newman: A method for protocol-based COLREGS collision avoidance navigation between unmanned marine surface craft, J. Field Robotics **23**(5), 333–346 (2006)

23.8 P.M. Newman: *MOOS – A Mission Oriented Operating Suite, Tech. Rep. OE2003-07* (MIT Department of Ocean Engineering, Cambridge 2003)

23.9 M. Benjamin, H. Schmidt, P. Newman, J. Leonard: Nested autonomy for unmanned marine vehicles with MOOS-IvP, J. Field Robotics **27**(6), 834–875 (2010)

23.10 M. Benjamin: Multi-objective helming with interval programming on autonomous marine vehicles, Proc. IEEE/RJS IROS 2006 Workshop Multi-Objective Robotics (IROS-MOR 2006) (2006)

23.11 L. Elkins, D. Sellers, W.R. Monach: The autonomous maritime navigation (AMN) project: Field tests, autonomous and cooperative behaviors, data fusion, sensors, and vehicles, J. Field Robotics **27**(6), 790–818 (2010)

23.12 M. Benjamin, J. Curcio, J. Leonard, P. Newman: Protocol-based COLREGS collision avoidance navigation between unmanned marine surface craft, J. Field Robotics **23**(5), 333–346 (2006)

23.13 J. Curcio, J. Leonard, A. Patrikalakis: SCOUT – a low cost autonomous surface platform for research in cooperative autonomy, MTS/IEEE Proc. OCEANS (2005) pp. 725–729

23.14 M. Benjamin, H. Schmidt: Massachusetts Institute of Technology Graduate Level Course 2.S998, Marine Autonomy, Sensing and Communications, 2012.

23.15 PUBLIC LAW 112–95–FEB. 14, 2012, 112th Congress, 126 Stat 72, Title III, Unmanned Aircraft Systems, Section 332, Integration of civil unmanned aircraft systems into national airspace system.

23.16 Next Generation Air Transportation System, Joint Planning and Development Office: *NextGen UAS Research Development and Demonstration Roadmap*, Version 1.0 (JPDO, Washington 2012)

23.17 Department of Defense: *Unmanned Systems Integrated Roadmap FY2011-2036* (2011)

23.18 D.J. Bederman: The future of maritime law in the federal courts: A faculty colloquium, J. Mar. L. Commer. **31**, 189 (2000)

23.19 1 U.S.C.S. §3 (2006) (definition of "vessel")

23.20 28 U.S.T. 3459 (Oct. 20, 1972)

23.21 2 *American Jurisprudence* 2d §4, Admiralty p. 722

23.22 2 *American Jurisprudence*2d §33, Admiralty p. 740

23.23 70 *American Jurisprudence* 2d edition, shipping, p. 456

23.24 *American Jurisprudence*2d §33 p. 740

23.25 M. Benjamin, J. Curcio: COLREGS-Based Navigation of Autonomous Underwater Marine Vehicles, Proc. Inst. Electrical Electronics Eng. (IEEE) Conf. Autonomous Unmanned Vehicles (1994) pp. 32–39

23.26 C.H. Allen: The seabots are coming here: Should they be treated as vessels?, J. Navig. **65**, 749–752 (2012)

23.27 A.H. Henderson: Murky waters: The legal status of unmanned undersea vehicles, Naval Law Review **53**, 55–72 (2006)

23.28 Code of Federal Regulations, Title 33, Chapter 1, Sub-chapter E, Part 83, Subpart A, Section 83.03, General Definitions (Rule 3g).

24. Autonomy: Risk Assessment

Mario P. Brito, Gwyn Griffiths

Oceanography and ocean observation in general are ever trending toward both automated in situ observation and working in extreme environments. These goals can only be met by de-risking the technology and deployment practices to acceptable levels of risks. A number of industries have standardized risk management processes to support the design and development of their systems. The lack of formal risk assessment of autonomous ocean vehicles has hindered the potential for true autonomy, which is required for exploring unstructured and unexplored environments. When discussing risks, different stakeholders may have different consequences foremost in mind. For example, the vehicle owner may be interested in risk of loss, whereas the user is interested in risk of vehicle unavailability. Other risks, such as legal risks and risk of collision, affect all stakeholders. This chapter presents a risk management process using several methods tailored to autonomous ocean vehicles in which risk assessment is a key component.

24.1 **Risk Management Process for Autonomous Ocean Vehicles** 528
24.2 **Risk of Failure** 529
24.2.1 Reliability Estimation 529
24.2.2 Reliability Modelling 530
24.3 **Risk of Collision** 532
24.3.1 Risks of Collision on or Near the Surface 532
24.3.2 Risks of Collision Underwater 533
24.4 **Risk of Unavailability** 534
24.5 **Risk of Loss** 535
24.5.1 Expert Judgment Elicitation Processes 536
24.5.2 Survival Prediction 540
24.6 **Legal Risks** 541
References 542

Our human curiosity to understand unexplored and hostile environments has led us to develop state of the art automated technology capable of meeting our measurement needs [24.1–3]. Despite having identified the requirements for greater autonomous ocean vehicles (AOV) intelligence and autonomy, to date, physics prohibits the deployment of the truly intelligent autonomous underwater vehicle (AUV) [24.4]. Over the years, for practical implementations, the community has addressed risk by making the vehicles as simple as possible, the missions as modest as possible, and the level of supervision as high as possible [24.4–6].

Nevertheless, over the years there have been a number of vehicle losses. Among the most high-profile losses are the loss of Autosub2 under the Fimbulisen ice-shelf, Antarctica on 16 February 2005, during mission 383 [24.7] and more recently the loss of the Autonomous Benthic Explorer (ABE), during dive 222, off the coast of Chile, on 5 March 2010. Both losses are thought to have been caused by technical failure. In the case of Autosub2, a formal independent inquiry concluded that an *abort command* or a *loss of power* were equally likely to have caused the vehicle loss. There was no formal independent inquiry into the loss of ABE; following the accident, the design and operation team concluded that ABE suffered a catastrophic implosion of a glass sphere used for providing buoyancy causing instant destruction of on-board systems [24.8]. Anecdotal evidence exists of losses of smaller vehicles such as of Remus 100s and undersea gliders but these have never been formally reported.

Prior to the work presented in this chapter, the risk of AOV loss may have been estimated by the principal engineer during design and deployment. From anecdotal evidence this was often pursued on an informal basis. Such an unstructured approach is rarely, if ever, found entirely satisfying to all parties. Neither is it likely to be immune to criticism, from one side or another. Estimating the likelihood of loss depends on a number of factors such as the vehicle's intrinsic reli-

ability, the effects of the operational environment, the quality of the maintenance program, and the experience and competence of the deployment team. Assessing this risk requires a formal process that, in addition to providing a quantitative assessment, is transparent and able to be followed and replicated by others.

In 2007, *Griffiths* and *Trembanis* [24.9] introduced a risk management process tailored to the operation of AUV. The approach was initially used for managing AUV risk of loss [24.10]. Different stakeholders have different interests in risk, for example, the AUVs owner is interested in the safe recovery of the vehicle, whereas a scientist is mostly interested in the recovery of data or in the vehicles' availability at a given time. The process for AUV risk management can be applied for managing other risks, not only the risk of loss. In this chapter, we present a risk management process tailored to AOVs. Different methods for assessing different risks are discussed in detail.

Most examples are given for propeller-based AUVs but these are equally applied to vehicles using buoyancy change engines, typically denoted as gliders. We also look at aspects of AUV design and deployment affecting risk.

24.1 Risk Management Process for Autonomous Ocean Vehicles

While aspects of reliability had featured within papers on the use of AOVs, reliability had been the topic of very few specific studies until the mid-2000. The provocative, anecdotal evidence in [24.11] gave rise to discussions within the community of users on how to improve reliability and reduce risk at a time when vehicle operations were just starting to be independent of deploying vessels. Spurred by these considerations, and especially by future requirements to operate under ice, *Griffiths* et al. showed that simple statistical methods could be applied to estimate risk of loss of autonomous vehicles operating in various environments [24.12]. Subsequently, *Podder* et al. extended this approach to look at reliability growth [24.13], which is the main desired outcome. However, these studies looked only at revealed reliability, documenting what problems were emerging; they were not set within a framework for risk management.

The loss of the Autosub2 vehicle under the Fimbulisen in 2005, and the recommendations of a subsequent board of inquiry led to the development of a holistic risk management process for AUVs (Fig. 24.1) [24.9]. By simple extension of differently described operating environments, it is applicable to the wider class of AOVs. This remains the only published structured procedure for risk management of these vehicles.

Taking the steps in sequence, first a responsible owner is identified and they state an acceptable probability of loss for the campaign under consideration. The owner may factor in the importance of the vehicle's mission, the value of the data it would return, the future program for the vehicle, and other considerations, into the acceptable probability of loss. Independently, the principal investigator, or user, sets out the requirements of the campaign in terms of number of missions, their duration, and the environment characteristics, such as surface or subsurface, coastal, open ocean, under ice, or in areas of high traffic. The next step is for the technical team to assess the probability of loss in light of the campaign just described. This is a difficult step, and much of the remainder of this chapter is given over to methods used to make this estimate.

Accepting for now that this estimate can be made, it is then compared with the responsible owner's acceptable risk. If the estimated risk of loss is less than the owner is willing to accept, there is a need to demonstrate that the estimated risk is realistic before the campaign can proceed. This demonstration could take the form of reliability trials, where the vehicle would be run in a benign environment over a duration or range commensurate with the planned missions and the outcomes evaluated. If the estimated risk of loss is greater than the owner's acceptable risk, several feedback paths are followed.

The first activity is to rank the risk factors identified by the analysis of fault history, or fault projections. Those that can be mitigated, without question, are dealt with. This demands that the causes of failure are fully understood and the corrective measures are known, able to be implemented, and tested, before the campaign. Belief in ability to mitigate faults, or assertion alone, is not sufficient.

In parallel, the user may reassess the missions required and the responsible owner may reassess the acceptable risk, following which the test for acceptable risk is repeated. Two or more iterations may be needed, or indeed, the decision may be to postpone or cancel the proposed campaign.

Primarily, this process has been used for campaigns with a significant risk of vehicle loss in the Polar Regions. *Brito* et al. described its use for a campaign of six Autosub3 missions under Pine Island Glacier, Antarctica [24.10]. Here, the initial estimated risk was sufficiently high that reliability-proving trials in a Nor-

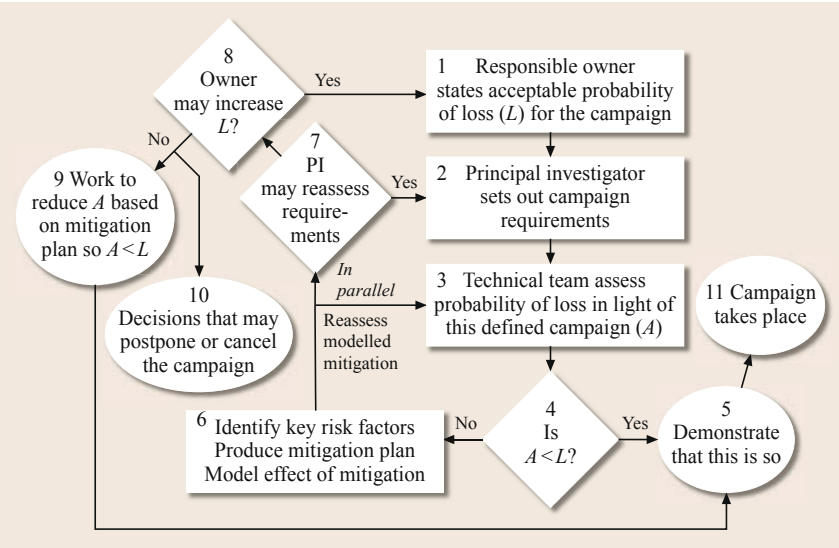


Fig. 24.1 A flowchart representing the risk management process proposed by *Griffiths* and *Trembanis* (after [24.9])

wegian Fiord were an essential precursor. While there were incidents on the Pine Island Glacier missions, the vehicle survived the campaign. More formal estimates of the likelihood of successful fault mitigation were made when the process was applied to an International Submarine Engineering (ISE) Explorer AUV for use in under sea ice missions in the Arctic [24.14]. The process has also been used in high-risk open water areas, such as for Autosub3 missions in the high traffic Bosporus Strait.

In practice, the methods used for quantifying the actual risk are directly linked with the acceptable level of risk. Risk is a measure of uncertainty. If the acceptable risk is high, for example, if the acceptable risk of AOV loss is 90%, there is little reward in applying formal methods for quantifying the actual risk. This is the case for platforms that are considered expandable, which may be deployed to study very rare events, such as a hurricane or the eruption of an underwater volcano. If the acceptable risk of AOV loss is low, then it becomes important to quantify the actual risk. This will give the user more confidence that the acceptable risk is met. This is the case where the platform is financed for long-term use.

24.2 Risk of Failure

The discipline of estimating a system's probability of failure or a system's reliability is well understood in many engineering fields. Reliability is defined as the probability that failure will not occur in the period of interest [24.15]. This probability figure can be estimated from system or component usage. However, in some applications it is not possible to obtain system or component failure history, particularly if the operating environment cannot be recreated in the laboratory or if a new system or component is being used. In such situations, reliability estimates can be obtained from expert subjective judgment – this is discussed in detail in Sect. 24.5.1.

Probabilistic modeling is key for the effective reliability estimation of a system composed of several components or subsystems. To facilitate the development of such models, graphical methods such as fault tree analysis can be applied. In this section, we present a summary of the techniques used for estimating AUV reliability. In Sect. 24.2.1, we present methods for estimating the reliability and reliability growth from operational data. In Sect. 24.2.2, we present how fault trees can be used for estimating the reliability of a complex system comprised of many other components.

24.2.1 Reliability Estimation

The number of failures that emerge during a vehicle, system, or component test can be used to estimate the mean time to failure for the vehicle. However, the mean time to failure alone is not sufficient to support decision-making with regard to AUV deployment. By fitting parametric models of reliability to AUV fault data it is possible to produce formulae that can be used to predict the probability of success of any mission in terms of its length. A number of distributions have been used for

modeling AUV failure distribution with time. The first study on AUV reliability estimation was conducted by *Griffiths* et al. [24.16]. Failures of Autosub1 AUV were collected from its first mission within Empress Dock, in Southampton in June 1996, to a science campaign on board the RV *Calanus* at Oban in November 1999. During this time, the vehicle conducted 216 missions covering 2125 km. Fifty failures were recorded, giving a probability of failure per mission of 0.231. Pareto analysis showed that human error, acoustic telemetry failure, and failure to dive were the top three failure modes in Autosub1's failure history. This report used statistical models for predicting the failure probability as function of mission length. The dataset used for this analysis consisted of the distance traveled for each mission and whether the mission ended due to a fault. Each entry was classed as censored or not censored. A censored observation is one where the mission was terminated not due to failure, whereas a noncensored data captures the distance at which a failure took place. Two types of functions were defined directly from the data [24.12]. The first function, the probability of failure $F(x)$, where x stands for an instantiation of distance X. In mathematical terms, this is captured by $F(x) = P(X < x)$, where $P(X < x)$ is obtained directly from the probability function that is fitted to the historic data and lower case x is an instantiation of X. The second function of interest is the reliability, otherwise known as survival function, is the probability of survival, $R(x)$. The reliability is the complement of the probability of failure, it is mathematically defined as $R(x) = 1 - F(x) = 1 - P(X < x)$. It stands for the probability of the system surviving without failure, the distance x.

Griffiths et al. fitted three well-known distributions to the historic data: *Weibull*, *log logistic*, and *log normal*. Results showed that the probability of a fault occurring in a 100 km mission was 0.352 using the *Weibull*, 0.334 for the *log logistic*, and 0.309 for the *log normal*. Software reliability was studied separately using two independent software reliability models: the Poisson model and the Littlewood model. Both models lead to the same reliability estimate for a 100 km mission of 0.93 for the software.

The dataset was later updated to include Autosub2 science missions on MV *Terschelling* at three sites on the west coast of Scotland in March and April 2000, equipment trials at Plymouth and a science campaign in the Strait of Sicily in June 2000. A total of 869 km were covered in 24 missions [24.12]. Here the authors fitted six different distributions to the data: extreme value, normal, logistic, and their logarithmic versions: Weibull, log normal, and log logistic. The logarithmic distributions provided a better fit to the data than their linear counterparts. The Weibull distribution showed the highest failure rate, and, to err on the conservative, was chosen as the preferred parametric model. Results were similar to those previously reported.

The first reported work on the effects of upgrades on the reliability growth for AUVs was presented in 2003 [24.17]. The dataset consisted of Autosub missions considered in [24.12] plus missions that were conducted in two engineering trials in June 2002 and September/October 2002. The purpose of these trials was to test new software that gave enhanced autonomy. The results of the analysis showed reliability growth prior to the missions carried out in June 2002. However, after the upgrades there was an increasing failure rate. These were worrying results because the vehicle was due for an under ice campaign in February–March 2003. The team used the reliability model to estimate the number of missions, with an acceptable number of faults, needed to ensure reliability growth. The analysis concluded that 10 missions were required with distances varying from 5 to 144 km in length, with only two high-impact faults allowed on the shorter missions.

A study in reliability growth was carried out for the Dorado AUV designed and operated by Monterey Bay Aquarium Research Institute (MBARI) in preparation for an Arctic campaign in the following summer [24.13]. The dataset consisted of mission data collected during operations in 2003 and 2004, where the vehicle had traveled approximately 1700 km in Monterey Bay, California. Several failures emerged; the analysis considered only failures that resulted in an abort – a total of 14. Basic trend analysis using the Laplace test yielded the conclusion that the AUV reliability exhibited an increasing trend. This was verified using two independent reliability growth models due to Duane and Crow. Both use graphical approaches to support reliability growth analysis. The estimation of the mean time between failures (MTBF) was conducted for different groups of failures; set according to their criticality. The estimated MTBF for the extremely critical faults was 167.8 h.

Reliability estimation based on revealed operational history helps those involved understand better the failure pattern of these autonomous vehicles, enabling the implementation of failure mitigation mechanisms during the operational lifetime of the AUV. However, for early lifetime, this failure history is not available. For the early lifetime of a vehicle, the reliability of a vehicle can be estimated using reliability modeling techniques. These are discussed in the next sections.

24.2.2 Reliability Modelling

Reliability modeling aims to estimate system's reliability based on its intended functionality, and design

architecture. The failure model end effect analysis (FMEA) is a well-established technique for identifying critical components or phases of the system design and operation [24.18]. However, this technique does not accommodate quantitative estimation of the system's reliability and therefore it will not be discussed in this chapter. Event trees and fault trees are two basic methods for system reliability quantification. Event trees use forward logic. They begin with an initiating event (an abnormal incident) and *propagate* this event through the system under study by considering all possible ways in which it can affect the behavior of the subsystems. Event trees are useful for accident sequence analysis in which the aim is to estimate the likelihood of a sequence of potential functioning or malfunctioning events. Thus event trees cannot be used for estimating the probability of failure of a system. This figure can be estimated with the support of fault trees. In a fault tree analysis, one attempts to develop a deterministic description of the occurrence of an event, the top event, in terms of the probability of occurrence of other (intermediate) events. Intermediate events are also described further until, at the finest level of detail, the basic events are reached. When the top event is failure of a system then the basic events are usually failures of components. A fault tree can be considered as an expression in a Boolean model, which implements predicate logic. A fault tree contains the same operators, for example, conjunctions and disjunctions. However, instead of operating with bits, where 0 corresponds to false and 1 corresponds to true, fault trees operates with probabilities – varying from 0 to 1 [24.20]. Figure 24.2 presents a fault tree devised by *Griffiths* and *Brito* for estimating the probability of failure during the deployment of an AUV through an ice hole [24.19]. Graphically a conjunction is represented using AND and a disjunction is represented using an OR gate. In Fig. 24.2, the operators named G_1, G_2, G_3, and G_4 are OR gates. The G_5 operator is an AND gate. A fault tree injects transparency into the fault analysis process. Failure modes are clearly presented and how they propagate in the system is visually displayed.

There is a unique probabilistic model associated with each fault tree, which enable us to calculate the probability of failure for the top event. The probability of failure to launch, $P10$, is calculated using the following expression:

$$\begin{aligned} P10 &= G_1 \\ &= 1-(1-G_2)\times(1-P3)\times(1-P4) \\ &\quad \times(1-G_3) \\ &= 1-((1-P1)\times(1-P2)) \\ &\quad \times(1-P3)\times(1-P4) \\ &\quad \times(1-(1-(1-G_4)\times(1-G_5)))\,. \end{aligned} \tag{24.1}$$

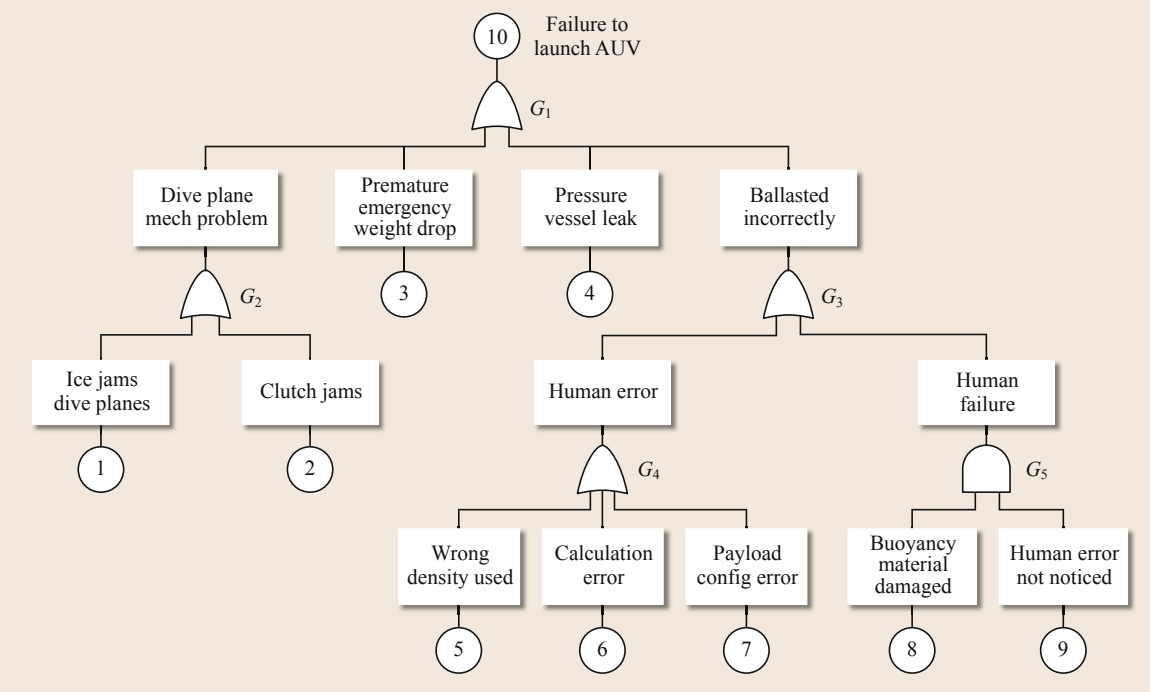


Fig. 24.2 Probability of failure during AUV deployment through an ice hole. Fault tree presented in [24.19]

and

$$G_4 = 1 - (1 - P5) \times (1 - P6) \times (1 - P7) \,, \quad (24.2)$$

$$G_5 = 1 - (1 - P8) \times (1 - P9) \,. \quad (24.3)$$

The probability of failure for each basic event can be estimated using statistical modeling discussed in the previous subsection in which a probability function is fitted to observed data. Alternatively, if operational data is nonexistent, the probability of failure for the base events can be estimated using expert subjective judgment. In Sect. 24.5, we provide more details on the formal methods that can be used to elicit expert judgment.

24.3 Risk of Collision

The risks of collision for AUVs are very real. However, as unregulated vehicles, no figures for the number of losses due to collisions can be given; there is no requirement to record losses for these unmanned vehicles. Through anecdote, the incidence of collision for underwater or surface vehicles is sufficiently high that any group operating several vehicles is likely to encounter this risk regularly. Sensitivity of many operators to the details of collisions also means that quantitative, or attributable, information is rare. The examples given here are based on our knowledge of real incidents.

24.3.1 Risks of Collision on or Near the Surface

For surface vehicles, and when underwater vehicles operate on or near the surface, the main collision risks are:

- *The shore, or water shallower than the minimum operating depth of the vehicle*: At least one autonomous vehicle has run ashore because parts of the chart from which waypoints were taken were surveyed in the 19th century, and the position error for the shoreline was substantial. Operating near a coast with very gentle beach gradients without full knowledge of, and accounting for, the tides can give rise to grounding.
- *Manned surface craft*: Collision with manned surface craft has probably been responsible for many autonomous vehicle losses. In some cases, the manned craft concerned has been the craft deploying the autonomous vehicle. Mistakes and miscommunication during launch and recovery, especially, can lead to collision with the support vessel. Collision with the sides of the vessel may inflict little damage, but collision with propellers has on more than one occasion led to damage to the pressure vessels of underwater vehicles and immediate loss.
  Where statistics of vessel traffic are available, probabilistic models may be used to assist in estimating the risk of collision. *Merckelbach* has developed such a model, specifically for undersea gliders [24.21]. In principle, his approach is applicable to other autonomous vehicles. The model considers the vessel and vehicle speeds, the vessel traffic density, the statistics of vessel length, draft and breadth, the water depth, the operating parameters of the vehicle, for example, time on surface and within the draft of vessels, and a bow-wave factor. This last factor accounts for the pushing aside that can happen as collision is imminent; in effect the effective breadth of the ship is reduced by this factor.
  Mission planning to take account of surface vessels can make use of the automatic identification system required on *internationally voyaging ships* of 300 tonnes or more, all passenger ships, and, from 2014, the entire European Union (EU) fishing fleet of vessels over 15 m in length. Websites with real-time maps enable planning ahead, and also near-real time collision avoidance.
  For very intensive shipping areas, for example, Bosporus Strait with typically six large ships per hour, autonomous vehicles should avoid the surface.
- *Large flotsam and jetsam, such as logs and baulks of timber, lost freight containers*: These hazards are less amenable to quantifiable modeling than ship traffic. Local knowledge is likely to be important, for example, on the known areas where felled trees from boreal forests may be prevalent.
- *Floating nets and fisheries-related hazards*: Despite being banned by international agreement since 1992 in international waters, drift nets with surface floats that are meant to catch pelagic fish are still in use within the exclusive economic zones of many countries, for example, the USA. Other nations stipulate maximum lengths (2.5 km for the EU). These nets are a hazard for AOVs on the surface. Local knowledge may suggest where such fishing happens. Avoidance using telepresence from the vehicle to the command center of video or radar imagery may assist in spotting the surface floats.
- *Offshore structures, for example, oil and gas installations, wind farms*: The positions of these surface

hazards are usually well known, and fixed. The onus is on the autonomous vehicle operator to plan missions with sufficient contingency and allowance for navigational error such that a vehicle does not infringe the exclusion zones around these structures.
- *Coastal structures such as breakwaters, or moles*: While there may be no exclusion zones around these hazards, their positions are fixed, and appropriate prior planning should be adequate to mitigate the risks.

24.3.2 Risks of Collision Underwater

Apart from the universal case of collision with the seabed, the other risks of collision when underwater depend on the environment of the missions. Near-shore missions may encounter obstacles such as piers and other fixed structures, vessels at anchor, and the submerged parts of vessels underway. In the open ocean, apart from with the seabed, collision when submerged is likely to be a rare event. Under ice, collision is likely with ice at the sea surface, with projecting keels of sea ice, and with icebergs. While grounded icebergs, ice shelves and glaciers and sea ice that is locked to the shore or a shoal may be stationary, other forms of ice should be considered as moving obstacles.

Collision, or obstacle, avoidance for AUVs is a well-studied topic, combining as it does the challenges of sensing, interpretation, and action. The effectiveness of obstacle avoidance is one indicator of the degree of autonomy of a vehicle. *Horner* et al. [24.22] summarized these challenges, and provided examples of how a forward-look sonar array can be integrated with an autopilot control algorithm that minimizes the cross-track error from the desired path. That is, the system provides for adequate, but minimum, deviation from the desired path to avoid the obstacle. The interpretation of the obstacle characteristics in this implementation uses image information from the sonar array, borrows techniques from computer vision, and can be made robust against false targets. This approach is relevant for avoiding the seabed, but is also sufficiently powerful to provide obstacle avoidance in more complex situations, such as near piers and other man-made structures.

There are situations where a simpler approach may give sufficient capability, for example, where the seabed is the only obstacle likely to be encountered. Using a single beam mechanically scanned sonar and a horizon-tracking algorithm, *McPhail* et al. [24.23] showed that effective seabed avoidance could be delivered when terrain following at altitudes down to 10 m on the flanks of a seamount, and down to 3 m on the flatter, but rocky, summit. Vehicle reactions included simple avoidance in the vertical plane, and turn-around and retry.

Under ice, information is needed on obstacles ahead, below, and above the vehicle. One approach, devised for the DEPTHX (deep phreatic thermal explorer) vehicle [24.24] proposed a suite of 24 narrow beam sonars and 30 imaging sonars. In the subsequent Endurance vehicle, 64 pencil beam sonars provided a 3.5 steradian view of the environment within ice-covered Lake Bonney in Antarctica. A far simpler approach is to use for obstacle avoidance purposes, sonar information from other instruments on the vehicle. *Pebody* has described how, for an under ice application, a forward-looking sonar may be augmented using range to reflector information from the four beams of an upward-looking and a downward-looking acoustic Doppler velocity log [24.25]. This system was tested using an air curtain behind a ship, to simulate an iceberg, and used under fast ice off Greenland, and under Fimbulisen and Pine Island Glacier in Antarctica to good effect.

As a backup to collision avoidance, some operators include rubber fenders or other protection on the nose and undersides of their vehicles. Incidents where collision avoidance has failed include repercussions of hardware failures and failures of the control systems to properly appreciate the form of the environment and hence failure to take proper action. Drawing on examples with Autosub AUVs, a sternplane fault led to the vehicle diving far too steeply, such that its bottom avoidance sonar was at the incorrect angle, the sonar could not see the seabed, the vehicle nose collided with the seabed but the vehicle continued to make slow forward progress until the nose became full of sediment. For an example of inappropriate response to an unexpected environment, when terrain following at low altitude up a steep cliff off Sicily, Autosub2 encountered an overhang. The appropriate response would have been to reverse out, but this behavior was not preprogrammed, and instead the vehicle dropped its abort weight and had to be retrieved using an remotely operated vehicle. These lessons also serve to show that not all collisions are fatal to a vehicle.

24.4 Risk of Unavailability

System availability is defined as the probability of the system being available given that the system is needed at a given time. System unavailability is the complement of this figure. The deployment of an AUV consists of several phases. A series of tasks are carried out in each phase of the deployment. These human- or machine-related tasks are not immune to error. Estimating the availability for the AUV being at a given phase is only possible if we take into account the sequence of phases that precede that phase. This problem can be mathematically modeled using a probability approach denoted as Markov theory.

For a brief description of Markov theory, consider a probability problem with a set of outcomes of interest $E_1, E_2, \ldots, E_k$. Given that there is a probability p_j associated with each event the joint probability for a given sequence is defined by the multiplicative property, thus $P(E_1, E_2, \ldots, E_k) = p_1 p_2 \ldots p_k$. The Markov chain theory introduces an assumption that simplifies this expression; it considers that the outcome of any trial depends on the outcome of the preceding trial and only on it [24.26]. Therefore, if event E_1 precedes event E_2 and event E_2 precedes event E_3 and so on for the remaining events, then instead of associating a probability to an event E_j, it uses a transition probability p_{jk} for every pair of events (E_j, E_k), where p_{jk} is the probability of E_k occurring given that E_{k-1} occurred in the previous trial. A Markov model may have more than one sequence of events; E_j may have two or more posterior states. Therefore j and k are not necessarily adjacent. In addition to p_{jk}, one must also define the probability of E_j occurring at the initial trial, a_{j0}. Therefore for the initial trial, $P(E_j, E_k) = a_{j0} \times p_{jk}$. For the general case, considering a sequence of several transitions, given that event E_{j0} precedes E_{j1} which precedes E_{j2} and so on for the remaining events, the joint probability distribution is computed using the expression in (24.4)

$$P_r(E_{j0}, E_{j1}, \ldots, E_{jn}) = a_{j0} p_{j0j1} p_{j1j2} \ldots p_{jn-2jn-1} P_{jn-1jn} \,. \tag{24.4}$$

It is not unusual to find problems where one state has more than one potential preceding state. When this is the case, the mathematical calculation cannot be performed using the simple equation presented above. Instead the calculation is performed using matrix operations. The transition probabilities are arranged in a matrix denoted as transition matrix or stochastic matrix. The transition probability together with the initial state vector completely defines the Markov chain. The availability after n transitions can be calculated by the product between the transition matrix to the power of n and the initial state vector.

Brito and *Griffiths* [24.27] used Markov chains for modeling and estimating the availability of Autosub3. The Autosub3 deployment sequence was modeled as a Markov chain in which each state corresponds to one phase of the vehicle deployment. A key assumption in the proposed model is that the Autosub is deployed from a stationary vessel. The model can be altered to capture the case where the AUV is deployed from an ice hole or from the coast. This would consist of adding a state for which the transition probabilities would depend on the failure modes that can emerge in the different type of deployment. *Griffiths* and *Brito* in [24.19] give an example of when the approach was adopted for deployments from ice holes.

Here we briefly describe the model proposed in [24.27] for estimating AUV availability based on surface vessel deployments. The Markov chain model for AUV availability consists of 11 phases, these states and the transitions are presented in Fig. 24.3. The first state captures the phase when the vehicle is on board of the vessel, switched on, ready for testing (*Dp*). A series of communication, actuation, and navigation checks are carried out during this phase. Having passed the onboard checks, the deployment moves to vehicle ready for deployment phase (*Dr*). Here the vehicle is waiting to be deployed overboard. While overboard there is the risk that the vehicle may run under the vessel causing severe damage, in which case the vehicle may need to be salvaged. This scenario is captured by the transition from phase overboard (*O*) to salvage (*S*). Once overboard and at a safe distance, communications checks are carried out before the vehicle starts diving. The vehicle is then set to stop diving when it reaches a predefined depth and enters a holding pattern (*Sh*). During this period, the vehicle is still within acoustic range, more checks are carried out before committing to the mission. Once underway (*U*), there are two possible outcomes: the vehicle can be recovered (*R*), or it can be lost (*L*). If historic operational data exist then the probability of AUV loss can be computed using reliability modeling techniques combined with expert judgment – this method is described in the Sect. 24.5. Alternatively, if there is no historic data of AUV operation then the probability of loss can be computed using fault tree modeling that takes into account expert judgment with regard to the consequence of each failure mode.

The AUV can be recovered while it is in the overboard phase, during diving, in holding pattern or underway. The vehicle can also be lost from any of these

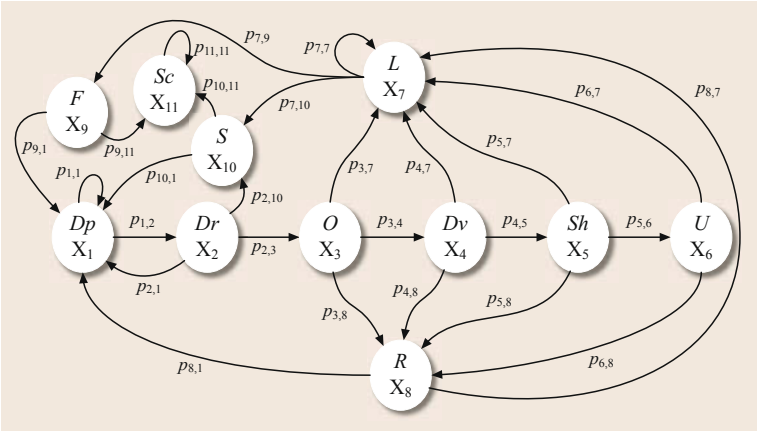


Fig. 24.3 Markov state-space model capturing the sequence of events undertaken during AUV deployment and operation. A directional arrow from state i to state j means that the process can move from state i to state j

Table 24.1 Aggregated expert judgments for the transition probabilities

Transition stimuli	Transition probability	Transition stimuli	Transition probability	Transition stimuli	Transition probability
$P_{1,1}$	$1-p_{1,2}$	$P_{4,5}$	0.9565	$P_{7,9}$	0.33
$P_{1,2}$	0.875	$P_{4,7}$	0.0085	$P_{7,10}$	0.33
$P_{2,1}$	0.055	$P_{4,8}$	0.035	$P_{8,1}$	0.998
$P_{2,3}$	0.94	$P_{5,6}$	0.98	$P_{8,7}$	0.002
$P_{2,10}$	0.005	$P_{5,7}$	0.0055	$P_{9,1}$	0.75
$P_{3,3}$	0.0195	$P_{5,8}$	0.0145	$P_{9,11}$	0.25
$P_{3,4}$	0.925	$P_{6,7}$	$1-p_{6,8}$	$P_{10,1}$	0.7
$P_{3,7}$	0.0495	$P_{6,8}$	p_{survival}	$P_{10,11}$	0.3
$P_{3,8}$	0.006	$P_{7,7}$	$1-(p_{7,9}+p_{7,10})$	$P_{11,11}$	1

phases. If a vehicle is lost, the deployment can be maintained in a permanent loss state (L), captured by the transition from $p_{7,7}$ or it can move to a state of Salvage (S) or Found (F). From these states, the AUV can be declared fit for re-deployment, this is captured with the transition from state (S) to (Dp) and (F) to (Dp), or declared scrapped (Sc). A deployment can reach a state of AUV scrapped from overboard (O), dive (Dv), holding pattern (Sh), underway (U), and recovery state (R).

For Autosub3 availability analysis, the transition probabilities were calculated based on expert judgments. Three experts with more than 60 years combined experience on AUVs have estimated the probability for each transition. The transition probabilities are a result of the un-weighted linear pool aggregation of the expert judgments. The transition probabilities for the availability of Autosub3 are in Table 24.1.

The Markov property states that the sum of the probabilities leaving any given state must be equal to unity. If the sum of all transitions leaving a state is constant c, where c is lower than 1, then the probability of the process remaining in the same state in the next transition is $1-c$. The Markov condition was applied to calculate transition probabilities $p_{1,1}$, $p_{6,7}$ and $p_{7,7}$.

The availability of Autosub3 from ready for test, while on deck, to underway was calculated to be 0.75. This estimate is similar to that obtained for unmanned aerial vehicles (UAVs). The United States Office of the Secretary of Defense published the availability figures for five UAVs: Predator RQ-1A (concept demonstrator) (0.40), Predator RQ-1B (early production) (0.93), Pioneer RQ-2A (1990–1991) (0.74), Pioneer RQ-2B (0.78), Hunter RQ-5 (reliability enhanced 1996–2001), (0.98), Average UAV (0.77).

24.5 Risk of Loss

The task of estimating AUV risk of loss is affected by many factors. Technical failure rate and the likelihood of a human error are two important factors in the assessment of AUV risk of loss but these factors alone are not sufficient. A failure has different impact depending on the AUV operational environment. For example, a failure leading to unexpected drop of the abort weight causing the vehicle to surface may have a low impact

in open-water operations; however, if the same failure takes place while the vehicle is under an ice shelf, sea ice, or busy coastal area it may lead to vehicle loss. The human element is also of great importance when it comes to failure mitigation, as highlighted by *Stokey* et al. [24.11], many technical failures in AUV operations can be mitigated by a more experienced team.

These factors are difficult if not impossible to capture deterministically, particularly if the number of previous deployments in the target operational environment is small or if the environment and operational conditions are impossible to predict. Autonomous ocean systems are not unique in this challenge, and many techniques have been developed and widely accepted, for risk quantification in areas such as nuclear, process, oil and gas, and aerospace safety. Common to these approaches is the formal use of expert judgment.

In this section, we present formal expert judgment elicitation methods focusing on their pragmatic characteristics for estimating AOV risk. When studying component or systems reliability, it is not sufficient to have the MTBF, designers and operators are often interested in knowing the component or system reliability as a function of time in operation or mission length. This is imperative for the implementation of corrective measures for risk reduction. Similarly, when operating AOVs the decision to go, or not go, for a mission is better informed by a risk profile that captures the probability of loss with mission time or length. The creation of a probability of loss profile is possible by integrating experts' subjective judgments with statistical survival techniques.

24.5.1 Expert Judgment Elicitation Processes

Governments and companies are increasingly relying on expert panels for providing risk assessments for high-risk or high-profile projects such as the safety assessment of a nuclear power station or the implementation of a major science program [24.28]. There are many dangers in eliciting expert judgments in an informal way. When providing assessments of events, people often follow one of a number of mental shortcuts, denoted as heuristics [24.29, 30]. Research has shown that when used incorrectly, these heuristics can lead to systematic and predictable bias [24.31]. In their work, *Tversky* and *Kahnemann* [24.29] have identified three predominant heuristics: representativeness, availability, and adjustment and anchoring. Representativeness is described as the tendency to judge the probability that A belongs to B by how representative A is to B. For example, if the description of Peter is highly representative of an engineer then people tend to judge the probability of Peter being an engineer as high regardless of the base rate, that is, the proportion of engineers in society. The base rate neglect is one type of bias that can be introduced by following the representativeness heuristics; other types of biases are, for example, insensitivity to sample size, insensitivity to predictability, and misconceptions of chance and regression [24.29]. Availability heuristics is the tendency to judge a particular event by the ease with which the instances and events can be brought to mind. For example, one may estimate the risk of heart attack among middle-aged men by remembering specific personal examples. Here bias can be introduced due to exposure to negative outcomes; they tend to be easier to remember than positive ones. Anchoring and adjustment is a tendency to anchor probability estimates at an initial estimate. Insufficient adjustment results in biases of underestimation or overestimation.

Formal judgment elicitation has been proposed as a solution to reduce biases and facilitate reproducibility of the results [24.32]. A formal expert judgment elicitation is a structured process that consists of a number of phases; a number have been proposed in the literature [24.33]. Below we describe the generic structure of a formal judgment elicitation process tailored to AOVs:

1. *Set out the issues*: In this phase, the issue that is to be addressed is described in general terms. This is a scope definition stage; it defines the nature and direction of the analysis, the choice of questions to be asked and issues to be considered. For example, the issue of interest is to estimate the probability of vehicle loss.
2. *Selecting the experts*: An expert is someone with specialist knowledge in the task in hand who is also capable of decomposing a complex problem into subsidiary problems. In vehicle risk assessment, an expert should be someone with experience in the type of fault scenarios and in AOV operations.
3. *Clearly define the issues*: The issue in hand is what is the probability of failure F leading to loss (L) in operational environment E? This is mathematically represented as $P(L \mid F, E)$.
4. *Training the experts and eliciting judgments*: This is the most important task in the elicitation process. Different formal judgment elicitation processes follow different approaches. For some, training is quite informal it consists of providing some examples of typical assessments and an explanation of basic principles of probability theory. Other judgment elicitation processes are stricter and encourage the facilitator to elicit expert assessments for a number of seed questions. These are questions for which the facilitator knows the answer but the expert is not familiar with the problem or question [24.32], and

where responses can be used to calibrate individual assessments.

5. *Analyzing and aggregating*: This is a critical phase of the assessment. Here the decision maker uses a strategy for combining the expert judgments into a single assessment that represents the group's view. This can be done mathematically or behaviorally. A mathematical aggregation uses analytical functions for combining the expert judgments. The analysis of expert judgments must be conducted prior to the aggregation to ensure that any misunderstandings and bias are removed from the assessments.
6. *Complete analysis and write-up*: The assessments provided by the experts are documented, and disagreements between experts are recorded. A report is submitted to the experts for review. Following this review, a report is submitted to the decision maker informing the results of the risk assessment exercise.

Mathematical Aggregation

The use of mathematical methods for consensus building has the benefit of facilitating the application of the elicitation process because experts are not asked to agree on the final judgment; in fact experts do not have to be in the same room during the elicitation. Supporters of mathematical methods have argued that behavioral methods, in addition to being time-consuming, can lead to systematic biases caused by group polarization [24.34].

The analytical function for combining individual judgments can take into account the level of expertise in the topic and the effectiveness of the calibration. These factors are captured in a weight w_i, where $i = 1, \ldots, n$ or in an a priori distribution judgment. Methods for mathematically aggregating expert judgments were developed based on a set of widely accepted axioms [24.35]. Perhaps the most appealing method for combining expert judgments is the *linear opinion pool* [24.36]. Here the expert judgments are aggregated by taking the weighted average of the assessments provided by the experts. The *linear opinion pool* complies with the property of *consistent marginalization* [24.37]. A different mathematical aggregation method, the *logarithmic opinion pool*, uses a multiplicative averaging of the expert assessments [24.36]. The *logarithmic opinion pool* does not comply with the marginalization principle but unlike the *linear opinion pool* it complies with the *principle of external Bayesianity*. This can be explained as follows: suppose that the decision maker or facilitator has reached an aggregated $p(\theta)$ but has subsequently learned new information relevant to θ. One way to update the judgment $p(\theta)$ in light of the new information is by updating the individual expert judgments $p_i(\theta)$ and then aggregate all judgments using an analytical function. A second way is to update the original $p(\theta)$ without updating the individual expert judgments, using Bayes theory. A mathematical aggregation method that complies with the *principle of Bayesianity* will produce the same result regardless of the process used for creating the final judgment.

Other mathematical aggregation methods are more complete from the axiomatic viewpoint. But this comes at a price of being more difficult to implement. *Bayesian aggregation methods* adopt a completely different prerogative to the *linear and logarithmic pools*. Here, the decision maker begins by defining his own prior distribution $f(\theta)$. The expert judgments are then incorporated, using Bayes' rule, to form the decision maker's *posterior* distribution $f(\theta \mid D)$, where $D = \{f_1(\theta), \ldots, f_n(\theta)\}$ is the set of experts' elicited distributions. In this context, according to Bayes theorem, the $f(\theta \mid D)$ is proportional to $f(\theta)$ multiplied by the likelihood term $f(D \mid \theta)$. The problem with Bayesian methods is that the decision maker must specify his own prior belief about the risk $f(\theta)$ and the knowledge and beliefs about the experts $f(D \mid \theta)$ [24.35].

In 2008, in preparation for the Autosub3 deployment under the Pine Island glacier, a risk model for Autosub3 was developed based on the assessments provided by eight independent experts. Experts were asked to assess the probability of AUV loss given that a fault X emerges during the deployment [24.10]. In addition to providing an estimated probability of loss, experts provided a weight, from 1–5, capturing their confidence in the assessment. A weight of 1 represented little confidence in the assessment, while a weight of 5 meant that the expert was very confident. The experts considered 63 faults, in four environments: open water, coastal water, under sea ice, and under ice shelf. A detailed description of each environment was provided prior to the assessment. The faults were collected during six Autosub3 campaigns from mid 2005 to mid 2008.

Figure 24.4a shows the relative frequency of the average of the weights used by the experts. The figure shows that on average, for 70% of the assessments, experts assigned a weight of 3 or higher, for open water, coastal waters, and ice shelf environments. Experts were least confident with the sea ice environment.

Figure 24.4b shows the un-weighted linear pool probability of loss, for the five most critical failures in the four different environments. This makes it clear that, in these experts' opinion, faults that would have a low probability of leading to loss in open or coastal water would have a high probability of loss under an ice shelf, but not through simple scaling.

Expert judgment has also been applied to estimating the risk of loss of commercial AUVs, including two Remus 100 vehicles in coastal areas and under

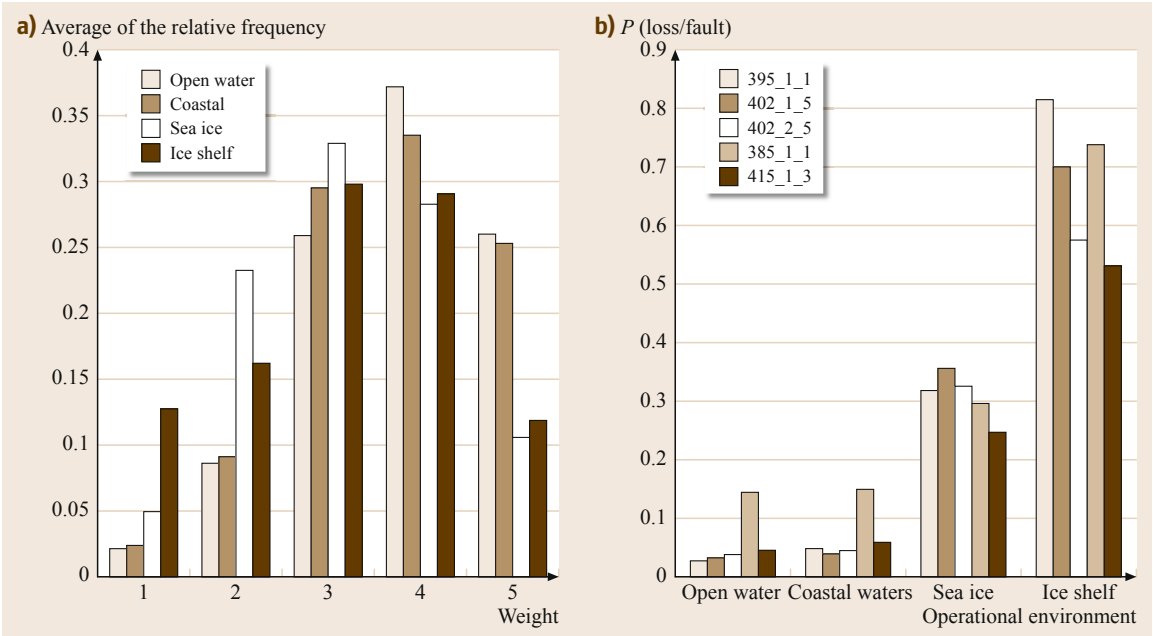


Fig. 24.4 **(a)** Relative frequency of the average weights used by experts Pareto plot of the faults collected for the expert judgment elicitation. **(b)** Un-weighted linear pool, probability of loss, for the five most critical failures

sea ice. These vehicles were operated by the Center for Coastal Marine Sciences at California Polytechnic State University (CalPoly). They were used for a range of missions to better characterize and improve understanding of coastal waters [24.38]. The operational data set contains faults and incidents recorded from 186 missions between July 2001 and February 2009.

For this elicitation experts were asked to specify an unimodal distribution for the probability of loss given a fault. A narrow distribution would show that the expert is confident about the assessment; a widely spread distribution would show that the expert was very uncertain about the assessment. Expert specified this distribution using five quantities:

1. The lower bound, L. The minimum value that P(loss) can take.
2. The upper bound, U. The maximum possible value of P(loss).
3. The median, M. The value for which there is 50% chance of P(loss) being above or below it.
4. The lower quartile, LQ. The value for which there is 25% chance that P(loss) is between L and LQ and 25% chance that it is between LQ and the median, M.
5. The upper quartile, UQ. The value for which the expert is 25% confidence that P(loss) is between the median and UQ and 25% confidence that it is between UQ and the upper bound.

The experts' judgments were aggregated using the un-weighted linear pool, but in two separate groups, the optimists and the pessimists; results are presented in [24.39]. A key conclusion was that these vehicles were being knowingly operated with very low probability of loss vulnerabilities, which, on a statistical basis could become important given the large number of missions undertaken.

The limitation with mathematical aggregation approaches is that experts do not have the opportunity to discuss different views and thus to remove bias from their assessment. The aggregation of expert judgments in two different groups helped solve this potential bias issue. However by doing so, the decision maker needs to decide whether he or she is going to be an optimist or pessimist. One way to mitigate the risk of bias is by having experts in the same room, allowing them to discuss their views before agreeing on a judgment. Such an elicitation method is denoted a behavioral elicitation method.

Behavioral Aggregation

The decision of whether to aggregate expert judgments mathematically or behaviorally ultimately rests with the decision maker. Research has shown that a group consensus will never outperform the assessment of the best expert in the group. The difficulty is in identifying the best expert [24.40]. Behavioral aggregation brings the experts together as a group, provides a structured

process so they can share their knowledge and allow persuasive arguments to change their views, and help them to generate a consensus distribution [24.41]. This approach is often more transparent and more immune to criticism than mathematical aggregation methods.

A behavioral expert judgment elicitation was conducted to build a risk model for two ISE Explorer AUVs, operated by Defence Research Development Canada (DRDC) [24.19]. As these vehicles were to be used on data gathering of national importance in the high Arctic [24.42], a transparent and justifiable process was needed for how the risk of vehicle loss was evaluated; hence a behavioral aggregation approach was taken. Analysis of the outcomes of the discussions central to behavioral aggregation showed that there were seven classes of fault assessments:

- Assessments where the panel reached unanimity that the fault would inevitably lead to loss under ice, which included five faults where experts set all parameters of the distribution to 1.
- Assessments where the panel reached unanimity that the fault would have no impact at all on survivability, such as failure of a component or subsystem that would not be present for the Arctic.
- Assessments for faults where the experts considered that the phase of the mission may affect the consequence. This class of faults, typified by a failure in the vehicle control computer, resulted in vigorous discussion by the experts on the probability of loss, as some experts considered the outcome to be strongly dependent on the phase of the mission during which the fault occurred.
- Assessments of faults where individual experts shared particular insights affecting the aggregated outcome.
- Assessments where there was an agreement that the fault leads to a wide range of probability of loss.
- Assessments of faults that provided insights into instances of where a fault implied a consequential vulnerability.
- Assessments where the panel reached an agreement to use heuristic shortcuts. The group agreed collectively that they would spend little time on those faults that had a very low, but nonzero, consequence for the risk of loss. For these, they agreed on a standard distribution with a lower limit of 0, a median of 6.2×10^{-8}, and an upper limit of 10^{-6}.

A novel aspect of this study was the quantification of the risk mitigation activities. For each fault, experts were asked to assess of the impact of the risk mitigation plan, setting P_M to 0 if they believed that the mitigation strategy would not mitigate the fault and to 1 if the mitigation plan would completely mitigate the fault, with intermediate values reflecting intermediate belief in mitigation effectiveness. When these judgements were plotted in a histogram, three distinct distributions were observed (Fig. 24.5).

One distribution, with a mode at zero, covered faults for which the experts agreed that the cause of the fault was unknown or unproven and thus experts were unconvinced that the proposed mitigation strategy would prove effective. The second distribution, with a sharp mode at 0.5, represented those faults the experts considered that, although the proposed solution was appropriate, the mitigation strategy had not been sufficiently tested or proven in field trials, or where a recurrence of a similar fault could not be ruled out. The third distribution had a mode at over 0.9, indicating a high to very high level of confidence by the experts that the causes of faults were well understood, the solutions known and tested.

The resulting probability of loss given the mitigation is calculated using the following expression $P(L \mid F, E, M) = P(L \mid F, E) \times (1 - P_M)$. Combining the assessments on $P(L \mid F, E)$ with P_M identifies those faults where $P(L \mid F, E)$ is high but P_M is low. These form an important subset of faults for the engineers to address. Most critical was a vehicle control computer configuration problem, where $P(L \mid F, E)$ was 1 and P_M was 0. All other faults where P_M was less than 0.1, $P(L \mid F, E)$ was less than 0.01, consequently, the need for effort into improving the understanding of the mitigation required was far less important. Of the 14 faults where $0.4 <$

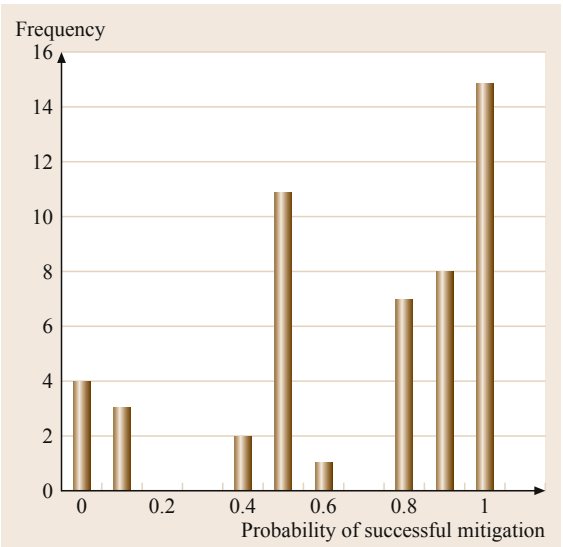


Fig. 24.5 Histogram of the assignments of probability of successful mitigation for the 51 faults, showing three distinct distributions

$P_M < 0.6$, 8 were assessed with $P(L \mid F, E) > 0.5$. This was the most important set of faults for further investigation and improvement in P_M.

24.5.2 Survival Prediction

The risk assessments discussed in the previous section can be used for directing engineers toward where to put more effort in fault mitigation. However, the AUV risk of loss will vary with mission time or length. The decision to deploy an AUV in an extreme environment is better informed by a risk profile that captures the variation of risk with mission distance or time. The creation of such risk profile is possible by integrating the expert judgments of probability of loss with statistical survival models.

Survival modeling is very popular in medical statistics, where the survival function $S(t)$ is defined as the ratio of the number of individuals with survival times $\geq t$ to the total numbers of individuals in the study. These methods have been adopted for modeling reliability of systems and components [24.43].

An adequate and representative dataset is central to the application of statistical survival techniques. Some entries will consist of recorded time of failures; however there may be missions with no failures. When this is the case an entry is denoted as right-censored data.

There are several mathematical models for representing censored data; these are divided into parametric and nonparametric models. Parametric models assume that the failure history follows a particular shape whereas nonparametric models make no assumption with respect to the shape of distribution. In [24.12] the authors used the Weibull parametric model and the Kaplan–Meier nonparametric estimator to model the probability of Autosub2 loss. For this analysis, faults were discriminated between low-impact faults and high-impact faults. Only high-impact faults were considered in the analysis as these were considered to lead to loss in the target operational environment. In their analysis the Kaplan–Meier estimator was used in its usual form [24.12, 44]

$$\hat{S}(r) = \prod_{r_i < r} \frac{n_i - d_i}{n_i} \,. \tag{24.5}$$

The survivor function $S(r)$ is defined as function of range r, where n_i is the number (of missions) at risk immediately prior to range r_i and d_i is the number of high-impact faults (or *deaths*) at range r_i. In the previous section, we have shown that the uncertainty over whether a fault can lead to loss should, and can, be captured through expert subjective judgments. Thus instead of having a number of *deaths* at time i, the statistical estimator must operate with the probability of loss given a fault in a given environment ($P(L \mid F_i, E)$). *Brito* et al. [24.10] derived an extended version of the Kaplan–Meier estimator to account for this uncertainty

$$\hat{S}(r) = \prod_{r_i < r} \left[1 - \frac{1}{n_i} P(L|F_i, E)\right] . \tag{24.6}$$

The expressions for the variance for this estimator were also derived

$$\hat{V}(r) = \frac{1}{\left[\log \hat{S}(r)\right]^2} \sum_{r_i < r} \frac{p_i}{n_i(n_i - p_i)} + \hat{V}_{EEJ_r} + 2\frac{1}{\left[\log \hat{S}(r)\right]^2} \sum_{r_i < r} \frac{p_i}{n_i(n_i - p_i)} \hat{V}_{EEJ_r} \,. \tag{24.7}$$

Where p_i is equal to $P(L \mid F_i, E)$, p_i is used above only to simplify the expression is the variance in the expert judgements.

This estimator was used for calculating the risk of losing Autosub3 under the Pine Island glacier in January–February 2009. Before the campaign, risk estimates based on prior history and the extended Kaplan–Meier approach helped devise a mitigation strategy based on a monitoring distance – reducing the risk to an acceptable level. In addition to a test mission prior to the start of the campaign, each mission beneath the glacier included an initial 5 h in open water where progress could be monitored [24.10]. The logic being that faults

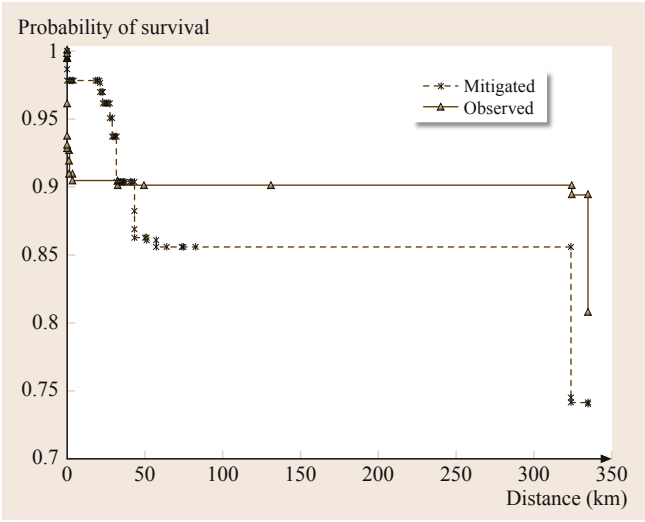


Fig. 24.6 Kaplan–Meier probability of survival against distance for the data set as considered by the experts with the pre-campaign assessment of individual fault mitigation included (*dashed* line). *Solid* line is the observed risk during an ISE Explorer Arctic campaign

emerging during this time could be discovered while the vehicle was in open water, and recovery effected. The appropriate monitoring period is determined from the initial reduction in probability of survival in the Kaplan–Meier plot (Fig. 24.6).

More recently, the same approach was used for estimating the probability of losing an ISE Explorer vehicle in the Arctic. The risk profile for the ISE Explorer under sea ice is presented in Fig. 24.6. Post-campaign a risk assessment exercise was carried out to calculate the observed risk. This profile is presented in full line in Fig. 24.6. Results show that following the mitigation, the maximum error between the predicted risk and the observed risk was 9% [24.14].

24.6 Legal Risks

The international rules for the law of the sea, as embedded in conventions such as the UN Convention on the Law of the Sea (UNCLOS) and the Convention on the International Regulations for Preventing Collisions at Sea (COLREGS) are clear with regard to the obligations of *ships* or *vessels*. These legal instruments were written and agreed before the use of AOVs became widespread in the academic, military and commercial arenas. It is not surprising therefore that it is not clear at all whether, in law, AOVs are indeed *ships* or *vessels*. The situation is further muddied by the possibility for a vehicle to be a *ship* for the purposes of one law, while not being a *ship* for another law. Consequently, the question, *Is an autonomous ocean vehicle a ship*? needs to be asked for each international Convention or national law. Only after the question has been answered each time can the legal issues, and the legal risks, within the scope of that Convention or law be fully understood. Unfortunately, as made clear in the exhaustive analysis in *Brown* and *Gaskell* [24.45] there is rarely a definitive answer to this question.

Given this situation, the vehicle owner or operator may face legal risks following incidents if their interpretation of laws or Conventions is found to be at variance with the opinions of courts. For example, if a court took a view that AOVs were vessels for the purposes of the Standards of Training, Certification and Watchkeeping (STCW) Convention 1995 their autonomous use on the surface would probably be proscribed [24.45]. It is not surprising therefore that action is being taken to prevent wholesale blocks on the use of unmanned systems in the ocean. These actions may be legal, procedural, or technical.

First, examining the legal context, *James Kraska*, a lawyer with the US Judge Advocate General's office, considers that the way forward is to recognize AOVs as *vessels* [24.46]. His opinion is that the *liberal legal architecture* of the key international conventions would allow for inclusion of autonomous systems. If he were proved correct then they would indeed gain benefit from being considered *vessels*. For example, the need to register an autonomous ocean vehicle as a vessel would provide for state protection. Rights in international law are given to *international persons*, such as states, and not to individual artifacts, such as an AOV. But a vehicle could benefit from those rights if it was clearly linked to a state, and hence the usual practice of registry under a particular state and flag. *Henderson* reinforced this point for unmanned underwater vehicles (UUV) [24.47]:

> *It is in the interest of the United States to establish the sovereignty of its UUVs to protect them from foreign seizure.*

Bork et al. [24.48] agree with Kraska that there is a need for action, and that a proper legal regime would give protection to autonomous (and free drifting) ocean vehicles, but they are not as sanguine that the existing framework of international maritime law can provide what is needed. They call for a new legal regime, but *it is not probable that one will soon come about* [24.48].

One aspect of international law that is being followed by many is that of seeking permits for use of AOVs in the exclusive economic zones of foreign nations, in accordance with UNCLOS Part XIII. *Showalter* gives an accessible summary of these issues [24.49]. Notable precedents are being set particularly relevant to the command and control of these vehicles. For example, the UK Foreign and Commonwealth Office has accepted that it is where the vehicle is piloted which determines which nation applies for diplomatic clearance from the foreign nation, not the nationality of the vessel making the deployment. Of course, had the vehicle the benefit of registration and a flag state in its own right, this question would be moot, but at least this acceptance is coherent with the outcome if indeed the vehicle did have a flag state.

Given strict interpretations of international maritime conventions, where autonomous vehicles would be considered as vessels, they would be hard to adhere to, consequently, many organizations and users are preferring self-regulation through collective codes of practice. An early attempt was the code of practice for AUVs produced by the AUV legal working group of the Society for Underwater Technology in 2000 [24.50]. With representation from military, indus-

try, and academic AUV users, the Code of Practice was a practical document that dealt with immediate issues of water-space management, environmental protection, operational considerations and training as well as vehicle marking and other matters. It was updated in 2008 with additional sections, among others, on deployment in congested waters, rules on salvage, protection of data, and damage to third parties, and broadened its coverage to AOVs [24.51]. A succinct code of practice specifically for commercial use of AUVs concentrating on safe operations, mission planning and launch and recovery has been produced by the International Marine Contractors Association [24.52]. Research supported by the European Defence Agency (EDA), through the Safety and Regulations for European Unmanned Maritime Systems (SARUMS) program, has led to an extensive draft best practice guide on handling, operations, design and regulations connected with unmanned maritime systems. Directed toward military use of these systems, the starting position in this paper is that these autonomous systems are vessels, and that *the main content of the COLREGS applies*. An earlier EDA study on maritime unmanned surface vehicles summarized *what additional rules and guidelines that should be developed*, which included adopting a classification scheme, sponsoring research into systems to enable the vehicles to comply with the COLREGS, and promoting cooperation with US agencies to examine whether these vehicles could become a specific class for the purposes of COLREGS.

There is active research into providing autonomous marine vehicles with both a rules-based approach to preventing collisions and with a common sense capability. *Benjamin* et al. examined a behavior-based control framework, and demonstrated the system in operation on an autonomous surface platform [24.53]. An alternative, path planning, approach has been investigated and demonstrated by *Naeem* et al. focusing on maneuvering and preventing head-on collisions [24.54].

There is a role for insurance in mitigating or limiting the financial consequences of legal risks, especially risks to third parties. The practical issues have been set out by *Edwards* [24.55], and *Griffiths* et al. provide case studies from scientific and industry users that have made insurance claims connected with AOVs [24.56].

References

24.1 J.G. Bellingham, K. Rajan: Robotics in remote and hostile environments, Science **318**(5853), 1098–1102 (2007)

24.2 J.A. Dowdeswell, J. Evans, R. Mugford, G. Griffiths, S.D. McPhail, N. Millard, P. Stevenson, M.A. Brandon, C. Banks, K.J. Heywood, M.R. Price, P.A. Dodd, A. Jenkins, K.W. Nicholls, D. Hayes, E.P. Abrahamsen, P.A. Tyler, B.J. Bett, D.O.B. Jones, P. Wadhams, J.P. Wilkinson, K. Stansfield, S. Ackley: Autonomous underwater vehicles (AUVs) and investigations of the ice-ocean interface: Deploying the autosub AUV in antarctic and arctic waters, J. Glaciol. **54**(187), 661–672 (2008)

24.3 A. Jenkins, P.D. Stanley, S. Jacobs, S.D. McPhail, J.R. Perrett, A.T. Webb, D. White: Observations beneath Pine Island Glacier in West Antarctica and implications for its retreat, Nat. Geosci. **3**, 468–472 (2010)

24.4 T.B. Curtin, D.M. Crimmins, J. Curcio, M. Benjamin, C. Roper: Autonomous underwater vehicles: Trends and transformations, Mar. Technol. Soc. J. **39**(3), 65–75 (2005)

24.5 D. Yoerger, A. Bradley, M. Jakuba, C.R. German, T. Shank, M. Tivey: Autonomous and remotely operated vehicle technology for hydrothermal vent discovery, exploration, and sampling, Oceanography **20**(1), 152–161 (2007)

24.6 M.V. Jakuba, C.N. Roman, H. Singh, C. Murphy, C. Kunz, C. Willis, T. Sato, R.A. Sohn: Long-baseline acoustic navigation for under-ice autonomous underwater vehicle operations, J. Field Robotics **25**(11/12), 861–879 (2008)

24.7 J.E. Strutt: *Report of the Inquiry Into the Loss of Autosub2 Under the Fimbulisen*, National Oceanography Centre Southampton Research and Consultancy Report, Vol. 12 (National Oceanography Centre, Southampton 2006)

24.8 L. Lippsett: R.I.P. A.B.E. – The pioneering autonomous benthic explorer is lost at sea, Oceanus **48**(1), 42 (2010)

24.9 G. Griffiths, A. Trembanis: Towards a risk management process for autonomous underwater vehicles, Proc. Mastercl. AUV Technol. Polar Sci., ed. by G. Griffiths, K. Collins (Society for Underwater Technology, London 2007) pp. 103–118

24.10 M.P. Brito, G. Griffiths, P. Challenor: Risk analysis for autonomous underwater vehicle operations in extreme environments, Risk Anal. **30**, 1771–1788 (2010)

24.11 R. Stokey, T. Austin, C. von Alt, M. Purcell, R. Goldsborough, N. Forrester, B. Allen: AUV bloopers or why Murphy must have been an optimist, Proc. 11th UUST (1999) pp. 32–40

24.12 G. Griffiths, N.W. Millard, S.D. McPhail, P. Stevenson, P.G. Challenor: On the reliability of the autosub autonomous underwater vehicle, Underw. Technol. **25**, 175–184 (2003)

24.13 K. Podder, M. Sibenac, H. Thomas, W. Kirkwood, J.G. Bellingham: Reliability growth of autonomous underwater vehicle – Dorado, Proc. IEEE OCEANS'04 (2004) pp. 856–862

24.14 M.P. Brito, G. Griffiths, J. Ferguson, D. Hopkin, R. Mills, R. Peterson, E.A. McNeil: A behavioural probabilistic risk assessment framework for man-

aging autonomous underwater vehicle deployments, J. Atmos. Ocean. Tech. **29**(11), 1689–1703 (2012)
24.15 P.D.T. O'Connor: *Practical Reliability Engineering* (Wiley, Chichester 1995)
24.16 G. Griffiths, N. Millard, S.D. McPhail, P. Stevenson, P. Challenor: *On the Reliability of the Autosub Autonomous Underwater Vehicle*, Southampton Oceanography Centre Research and Consultion by Report 48 (SOC, Southampton 2000)
24.17 G. Griffiths, N.W. Millard, S.D. McPhail, J. Riggs: Effect of upgrades on the reliability of the autosub AUV, Proc. 13th UUST (2003)
24.18 J.D. Andrews, T.R. Moss: *Reliability and Risk Assessment* (Longman, Harlow 1993)
24.19 G. Griffiths, M.P. Brito: Risk management for autonomous underwater vehicles operating under ice, Proc. OTC Arct. Technol. Conf., Houston, Texas (2011), OTC-22162-MS
24.20 M.V. Frank: Reentry safety: Probability of fuel release, Proc. Eur. Saf. Reliab. Conf., Rotterdam, 1999 (Taylor Francis, Rotterdam 1999)
24.21 L. Merckelbach: On the probability of underwater glider loss due to collision with a ship, J. Mar. Sci. Technol. **18**(1), 75–86 (2013)
24.22 D.P. Horner, A.J. Healey, S.P. Kragelund: AUV experiments in obstacle avoidance, Proc. MTS/IEEE OCEANS'05 (2005) pp. 1464–1470
24.23 S.D. McPhail, M. Furlong, M. Pebody: Low-altitude terrain following and collision avoidance in a flight-class autonomous underwater vehicle, Proc. Inst. Mech. Eng. M J. Eng. Marit, Environ. **224**(4), 279–292 (2010)
24.24 W. Stone, N. Fairfield, G.A. Kantor: Autonomous underwater vehicle navigation and proximity operations for deep phreatic thermal explorer (DEPTHX), Proc. Int. Mastercl. AUV Technol. Polar Sci., National Ocean. Cent., Southampton, ed. by G. Griffiths, K. Collins (Society for Underwater Technology, London 2007) pp. 127–142
24.25 M. Pebody: Autonomous underwater vehicle collision avoidance for under-ice exploration, Proc. Inst. Mech. Eng. M J. Eng. Marit, Environ. **222**(2), 53–66 (2008)
24.26 W. Feller: *An Introduction to Probability Theory and Its Applications*, Vol. 1 (Wiley, New York 1950)
24.27 M. Brito, G. Griffiths: A Markov chain state transition approach to establishing critical phases for AUV reliability, IEEE-JOE **36**(1), 139–149 (2011)
24.28 R.L. Keeney, D. Winterfeldt: Eliciting probabilities from experts in complex technical problems, IEEE Trans. Eng. Manag. **38**(3), 191–201 (1991)
24.29 A. Tversky, D. Kahnemann: Judgment under uncertainty: Heuristics and biases, Science **185**(4157), 1124–1131 (1974)
24.30 D. Kahnemann, A. Tversky: Subjective probability: A judgment of representativeness, Cogn. Psychol. **3**(3), 430–454 (1972)
24.31 M.L. Finucane, A. Alhakami, P. Slovic, S.M. Johnson: The effect of heuristic in judgments of risk and benefits, J. Behav. Decis. Mak. **13**(1), 1–17 (2000)
24.32 M. Kynn: The 'heuristics and biases' bias in expert elicitation, J. Royal Stat. Soc. A **171**(1), 239–264 (2008)
24.33 R.M. Cooke, L.H.J. Goossens: Expert judgement elicitation for risk assessments on critical infrastructures, J. Risk Res. **7**(6), 643–656 (2004)
24.34 D.J. Isenberg: Group polarization: A critical review and meta-analysis, J. Pers. Soc. Psychol. **50**(6), 1141–1151 (1986)
24.35 P.A. Morris: An axiomatic approach to expert resolution, Manag. Sci. **29**(1), 24–32 (1983)
24.36 R.T. Clemen, R.L. Winkler: Combining probability distributions from experts in risk analysis, Risk Anal. **19**(2), 187–203 (1999)
24.37 K.J. McConway: Marginalisation and linear opinion pools, J. Am. Stat. Assoc. **76**, 410–414 (1981)
24.38 M.A. Moline, S.M. Blackwell, C. Von Alt, B. Allen, T. Austin, J.F. Case, N. Forrester, R. Goldsborough, M. Purcell, R. Stokey: Remote environmental monitoring units: An autonomous vehicle for characterizing coastal environments, J. Atmos. Ocean Tech. **22**, 1797–1808 (2005)
24.39 G. Griffiths, M. Brito, I. Robbins, M. Moline: Reliability of two REMUS-100 AUVs based on fault log analysis and elicited expert judgment, Proc. Int. Symp. Unmanned Untethered Submers. Technol., Durham, New Hampshire (2008)
24.40 A. O'Hagan, C.E. Buck, A. Daneshkhah, J.R. Eiser, P.H. Garthwaite, D.J. Jenkinson, J.E. Oakley, T. Rakow: *Uncertain Judgments: Eliciting Experts' Probabilities* (Wiley, Chichester 2003)
24.41 L.D. Phillips, M.C. Phillips: Facilitated work groups: Theory and practice, J. Oper. Res. Soc. **44**, 533–549 (1993)
24.42 T. Crees, C. Kaminski, J. Ferguson, J.M. Laframboise, A. Forrest, J. Williams, E. MacNeil, D. Hopkin, R. Pederson: UNCLOS under ice survey – An historic AUV deployment in the Canadian high arctic, Proc. IEEE OCEANS'10 (2010) pp. 1–8
24.43 R.L. Prentice, J.D. Kalbfleisch: *The Statistical Analysis of Failure Time Data* (Wiley, New York 2002)
24.44 E.L. Kaplan, P. Meier: Nonparametric estimation from incomplete observations, J. Am. Stat. Assoc. **53**(282), 457–481 (1958)
24.45 E.D. Brown, N.J.J. Gaskell: *Questions and Answers on the Law, State Practice and Procedure Relating to Autonomous Underwater Vehicles* (Society for Underwater Technology, London 2001)
24.46 J. Kraska: The law of unmanned naval systems in war and peace, J. Ocean Tech. **5**, 43–68 (2010)
24.47 A.H. Henderson: Murky waters: The legal status of unmanned undersea vehicles, Naval L. Rev. **55**, 1–17 (2006)
24.48 K. Bork, J. Karstensen, M. Visbeck, A. Zimmermann: The legal regulation of floats and gliders: In quest of a new regime?, Ocean Dev. Int. Law **39**, 298–328 (2003)
24.49 S. Showalter: The legal status of autonomous underwater vehicles, Mar. Technol. Soc. J. **38**, 80–83 (2004)
24.50 SUT: *The Operation of Autonomous Underwater Vehicles – Recommended Code of Practice*, Vol. 1

(Society for Underwater Technology, London 2001)
24.51 SUT: *Recommended Code of Practice for the Operation of Autonomous Marine Vehicles*, 2nd edn. (Society for Underwater Technology, London 2009)
24.52 IMCA: *Considerations for the Safe Operation of Autonomous Underwater Vehicles (AUVs)* (International Marine Contractors Association, London 2009)
24.53 M.R. Benjamin, J.J. Leonard, J.A. Curcio, P.M. Newman: A method for protocol-based collision avoidance between autonomous marine surface craft, J. Field Robotics **23**, 333–346 (2006)
24.54 W. Naeem, G.W. Irwin, A. Yang: COLREGs-based collision avoidance strategies for unmanned surface vehicles, Mechatronics **22**(6), 669–678 (2012)
24.55 S.R.P. Edwards: AUVs. How do you insure them?, Int. Ocean Sys. Des. **4**, 10–12 (2000)
24.56 G. Griffiths, N. Bose, J. Ferguson, R. Blidberg: Insurance for autonomous underwater vehicles, Underw. Technol. **27**, 43–48 (2007)

Part C Coastal Design

Part C Coastal Design

Ed. by Jennifer L. Irish, James M. Kaihatu

25 **Physical Characteristics of Coastal Hazards**
Jennifer L. Irish, Blacksburg, USA
Robert Weiss, Blacksburg, USA
Donald T. Resio, Jacksonville, USA

26 **Statistical Characterization of Hazards and Risk in Coastal Areas**
Donald T. Resio, Jacksonville, USA
Mark A. Tumeo, Jacksonville, USA
Jennifer L. Irish, Blacksburg, USA

27 **Modeling of Coastal Waves and Hydrodynamics**
Patrick J. Lynett, Los Angeles, USA
James M. Kaihatu, College Station, USA

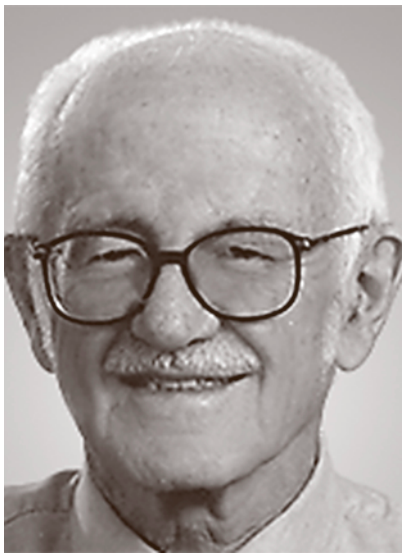
28 **Modeling of Coastal Morphological Processes**
J.A. Dano Roelvink, Delft, Netherlands
Dirk-Jan R. Walstra, Delft, Netherlands
Mick van der Wegen, Delft, Netherlands
Roshanka Ranasinghe, Delft, Netherlands

29 **Beach Nourishment**
Robert G. Dean, Gainesville, USA
Thomas J. Campbell, Boca Raton, USA

30 **Storm Hazard Mitigation Structures**
David R. Basco, Norfolk, USA

31 **Port and Harbor Design**
Andrew Cairns, New York, USA
John M. Carel, Estero, USA
Xiao Li, Livingston, USA

32 **Marine Outfalls**
Peter M. Tate, Parramatta, Australia
Salvatore Scaturro, Sydney, Australia
Bruce Cathers, Manly Vale, Australia



Dr Robert G. Dean (1930–2015)

Dedication. Part C Coastal Design is dedicated to Dr Robert G. Dean (1930–2015), a world leader in coastal engineering – the practice of civil engineering at the coast – and Professor Emeritus at the University of Florida. During Dr Dean's long and distinguished career, he made important contributions in multiple facets of coastal engineering. These theoretical and practical contributions inform ongoing coastal engineering practices and appear throughout the chapters in Part C, as well as in some of the foundational chapters in Part A. Dr Dean's early work was in water wave mechanics, with an emphasis on the novel use of stream function theory to facilitate computational analysis of wave mechanics problems. Dean and others successfully applied stream function theory to a range of coastal engineering problems spanning from wave propagation on a shear current to wave damping by emergent vegetation. Dr Dean and colleagues also made important advances in our understanding of depth-limited wave breaking. His textbook with Dr Robert Dalrymple is one of the most widely used books on water wave mechanics.

Dr Dean may be most widely known for his work in coastal processes and beach nourishment design, with the latter being the subject of his Chap. 29 in Part C. Dr Dean pioneered the use of equilibrium profile concepts in understanding shoreline change and in the design of beach nourishment. After Dr Per Bruun first introduced equilibrium beach profiles in 1954, Dr Dean validated and popularized the use of the well-known two-thirds power law for equilibrium beach profiles for coastal engineering applications. Throughout his career, Dr Dean applied equilibrium profile concepts to advance current understanding of beach processes.

Since the 1970s, beach nourishment has gained popularity as an economical, aesthetically pleasing, and natural alternative to traditional hard-structure-based erosion control measures. Dr Dean played a major role in establishing beach nourishment design criteria over this period (Chap. 29). His work in this area spanned selecting beach fill sediments, estimating beach fill design volumes using equilibrium profile concepts, and assessing project performance.

The broad impact of Dr Dean's accomplishments and subsequent implementation in the field of coastal engineering are evidenced by the varied collaborations he engaged in over the last five years. Dr Dean and his colleagues reported new insight on storm surge dynamics, river diversions, benthic water flux induced by waves, and levee erosion by wave overtopping. His recent controversial work on sea-level rise with Dr James Houston raised awareness of the large uncertainty in projecting future sea-level change.

Part C Coastal Design is organized such that the first four chapters emphasize present-day knowledge and practices for quantifying the physical coastal environment, while the last four chapters discuss contemporary practices within four specific branches of coastal engineering design. Dr Dean's contributions appear in five of these eight chapters.

Chapter 25, Physical Characteristics of Coastal Hazards, describes the physical features of hurricane surge, tsunamis, and other severe coastal phenomena. Included are the impacts of climate change, sea-level rise, and anthropogenic effects on these hazards. The current understanding of these coastal phenomena is detailed. Straightforward methods for initial interpretation of these hazards are described. A historical perspective is also used, so the evolution of the understanding of these phenomena and the corresponding human response to them are evident.

Chapter 26, Statistical Characterization of Hazards and Risk in Coastal Areas, outlines the probabilistic nature of coastal hazards and details the implementation of this in coastal design. The impacts of combined forcing mechanisms, such as storm surge combined with rainfall, on the evaluation of the risk of flooding are discussed. As with the previous chapter, a historical perspective is given to the discussion of risk evaluation, which illuminates the transition from

early deterministic methods to more recent probabilistic approaches.

Chapter 27, Modeling of Coastal Waves and Hydrodynamics, covers the details of numerical models of wave propagation. Both phase-averaged (spectral) models and phase-resolved (Boussinesq time-domain, frequency-domain) models are covered, and an indication of the typical domains of applicability, and associated model computation times, are discussed. The appropriate modeling strategies for wind waves, swell, and long waves (tsunamis) are detailed. The use of Navier–Stokes type models for breaking waves and turbulence modeling is introduced, along with strategies to couple with time-domain wave models.

Chapter 28, Modeling of Coastal Morphological Processes, addresses the various types of models used to simulate the evolution of the shoreline and beach face due to waves and nearshore currents. Simple analytical and semi-empirical models are introduced first. The drawback of such models is the tendency toward an equilibrium solution that may not be realistic. Process-based models are then discussed, which eliminate any consideration of an equilibrium solution and allow the beach to evolve freely according to the forcing and resultant feedback. Details of these processed-based models are described, and their own development toward equilibrium states discussed. Case studies showing the application of these models are outlined.

Chapter 29, Beach Nourishment, describes the design, construction and analysis of beach fills, and shoreline stabilization. Design considerations for successful beach nourishment projects are discussed, with an emphasis on the quality of the sand and its effect on the success of the nourishment. The effects of coastal structures on beach nourishment are described. Analytical solutions for shoreline evolution are outlined as possible design tools for relatively quick analysis, with numerical models of shoreline evolution used for more detailed studies. Legacy beach nourishment projects are covered, and the reasons for their successful performance detailed.

Chapter 30, Storm Hazard Mitigation Structures, covers design issues regarding coastal structures and their implementation as part of a coastal system. Design conditions (waves, water levels) and their translation to forces on structures are described. Methods for quantifying wave runup and overtopping of structures are discussed. Vertical caisson structures and rubble mound structures are covered; the discussion of rubble mound structures also includes armor layer design, concrete armor unit design, and cross-section design. Life-cycle cost analysis and risk are also discussed. In addition, structures to enhance shoreline stabilization are addressed.

Chapter 31, Overview of Port and Harbor Design, describes the various components of port and harbor layout and their respective designs. There is an emphasis on the anticipated marine traffic and associated accommodation requirements in the design discussion. Types of port and harbor structures and their design criteria are described, including some aspects of their construction. Various kinds of piers and wharves, and the advantages and disadvantages of each, are outlined. Mooring and berthing operations are also described, including evaluation of the load on piers due to berthing and the associated fender design.

Chapter 32, Marine Outfalls, covers the design, analysis and environmental monitoring of outfalls used to discharge treated wastewater into the marine environment. Various drivers for marine outfall design are covered, including social, political, and health concerns. Treatment criteria for wastewater prior to discharge are outlined. Numerical modeling considerations for the outfall operation are discussed, and the impact of the results on the design outlined. The hydraulic characteristics of outfalls are addressed, including nozzle velocity, head losses, and anticipated mixing. Various configurations of outfalls are described and their impact on both treatment plant design and discharge performance addressed. Outfall construction and environmental monitoring techniques are also detailed.

25. Physical Characteristics of Coastal Hazards

Jennifer L. Irish, Robert Weiss, Donald T. Resio

Coastal hazards are among the world's most threatening hazards. With half of the world's population living near the coast, an immense threat is posed by these hazards to life and health, livelihood, and the economy. In this chapter, characteristics of coastal hazards are described, first by presenting a historical context, then by laying out our current understanding of the physical problems of interest for coastal design, and finally by presenting simple methods for initial interpretation of the hazard. A number of natural processes and anthropogenic activities adversely impact coastal communities, critical infrastructure, and port facilities along the coast. Natural coastal hazards may generally be classified into those that are episodic and those that occur over long periods of time. Episodic hazards include coastal storms and tsunamis while long-term hazards include, for example, sea-level rise. Many anthropogenic activities also result in long-term impacts at the coast. For example, inland dam construction produces long-term impacts on coastal sediment supply. In the following, we describe the most common coastal hazards and then discuss methods for evaluating their impacts at the coast.

25.1 **Types of Coastal Hazards** 549
25.1.1 Coastal Storms 549
25.1.2 Tsunamis 551
25.1.3 Climate Variability, Climate Change, and Sea-Level Rise 553
25.1.4 Anthropogenic Activities 554
25.2 **Coastal Impacts** 555
25.2.1 Winds 555
25.2.2 Wind-Generated Waves 556
25.2.3 Flood Levels and Runup 557
25.2.4 Environmental Impacts 560
25.3 **Summary** 561
25.4 **Nomenclature** 561
References 562

25.1 Types of Coastal Hazards

25.1.1 Coastal Storms

Coastal storms are categorized either as tropical cyclones, for example, hurricanes, typhoons, or as extratropical cyclones, for example, Nor'easters in the United States. While both storm types generate high winds, surges, waves, and precipitation, there are several key differences between the two.

Tropical Cyclones

Tropical cyclones, which by definition form in the tropics, are generally well organized and compact. Thus, tropical cyclones are characterized by moderately strong to severe winds with the capacity to generate very large waves and surges. Because they are compact relative to other storms, the coastal region impacted by high winds and large surges, and to a lesser extent by large waves, is typically confined to a distance of the order of hundreds of kilometers. These storms also tend to pass fairly quickly across the continental shelf and coastal region, meaning that the period of high winds and large surges is generally of the order of hours, while large waves may persist for several days.

Historically, tropical cyclones have taken hundreds of thousands of lives and caused extensive destruction. The National Oceanic and Atmospheric Administration's (NOAA) local forecast official *Isaac Cline* (1900) commented on his firsthand experience during the Galveston Hurricane of 1900 [25.1]:

> *Sunday, September 9, 1900, revealed one of the most horrible sights that ever a civilized people looked upon. About three thousand homes, nearly half the residence portion of Galveston, had been completely swept out of existence, and probably more than six thousand persons had passed from life to death during that dreadful night. The correct number of those who perished will probably never be known, for many entire families are missing. Where* 20 000 *people lived on the 8th not a house*

remained on the 9th, and who occupied the houses may, in many instances, never be known.

The 1900 hurricane, which made landfall in Galveston, TX, USA, remains the most deadly natural disaster on record in the United States. Some of the largest death tolls associated with tropical cyclones have been those passing through the Bay of Bengal and impacting India and Bangladesh. The 1970 tropical cyclone, which made landfall near high tide, generated flood levels in excess of 6 m and claimed as many as a half million lives along the Bay of Bengal [25.2–4]. This area was impacted again in 1991, when another tropical cyclone with winds exceeding 130 km/h generated similar flood elevations, claiming over 100 000 lives and causing more than 1.5 billion US dollars in damages [25.4].

The United States also has a history of experiencing extensive damages as high winds, flooding, and large waves impact the coast during tropical cyclones. Over the last decade, the US has seen some of its highest storm surges on record as several significant hurricanes made landfall along the northern Gulf of Mexico coastline. Hurricane Katrina made landfall near New Orleans in 2005, generating flood elevations of more than 8 m along the Louisiana and Mississippi coasts [25.5], and became the most destructive hurricane in US history. Estimated economic damages from this event are in excess of 100 billion US Dollars [25.6]. The death toll caused by Hurricane Katrina was just under 2000, the third most deadly hurricane in US history [25.7]. Just three years later, Hurricane Ike made landfall in Galveston, TX, USA. While a moderate hurricane wind event, registering as a Category 2 hurricane at landfall [25.8], this very large storm generated flood elevations in excess of 4 m along much of the Texas and Louisiana coasts [25.9]. This storm became the second most costly storm in US history, with damages estimated to be around 30 billion US Dollars. Both of these events dramatically reshaped the natural and built landscape. Hurricane Katrina severely damaged the Chandeleur Islands and coastal wetlands of Louisiana (Fig. 25.1) while Hurricane Ike reshaped Texas's Bolivar Peninsula (Fig. 25.2). Most recently, Hurricane Sandy in 2012 wreaked havoc along the US eastern seaboard, causing widespread coastal flooding and damaging densely populated New York City and surrounding areas; while recovery is still ongoing, damages from this storm are currently expected to total 75 billion US dollars.

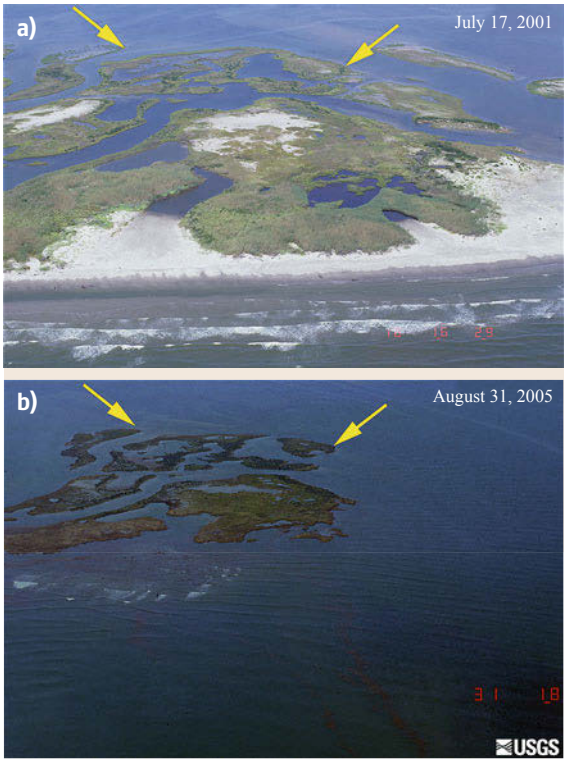


Fig. 25.1a,b Aerial imagery of a portion of the Chandeleur Islands, Louisiana, USA taken in 2001 **(a)** and shortly after Hurricane Katrina made landfall in 2005 **(b)**. The significant surge and waves associated with this tropical cyclone resulted in significant changes to the islands. Photos courtesy of the US Geological Survey

Extratropical Cyclones

In contrast to tropical cyclones, extratropical cyclones are generally less well organized and usually span a large region. Because these storms tend to be much larger than tropical cyclones, the winds within them tend to be weaker overall, and associated surge magnitudes are generally smaller. Yet, since extratropical cyclones are large and oftentimes relatively move slower than tropical cyclones, the duration and extent of high winds and of large surges and waves are much greater. It is not uncommon for adverse conditions from a single extratropical cyclone to persist over several days to a week and for coastal shoreline distances of the order of thousands of kilometers to be impacted.

An example of a historic extratropical storm occurred in 1953, when a large extratropical depression moved across the North Sea. This storm caused widespread damages across The Netherlands and The United Kingdom, with hurricane-force winds and flood elevations exceeding 2 m above normal levels in many locations [25.10]. It is reported that nearly 1800 people died in the floods associated with overtopping and failure of many of the Dutch sea dikes, which protected communities below sea level [25.11].

Fig. 25.2a,b Aerial imagery of a portion of the Bolivar Peninsula, Texas, USA taken in 2008, shortly before **(a)** and shortly after **(b)** Hurricane Ike made landfall. The significant surge and waves associated with this tropical cyclone resulted in significant erosion across the peninsula. Photos courtesy of the US Geological Survey

In summary, coastal storms are among the most deadly and costly of coastal hazards. Understanding the conditions that lead to, as well as the likelihood of, high winds, flooding, and waves is essential for effective coastal design.

25.1.2 Tsunamis

Unlike storms, tsunamis occur unpredictably, and the response time to an event might be very short. Therefore, tsunami resilience strongly depends on the quick evaluation of the tsunami characteristics as soon as possible after the generation and on educating the population at risk to increase tsunami awareness and preparedness.

Earthquake-Generated Tsunamis

Earthquakes are the most common source of tsunami waves. The characteristics of earthquake-generated tsunami waves depend on the area and magnitude of the deformed seafloor, which is linked to the amount of energy released and depths of the focus underneath the seafloor. Because the magnitude of the vertical seafloor displacement is small compared to the water depth, and the area of displacement is large compared to the water depth, the generated tsunami waves are well described by linear wave theory while propagating in deep waters. This also means that the tsunami will not lose a significant amount of energy during the propagation in deep water and will bring its potentially destructive energy to coastlines hundreds of kilometers away. Tsunamis from very large earthquakes can even have a global reach as demonstrated by *Titov* et al. [25.12] for the 2004 Sumatra tsunami. The National Geographic (2005) reported on the 2004 tsunami [25.13]:

> *Giant forces that had been building up deep in the Earth for hundreds of years were released suddenly on December 26, shaking the ground violently and unleashing a series of killer waves that sped across the Indian Ocean at the speed of a jet airliner. By the end of the day more than* 150 000 *people were dead or missing and millions more were homeless in 11 countries, making it perhaps the most destructive tsunami in history.*

Throughout history, underwater earthquakes have caused widespread destruction not only by the tsunamis generated locally at the earthquake's epicenter, but also by causing landslides which in turn generate more localized tsunamis.

Landslide-Generated Tsunamis

Landslides can also generate tsunamis. Whether or not submarine and subaerial landslides generate tsunamis depends on the slope where the landslide occurs and the volume that is triggered to move. Subaerial landslides are linked to very steep slopes, such as in Alaska for Lituya Bay. The fjords in Norway are also very steep, and settlements and harbors are located along the shorelines [25.14–16]. Because the fjords generally become narrower going inland from the fjord mouth, waves will become larger and larger, posing a significant and increasing threat to low-lying communities (see [25.14] and references therein). For submarine landslides, slopes are usually much more gentle. Submarine landslides can occur on the continental shelf, but also at trenches of subduction zones. The initial waves from submarine landslides can be large, but the relatively short period causes the waves to dissipate while traveling toward the shoreline. Where the distance from the trigger area to the shoreline is short, tsunami waves can be very lethal. An example is the 1998 Papua New Guinea tsunami. An earthquake of about Mw 7 [25.17,

18] triggered a submarine landslide that then caused the death of 3000 people along a 30 km stretch of the Papua New Guinea coast.

The 1958 Lituya Bay, Alaska tsunami, a landslide-generated tsunami, produced a maximum runup of about 524 m, and is the largest tsunami runup ever measured. In the case of Lituya Bay, the rockslide started above the water column, accelerated, and hit the water surface with a velocity of about 110 m s^{-1} [25.19]. In a scaled model of the Lituya Bay rockslide and tsunami, *Fritz* et al. [25.19] demonstrated that Froude number scaling is applicable for the slide problem. The Froude number for this problem is defined as the ratio between the slide velocity at impact and $\sqrt{gh}$, where h is the average water depth in the vicinity of the impact location. The Lituya Bay rockslide exhibited a large Froude number. The impact of the slide body on the surface generates a cavity whose collapse generates a wave pulse that reached about 170 m in the experiments in *Fritz* et al. [25.19] and in simulations by *Weiss* et al. [25.20]. For slides featuring a small Froude number, the generated waves are much smaller with wave amplitudes of the order of the slide thickness.

Submarine slides generate very different waves. The wave characteristics of the generated tsunamis depend on the initial slide volume and its initial acceleration. In general, tsunami waves generated by landslides are steeper than those generated during earthquakes, but exhibit a smaller steepness than waves generated by high-Froude number subaerial landslides. The steeper the waves, the more effectively nonlinear processes work to dissipate and shift energy from longer to shorter frequencies. While early work on landslide-generated tsunamis suggested that a transoceanic propagation is possible [25.21], it is now believed that the extent of landslide-generated tsunamis are limited to locations near the source, and as such only have a regional effect. Furthermore, *Raichlen* and *Synolakis* [25.22] carried out experiments, which explored the influence of submergence (distance of the slide top to the water surface) and slide geometry on the tsunami runup. The slide body did not deform, but it could be shown that for slides triggered in shallower waters, the runup is more influenced by the shape of the slide than for slides that start in deeper water. Again these experiments were performed with a solid slide body. It can be expected that deformable landslides respond differently (for a preliminary discussion, see *Weiss* et al. [25.23]).

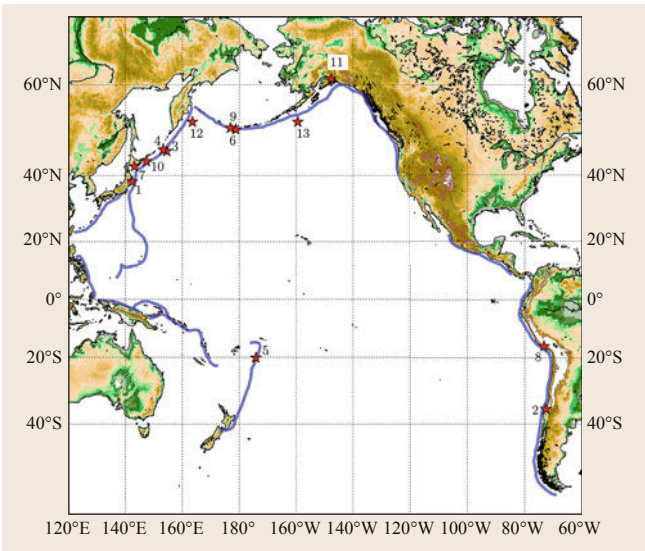


Fig. 25.3 Map of the Pacific Ocean. The *lines* represent the locations of subduction zones where large tsunamigenic earthquakes can occur. The *stars* represent the locations of earthquakes that generated significant tsunamis: 1 – Japan (2011), 2 – Chile (2010), 3 – Kuril Islands (2007), 4 – Central Kuril Island (2006), 5 – Tonga (2006), 6 – Rat Island (2003), 7 – Hokkaido (2003), 8 – Peru (2001), 9 – Anderanof (1996), 10 – Kuril Islands (1994), 11 – Alaska (1964), 12 – Kamchatka (1952), 13 – Unimak (1946)

Volcanic Eruptions and Meteorite Impacts

Tsunamis from volcanic eruptions and meteorite impacts are very rare; however, both events have the potential to generate very large initial waves. While generation of tsunamis from volcanic eruptions often reduces to a slide problem, other processes during eruptions, such as explosions and a caldera collapse are not very well understood. A larger body of knowledge exists for processes that lead to the generation of tsunamis during meteorite impacts into oceans. Understanding of impact-generated tsunamis requires numerical simulation techniques in order to provide the necessary flexibility for the multiphysics processes that govern meteorite impacts to be simulated adequately. For more details, see *Weiss* and *Wuennemann* [25.24] for wave generation and *Wuennemann* et al. [25.25] for wave attenuation as a function of meteorite size and water depth in the impact area.

Meteorite impacts in oceans can occur anywhere. Their hazard is large because they generate very large waves; therefore, any low-lying coastal area may face a significant threat. However, the likelihood of a meteorite impact to happen is very low. On the other hand, earthquakes, landslides, and explosive volcanic eruptions are more common and are linked to certain geographic areas. Strong earthquakes and explosive volcanic eruptions are directly linked to the subduction processes of the oceanic crust underneath the continental crust. The Pacific Ocean is surrounded by sub-

duction zones. Figure 25.3 depicts the locations of the earthquakes that generated significant tsunamis; also marked in Fig. 25.3 are the locations of island arcs that feature explosive volcanism.

The eruptions of Krakatau between 26th and 27th of August 1883 generated tsunamis in the Indian Ocean during different stages of the eruption (explosion, caldera collapse, etc. [25.26]). While the main explosion and caldera collapse are direct ways to generate tsunamis, oftentimes associated flank collapse either during an eruption or later, such as predicted for Cumbre Vieja Volcano (La Palma, Canary Islands, [25.27]), can be approached by considering the slide problem.

Tsunami Source Frequency

Earthquakes occur frequently and have been recorded for about 100 years. The distribution of earthquakes obeys a power-law distribution, indicating that smaller earthquakes occur more frequently than larger ones. A similar power-law distribution can be inferred from meteorite impact craters on the Moon and other planets [25.28, 29]. Like for earthquakes, the size distribution influences respective hazard assessment in that smaller impacts are more frequent than larger ones. Unfortunately, there is no direct measurement and specific distribution of submarine landslides. The main problem with defining appropriate frequency distributions originates from the fact that submarine landslides can be triggered not only by earthquakes, but also by sea-level rise, excess pore pressure, weak layers, and tectonic over-steepening of the local slope.

25.1.3 Climate Variability, Climate Change, and Sea-Level Rise

A changing climate impacts conditions at the coast via two main processes:

1. Changes in weather patterns, such as location and severity, and
2. Changes in the mean sea level.

Additionally, climate varies on multiple scales. We will only consider long-term changes; however, it should be noted that seasonal and decadal effects are important, for example, for determining annual periods of storm activity, high wave energy, etc. Consequently, even though our focus here is on secular trends, we will include some discussion of the impact of specific cycles.

Sea-Level Rise

Long-term changes in the climate are often associated with global trends. Recent literature supports a global warming trend [25.30], which in turn has implications on processes at the coast. Eustatic, or global sea-level rise (SLR), arises from the thermal expansion of the ocean and ice-sheet melting. Observations of eustatic SLR in response to global warming indicate an accelerating SLR trend [25.30–32]. Observations made in recent years indicate that eustatic SLR has accelerated at a rate of 3 mm/yr, with respect to an average rate of 1.7 mm/yr over the last century [25.30]. The Intergovernmental Panel on Climate Change (IPCC) [25.30], by considering a range of future climate scenarios, from low to high carbon emission rates resulting in a 1.1–6.4 °C sea-surface-temperature (SST) rise over the next century, estimated that eustatic SLR is likely to continue accelerating, reaching rates as high as 6.0 mm/yr when major ice-sheet melting is not considered. Literature published since the 2007 IPCC report, which considers major ice-sheet melting, largely supports a substantially larger upper limit on eustatic SLR projections, up to 20 mm/yr over the next century [25.33–36]. Currently, projections of eustatic SLR over the next century range from a low of 1.8 to a high of 20 mm/yr.

Regional influences also contribute to relative, or local, SLR at a particular location. These regional impacts are due to processes including subsidence and tectonic activity. The Permanent Service for Mean Sea Level reports relative sea-level trends for locations worldwide (Table 25.1), and shows that in many locations relative SLR rates exceed the eustatic rate. For example, the Virginia and Louisiana coasts of the United States exhibit 30-year relative SLR rates that are 30–300% larger than the eustatic rate; the Shanghai, China region exhibits a 30-year relative SLR rate that is also about 300% larger than the eustatic rate.

In addition to the obvious impact of raising the mean water level, SLR causes shoreline retreat and land-cover changes, allows for larger waves to propagate farther inland, alters tide levels and currents, alters storm surge magnitudes and as noted in the previous section induces landslide-generated tsunamis [25.37–41].

Storm Climate

Global warming trends have also been shown to impact coastal storm activity, with the potential to alter both storm intensity and frequency, leading to changes in wave climate and coastal flooding. In response to the SST rise, the coastal storm intensity and rate of occurrence is expected to change over the next century. A number of recent studies support an increase in tropical cyclone intensity with SST rise [25.43–47], where wind intensity is expected to increase by 2–11% over

Table 25.1 Observed sea-level rise rates at selected locations (adapted from [25.42])

Location	Time period	Sea-level rise rate [mm/yr]
Dublin, Ireland	1938–2001	0.21
Aberdeen, Great Britain	1932–2010	1.01
Helsinki, Finland	1900–2010	−2.08
Oslo, Norway	1916–2010	−3.75
Stockhold, Sweden	1900–2010	−3.80
Kiel, Germany	1956–2010	1.45
Marseille, France	1900–2009	1.19
Venice, Italy	1909–2000	2.40
Karachi, Pakistan	1937–1986	0.41
Mumbai, India	1900–2006	0.83
Gangra, India	1974–2006	1.19
Quinhon, Vietnam	1977–2006	−1.27
Hong Kong, China	1963–2010	2.71
Shanghai, China	1961–2010	5.01
Pusaon, Korea	1961–2009	2.03
Tokyo, Japan	1930–2010	3.60
Nagasaki, Japan	1965–2010	2.19
Singapore	1954–2010	−0.91
Brisbane, Australia	1966–2010	0.21
Sydney, Australia	1915–2010	0.90
Auckland, New Zealand	1904–1998	1.26
Victoria, Canada	1910–2010	0.61
Quebec, Canada	1911–2010	−0.38
Los Angeles, USA	1924–2010	0.86
San Francisco, USA	1900–2010	1.91
New York, USA	1900–2010	3.03
Washington DC, USA	1931–2010	2.99
Juneau, USA	1936–2010	−13.05
Houston, TX	1958–2010	6.73
Buenos Aires, Argentina	1905–1987	1.54
Montevideo, Uruguay	1954–2009	0.92
La Libertad, Ecuador	1950–2002	−1.31
Balbao, Panama	1908–2006	1.64

the next century [25.48]. The impact of SST on future tropical cyclone rate-of-occurrence is not as well understood. Various analyses of the historical record reached conclusions ranging from no significant change [25.47] to an increase [25.49, 50] in tropical-cyclone activity with the SST rise. *Bender* et al. [25.51] summarized that for a mid-range carbon emissions scenario, climate models showed a decrease in Atlantic tropical cyclone activity, of −4 to −49%, by the 2100s, while most models showed an change for the most intense cyclones, between −53 and +110%.

A future increase in storm intensity means that, when storms do occur, storm wave and surge heights will likely be larger [25.41, 52–58]. Thus, we can expect future coastal storm impacts, including inundation elevation, and extent, wave and wind loading, erosion, and damages to be more severe [25.59–63].

25.1.4 Anthropogenic Activities

Coastal urbanization can lead to unintended, unmitigated increases in coastal hazard risk, exacerbating the need for improved coastal planning to mitigate such hazard impacts. Consider, for example, the low-lying and flood-prone New Orleans region in the United States. To stabilize the Mississippi River mouth, an important navigation waterway connecting the port of New Orleans with much of the United States, and to manage coastal flooding, a complex system of levees and polders was designed and constructed throughout the region over the last century. Recent studies show that under certain storm conditions, levees used to control the river may enhance and trap storm surge, thereby increasing flood-related damages [25.64]. As another example of unintended consequences in this area, the population of New Orleans grew more than 60% during the 20th century [25.65], with many residents now settling in areas behind levees, at elevations below the sea level. During Hurricane Katrina (2005), several levees breached, leading to widespread flooding of these populated, low-lying areas. While the devastating impacts of levee failure during Hurricane Katrina can be attributed to a combination of factors, including design load exceedance, this event made clear the need to better plan urban development to avoid urban growth in vulnerable locations.

Another example of the impact of development at the coast is the unintended consequences of construction of inland dams, mining of sediments from inland rivers and waterways, and urbanization that hardens the landscape. All of these processes impede the natural movement of sediments to the coast [25.66], resulting in a depletion of the natural sediment supply and contributing to widespread shoreline recession. Similarly, coastal erosion and navigation measures put in place without considering that the downdrift impacts often lead to unnecessary shoreline erosion. Recently, regional sediment management techniques have been used to manage coastal and inland sediment zones holistically [25.67].

There is also the danger of a major industrial accident, which leads to adverse environmental impacts at the coast. Recently, the Deepwater Horizon offshore platform explosion in 2010 released a 35−km-long plume of oil into the Gulf of Mexico [25.68]. Oil made its way to the coast, damaging the coastal ecosystem. This event directly impacted both tourism and commercial fishing in the region. The proliferation of oil refineries, nuclear power plants, chemical manufacturing plants, and other such major industrial developments in coastal areas must be carefully considered within this fragile and dynamic environment.

In summary, existing development and future anthropogenic activities may themselves aggravate coastal flooding and erosion. Combined with other natural threats, including sea-level rise, it is prudent consider long-term and regional impacts during all phases of coastal development, planning, and design.

25.2 Coastal Impacts

25.2.1 Winds

Winds associated with both tropical and extratropical cyclones must be considered during coastal development. High winds, which vary temporally and spatially, can damage residential and commercial infrastructure and can lead to route closures during evacuation. From a coastal engineering perspective, an understanding of storm winds is also relevant as they are the primary forcing mechanism for wave and surge generation. Field observations via satellite, point measurement, and real-time aircraft-based measurement transects have been used recently to successfully reconstruct cyclone wind fields [25.69, 70] and to provide real-time information on wind speeds. For cyclonic winds, the momentum balance yields the following equation for wind speed (U) in a moving reference frame relative to the cyclone's center

$$U = \frac{1}{\rho_a f}\frac{\partial p}{\partial r} + \frac{U^2}{f\hat{r}_c}\,, \tag{25.1}$$

where ρ_a is air density, f is the Coriolis parameter, p is the barometric pressure, r is the radial distance from the cyclone's center, and $\hat{r}_c$ is the radius of curvature of the isobars. The actual wind velocity relative to earth ($\boldsymbol{U}_a$) must include the storm's translational, or forward, velocity ($\boldsymbol{V}_f$). A simple approximation of the actual wind velocity is the vector sum: $\boldsymbol{U}_a = \boldsymbol{U} + \boldsymbol{V}_f$. Thus, the strongest winds in a cyclone are typically in the forward right (left) quadrant of the storm for storms in the Northern (Southern) hemisphere.

Because they tend to be less well organized, wind fields from extratropical cyclones are not as readily parameterized, and wind-field reconstructions for individual storms are typically developed from observations [25.71]. On the other hand, because tropical cyclones tend to be well organized, several vortex-based models exist [25.72–75] that may be used to estimate the cyclone's wind field for use in coastal hazard analysis and engineering design. The class of models known as planetary boundary layer (PBL) models generally assume that the tropical cyclone's barometric pressure field (p) is described as follows,

$$\frac{\partial p}{\partial r} = \frac{B(p_{far} - p_c)}{r}\left(\frac{R_{max}}{r}\right)^B \exp\left[-\left(\frac{R_{max}}{r}\right)^B\right], \tag{25.2}$$

where r is the radial distance from the cyclone's center, B is *Holland*'s [25.72] pressure profile shape parameter, p_{far} is the far-field barometric pressure, p_c is the cyclone's central pressure, and R_{max} is the radial distance from the cyclone's center to the location of the maximum wind speed. Equations (25.1) and (25.2) may be solved numerically to develop the tropical cyclone's wind field. In the centermost portion of the tropical cyclone, the second term in (25.1) approaches zero, allowing analytical estimation of the maximum relative wind speed parameter, U_{max}

$$U_{max} = \sqrt{\frac{B}{\rho_a e}(p_{far} - p_c)}\,, \tag{25.3}$$

where e is the base of the natural logarithm, 2.718. Figure 25.4 shows $U_{a\text{-}max} = U_{max} + V_f$ versus $p_{far} - p_c$ for three typical V_f values. It should be noted here that U_{max} is used in these models to represent winds that are in approximate balance with the slowing varying pressure gradients. They are not the characteristic of a particular maximum value over a fixed time interval similar to some wind statistics (such as the fastest 1-min wind speed in an hour typically given in tropical cyclone weather reports). As a rough approximation, this parameter varies suffi-

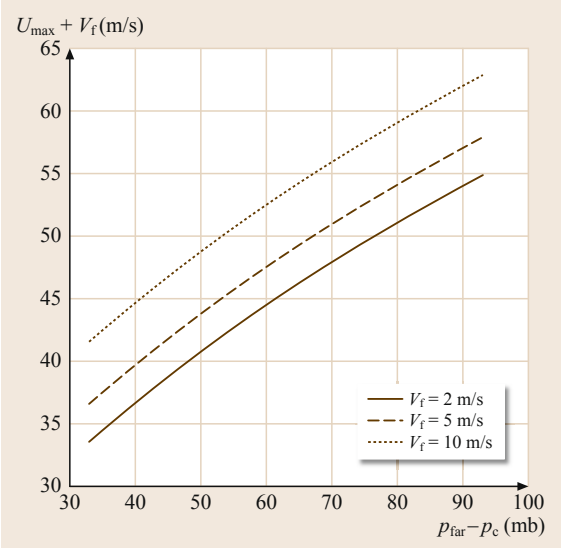


Fig. 25.4 Estimated maximum wind speed in a tropical cyclone using (25.3) and assuming Holland $B = 1$

ciently slowly to be considered constant over about 15–30 min.

In summary, storm winds in and of themselves pose a wind load threat to the built environment at the coast, for example critical infrastructure, such as roadways and utilities, as well as residential and commercial buildings. The above equations may be used to get a general sense of wind strength as it depends on storm meteorological parameters, but it is recommended that numerical simulation and assimilation techniques be employed for the rigorous assessment of winds and wind forcing. Lastly, these storm winds induce large storm waves and surges that are often the design condition of interest in coastal design, as discussed in the following sections.

Example 25.1

At landfall near the Louisiana-Mississippi border, US, Hurricane Katrina was observed to have a forward speed and central pressure of $V_f = 7\,\mathrm{m\,s^{-1}}$ [25.76] and $p_c = 920\,\mathrm{mb}$ [25.77]. Hurricane Katrina's maximum relative wind speed at landfall may be approximated using (25.3) or Fig. 25.4, where we assume $p_{far} = 1013\,\mathrm{mb}$ (standard atmospheric pressure), $B = 1.0$ (typical value), and $\rho_a = 1.23\,\mathrm{kg\,m^{-3}}$. The relative maximum wind speed is then $U_{max} = 53\,\mathrm{m\,s^{-1}}$. The actual maximum wind speed is estimated by summing the relative maximum speed and the tropical cyclone's forward speed, giving $U_{a\text{-}max} = 60\,\mathrm{m\,s^{-1}}$. Observed actual maximum 15-min wind speed for Hurricane Katrina at landfall was $U_{a\text{-}max} = 44\,\mathrm{m\,s^{-1}}$ [25.77]; thus, for this storm, (25.3) provides a maximum wind speed estimate that is about 35% larger than the observed value. This example illustrates the errors that can be introduced by using a simplified wind model close to the coastline, where the wind field is significantly altered by the storm's interaction with land.

25.2.2 Wind-Generated Waves

Large wind-generated waves often govern coastal engineering design. Understanding the magnitude and frequency of large waves is essential for quantifying impact loading on and overtopping of coastal structures and for quantifying storm erosion. These waves develop during far-field and near-field storm events. The reader is referred to Chap. 4 for a discussion of nearshore wave evolution and processes and Chap. 27 for a discussion of numerical modeling of coastal waves. While a thorough analysis of the design wave climate involves analysis of wave-field observations and computational modeling (Chap. 27), in this section we discuss simple methods for identifying general storm wave climate characteristics by considering tropical cyclone waves at the coast.

Wind-wave generation is governed by the strength, extent, and duration of the winds, as well as topographic conditions that limit energy transfer and growth by reducing fetch distance and depth, respectively. The empirical model developed by [25.78] from the interpretation of numerical simulations has been used extensively [25.79, 80] for approximating a tropical cyclone's maximum nonbreaking significant wave height ($H_{s\text{-}max}$)

$$\frac{gH_{s\text{-}max}}{U^2_{a\text{-}max}} = a\sqrt{\frac{gX}{U^2_{a\text{-}max}}}\,, \tag{25.4}$$

where $U_{a\text{-}max}$ is taken at a 10-m elevation, a is an empirical fit coefficient, and X is an equivalent fetch distance. The empirical coefficient is $a = 0.0016$, and is based on the earlier work of *Hasselmann* et al. [25.81] on fetch-limited spectral wave growth. *Taylor* [25.82] found that the formulation of *Alves* et al. [25.83] for X provides less scatter and bias for nearshore wave estimation, with respect to the original formulation of *Young* [25.78]. The *Alves* et al. [25.83] definition is

$$X = b\left[R_{max}\left(1 + d\frac{V_f}{c_{g\text{-}max}}\right)\right]^m\,, \tag{25.5}$$

where $c_{g\text{-}max}$ is the dominant wave's group speed at the location of interest (Part A, Chap. 4) and b, d, and m are empirical fit coefficients. *Alves* et al. [25.83] found $b = 1.11$, $d = 2.3$, and $m = 0.38$, based on numerical simulations with WAVEWATCH III, where R_{max} is in meters and V_f and $c_{g\text{-}max}$ are in $\mathrm{m\,s^{-1}}$. In shallower depths, depth-limited breaking should be considered and the wave-height prediction by (25.4) and (25.5) reduced accordingly. *Alves* et al. [25.83] suggested the following approximation for the peak wave period (T_p)

$$T_p = \frac{0.123U_{a\text{-}max}}{g}\,. \tag{25.6}$$

Representative calculations using (25.4) through (25.6) to estimate $H_{s\text{-}max}$ are presented in Fig. 25.5.

In summary, while it is recommended that computational simulations be used when possible to accurately quantify wave design conditions, the above empirical formulae provide a means for initially assessing extreme wave conditions at the coast.

Example 25.2

Continuing with the information provided in Example 25.1, where the 15-min maximum wind speed and the radius to maximum wind for Hurricane Katrina at

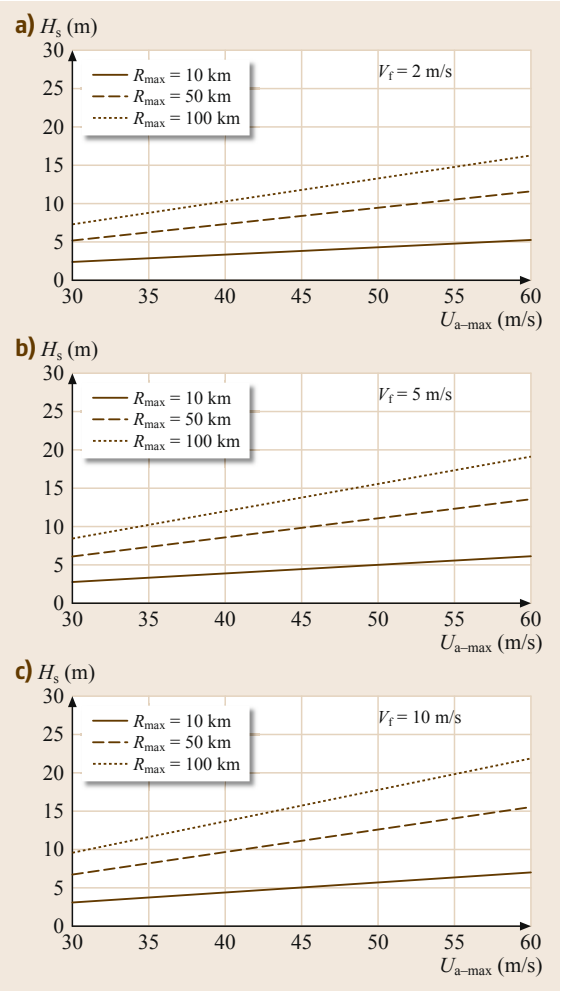


Fig. 25.5a–c Estimated maximum significant wave height for tropical cyclones by (25.4) through (25.6), when mean water depth is 25 m when $V_f = 2\,\mathrm{m/s}$ **(a)**, 5 m/s **(b)**, and 10 m/s **(c)**

landfall were estimated to be $U_{a\text{-}max} = 44\,\mathrm{m\,s^{-1}}$ and $R_{max} = 65\,\mathrm{km}$, respectively [25.77]. Using (25.4) and (25.5) or Fig. 25.5, and estimating $c_{g\text{-}max}$ using linear wave theory (Part A, Chap. 4), the maximum significant wave height at a 25-m depth may be estimated to be $H_{s\text{-}max} = 11\,\mathrm{m}$.

25.2.3 Flood Levels and Runup

Extreme water levels are oftentimes the critical design criteria in coastal engineering. At the coast, these extreme water levels are typically associated with coastal storms and tsunamis. In the following section, each of these phenomena is described.

Storm Surge

Storm surge is one of the most deadly and destructive components of coastal storms, often resulting in widespread inundation of roadways, homes, and critical infrastructure. The extent and magnitude of coastal flooding arise from the combination of the meteorological forcing that induces large waves and wave setup (ζ_{wave}), wind surge (ζ_{wind}), barometric surge (ζ_{baro}), and Ekman setup (ζ_{Ek}), and daily water-level conditions including astronomical tide levels (ζ_{tide}). To accurately estimate storm water levels and inundation along real coasts, high-resolution computational simulation of the shallow-water equations should be employed [25.84, 85] (Chap. 27). In this section, we will discuss the general behavior and characteristics of storm surges. While nonlinearities certainly arise during surge generation in areas with complex terrain and strong tidal signals, total storm flood elevation (ζ_{storm}) may be approximated as the linear sum

$$\zeta_{storm} = \zeta_{baro} + \zeta_{wind} + \zeta_{wave} + \zeta_{Ek} + \zeta_{tide}\,. \qquad (25.7)$$

Neglecting tidal effects, the shallow-water momentum equation for storm surge, in its linearized, depth-integrated form, is

$$-\frac{1}{g}\frac{\partial v_x}{\partial t}\frac{\partial \zeta_{storm}}{\partial x} = -\frac{1}{\rho g}\frac{\partial p}{\partial x} + \frac{\tau_{s\text{-}x}}{\rho g h} + \frac{\tau_{wave\text{-}x}}{\rho g h} + \frac{f v_y}{g} - \frac{\tau_{b\text{-}x}}{\rho g h}\,, \qquad (25.8a)$$

$$-\frac{1}{g}\frac{\partial v_x}{\partial t}\frac{\partial \zeta_{storm}}{\partial y} = -\frac{1}{\rho g}\frac{\partial p}{\partial y} + \frac{\tau_{s\text{-}y}}{\rho g h} + \frac{\tau_{wave\text{-}y}}{\rho g h} - \frac{f v_x}{g} - \frac{\tau_{b\text{-}y}}{\rho g h}\,, \qquad (25.8b)$$

where v is water velocity, τ_s is the surface wind stress, τ_{wave} is the wave radiation stress gradient, h is the total water depth, f is the Coriolis parameter, and τ_b is the bottom stress. The first four terms on the right-hand side of (25.8) respectively induce the barometric surge, the wind surge, wave setup, and Ekman setup.

The barometric surge arises from the storm's barometric pressure change, or gradient as seen in the first right-hand term in (25.8). The maximum contribution by barometric surge occurs at the location of the lowest pressure, at the storm's center, and may be approximated by assuming hydrostatic pressure

$$\zeta_{baro} = \frac{(p_{far} - p_c)}{\rho g}\,. \qquad (25.9)$$

Barometric surge tends to be of the order of 1 m or less; for example, an intense tropical cyclone central

pressure deficit of $(p_{far} - p_c) = 100\,\text{mb}$ yields a barometric surge of about 1 m at the storm's center.

The wind surge arises from the flux of atmospheric momentum into the water column. In the absence of any other forcing, (25.8) becomes for steady-state wind surge

$$\frac{\partial \zeta_{wind}}{\partial x} = \frac{\tau_s - \tau_b}{\rho g h} , \tag{25.10}$$

where x is in the wind direction and it is assumed that τ_s and τ_b are aligned.

Since nearshore wave conditions vary over a narrower range during storms, wave setup contributes of the order of 1 m to the total water level during most storms. The exact amount of wave setup will depend on incident wave and bathymetric conditions. In regions with steep topography, such as that characteristic of volcanic islands, wave setup is typically the dominant contributor to storm flood levels. See *Resio* and *Westerink* [25.86] for a discussion of wave setup during coastal storms.

Ekman setup is developed as a result of the Coriolis force acting on strong alongshore currents, driven by winds, waves, or tides. The result is a rise in water level at the coastline. In the absence of any other forcing, (25.8) becomes the steady-state Ekman setup

$$\frac{\partial \zeta_{Ek}}{\partial x} = \frac{f v_y}{g} , \tag{25.11}$$

where x is in the cross-shore direction and the water velocity v_y is in the longshore direction. While not always significant, the Ekman setup has been found to contribute up to several meters of surge under certain conditions. For example, *Kennedy* et al. [25.87] showed through observations and numerical simulation that as Hurricane Ike (2008) approached the Gulf of Mexico coastline, wind-generated alongshore currents yielded an Ekman setup of 3 m along the coasts of Texas and Louisiana (USA), well in advance of landfall.

Of the four meteorologically driven surge components, wind setup typically is the largest in shallow-sloping coastal areas. To aid in understanding the wind surge response for simple cases, (25.10) may be solved analytically by further assuming the still-water depth (h_o) is $h_o \gg \zeta_{wind}$, $\tau_b \ll \tau_a$, and a quadratic wind stress formulation

$$\tau_s = \rho_a c_d U^2 , \tag{25.12}$$

where c_d is the wind drag coefficient. The solution to (25.10) is then

$$\zeta_{wind} = \left(\frac{\rho_a}{\rho}\right) \frac{c_d U^2}{g} \int_o^{\tilde{x}} \frac{dx}{h_o} , \tag{25.13}$$

where $\tilde{x}$ is distance to the point of interest, measured in the wind direction. Equation (25.13) reveals that the wind surge gradient is proportional to the square of the wind speed, yet is inversely proportional to the depth. Thus, relatively larger surges are expected for relatively stronger wind conditions and for relatively longer distances over which the wind acts, and relatively smaller surges are expected in locations with relatively deeper depths.

For the case of wind surge generated by an onshore wind along the open coast, X is the cross-shore distance to the location of interest measured from a point deep enough for surge generation to be negligible ($x = 0$). The bathymetric profile of the continental shelf fronting many storm-prone coasts may be roughly approximated as an exponential function [25.88], such that (25.13) gives wind surge at the coast as

$$\zeta_{wind} = a \left(\frac{\rho_a}{\rho}\right) \frac{c_d U^2}{g} \left[\exp\left(b X_{shelf}\right) - 1\right] , \tag{25.14}$$

where a and b are the fit constants specific to the local continental shelf shape, and X_{shelf} is the local width of the continental shelf.

For the case of wind surge generation in an enclosed basin of uniform depth, analogous to a lake or coastal bay with restricted exchange with the ocean, the solution to (25.13) for the maximum wind surge is

$$\zeta_{wind} = \left(\frac{\rho_a}{\rho}\right) \frac{c_d U^2}{g} \frac{X_{basin}}{2 h_o} , \tag{25.15}$$

where X_{basin} is the width of the basin.

In real storms, the surge response is more complex. The wind forcing, for example, varies spatially and temporally with storm size and propagation speed. Likewise, the surge response varies from location to location based on local shoreline and bathymetry configuration. It was not until Hurricane Katrina made landfall near New Orleans, LA, USA in 2005 that the importance of the role of storm size on surge generation became widely recognized [25.89]. The Biloxi, Mississippi's Sun Herald [25.90] noted that following Hurricane Katrina:

> *an oft heard refrain ... is Hurricane Camille killed more people in 2005 than it did in 1969. Many officials and locals believed those ... who had survived what was then the strongest recorded hurricane were lulled into a false sense of security that kept them in harm's way.*

This sentiment reflects the disparity between storm intensity (central pressure, maximum wind speed) and surge generation from Hurricanes Camille and Katrina, both of which made landfall near New Orleans.

While Hurricane Camille was by far the more intense of the two, Hurricane Katrina's storm surge at 8.5 m was significantly larger than Camille's 6.9 m. The reason Hurricane Katrina generated such high surges was because of the large extent of hurricane-force winds, more than twice that for Hurricane Camille. It is now well accepted that tropical cyclone surge magnitude is primarily influenced by the regional bathymetric slope and the storm's intensity and size [25.89, 91–93], while the cyclone's forward speed and track orientation with respect to the coastline are generally less influential [25.85, 94, 95]. *Irish* and *Resio* [25.88] proposed a quasi-empirical model for estimating maximum tropical cyclone surge (excluding astronomical tide) on continuous coasts

$$\zeta_{\text{storm-max}} \cong n\frac{(p_{\text{far}}-p_{\text{c}})}{p_{\text{atm}}}X_{\text{shelf-30}}\psi\left(\frac{R_{33}}{X_{\text{shelf-30}}}\right)+q\,, \quad (25.16a)$$

$$\psi\left(\frac{R_{33}}{X_{\text{shelf-30}}}\right)= \begin{cases} \left(\dfrac{R_{33}}{X_{\text{shelf-30}}}\right) & \text{when } \left(\dfrac{R_{33}}{X_{\text{shelf-30}}}\right)\le 1\,, \\ 1 & \text{when } \left(\dfrac{R_{33}}{X_{\text{shelf-30}}}\right)>1\,, \end{cases} \quad (25.16b)$$

where n and q are empirical fit coefficients, p_{atm} is standard atmospheric pressure, R_{33} is the storm radius to $U_{\text{a}} = 33\,\text{m}\,\text{s}^{-1}$ (hurricane-force winds), and $X_{\text{shelf-30}}$ is cross-shore distance from the shoreline to the 30 m depth contour. Based on the observations from continental US hurricane strikes between 1941 and 2008, *Irish* and *Resio* [25.88] determined $n = 4.51\times10^{-4}$ and $q = 1.84\,\text{m}$. Figure 25.6 shows maximum surge estimates by (25.16) for selected storm parameters.

Example 25.3

Continuing with the information provided in Examples 25.1 and 25.2, we may use (25.16) or Fig. 25.6 to estimate the maximum storm surge from Hurricane Katrina, where the central pressure and radius to hurricane force winds at landfall were $p_{\text{c}} = 920\,\text{mb}$ and $R_{33} = 217\,\text{km}$, respectively [25.77]. At the landfall location, the cross-shore distance to the 30 m depth contour is $X_{\text{shelf-30}} = 140\,\text{km}$ [25.88]. Thus, (25.16) yields a maximum surge estimate of $\zeta_{\text{storm-max}} = 7.6\,\text{m}$. Maximum observed flood elevations [25.5] for Hurricane Katrina are between 7.5 and 8.5 m, indicating that (25.16) performs well for this storm.

In summary, storm surge is a complex process whose magnitude varies both as a function of topographic conditions and storm conditions. While it is recommended that computational simulations be used for detailed assessment of storm surge, the above simplified scenarios provide a means for understanding the general behavior of flood response during coastal storms.

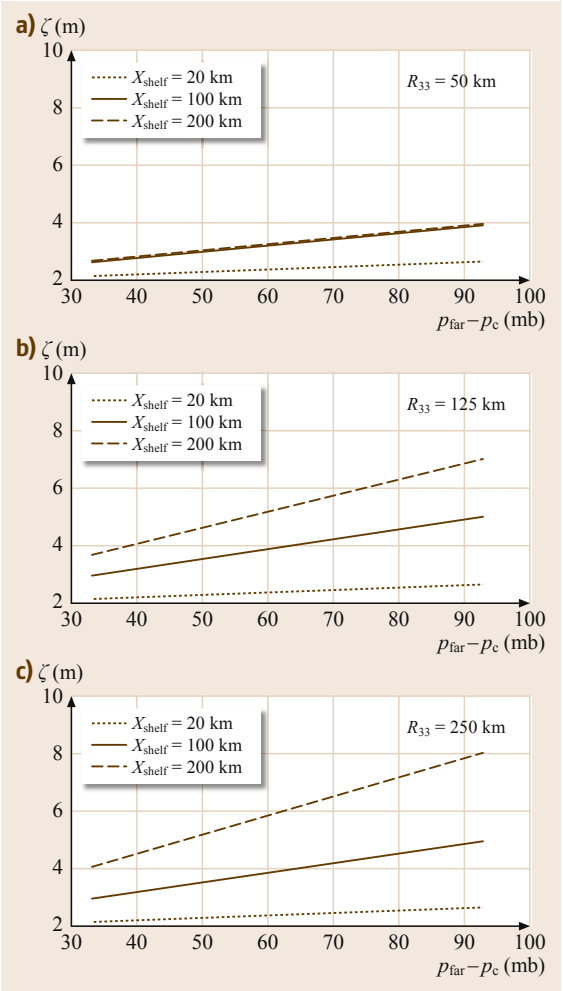


Fig. 25.6a–c Estimated maximum surge by (25.13) when $R_{33} = 50\,\text{km}$ **(a)**, 125 km **(b)** and 250 km **(c)**. In **(a)**, the result when $X_{\text{shelf}} = 200\,\text{km}$ matches the result when $X_{\text{shelf}} = 100\,\text{km}$

Tsunamis

It is a well-known fact that tsunami waves are capable of traveling very long distances to bring their destructive force to far-field coastlines. Depending on the tsunami source, the runup magnitude and the distribution of runup along the coastline can vary drastically. Tsunami events provide the important opportunity to study both runup magnitude and distribution. For example, the runup of the 2004 Sumatra tsunami was surveyed in Indonesia and Thailand, Sri Lanka and In-

dia, but also in Somalia, Madagascar, and Kenya [25.96, and references therein].

From a general point of view, there are very few simple rules that can be applied to quickly estimate the tsunami runup. The most successful rule is the Plafker rule of thumb, which states that the runup on a straight beach does not exceed twice the amount of slip on the fault. Numerical modeling is able to predict the runup magnitude and distribution during an event in a forecast framework [25.98] and is benchmarked with survey data in the aftermath of events. Commonly depth-integrated equations are solved numerically and provide a very flexible computational framework (Chap. 27). While theoretical analyses are restricted to the canonical problem for the long-wave runup, analytical treatment of the shallow-water equations provides not only crucial verification for numerical models, but also general information on the runup magnitude (R) as a function of the average nearshore and onshore topographic slope and wave characteristics. The canonical problem of long-wave runup is made up of constant-depth part and a sloping beach. Originally, solitary waves were employed as wave forcing. *Synolakis* [25.99, 100] found the long-sought analytical solution of the shallow water equations, which resulted in a runup law as a function of the slope ($\tan\beta$) and offshore wave amplitude (H, where the wave amplitude is normalized by the water depth in the constant depth portion, h). Synolakis' runup law is

$$\frac{R}{h} = 2.831\,(\tan\beta)^{-\frac{1}{2}}\left(\frac{H}{h}\right)^{\frac{5}{4}} . \qquad (25.17)$$

The canonical problem was extended by *Tadepalli* and *Synolakis* [25.97, 101] to consider N-shaped waves (hereafter N-waves), which are geophysically more realistic. The respective runup law for the N-wave is

$$\frac{R}{h} = \xi\,(\tan\beta)^{-\frac{1}{2}}\left(\frac{H}{h}\right)^{\frac{5}{4}} \qquad (25.18)$$

with the constant $\xi = 3.86$ for leading elevation and $\xi = 4.55$ for leading depression N-waves. Figure 25.7 shows the runup of N-waves compared to solitary waves. In general, the runup of N-waves is larger than that for solitary waves; therefore, runup estimated from (25.17) may underestimate the maximum runup. Furthermore, *Madsen* et al. [25.102] compared different types of tsunami problems and evaluated the performance of the solitary waves with the results that question the canonical nature of the solitary wave.

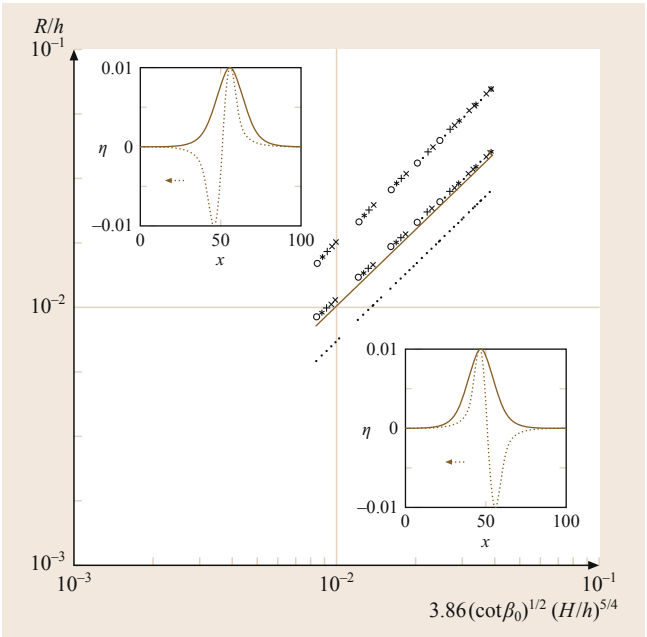


Fig. 25.7 The maximum runup of N-waves of both types: leading elevation and leading depression. The leading-depression N-waves produces larger runup than the leading-elevation N-waves. Both type of N-waves have a larger runup than solitary waves (*dotted line*) (adapted from [25.97])

With the help of numerical simulation, *Okal* and *Synolakis* [25.103] were able to confirm the Plafker rule of thumb with quantitative arguments, and they showed that tsunamis generated by earthquakes differ from those generated by landslides not only in runup magnitude but also in the runup distribution, which can be ascribed to the length scale of the different sources (Fig. 25.8). The runup distribution for landslide-generated tsunamis is narrow and may have a maximum in the runup distribution that can be four to five times larger that those from earthquake-generated tsunamis.

25.2.4 Environmental Impacts

Environmental impacts of natural hazards span from episodic and long-term changes to the natural landscape to episodic and long-term damage to critical infrastructure and industrial complexes, and the surrounding areas. In response to rising sea levels, extreme waves, and coastal flooding, the nearshore and onshore coastal zone is reworked, changing the topography and land cover. Long-term loss of intertidal habitat, for example wetlands and mangroves and the rich ecosystems they support, is anticipated as sea levels rise [25.104–107] and as coastal development continues. It is not yet clear what the broader environmental, social, and economic consequences will be of such changes to the coastal landscape.

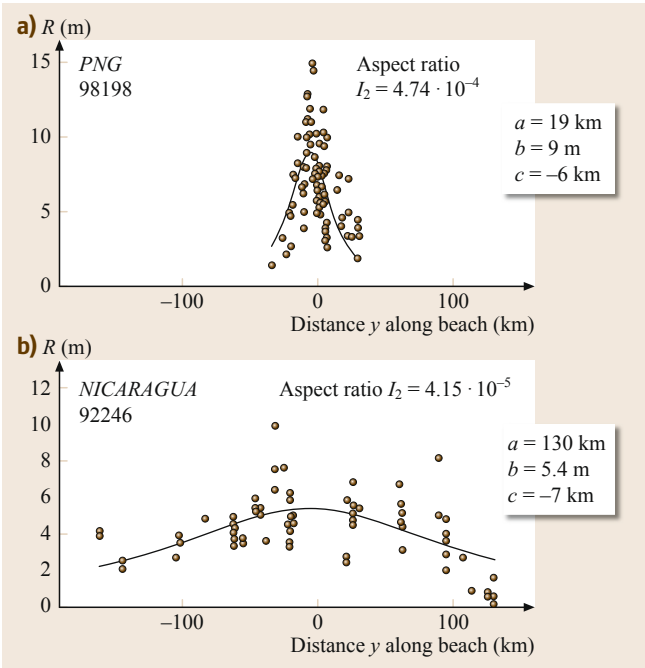


Fig. 25.8a,b Runup distributions along the shoreline for **(a)** the 1998 Papua New Guinea tsunami and **(b)** the 1992 Nicaragua tsunami. The narrow distribution in **(a)** indicates the runup of a landslide-generated tsunami, and the wide distribution in **(b)** is typical for earthquake-generated tsunamis (adapted from [25.103]) ►

Sea-level rise, episodic coastal flood events, and increasing drought events with global warming may also push salt water farther inland along rivers and estuaries, and in groundwater. Such processes will shift ecosystem habitats that are sensitive to salinity changes. In addition, there is the potential for direct human impact. The 2012 droughts in the United States, along with observed high rates of the relative sea-level rise in New Orleans, led to severe salt-water intrusion in the Mississippi River [25.108]. The salt-water intrusion was so severe that freshwater intakes for the region's drinking water supply were contaminated.

The impact of a major coastal inundation event on critical infrastructure can be catastrophic. In Japan, the March 2011 earthquake and tsunami lead to a meltdown at the Fukushima Daiichi nuclear power plant. The subsequent release of radioactivity resulted in the mandatory evacuation of more than 200 000 people, and long-term human health and environmental impacts are not yet known [25.109].

In summary, episodic and long-term hazards at the coast induce changes to the natural environment and can lead to human health threats. The feedback between the hazards and changes induced to the natural environment is currently not well understood. Yet, possible environmental consequences must be considered in all phases of coastal design.

25.3 Summary

In this chapter, we have outlined the various hazards threatening coastal communities and the surrounding environment, and we have presented a series of simple tools for assessing the general behavior of the physical hazard. The primary physical threats include flooding, wave impacts, erosion, loss of natural habitat, and environmental contamination and human health consequences. Coastal hazards are a complex problem, and coastal design often calls for an integrated, interdisciplinary team effort, in order to accurately assess the physical hazard and the threats it poses. Project teams often include a combination of engineers, physical scientists, ecosystem scientists, economists, and social scientists, and the hazard is quantified in terms of the risk it poses. In Chap. 26, hazard probability and risk concepts for coastal design are discussed.

25.4 Nomenclature

a, b, d, m, n, q	Coefficients
B	Holland's pressure profile shape parameter
c_d	Wind drag coefficient
$c_{g\text{-max}}$	Dominant wave's group speed
h	Water depth
h_o	Still-water depth
H	Offshore wave amplitude
$H_{s\text{-max}}$	Cyclone's maximum significant wave height
f	Coriolis parameter
p	Barometric pressure
p_{atm}	Standard atmospheric pressure
p_c	Cyclone central pressure

p_{far}	Far-field barometric pressure
r	Radial distance from cyclone center
$\hat{r}_c$	Radius of curvature of the isobars
R	Tsunami runup
R_{33}	Cyclone radius to $U_a = 33\,m\,s^{-1}$
R_{max}	Cyclone radius to maximum wind speed
SLR	Sea-level rise
T_p	Peak wave period
U	Wind velocity
U_a	Wind velocity relative to earth
$U_{a\text{-}max}$	Cyclone's maximum wind speed
U_{max}	Cyclone maximum wind speed, relative to the cyclone's center
v	Water velocity
V_f	Cyclone forward velocity
$\tilde{x}$	Distance to point of interest
X	Fetch distance
X_{basin}	Width of the basin
X_{shelf}	Local width of continental shelf
$X_{shelf\text{-}30}$	Cross-shore distance from shoreline to 30 m depth contour
$\cot\beta$	Bottom slope
ζ_{baro}	Barometric surge
ζ_{Ek}	Ekman setup
ζ_{storm}	Total storm flood elevation
ζ_{tide}	Astronomical tide level
ζ_{wind}	Wind surge
ζ_{wave}	Wave setup
ξ	Coefficient
ρ_a	Air density
τ_b	Bottom stress
τ_s	Surface wind stress
τ_{wave}	Wave radiation stress gradient

References

25.1 I.M. Cline: *Special Report of the Galveston Hurricane of September 8, 1900* (United States Weather Bureau Office, Galveston 1900)

25.2 N.L. Frank, S.A. Husain: Deadliest tropical cyclone in history?, Bull. Am. Meteorol. Soc. **52**(6), 438–445 (1971)

25.3 G.R. Flierl, A.R. Robinson: Deadly surges in bay of Bengal – Dynamics and storm-tide tables, Nature **239**(5369), 213–215 (1972)

25.4 National Oceanic and Atmospheric Administration: NOAA's top global weather, water and climate events of the 20th century, http://www.noaanews.noaa.gov/stories/images/global.pdf

25.5 US Army Corps of Engineers: *Performance Evaluation of the New Orleans and Southeast Louisiana Hurricane Protection System*, Draft final report of the Interagency Performance Evaluation Task Force (US Army Corps of Engineers, Washington 2006)

25.6 N. Lott, T. Ross: *Tracking and Evaluating US Billion Dollar Weather Disasters, 1980–2005* (National Climatic Data Center, Asheville 2006), database updated in 2012

25.7 E.S. Blake, C.W. Landsea, E.J. Gibney: *The Deadliest, Costliest, and Most Intense United States Hurricanes from 1851 to 2010 (and Other Frequently Requested Hurricane Facts)*, Tech. Memo. NWS, Vol. NHC-6 (NOAA, Miami 2011)

25.8 R.H. Simpson: The hurricane disaster-potential scale, Weatherwise **27**, 169, 186 (1974)

25.9 J.W. East, M.J. Turco, R.R. Mason Jr.: Monitoring inland storm surge and flooding from Hurricane Ike in Texas and Louisiana, September 2008, http://pubs.usgs.gov/of/2008/1365/ (2008), US Geological Survey Open-File Report 2008–1365

25.10 P.J. Baxter: The east coast big flood, 31 January–1 February 1953: A summary of the human disaster, Philos. Trans. R. S. Math. Phys. Eng. Sci. **363**(1831), 1293–1312 (2005)

25.11 H. Gerritsen: What happened in 1953? – The big flood in the Netherlands in retrospect, Philos. Trans. R. S. A Math. Phys. Eng. Sci. **363**(1831), 1271–1291 (2005)

25.12 V.V. Titov, A.B. Rabinovich, J.O. Mofjeld, R.E. Thomson, F.I. Gonzalez: The global reach of the 26 Decemeber 2004 Sumatra tsunami, Science **390**, 2045–2048 (2005)

25.13 National Geographic: The deadliest tsunami in history? National Geographic News, January 7, http://news.nationalgeographic.com/news/2004/12/1227_041226_tsunami.html 2005

25.14 D.G. Masson, C.B. Harbitz, R.B. Wynn, G. Perdersen, F. Lovholt: Submarine landslides: Processes, triggers and hazard prediction, Phil. Trans. R. Soc. A **364**, 2009–2039 (2006)

25.15 C.B. Harbitz, G. Pedersen, B. Gjevik: Numerical simulations of large water waves due to landslides, J. Hydraulic Eng. **119**(12), 1325–1342 (1993)

25.16 C.B. Harbitz: Model simulations of tsunamis generated by the Storrega slides, Mar. Geophys. **105**, 1–21 (1992)

25.17 E.A. Okal: T-waves from the 1998 Papua New Guinea earthquake and its aftershocks: Timing of the tsunamigenic slump, Pure Appl. Geophys. **160**, 1843–1863 (2003)

25.18 C.E. Synolakis, J.-P. Bardet, J.C. Borrero, H.L. Davis, E.A. Okal, E.A. Silver, S. Sweet, D.R. Tappin: The slump origin of the 1998 Papua New Guinea tsunami, Proc. Roy. Soc. A **458**, 763–789 (2002)

25.19 H.M. Fritz, W.H. Hager, H.-E. Minor: Lituya Bay case: Rockslide impact and wave run-up, Sci. Tsunami Hazard **19**(1), 3–22 (2001)

25.20 R. Weiss, H.M. Fritz, K. Wuennemann: Hybrid modeling of the mega tsunami run-up in Lituya

Bay after half a century, Geophys. Res. Lett. **36**, L09609 (2009)
25.21 G.J. Fryer, P. Watts, L.F. Pratson: Source of the great tsunami of 1 April 1946: A landslide in the upper Aleutian forearc, Mar. Geol. **203**, 201–218 (2004)
25.22 F. Raichlen, C. Synolakis: Waves and runup generated by a three-dimensional sliding mass. In: *Submarine Mass Movements and Their Consequences*, Advances in Natural and Technological Hazards Research, ed. by J. Locat, J. Mienert, L. Boisvert (Kluwer, Dordrecht 2003) pp. 113–119
25.23 R. Weiss, C.E. Synolakis, J. O'Shay: Initial waves from deformable submarine landslides, Proc. NSF CMMI Eng. Res. Innov. Conf. (2011)
25.24 R. Weiss, K. Wuennemann: Large waves caused by oceanic impacts of meteorites. In: *Tsunamis and Nonlinear Waves*, ed. by A. Kundu (Springer, Heiderlberg 2007) pp. 235–260
25.25 K. Wuennemann, R. Weiss, K. Hofmann: Characteristics of impact-induced large waves – Re-evaluation of the tsunami hazard, Meteorit. Planet. Sci. **72**, 1–11 (2007)
25.26 G. Pararas-Carayamis: Near-field and far-field effects of tsunami generated by the paroxysmal eruption, explosions, caldera collapses and massive slope failures of the Krakatay volcano in Indonesia on August 26–27 1883, Sci. Tsunami Hazard **21**(4), 191–201 (2003)
25.27 S.N. Ward, S.J. Day: Cumbre Vieja volcano – Potential collapse and tsunami at La Palma, Canary Islands, Geophys. Res. Lett. **28**, 397–400 (2001)
25.28 S.N. Ward: Planetary cratering: A probabilistic approach, J. Geophys. Res. **107**(E40), 7-1–7-11 (2002)
25.29 R.B. Baldwin: On the history of lunar cratering: The absolute scale and the origin of planetesimals, Icarus **14**, 36–52 (1971)
25.30 Intergovernmental Panel on Climate Change: *Intergovernmental Panel on Climate Change Fourth Assessment Report Working Group 1 Report: The Physical Science Basis* (Cambridge Univ. Press, Cambridge 2007)
25.31 N.J. White, J.A. Church, J.M. Gregory: Coastal and global averaged sea level rise for 1950 to 2000, Geophys, Res. Lett. **32**, L01601 (2005)
25.32 L. Miller, B.C. Douglas: Mass and volume contributions to twentieth-century global sea level rise, Nature **428**(6981), 406–409 (2004)
25.33 S. Jevrejeva, J.C. Moore, A. Grinsted: How will sea level respond to changes in natural and anthropogenic forcings by 2100?, Geophys. Res. Lett. **37**, L07703 (2010)
25.34 M. Vermeer, S. Rahmstorf: Global sea level linked to global temperature, Proc. Natl. Acad. Sci. USA **106**(51), 21527–21532 (2009)
25.35 W.T. Pfeffer, J.T. Harper, S. O'Neel: Kinematic constraints on glacier contributions to 21st-century sea-level rise, Science **321**(5894), 1340–1343 (2008)
25.36 R. Horton, C. Herweijer, C. Rosenzweig, J.P. Liu, V. Gornitz, A.C. Ruane: Sea level rise projections for current generation CGCMs based on the semi-empirical method, Geophys. Res. Lett. **35**(2), L02715 (2008)
25.37 R.E. Flick, J.F. Murray, L.C. Ewing: Trends in United States tidal datum statistics and tide range, J. Waterw. Port Coast. Ocean Eng. **129**(4), 155–164 (2003)
25.38 L.R. Kleinosky, B. Yarnal, A. Fisher: Vulnerability of Hampton Roads, Virginia to storm-surge flooding and sea-level rise, Nat. Hazards **40**(1), 43–70 (2007)
25.39 P. Kirshen, K. Knee, M. Ruth: Climate change and coastal flooding in Metro Boston: Impacts and adaptation strategies, Clim. Change **90**(4), 453–473 (2008)
25.40 D.R. Cayan, P.D. Bromirski, K. Hayhoe, M. Tyree, M.D. Dettinger, R.E. Flick: Climate change projections of sea level extremes along the California coast, Clim. Change **87**, S57–S73 (2008)
25.41 P. Ruggiero, P.D. Komar, J.C. Allan: Increasing wave heights and extreme value projections: The wave climate of the US Pacific Northwest, Coast. Eng. **57**(5), 539–552 (2010)
25.42 S.J. Holgate, A. Matthews, P.L. Woodworth, L.J. Rickards, M.E. Tamisiea, E. Bradshaw, P.R. Foden, K.M. Gordon, S. Jevrejeva, J. Pugh: New data systems and products at the permanent service for mean sea level, J. Coastal Res. **29**(3), 493–504 (2013)
25.43 J.B. Elsner, J.P. Kossin, T.H. Jagger: The increasing intensity of the strongest tropical cyclones, Nature **455**(7209), 92–95 (2008)
25.44 K. Emanuel, R. Sundararajan, J. Williams: Hurricanes and global warming – Results from downscaling IPCC AR4 simulations, Bull. Am. Meteorol. Soc. **89**(3), 347–367 (2008)
25.45 T.R. Knutson, R.E. Tuleya: Impact of CO_2-induced warming on simulated hurricane intensity and precipitation: Sensitivity to the choice of climate model and convective parameterization, J. Clim. **17**(18), 3477–3495 (2004)
25.46 G.A. Vecchi, B.J. Soden: Effect of remote sea surface temperature change on tropical cyclone potential intensity, Nature **450**(7172), 1066–1070 (2007)
25.47 P.J. Webster, G.J. Holland, J.A. Curry, H.R. Chang: Changes in tropical cyclone number, duration, and intensity in a warming environment, Science **309**(5742), 1844–1846 (2005)
25.48 T.R. Knutson, J.L. McBride, J. Chan, K. Emanuel, G. Holland, C. Landsea, I. Held, J.P. Kossin, A.K. Srivastava, M. Sugi: Tropical cyclones and climate change, Nat. Geosci. **3**(3), 157–163 (2010)
25.49 G.J. Holland, P.J. Webster: Heightened tropical cyclone activity in the North Atlantic: Natural variability or climate trend?, Philos. Trans. R. Soc. A Math. Phys. Eng, Sci. **365**(1860), 2695–2716 (2007)
25.50 M.E. Mann, T.A. Sabbatelli, U. Neu: Evidence for a modest undercount bias in early historical Atlantic tropical cyclone counts, Geophys. Res. Lett. **34**(22), L22707 (2007)

25.51 M.A. Bender, T.R. Knutson, R.E. Tuleya, J.J. Sirutis, G.A. Vecchi, S.T. Garner, I.M. Held: Modeled impact of anthropogenic warming on the frequency of intense atlantic hurricanes, Science **327**(5964), 454–458 (2010)

25.52 A. Ali: Vulnerability of Bangladesh to climate change and sea level rise through tropical cyclones and storm surges, Water Air Soil Pollut. **92**(1/2), 171–179 (1996)

25.53 A. Ali: Climate change impacts and adaptation assessment in Bangladesh, Clim. Res. **12**(2/3), 109–116 (1999)

25.54 J.A. Lowe, J.M. Gregory, R.A. Flather: Changes in the occurrence of storm surges around the United Kingdom under a future climate scenario using a dynamic storm surge model driven by the Hadley Centre climate models, Clim. Dyn. **18**(3/4), 179–188 (2001)

25.55 M. Danard, A. Munro, T. Murty: Storm surge hazard in Canada, Nat. Hazards **28**(2/3), 407–431 (2003)

25.56 G. Gonnert: Maximum storm surge curve due to global warming for the European North Sea Region during the 20th–21st century, Nat. Hazards **32**(2), 211–218 (2004)

25.57 M.E. Mousavi, J.L. Irish, A.E. Frey, F. Olivera, B.L. Edge: Global warming and hurricanes: The potential impact of hurricane intensification and sea level rise on coastal flooding, Clim. Change **104**(3/4), 575–597 (2011)

25.58 N. Lin, K. Emanuel, M. Oppenheimer, E. Vanmarcke: Physically based assessment of hurricane surge threat under climate change, Nat. Clim. Change **2**(6), 462–467 (2012)

25.59 P. Bruun: Sea level rise as a cause of shore erosion, J. Waterw. Harb. Div. **88**(1–3), 117–130 (1962)

25.60 K.Q. Zhang, B.C. Douglas, S.P. Leatherman: Global warming and coastal erosion, Clim. Change **64**(1/2), 41–58 (2004)

25.61 J.L. Irish, A.E. Frey, J.D. Rosati, F. Olivera, L.M. Dunkin, J.M. Kaihatu, C.M. Ferreira, B.L. Edge: Potential implications of global warming and barrier island degradation on future hurricane inundation, property damages, and population impacted, Ocean Coast Manag. **53**(10), 645–657 (2010)

25.62 A.E. Frey, F. Olivera, J.L. Irish, L.M. Dunkin, J.M. Kaihatu, C.M. Ferreira, B.L. Edge: Potential impact of climate change on hurricane flooding inundation, population affected and property damages in Corpus Christi, J. Am. Water. Resour. Assoc. **46**(5), 1049–1059 (2010)

25.63 A.J. Condon, Y.P. Sheng: Evaluation of coastal inundation hazard for present and future climates, Nat. Hazards **62**(2), 345–373 (2012)

25.64 P.C. Kerr, J.J. Westerink, J.C. Dietrich, R.C. Martyr, S. Tanaka, D.T. Resio, J.M. Smith, H.J. Westerink, L.G. Westerink, T. Wamsley, M. van Ledden, W. de Jong: Surge generation mechanisms in the lower Mississippi River and discharge dependency, J. Waterw. Port Coast. Ocean Eng. **139**(4), 326–335 (2013)

25.65 E. Fussell: Constructing New Orleans, constructing race: A population history of New Orleans, J. Am. Hist. **94**, 846–855 (2007)

25.66 O.T. Magoon, B.L. Edge, K.E. Stone: The impact of anthropogenic activities on coastal erosion, Int. Conf. Coast. Eng. (2000) pp. 3934–3940

25.67 L.R. Martin: *Regional Sediment Management: Background and Overview of Initial Implementation* (US Army Corps of Engineers Institute for Water Resources, Washington 2002)

25.68 R. Camilli, C.M. Reddy, D.R. Yoerger, B.A.S. Van Mooy, M.V. Jakuba, J.C. Kinsey, C.P. McIntyre, S.P. Sylva, J.V. Maloney: Tracking hydrocarbon plume transport and biodegradation at deepwater horizon, Science **330**(6001), 201–204 (2010)

25.69 M.D. Powell, S. Murillo, P. Dodge, E. Uhlhorn, J. Gamache, V. Cardone, A. Cox, S. Otero, N. Carrasco, B. Annane, R.St. Fleur: Reconstruction of Hurricane Katrina's wind fields for storm surge and wave hindcasting, Ocean Eng. **37**(1), 26–36 (2010)

25.70 M.D. Powell, S.H. Houston: Surface wind fields of 1995 Hurricanes Erin, Opal, Luis, Marilyn, and Roxanne at landfall, Mon. Weather Rev. **126**(5), 1259–1273 (1998)

25.71 V.J. Cardone, A.J. Broccoli, C.V. Greenwood, J.A. Greenwood: Error characteristics of extratropical-storm wind fields specified from historical data, J. Petrol Technol. **32**(5), 872–880 (1980)

25.72 G.J. Holland: Tropical cyclone structure in the Southwest Pacific, Bull. Am. Meteorol. Soc. **61**(9), 1132–1132 (1980)

25.73 E.F. Thompson, V.J. Cardone: Practical modeling of hurricane surface wind fields, J. Waterw. Port Coast. Ocean Eng. **122**(4), 195–205 (1996)

25.74 P.J. Vickery, D. Wadhera, M.D. Powell, Y.Z. Chen: A hurricane boundary layer and wind field model for use in engineering applications, J. Appl. Meteorol. Clim. **48**(2), 381–405 (2009)

25.75 H.E. Willoughby, R.W.R. Darling, M.E. Rahn: Parametric representation of the primary hurricane vortex. Part II: A new family of sectionally continuous profiles, Mon. Weather Rev. **134**(4), 1102–1120 (2006)

25.76 National Oceanic and Atmospheric Administration: Hurricane Sandy advisory archive, http://www.nhc.noaa.gov/archive/2012/SANDY.shtml?

25.77 M.D. Powell, T.A. Reinhold: Tropical cyclone destructive potential by integrated kinetic energy, Bull. Am. Meteorol. Soc. **87**, 513–526 (2007)

25.78 I.R. Young: Parametric hurricane wave prediction model, J. Waterw. Port Coast. Ocean Eng. **114**(5), 637–652 (1988)

25.79 V.S. Kumar, S. Mandal, K.A. Kumar: Estimation of wind speed and wave height during cyclones, Ocean Eng. **30**(17), 2239–2253 (2003)

25.80 J.K. Panigrahi, S.K. Misra: Numerical hindcast of extreme waves, Nat. Hazards **53**(2), 361–374 (2010)

25.81 K. Hasselmann, T.P. Barnett, E. Bouws, H. Carlson, D.E. Cartwright: Measurements of wind-wave growth and swell decay during the Joint

North Sea Wave Project (JONSWAP), Ergänz. Dtsch. Hydrogr. Z. A **8**(12) (1973)
25.82 S.A. Tayor: Parameterization of Maximum Waveheights Forced by Hurricanes: Application to Corpus Christi, Texas, Master Thesis (Texas AM University, Texas 2012)
25.83 J.H.G.M. Alves, H.L. Tolman, Y.Y. Chao: Forecasting hurricane-generated wind waves at NOAA/NCEP. JCOMM Tech. Rep. 29, WMO/TD-No 1319 (2004)
25.84 J.C. Dietrich, M. Zijlema, J.J. Westerink, L.H. Holthuijsen, C. Dawson, R.A. Luettich, R.E. Jensen, J.M. Smith, G.S. Stelling, G.W. Stone: Modeling hurricane waves and storm surge using integrally-coupled, scalable computations, Coast. Eng. **58**(1), 45–65 (2011)
25.85 R.H. Weisberg, L.Y. Zheng: Hurricane storm surge simulations for Tampa Bay, Estuaries Coasts **29**(6), 899–913 (2006)
25.86 D.T. Resio, J.J. Westerink: Modeling the physics of storm surges, Phys. Today **61**, 33–38 (2008)
25.87 A.B. Kennedy, U. Gravois, B.C. Zachry, J.J. Westerink, M.E. Hope, J.C. Dietrich, M.D. Powell, A.T. Cox, R.A. Luettich, R.G. Dean: Origin of the Hurricane Ike forerunner surge, Geophys. Res. Lett. **38**, L08608 (2011)
25.88 J.L. Irish, D.T. Resio: A hydrodynamics-based surge scale for hurricanes, Ocean Eng. **37**(1), 69–81 (2010)
25.89 J.L. Irish, D.T. Resio, J.J. Ratcliff: The influence of storm size on hurricane surge, J. Phys. Oceanogr. **38**(9), 2003–2013 (2008)
25.90 J. Norman: Katrina's dead. Sun Herald, A1, A8-9 17 February 2006
25.91 R.A. Hoover: Empirical relationships of the central pressure in hurricanes to the maximum surge and storm tide, Mon. Weather Rev. **85**, 167–174 (1957)
25.92 W.C. Conner, R.H. Kraft, D.L. Harris: Empirical methods for forecasting the maximum storm tide due to hurricanes and other tropical storms, Mon. Weather Rev. **85**, 113–116 (1957)
25.93 Y.K. Song, J.L. Irish, I.E. Udoh: Regional attributes of hurricane surge response functions for hazard assessment, Nat. Hazards **64**(2), 1475–1490 (2013)
25.94 J.L. Rego, C.Y. Li: On the importance of the forward speed of hurricanes in storm surge forecasting: A numerical study, Geophys. Res. Lett. **36**, L07609 (2009)
25.95 A.W. Niedoroda, D.T. Resio, G.R. Toro, D. Divoky, H.S. Das, C.W. Reed: Analysis of the coastal Mississippi storm surge hazard, Ocean Eng. **37**(1), 82–90 (2010)
25.96 C.E. Synolakis, L. Kong: Runup measurements of the December 2004 Indian Ocean tsunami, Earthq. Spectra **22**(S3), S67–S91 (2006)
25.97 S. Tadepalli, C.E. Synolakis: The run-up of N-waves on sloping beaches, Proc. R. Soc. A **445**, 99–112 (1994)
25.98 Y. Wei, E.N. Bernard, L. Tank, R. Weiss, V.V. Titov, C. Moore, M. Spillane, M. Hopkins, U. Kanoglu: Real-time experimental forecast of the Peruvian tsunami of August 2007 for US coastlines, Geophys, Res. Lett. **35**, L04609 (2008)
25.99 C.E. Synolakis: The Runup of Long Waves, Ph.D. Thesis (California Institute of Technology, Pasadena 1986)
25.100 C.E. Synolakis: The runup of solitary waves, J. Fluid Mech. **185**, 523–545 (1987)
25.101 S. Tadepalli, C.E. Synolakis: Model for the leading waves of tsunamis, Phys. Rev. Lett. **77**(10), 2141–2144 (1996)
25.102 P.A. Madsen, D.R. Fuhrman, H. Schaeffer: On the solitary wave paradigm for tsunamis, J. Geophys. Res. **113**, L12012 (2008)
25.103 E.A. Okal, C.E. Synolakis: Source discriminants for near-field tsunamis, Geophys. J. Int. **158**, 899–912 (2004)
25.104 H. Galbraith, R. Jones, R. Park, J. Clough, S. Herrod-Julius, B. Harrington, G. Page: Global climate change and sea level rise: Potential losses of intertidal habitat for shorebirds, Waterbirds **25**(2), 173–183 (2002)
25.105 D.M. Alongi: Mangrove forests: Resilience, protection from tsunamis, and responses to global climate change, Estuar. Coast. Shelf Sci. **76**(1), 1–13 (2008)
25.106 C. Craft, J. Clough, J. Ehman, S. Joye, R. Park, S. Pennings, H.Y. Guo, M. Machmuller: Forecasting the effects of accelerated sea-level rise on tidal marsh ecosystem services, Front. Ecol. Environ. **7**(2), 73–78 (2009)
25.107 M.L. Chu-Agor, R. Munoz-Carpena, G. Kiker, A. Emanuelsson, I. Linkov: Exploring vulnerability of coastal habitats to sea level rise through global sensitivity and uncertainty analyses, Environ. Model. Softw. **26**(5), 593–604 (2011)
25.108 D. Elliot: Saltwater from Gulf invades Mississippi River, http://www.npr.org/2012/08/21/159567048/saltwater-invades-mississippi-river (2012)
25.109 L.T. Dauer, P. Zanzonico, R.M. Tuttle, D.M. Quinn, H.W. Strauss: The Japanese tsunami and resulting nuclear emergency at the Fukushima Daiichi power facility: Technical, radiologic, and response perspectives, J. Nucl. Med. **52**(9), 1423–1432 (2011)

26. Statistical Characterization of Hazards and Risk in Coastal Areas

Donald T. Resio, Mark A. Tumeo, Jennifer L. Irish

We examine the foundation for hazard/risk assessment and its application to coastal problems. Historically, emphasis was on specifying expected values of wind waves and storm surges; however, as shown by the recent tsunamis in Southeast Asia in 2004 and in Japan in 2011, there are critical parts of the world where tsunamis represent the dominant threat to coastal communities. Recently, there has been an increased awareness of the combined effects of heavy rainfall and/or river discharge with surge levels and strong winds. This forcing combination played an important role in the flooding in southern Louisiana during Hurricane Isaac in 2012, where water levels exceeded the 500-year return interval levels. Such forcing combinations complicate both the modeling systems required for their simulation and the treatment of the multivariate probabilities that define the relative importance of their impacts.

We begin with a set of consistent hazards and risk definitions, along with comparative definitions from other fields. This should help readers who have focused primarily on traditional coastal hazards and risks understand the broader context of risk assessment and also allow readers with a broader perspective gain insight into the specific nature of coastal hazards and risk. Following this, we introduce the basic concepts used in estimating coastal hazards and risks. We then examine the historical perspective for the evolution of coastal risk assessment, beginning with early deterministic methods and culminating in a recent transition to probabilistic methods. The steady increase in the ability of probabilistic methods to deal with persistent problems such as the lack of data and uncertainty is documented as a part of this transition.

26.1 **Overview of Risk and Uncertainty** 567
26.1.1 Definitions of Basic Terms 568
26.2 **Quantifying Coastal Hazards/Risks** 570
26.3 **Historical Perspective** 574
26.3.1 The Development of Deterministic Methods 575
26.3.2 The Development of Probabilistic Methods 577
26.3.3 The Development of the Historical Storm Method for Estimating Coastal Extremes .. 581
26.3.4 The Development of Alternative Methods 583
26.3.5 Probabilistic Analyses of Extratropical Storms 586
26.3.6 Future Directions and Final Comments 586
26.4 **Summary** 587
26.5 **Nomenclature** 587
26.A **Appendix: Glossary of Probability and Risk Terms** 587
References 593

26.1 Overview of Risk and Uncertainty

Technically, *risk* is defined as the mathematical probability that a specific outcome will occur multiplied by the magnitude of the impact of that occurrence. However, there are many other definitions in use, and while a technical working definition such as the one presented is useful, it is more important that there is a contextual understanding of the concept.

While *risk* is now used in some manner across almost every discipline, the entire cultural framework upon which the concept is based is only a few hundred years old (and relatively young as a basis compared to other engineering and mathematical concepts). Most ancient cultures took the stance that the future was either fated (deterministic) or subject to the whims of

powers beyond our control or understanding. There is no concept of *risk* in such cultures beyond the fear of certain punishment should you anger the powers that be. The roots of our conception of risk arose in the Renaissance, as cultures in the west broke away from and challenged long-held beliefs. One of the critical turning points for the concept of risk as we know it was the development of the concept of *probability* – the idea that the future was not determined or unpredictable, but that there was the possibility to make predictions based on numbers, not oracles or soothsayers. The origin of this truly revolutionary concept is recounted by *Bernstein* as follows [26.1]:

> *In 1654, a time when the Renaissance was in full flower, the Chevailier de Mere, a French nobleman with a taste for both gambling and mathematics, challenged the famed French mathematician Blaise Pascal to solve a puzzle. The question was how to divide the stakes of an unfinished game of chance between two players when one of them is ahead. The puzzle had confounded mathematicians since it was posed some 200 years earlier by the monk Luca Paccioli ... Pascal turned for help to Pierre de Fermat, a lawyer who was also a brilliant mathematician. The outcome of their collaboration was intellectual dynamite. What might appear to have been a seventeenth-century version of the game of Trivial Pursuit led to the discovery of the theory of probability, the mathematical heart of the concept of risk.*

Hence, imbedded in the concept of risk is the inherent belief in predictability, at least in a probabilistic sense. This concept, however imbedded in our current culture, stands in stark contradiction to the well-known adage that one cannot accurately predict the future. In the modern world, this contradiction is resolved through the belief that if we knew enough, we could accurately predict the future. This belief in scientific determinism (the same concept from ancient times that was *overturned* by the concept of risk) was articulately stated by *Pierre-Simon Laplace* in his landmark 1812 treatise *Essai philosophique sur les probabilités* (*A Philosophical Essay on Probabilities* as translated by Truscott and Emory in 1902) [26.2]:

> *We may regard the present state of the universe as the effect of its past and the cause of its future. An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be uncertain and the future just like the past would be present before its eyes.*

We have therefore, in many ways, come full circle as a culture. Through the concept of risk we have moved away from the underpinning belief in determinism to accept that there is an unknown and unknowable future. Yet, through the application of mathematical probabilities, which gave rise to the concept of risk, we also have developed the idea of a scientific determinism, wherein if all lack of knowledge could be eliminated through scientific inquiry, we would once again have a certain and determined future.

The cultural concept of scientific determinism through the elimination of uncertainty and evaluation of undetermined but manageable future options are the heart of risk assessment. The purpose of the study of risk is to manage the future by using our best knowledge and current understanding to reduce uncertainty and quantify the remaining uncertainty through probabilistic prediction.

26.1.1 Definitions of Basic Terms

Because the concepts of risk and uncertainty are now so ingrained into our society, the use of terms associated with *risk and uncertainty*, including *hazard*, *probability*, *risk assessment*, *risk analysis*, and in fact the words *risk* and *uncertainty* themselves, are used frequently in differing and often confusing ways. When one adds to this situation the fact that risk and uncertainty inherently rely on other ill-defined and often misused concepts such as *error* and *variability*, it becomes apparent that before entering into a discussion of the application of hazard probability and risk assessment to coastal systems, it is necessary to ensure that there is a consistent and uniform understanding of terms. A full list of definitions for probability and risk related terms used in describing coastal hazards is given in the Sect. 26.A; however, several key definitions are retained here that are most pertinent to discussions within this chapter:

- *Probability*: In its most basic sense, probability is the chance or likelihood that something will happen. Qualitatively, the more likely an event is to happen, the more *probable* the event. Quantitatively, probability is a value between 0 and 1, with 1 representing absolute certainty of the event occurring. The probability of an event is typically measured as the ratio of the number of times an event occurred over the total number of times the event could have occurred. For example, if we consider the event to be the occurrence of precipitation on any given day,

we would collect information on whether it rained on any given day for a period of time (say 1 year). The probability of rain would then be

$$\text{probability of precipitation} = \frac{\text{number of days with precipitation}}{\text{Total number of days}} .$$

- *Probability distribution*: The basic definition of probability works for discrete events, but when the event or parameter being evaluated is continuous in nature (i. e., the maximum temperature on any given day, the stage of a river on any given day), then a single probability is not sufficient. Instead, possible values are grouped into discrete ranges and the number of occurrences in that range are counted, then divided by the total number of measurements.
- *Risk*: The potential for realization of unwanted, adverse consequences to human life, health, property, or the environment. The estimation of risk is usually quantified using the expected value of the conditional probability of the event occurring times the consequence of the event given that it has occurred. This definition is currently used by the Society of Risk Analysis. Mathematically, this is expressed as

$$R = P(A) * P(B|A) ,$$

where R – Risk (a probability from 0 to 1), $P(A)$ – probability of event (A) occurring, $P(B|A)$ – probability of a consequence (B) occurring given that event A occurred.
This technical definition is what will be used through this handbook. However, there are other definitions in other fields of which the knowledgeable practitioner should be aware (Sect. 26.A). Besides having multiple potential definitions, risk can also be differentiated by types. Different types of risk, even if they are of the same quantitative value, are often managed differently or even ignored. Some of the common risk types that greatly influence both how these risks are assessed, managed and communicated are defined below (Sect. 26.A for full list).
  - *Actual risk*: A scientifically verifiable risk. For example, it is well researched and documented that smoking places you at-risk for cancer [26.3–6]. Actual risk is sometimes referred to as *objective risk*, but whether risk is subjective or objective is related more to its ability to be measured than it is to its actual verifiability.
  - *Perceived risk*: Risk that is thought to exist by an individual or group that either is understated or is exaggerated. This often occurs in situations where the public is misinformed or in which media reports instill unnecessary panic. Food safety concerns often top the list of such events and lead to significant public policy debates, see, for example [26.7].
  - *Assumed risk*: Risk that is taken by choice. Assumed risk can be quantifiably large or small, and actual or perceived. For example, individuals who choose to partake in risky activities (skydiving, mountain climbing) choose to assume the relatively large risks associated with these activities, but choosing to drive a car, take medicines, or be involved in day-to-day activities all involve some assumed risk.
- *Imposed risk*: Risk that is forced upon an individual, either without the knowledge of the individual or if known, without consent. For example, second-hand smoke exposure is seen as an imposed risk [26.8–10]. Natural events such as earthquakes, tropical cyclones (e.g., hurricanes, typhoons), and extreme weather events are to a large extent *imposed* risks, but to some extent individuals assume that risk based on where they choose to live. For an interesting approach to this, see the work by [26.11].
- *Uncertainty*: Uncertainty arises from the fact that any model (including risk assessment models) cannot provide 100% *certainty* in its results. This uncertainty arises from two major sources.
  - *Epistemic uncertainty*: Epistemic uncertainty arises from the fact that there are things about the natural system we are analyzing that we do not know. This *lack of knowledge* can be because the knowledge is not yet scientifically available – and as such, it can be reduced (along with the resulting epistemic uncertainty) going forward through additional data, experimentation, theoretical development, and scientific inquiry. However, lack of knowledge also includes things that we do not even know that we do not know. This area of *lack of knowledge* is more difficult, because it is not easily identifiable and as such, becomes included in either variability, or is called *error* when comparing model results to reality.
  - *Alleatory uncertainty*: Variability is a range of potential values for a given parameter. Variability is an inherent characteristic of natural processes (also called natural variation). It is describable using probability distributions. The result of natural variation is sometimes called alleatory uncertainty. Variability in natural processes (and the resulting alleatory uncertainty) cannot be reduced, as it is an inherent property of the process itself.

26.2 Quantifying Coastal Hazards/Risks

A detailed treatment of all coastal hazards and risks is beyond the scope of the work presented here. Environmental/health hazards and risks such as those related to water quality, chemical spills, and salinity intrusions will not be treated here; however, it is hoped that the methods discussed in this section will be useful, at least in a general sense, to those working in these areas. Additionally, due to the fact that much of the recent research on coastal hazards/risks has been undertaken in response to Hurricane Katrina, much of the focus relating to recent developments will be on storm surges produced by tropical cyclones. Although extratropical storms play a very large role in many coastal areas, the lack of a recent catastrophic event, such as flooding the subway system in New York City before Hurricane Sandy (which was a hybrid mix of tropical cyclone and extratropical characteristics) has meant that methods used for evaluation of these hazards/risks have received far less attention than they probably deserve. A single section (Sect. 26.3.6) will be devoted here to specific issues related to extratropical storms, following our general treatment of extreme and related hazards/risks.

Before launching into our treatment of hazard probabilities, it should be recognized that both epistemic and alleatory uncertainties need to be considered within any thorough analysis of hazards and associated risks. Limitations in theoretical/numerical predictions come from many sources, including simplifications in our models to make nature fit our models, the lack of details in the information utilized by models, and shortcomings in our theoretical knowledge on which model are based. Similarly, the lack of sample size continues to be a serious problem in our estimations of extreme values. In some studies, many years of record are simulated and statements are made which might lead the reader to believe that a commensurate gain in information over the existing historical record has somehow been achieved. However, a detailed review of these studies reveals the obvious fact that alleatory uncertainty cannot be overcome by simply executing numerical models over many randomly posed simulations, since the information upon which the parent distribution is based remains the same.

Many available textbooks provide excellent formal reviews of statistics, derivations of distributions and applications specific to hazard/risk assessment [26.12–16]. Some recent examples of similar books which focus more on risk issues include [26.17, 18]. Here we will present only a summary of information which should allow a reader to gain sufficient familiarity with these basic statistical methods. As shown in Sect. 26.1, risk is associated with the probability of consequences. Since design thresholds are selected in a manner intended to limit consequences to an *acceptable* level, the study of natural hazards and associated risks tends to focus to either on extremely large or extremely small values in a distribution. Here we will focus on the upper end of the distribution, since these represent the major design consideration for most natural disasters in coastal areas.

We start here with the conventional definition of the probability density function (PDF) of a continuous variable x with the following properties

$$p(x) \geq 0 \,, \tag{26.1a}$$

$$\int_{-\infty}^{\infty} p(x)\,\mathrm{d}x = 1 \,, \tag{26.1b}$$

$$\int_{a}^{b} p(x)\,\mathrm{d}x = P(a < x < b) \,, \tag{26.1c}$$

where $p(x)$ is the probability density function for the variable x and $P(a < x < b)$ is the probability of x falling between the values of a and b. It is clear from the form of (26.1c) that the PDF is not a dimensionless function. Since $\mathrm{d}x$ has the same units as x, the dimensions of $p(x)$ must be x^{-1}. As an example if we allow the PDF to be given by

$$p(x) = k\mathrm{e}^{-qx} \,, \tag{26.2}$$

k and q must both have dimensions of x^{-1}; otherwise the integral would not produce results with the correct dimensions. A simple example with $q = 1$, $a = 1$ and $b = 3$ yields a value for (26.1c) equal to $\mathrm{e}^{-1} - \mathrm{e}^{-3} = 0.318$.

The cumulative distribution function (CDF) of x is equal to the probability that a value is less than or equal to a particular value of x,

$$F(x) = \int_{-\infty}^{x} p(x)\,\mathrm{d}x \,, \tag{26.3}$$

where $F(x)$ is the CDF of x. $F(x)$ is a dimensionless function since it is the product of two functions with dimensions x and x^{-1}; however, in engineering applications, it is very important to recognize that it becomes implicitly related to time when we convert to expected mean recurrence intervals. This is because, even though the time between samples does not appear explicitly in $F(x)$, it enters into the CDF implicitly via the time interval between samples. When we convert (26.3) into

a form for estimating the expected interval between exceedances of a particular value of x, we obtain a relationship of the form

$$T(x) = \frac{1}{\lambda(1-F(x))}\,, \tag{26.4}$$

where $T(x)$ is the mean return period (or interval), and λ is the sampling frequency. Even though (26.4) is often written without λ in the denominator, its omission can lead to a dimensional inconsistency. The units used to define λ (decade^{-1}, year^{-1}, month^{-1}, day^{-1}, etc.) provide the information which characterizes $T(x)$. In engineering, it is common practice to use years for $T(x)$, in which case the expected value annual probability of exceeding a particular value of x is given by $1/T(x)$. Thus, a 100-year mean return interval is associated with an event that has an annual exceedance probability (AEP) of 0.01.

Here we will use an example of some results taken from wave hindcasts (the reconstruction of wave conditions from simulations of past events) for a location in the southern Gulf of Mexico area to show the importance of considering the frequency term in our estimation of return periods. In this region, two different meteorological phenomena can generate large waves, tropical cyclones, and *Northers*. The latter type of wave generation occurs when a cold front crosses the Gulf of Mexico and produces very strong surface wind speeds behind it. For both hurricanes and Northers, the results are given for sites located in deep water and in nominal depths of 20 m. In Fig. 26.1, we plot $-\ln\{-\ln[F(x)]\}$ against x, using the 20 data points available in each specific sample. Points that plot along a straight line in such a graph are consistent with a Gumbel distribution.

From Fig. 26.1, it appears that Norther and hurricane hazards are roughly comparable in this region of the Gulf of Mexico; however, if we include the frequency parameter and estimate the mean return interval, we obtain Fig. 26.2. In this case, the hazards related to hurricanes are shown to be considerably larger than those related to the Northers.

Since the concept of a mean interval between occurrences (return period) is simpler for some people to grasp, it is often used instead of the AEP concept in applications; and for that reason we will use it in some of our figures and discussions. However, it is very important to recognize that a *100-year event* is not constrained to occur only once every 100 years. For example, we can define the probability of exceeding a given value of x_c as 1 minus the probability of that level not being exceeded in any year, in which case the probability of having no exceedances in n years is given by

$$\begin{aligned} &P(x < x_c|n \text{ years}) \\ &= 1 - P(x < x_c|\text{year } 1) \times P(x < x_c|\text{year } 2) \ldots \\ &\quad \times P(x < x_c|\text{year } n) \\ &= 1 - \prod_{i=1}^{n} P_i(x < x_c) = 1 - [1 - F(x_c)]^n\,, \end{aligned} \tag{26.5}$$

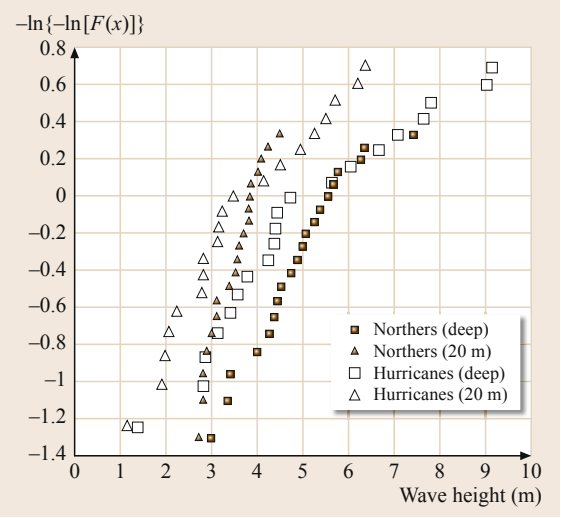


Fig. 26.1 Plot of the $-\ln\{-\ln[F(x)]\}$ versus x for waves generated at points in the southern Gulf of Mexico by hurricanes and *Northers* at sites in deep water and a depth of 20 m

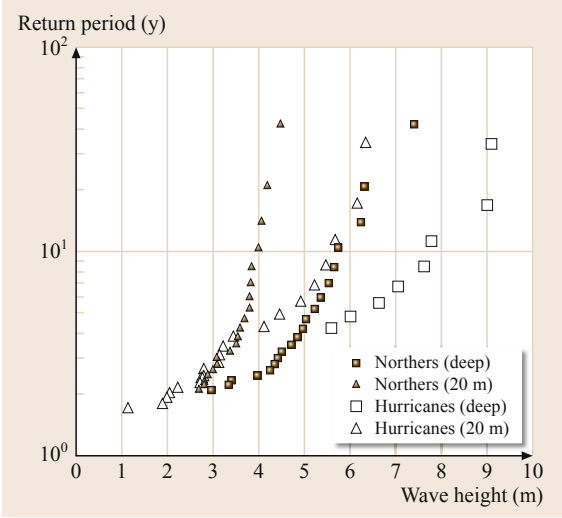


Fig. 26.2 Plot of the mean return period versus wave height for waves generated at points in the southern Gulf of Mexico by hurricanes and *Northers* at sites in deep water and a depth of 20 m

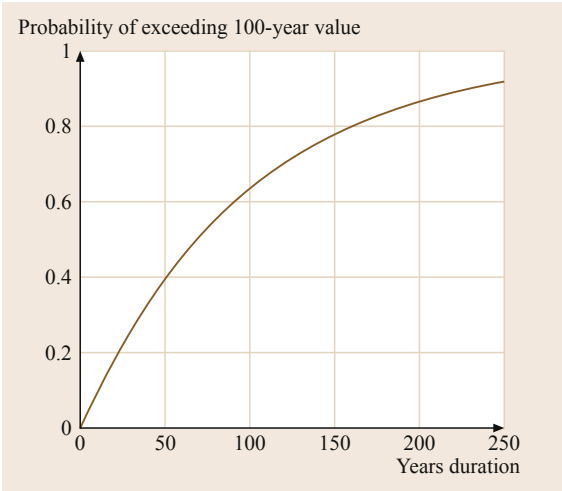


Fig. 26.3 Graph showing the probability of exceeding a value expected on the average only once every 100 years as a function of the number of years of duration

where n is the number of years and i is the year counter. Figure 26.3 gives the estimated encounter probability for an AEP of 0.01 as a function of the expected *design life*, the number of years being considered for the design. It can be seen that the actual exposure to hazards/risks is a clear function of the number of years considered for the design life. Over 30 years the probability of encountering a 100-year exceedance would be 0.2603; and over 50 years the probability of exceedance would be 0.3950.

Equation (26.5) is a bit tedious to use but is instructive regarding the reason why the probabilities become relatively large over a number of years. A second, perhaps simpler form for estimating the probability that a value, x_c, will not be exceeded in n years is to use a Poisson distribution with the integer value for occurrences equal to zero. In this case, if we define $\lambda =$ number of years over which the non-occurrence is being estimated divided by the mean recurrence interval for x_c, we can estimate the equivalent probability to (26.5) as $1-e^{-\lambda}$. For example, the equivalent value for the 30-year example above would be $\lambda = 0.3(= 30/100)$, or $1-e^{-0.3}$, which would equal 0.0259 and the equivalent value for the 50-year example would be $1-e^{-0.5}$, which would equal 0.3935. It should be noted that this would yield the same result for any combination of values of number of years over which the non-occurrence is being estimated and return value for x_c, such as 100 years of non-occurrence for a value with a 200-year recurrence interval for the latter example.

As an interesting side note to the above discussion, we can examine the common practice today of construction deterministic *flood zones* on maps for insurance purposes. If a property had an AEP for flooding of 0.009, it would lie outside of the primary flooding area covered under the program of the Federal Emergency Management Agency (FEMA) in the United States; yet the probability of flooding over 30 and 50 years would still be 0.2375 and 0.3637, respectively, which is only 8 to 9% different than for the property with the AEP of 0.01. Thus a definitive line on a flood map can be very misleading, with people on one side believing that their likelihood of flooding is much different than on the other side, which is certainly not the case.

Equation (26.5) shows that the accurate estimation of the CDF is fundamental to the estimation of hazards and risks. *Fuller* [26.19] credits George W. Rafter with initially pointing this out in 1896; however, the first published study directly linking hazards to a probability distribution in the United States (using the normal distribution) was that of [26.20]. Shortly afterward, *Hazen* [26.21] noted that logarithms of flood stages tended to provide more consistent estimates of CDF, essentially hypothesizing that floods followed a log-normal distribution. Since those early beginnings about a century ago, many studies have improved our understanding of the statistical behavior of extremes. Early studies typically attempted to fit specific distributions directly to empirically selected distributions. An excellent review of these early studies and the fitting methods used in them is contained in [26.22]. As shown there, many different distributions (for example, normal, lognormal, Poisson, generalized extreme value, and Pearson distributions) were developed, tested, and applied to a number of applications through the 1960s. The gist of the fitting methods was based on the concept that the CDF for each distribution could be related to the ranked data. For example, Table 26.1 shows the relationship between the estimated return period and the ranked order of the data, termed the plotting position, for a number of different distributions.

A detailed treatment of all of distributions used for estimating extremes is beyond the purpose of our discussion and Table 26.1 is only included to show how many different empirical approaches had been developed by the middle of the last century. For our purpose, it is sufficient to recognize that distributions can be classified into two general classes, parametric and nonparametric. Parametric distributions are characterized by assumed theoretical forms which include a small number of constants which must be determined by a statistical fitting technique; thus, such distributions constrain the number of degrees of freedom in the data. All of the distributions included in Table 26.1 are parametric. Nonparametric distributions are allowed to vary

Table 26.1 Plotting position formulae recommended by different investigators for the estimation of the return period

Name	Date	Formula[a] for T or $1/P(X \geq x)$
California	1923	$\frac{N}{m}$
Hazen	1930	$\frac{2N}{2m-1}$
Weibull/Gumbel	1939/1959	$\frac{N+1}{m}$
Beard[b]	1943	$\frac{1}{1-0.5^{1/N}}$
Chegodayev	1955	$\frac{N+0.4}{m-0.3}$
Blom	1958	$\frac{N+1/4}{m-3/8}$
Tukey	1962	$\frac{3N+1}{3m-1}$
Gringorten	1963	$\frac{N+0.12}{m-0.44}$

[a] N = total number of items; m = order number of the items arranged in descending magnitude, thus $m = 1$ for the largest item.
[b] This formula applies only to $m = 1$; other plotting positions are interpolated linearly between this and the value of 0.5 for the median event.

in a form consistent with the data elements; hence, they are not necessarily constrained to a small number of degrees of freedom.

Two different types of parametric distributions have been widely used in studies of natural phenomena, distributions which encompass the full range of data and distributions which focus only on the portion of the data with the largest observed values. Examples of coastal applications of the first type can be found in many studies of coastal waves in the 1960s through the 1980s, as discussed in [26.23, 24]. *Hogben* and *Lumb* [26.25] provide one of the earlier comprehensive collections of wave data around the world. Examples of subsequent studies using information covering the entire range of data measured can be found in [26.26–29]. However, here we will focus on the latter class of distributions, since extremes represent the primary variable range in most designs. This helps us avoid many of the issues with samples that have significant serial correlations, due to the shortness of the sampling interval. Most methods of the second type originally used a constant sampling interval (typically 1 year which implied a value of λ equal to one and the time units of years). However, as will be discussed in more detail in the historical perspective section, there are many cases where it is advantageous to allow the time interval between samples to vary. For example, if we are only interested in winds, waves and/or surges from tropical cyclones, the sampling time interval will vary and the λ in (26.4) would be equal to the mean frequency of tropical cyclone occurrence (or one over the mean interval between tropical cyclones). This parameter behaves analogously to the frequency term in the Poisson distribution and hence is often called the Poisson frequency, the average number of storms per year included in the analysis.

Before proceeding, it is important to note that wave extremes are usually posed in two different but related contexts: extremes related to the overall energy content in the local wave field (sometimes termed the sea state) and extremes related to individual waves. In the 1940s, it began to be recognized that the water surface was not well-described by simple unidirectional, monochromatic mathematical formulations. Instead, the wave field was better described as a random surface produced by the superposition of a continuous distribution of wave components in frequency and direction. The total energy within a unit area of water surface can be related to the variance of that surface around its mean position. The so-called *significant wave height* is approximately equal to the average of the one-third highest wave heights over an interval of time. More recently, it has been defined as $H_s = 4\sqrt{E_0}$, where H_s is the significant wave height and E_0 is the total variance of the water surface. The distribution of individual wave heights depends on the total energy and some additional factors which influence the degree of nonlinearity within the wave field, but these topics are beyond the main scope of what we will treat here. Significant wave heights tend to be used in applications dependent on time-averaged properties of waves (such as wave set-up, wave overtopping rates, beach erosion rates, etc.), while individual wave heights tend to be used in applications dependent on instantaneous forces (such as structural components in offshore facilities and some elements of ship design).

In nonparametric estimation of extremes, the probability of future events is obtained via methods that resample the available historical sample, under the assumption that future distributions will follow the same general form [26.30–34] This method is very powerful for application to extremes when the number of samples is large and the value range of interest falls within the range contained in the historical (initial) sample. In the context of coastal applications to extremes of waves, winds, and surges, probably the most widely used method of this kind is that of *Scheffner* et al. [26.35], which is termed the empirical simulation technique (EST). In this method, the CDF is estimated

in terms of the rank of the ordered values (m) and the total number of observations N,

$$F(x) = \frac{m}{N+1} , \tag{26.6}$$

which can be seen as an equivalent form to the Gumbel plotting position given in Table 26.1; the return period there is defined as the inverse of the nonexceedance probability. However, the manner in which such information is used in nonparametric methods is very different to its use in parametric methods.

To help demonstrate the basic premise of resampling, let us assume that our data includes 30 samples taken over 50 years. Thus, the range of the CDF covered would be $1/31(0.0323)$ to $30/31(0.9677)$, or more generally $1/(N+1)$ to $N/(N+1)$, where N is the number of samples. Based on (26.4), we would estimate that the return periods explicitly covered by this data cover the range 1.72–51.67 years, recalling that the value of λ would be $30/5$ for this example. If a specific design requirement fell into this range, the results from the EST could be used directly, without any extrapolation. However, two key points should be understood in applications of this methodology. First, outside of the CDF range covered by the data, it is necessary to extrapolate parametrically to obtain estimates of the CDF as a function of x; so in this example, for return periods larger than 51.67 years or less than 1.72 years, some fitting algorithm similar to that used in parametric methods must be utilized. Second, any sample drawn from nature cannot be regarded as the *parent* population constructed from of all possible samples.

Let us assume for simplicity that we have a fitting methodology, which extends the range of the CDF from 0 to 1. In this case, we can use a random number generator to construct a random sequence of values for both the number of storms in a particular year (based on a Poisson distribution) and the value of the variable of interest in all of the storms that occur. The latter of these is obtained by assuming that each random number is a CDF value and inverting this value (using the defined CDF-x relationship) to obtain an estimate of x. If we simulated 100 sets of 50 years using this method, we could obtain an estimate of the magnitude of the variability we might expect in a particular 50-year sample; however, it would still be assuming that the initial sample represented the population characteristics. In fact, the actual variability in a subsequent 50-year interval will include contributions from the potential deviation of the initial sample from the overall population sample plus the contribution of deviations due to sampling given an initial distribution.

26.3 Historical Perspective

It is important to recognize the role of thresholds in coastal hazards and their relevance to risk. For example, a flood over a particular value might overtop a levee, or a wave height greater than some limit might impact critical components of an offshore oil platform, and a toxin level over some threshold might endanger the environment. Any of these occurrences might have serious consequences on lives and property. Thus, the success or failure of engineering designs and decision making in general can be very sensitive to whether or not a particular threshold is exceeded. However, is it really this simple? Are we certain that any value lower than an exactly specified threshold (even by a minuscule amount) will not have serious consequences, while any value equal to or greater than the threshold will definitely have serious consequences? Given that we never have complete information about the processes and the probabilities of different combinations of environment conditions, it is intuitively obvious that such thresholds should be considered only as an approximation to a level at which consequences are initiated and our estimates of the probability of these thresholds being surpassed cannot be exact. Here we see that both alleatory and epistemic uncertainty affect our estimate of this threshold probability.

Traditionally, the role of conservatism is at least in part to compensate for uncertainties in our knowledge base, both alleatory and epistemic uncertainty, via some factor used to shift the design level in a manner that compensates for the lack of knowledge (i. e., an attempt to account for inaccuracies in our theories and models of nature and/or in our quantification of the probabilities of factors which influence these levels). In many engineering fields, it has been common practice to include the effects of such uncertainty through the application of a codified *safety factor* or some similar device; however, newer methods are beginning to evolve toward the consideration of risk in a more formal probabilistic sense.

Historically, two fundamentally different approaches have been used to estimate threshold values for the design of critical infrastructure: approaches based on the establishment of a deterministic estimate of an upper limit for a hazard level and approaches

based on the probabilistic estimation of a design level associated with a selected value for an acceptable hazard/risk level. Our review here will show that the field of engineering has been shifting through the years from the first approach to the second and that this trend is continuing.

Examples of the deterministic upper-limit approach can be found in early concepts used for designs of harbor protection, primarily based on somewhat subjective extrapolations of visual observations [26.36]. They have continued to play a dominant role in the development of design specifications throughout much of the twentieth century, including the estimation of environmental conditions for siting nuclear power plants in coastal areas, the formulation of design parameters for large dams, and the design of large coastal structures for protection from tsunamis. These methods tacitly assumed that there was no epistemic uncertainty in their estimated maximum values; hence, the estimated deterministic levels could be interpreted as absolute upper limits which could not be exceeded (i. e., the annual exceedance probability was zero or, conversely, the average return period was infinite). As will be discussed subsequently, applications of ultimate limits assuming no uncertainty can be under-conservative compared to probabilistic design levels; however, functioning structures based on this earlier approach still exist in many areas of the world.

Examples of the second approach include design specifications for most site-specific coastal structures around the world since the middle of the last century, when the combination of measurements and predictive tools attained sufficient skill to permit such estimates. Initially, these estimates were based predominantly on historical storms; but with time, there has been a slow transition from designs based only on events occurring in the historical record at a given site toward designs that consider all events that could occur at that site. Early during the development of the probabilistic approach, the accepted hazard level was taken to be associated with an annual exceedance probability of 0.01 (a mean interval between exceedances of 100 years). As shown earlier in (26.5) and the discussion following that equation, such a probability level does not represent a very low risk for many critical designs or applications, since there is a 1 in 3 chance that such a value will be exceeded in about 40 years.

26.3.1 The Development of Deterministic Methods

Before modern risk concepts, analytical tools, and computer codes existed, it was still necessary to estimate design parameters for various engineering applications. For example, in the 1950s, nuclear power began to be harnessed for commercial applications; and the issue of plant location quickly became extremely important to the overall scientific and engineering communities, as well as the public at large. Initially, a major consideration was given to issues related to national defense; and although the importance of this issue faded with time, it influenced the siting of many of the earliest nuclear power plants. In addition to economic and national defense issues, these plants were expected to be located in areas where natural hazards would not adversely affect performance and safe operation over their expected lifetime. The list of potential hazards considered was quite comprehensive and included the effects of earthquakes (both direct such as ground motions and indirect such as the loss of cooling water); atmospheric phenomena (lightning, tornadoes, tropical cyclones, etc.); flooding due to coastal storms, tsunamis and precipitation; and wave action superimposed on top flooding events.

A well-documented example of a deterministic design approach can be found in the methodology for siting of nuclear power plants in US coastal areas affected by hurricane storm surges. In 1959, the US Army Corps of Engineers (USACE) contracted the National Weather Service (NWS) to develop a hypothetical hurricane that could be used to design hurricane protection projects along the Gulf and Atlantic coasts of the United States. At that time the NWS, as part of its National Hurricane Research Project, set out to define *the most severe storm that is considered reasonably characteristic of a region*, defining a storm with such characteristics as the *Standard Project Hurricane* (SPH) [26.37]. Since their storm definition was related to nuclear power plants, the estimated maximum value possible was consistent with the concept of a very low allowable risk. *Graham* and *Nunn*'s report [26.37] contains a description of the derivation of the storm parameters for SPH along US Gulf and East coasts.

It is clear that all project designs do not require a level of acceptable risk as low as that used for nuclear power plants; so in 1979, the NWS Technical Report 23 [26.38] redefined the SPH in a fashion that would permit it to be more broadly applied. In this report, the SPH was defined as:

> *a steady state hurricane having a severe combination of values of meteorological parameters that will give high sustained wind speeds reasonably characteristic of a given region.*

This removed the idea that the SPH was related to the *most severe storm* expected in a particular area. Since many coastal projects are examined on an economic basis over a fixed interval of time, this re-definition of the SPH offered more general guidance on the relation

cost/benefits for a project. In fact, NWS goes on to say that this revised

> *concept of the SPH has been developed for Gulf and Atlantic coasts as a benchmark against which to judge the hazards for a particular community.*

Also in NWS 23 [26.37], the concept of a *probable maximum hurricane* (PMH) was introduced as:

> *a hypothetical steady-state hurricane having a combination of values of meteorological parameters that will give the highest sustained wind speed that can probably occur at a specified coastal location.*

It is clear from this definition that the PMH was intended to be an event that was much rarer than the SPH; but it is not clear that an objective definition can be gleaned from what is written in NWS 23. In its executive summary, NWS 23 states that the central pressure of the PMH *is simply the lowest sea-level pressure at the hurricane center*. Estimated central pressures for the PMH within the Gulf of Mexico had values ranging from about 887 mb at Port Isabel, Texas to about 891 in the vicinity of Apalachicola, Florida and then diminishing to about 885 at Fort Myers, Florida (Fig. 26.4). These are all values which are substantially lower than the lowest central pressure available in any observations within the Gulf of Mexico at that time and no methodology for their derivation is given. Thus, it is likely that a considerable amount of *expert judgment* was utilized in estimating these values. The treatments of the other parameters used to define the PMH (peripheral storm pressure p_0, radius to maximum winds R_{max}, forward speed of the storm V_f, storm heading θ_f) and the latitude of the eye of the storm φ_c) are all relatively straightforward; however, correlations among the parameters were not addressed.

Besides the apparent subjectivity in the SPH and PMH definitions, other problems with this deterministic approach have arisen. For example, the use of *sustained wind speed* as the fundamental parameter for specifying the combination of PMH characteristics implies that the highest sustained wind speed will always produce the highest storm surge and that other storm parameters are less significant to storm hazard levels and associated risk. However, *Irish* et al. [26.39, 40] have clearly shown that storm size can be equally important to the magnitude of the storm surges. Figure 26.5 from [26.39] shows that Hurricane Katrina, although a much weaker storm than Hurricane Camille at the time of landfall in Mississippi, produced much

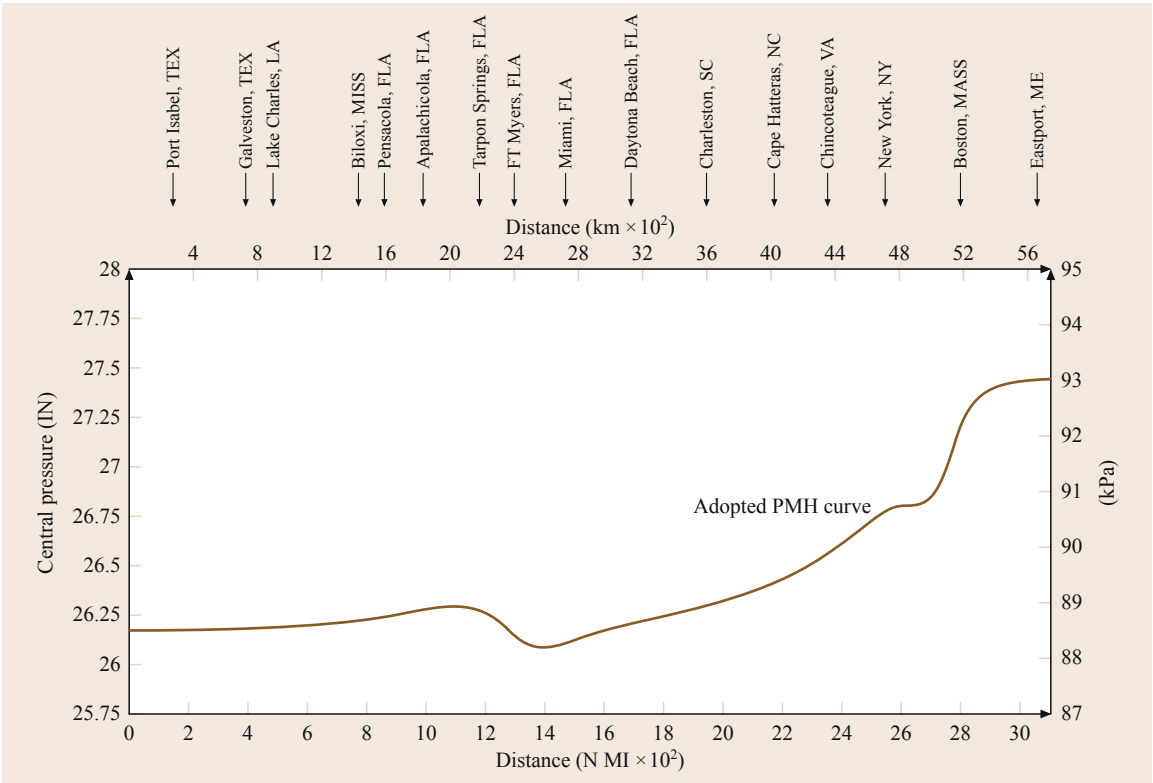


Fig. 26.4 Central pressure for PMH along US Gulf and East coasts (after [26.37])

higher surges there, consistent with results produced by idealized simulations. Additionally, storm approach direction can be important even along straight coastlines but can become exceptionally important on coasts with irregular shapes.

This particular example also affords a good opportunity to examine problems related to overly simplistic representations of the physics of processes in hazard/risk analyses. In the 1950s, the state of the art in surge prediction was quite crude and assumed that inland surges could be estimated via a two-step procedure, with the first step producing an estimate of the wind-driven surge levels at the coastline (neglecting the contributions of wave setup to these water levels) and the second step using hydrologic flow models to convey the water inland (neglecting both wind and wave contributions to inland surge levels). Subsequently, many studies demonstrated the need to estimate inundation in inland areas by including the effects of winds, topography, precipitation, and river discharge within an area [26.41–44]. Wind–wave contributions to coastal flooding have also been shown to be very significant in areas where the nearshore slope is high [26.41]. Based on all these studies, two primary results should be recognized. First, it is not possible to utilize a single factor to characterize waves and flooding in coastal and near-coast areas; and second, the older paradigm in which water levels were first estimated at the coast, using equations appropriate for surge prediction, and then allowed to propagate inland, using equations appropriate for hydrologic flow modeling, is inconsistent with today's state of the art.

There is an obvious analog between good computational models (Chap. 26) and good measurements in the estimation of extremes. No one would trust a design based on measurements known to have deficiencies leading to potentially large biases and random errors. In a similar vein, it is essential to obtain reliable, unbiased estimates from models of winds, waves, and surges for applications to hazard/risk assessment. Only after a modeling system is shown to be capable of generating accurate estimates is it worth expending the huge resources necessary to execute the number of computational simulations in such a system that would be required to provide reliable estimates of extremes.

26.3.2 The Development of Probabilistic Methods

During the latter half of the twentieth century two factors began to emerge which combined to allow significant progress in our capabilities for estimating coastal extremes: an improved theoretical foundation for the estimation of extremes and significantly increased quantities of data for analysis. The latter of these was associated with the development of automated measurement systems for waves and water levels in the 1960s and 1970s. For some purposes, such as daily operations at ports and coastal shipping routes, these direct measurements afforded an excellent source of quantitative information, since requirements for these purposes did not focus on extremes. For other applications, such as the design of breakwaters or the estimated recurrence of coastal flooding levels, the limited data collected from direct measurements proved to be insufficient, due to the short duration of available measurements.

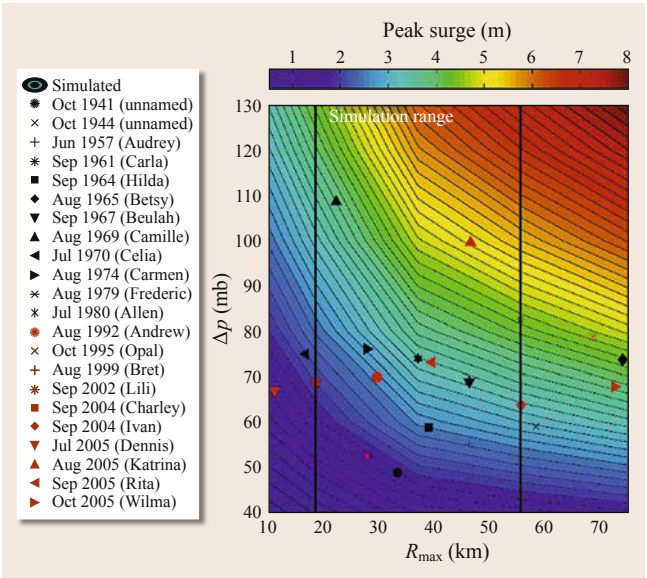


Fig. 26.5 Simulated peak surge as a function of hurricane size (R_{max}) and intensity (Δp) for a 1 : 10 000 bottom slope case. Historical R_{max} and Δp observations are superimposed on the numerical results to indicate peak surge potential of the historical storms (after [26.39], courtesy of American Meteorological Society)

An improved theoretical framework for extremes began to emerge when researchers started to question the application of empirical formulae to extremes based only on its *goodness-of-fit* to a given sample. Pioneers in this field such as Gumbel, Jenkinson, and Gringorten built upon earlier mathematical foundations of *Fisher* and *Tippett* [26.45], and *Gnedenko* [26.46] and began to develop improved methods for treating the behavior of the upper *tail* of the distribution of extremes. Following the work of these individuals, the concept of a generalized extreme value (GEV) distribution was formed. *Borgman* and *Resio* [26.47] offer an example of the application of GEV method to coastal hazards. More recent general reviews of the estimation of extremes from the sequences of random variables can be found in [26.16, 48].

The fundamental concept for the GEV is based on the analysis of asymptotic characteristics of probability functions. In this context, two CDFs are said to be of the same type if there exists constants $a > 0$ and b such that $F_1(ax + b) = F_2(x)$, where the subscripts denote the different CDFs. It was shown by *Fisher* and *Tippett* [26.45] that, excluding certain improper distributions characterized by discontinuities in the CDF, there are only three possible limiting distribution types which meet this constraint. *Jenkinson* [26.49] showed that all three of these asymptotic distributions could be written in a common form

$$F(x) = \exp - \left[1 + \varepsilon \left(\frac{(x-c)}{d}\right)^{\frac{-1}{\varepsilon}}\right] , \qquad (26.7)$$

where c, d, and ε are parameters of the distribution. For the case of $\varepsilon = 0$, this equation is defined over a range of x from $-\infty$ to $+\infty$ and reduces to the two-parameter Gumbel distribution, sometimes termed the Fisher–Tippett Type I distribution.

$$F(x) = \exp[-\exp(-y)], \quad \text{where } y = \frac{x-c}{d} . \qquad (26.8)$$

For $\varepsilon > 0$, the resulting distribution is a Frechet (or Fisher–Tippett Type II) distribution, which is bounded at the low end of the distribution but is unbounded for large x; and for $\varepsilon < 0$, the resulting distribution is a Weibull (or Fisher–Tippett Type III) distribution. The Weibull distribution has an explicit upper limit, but is unbounded in the negative direction.

Returning to Figs. 26.1 and 26.2, we see why a simple plot of x versus $-\ln\{-\ln[F(x)]\}$ will form a straight line if the sample follows a Gumbel distribution. If the points tend to curve upward on such a plot, the distribution is of the Fisher–Tippett Type III form; and if the points tend to curve downward, the distribution is of the Fisher–Tippet Type II form, which predicts larger values of x than the Gumbel distribution for a given value of the CDF.

Shortly after applications with GEV distributions began to become widespread, a second distribution with a somewhat different theoretical basis was introduced. This distribution, the generalized Pareto distribution (GPD), was first published by *Pickands* [26.50] and has been further developed by many subsequent researchers, with some of the more topical for our purposes being *Resnick* [26.14] and *Davison* and *Smith* [26.51]. Methods for fitting the GPD can be found in articles such as [26.52] and are now available in many statistical packages. The GPD can be written in the form

$$F(x|x_c) = 1 - \left(1 + \varepsilon' \frac{x_c}{\mu'}\right)^{\frac{-1}{\varepsilon'}} , \qquad (26.9)$$

which has three parameters (ε', μ', and x_c), similar to the GEV, and also similar to GEV; the GPD has three basic forms depending on ε. With $\varepsilon' = 0$, the GPD approaches an exponential form given by

$$F(x|x_c) = 1 - \exp\left(-\frac{x}{\mu'}\right) . \qquad (26.10)$$

Since the exponential and double exponential (Gumbel) distributions have been shown to converge for mean recurrence intervals greater than 7 years (based on annual data), significant deviations between the GPD and GEV for the case of $\varepsilon = \varepsilon' = 0$ tend toward zero for estimates of values with mean recurrence intervals greater than 7. When ε' is positive, the GPD will be similar to a Fisher–Tippett Type II and when ε' is negative, the GPD will be similar to a Fisher–Tippett Type III.

It is sometimes argued that the GPD is superior to the GEV, due to its formal incorporation of a threshold into its derivation; however, the essence of this distinction lies in the assumption that a sample of values from nature represents a *homogeneous population* in a statistical sense. As will be shown later in this chapter in Sect. 26.3.4, samples in nature often contain a mixture of populations. For example, large surges and waves can be generated along coasts by both extratropical and tropical storms. In smaller basins, squall lines and thunderstorms can become dominant wave and surge producers. Wherever possible it is advisable to try to avoid mixing populations from clearly different sources, but sometimes the ability to recognize more subtle variations can be obscured. A good example of this can be found in the work of *Resio* [26.53], who showed that what appeared to be a Fisher–Tippett II distribution for hindcast wave heights in the Cape Hatteras area was actually the sum of different types of storms definable in terms of how far off the coast they were when they passed Cape Hatteras. Estimated values within the record length are usually not affected by this mixed population problem however, it can become very significant when extrapolating to larger values.

Given the previous discussion, and hopefully without unduly disparaging the excellent theoretical work which has been accomplished in the field of extreme statistics, it should be recognized that given the nuances in the different theories upon which the GEV and GPD are based, it is difficult to argue definitively that one class of distribution is superior in applications to the other. Both distributions represent either a two-parameter distribution (for the limiting case $\varepsilon = \varepsilon' = 0$)

or a three-parameter distribution (for the case that either $\varepsilon \neq 0$ or $\varepsilon' \neq 0$). As noted here, samples of coastal winds, waves, and surges often contain very significant effects related to mixed populations, such as the obvious differences in statistical populations in surges generated by tropical cyclones making landfall some distance away from a site versus surges generated by a direct tropical-cyclone hits. These effects typically far outweigh any differences in theoretical justifications for using a GEV versus a GPD analysis. Because of this, it is probably better to develop a good level of familiarity with a tool of preference and use it appropriately (either the GEV or the GPD) rather than to be able to apply both distributions only marginally.

Returning to the problem of record lengths that are short relative to the mean recurrence interval of the design event, it is necessary to quantify the impact of randomness in the samples on the estimated values before we can assess the importance of this issue. As mentioned at the end of the last section, the quality of individual samples is an important factor in the estimation of extremes. Once routine measurements began to become available, there was an implicit expectation on the part of some that measurements would replace model estimates as the primary source of information used in the estimation of extremes. However, two sources of uncertainty can affect estimated extreme values: epistemic uncertainty in the sampled data (for example, theoretical unknowns and errors in model results) and alleatory uncertainty in the sampled data (natural randomness in the sample). Thus, there is a trade-off between using a short interval of very high quality measurements, removing most of the epistemic uncertainty, and using a long interval of lesser quality model results, which would reduce the alleatory uncertainty but would increase the epistemic uncertainty component.

Many studies in the 1960s and 1970s utilized a combination of theoretical formulations and computer simulations to establish guidelines for the standard deviations of estimates (sometimes referred to as confidence bands) around the value of x at a specific AEP. In most cases, the range of interest for design applications is beyond the range of observations. For example, the typical mean return interval was 100 years in most early design studies. However, in mapping flood zones for FEMA studies today, values up to the 500-year mean recurrence interval are required. For other applications, such as large offshore structures and coastal defenses in many areas of the world, the mean recurrence intervals of interest tend to be in the 1000s or even 10 000s of years, and probabilistic estimates for highly critical infrastructure such as nuclear power plants have been established to use mean recurrence intervals in the range of 1 000 000 to 10 000 000 years. Such estimates must consider the inherent uncertainties in extrapolations that are many times the length of the observational record. Estimates of standard deviations can be obtained by a number of methods based on the GEV or GPD forms [26.49, 54–56].

A useful form for the sampling uncertainty was given by *Gringorten* [26.57, 58]. Using a combination of theoretical characteristics and numerical testing, he showed that the expected root-mean-square (RMS) error of an estimated value as a function of mean recurrence interval in a Gumbel distribution could be estimated from the relationship

$$\sigma_T = \sigma \sqrt{\frac{1.1000y^2 + 1.1396y + 1}{N}} \,, \tag{26.11}$$

where σ is the distribution standard deviation; σ_T is the rms error at return period, T; N is the number of samples used to estimate the distribution parameters; and y is the reduced Gumbel variant given by $y = (x - \alpha)/\beta$; x is the value of the variable; and α and β are the parameters of the Gumbel distribution.

For a Gumbel distribution, the reduced variate and return period are related by

$$y = -\ln\left[\ln\left(\frac{T}{T-1}\right)\right] \,, \tag{26.12}$$

which for $T > 7$ years approaches an exponential form given by

$$\left(T - \frac{1}{2}\right) \to \mathrm{e}^y \,. \tag{26.13}$$

Equation (26.11) shows that the rms error at a fixed return period is related to the distribution standard deviation and the square root of a dimensionless factor involving the ratio of different powers of y (y^2, y^1, and y^0) to the number of samples used to define the parameters. In our previous description of the Gumbel distribution, we saw that the CDF for such a distribution has a double exponential form.

By the method of moments, the Gumbel parameters can be shown to be given by

$$\mu = c - \gamma d, \quad \sigma = \frac{\pi}{\sqrt{6}} d \,, \tag{26.14}$$

where γ is Euler's constant ($= 0.57721\ldots$), μ is the distribution mean, and σ is the distribution standard deviation.

Thus, the distribution standard deviation is related to the coefficient of variability for the distribution and

can be used for estimating the expected width of the confidence bands for a specified return period. The dependence of the confidence limits as a function of the mean recurrence interval and the presence of the square root of the sample number in the denominator show that these limits will become very large when one is extrapolating to several times the record length. Figure 26.6 gives an example for return periods greater than 100 years given a distribution standard deviation of only 0.25 m, which is somewhat smaller than found in many natural data sets. This figure shows the importance of record length to alleatory uncertainty (due to natural variation, see Sect. 26.A). The estimation of a 10% exceedance or 1% exceedance above the Gumbel value at a specific return period can be obtained by multiplying the estimated value of σ_T by 1.08 and 2.33, respectively, since the assumption of normality is inherent in this estimation method.

Although (26.11) through (26.14) were initially derived for applications to annual maxima, they can be adapted to any time interval for data sampling in a straightforward manner. For the case of tropical cyclones, the average interval between storms (the inverse of the Poisson frequency used in the compound Gumbel–Poisson distribution) can be used to transform (26.11) into the form

$$\sigma'_T = \sigma \sqrt{\frac{1.1000y'^2 + 1.1396y' + 1}{N'}} \,, \tag{26.15}$$

where σ is the distribution standard deviation; σ'_T is the rms error at return period, $T' = T/\hat{T}$, where $\hat{T}$ is the average years between hurricanes, and N' is the number of samples used to estimate the distribution parameters $(N/\hat{T})$.

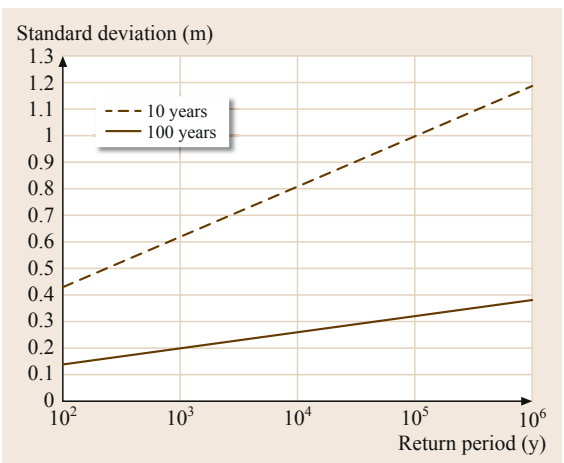


Fig. 26.6 Example standard deviation for confidence limits for a Gumbel distribution with a standard deviation of 0.25 m, shown here for sample lengths of 10 years and 100 years

Many decision-makers believe that the *best-fit* line for a distribution of extremes provides an unbiased estimate of the encounter probability of a value equal to or greater than a specific threshold; and this would be true if there were no alleatory or epistemic uncertainty in our sample. *Resio* et al. [26.44] have shown that uncertainty of both types can affect the expected encounter probabilities. In particular, their study shows that the estimated standard deviation provides a good means of quantifying the spread of the probabilities around the deterministic estimate; in this context, the probability of encountering a given surge value can be written in terms of an integral in two dimensions with a delta function to reduce it back to a single dimension

$$p(\eta) = \int_0^\infty \int_{-\infty}^\infty p[\hat{\eta}(T) + \varepsilon_\eta | \hat{x}] p(\hat{\eta}) p(\varepsilon_\eta) \times \delta(\hat{\eta} + \varepsilon_\eta - \eta)\, \mathrm{d}\varepsilon_\eta\, \mathrm{d}\hat{\eta}(T) \,, \tag{26.16}$$

where $\hat{\eta}(T_r)$ denotes the deterministic estimate of η for a given return period and ε_η denotes the deviation from the deterministic surge estimate.

In (26.6), the estimate for $p(\varepsilon_\eta)$ is taken as a Gaussian distribution with the mean value at $\hat{\eta}$ and the standard deviation taken from the equation for estimating confidence bands, which will have the form

$$p(\varepsilon_\eta | \hat{\eta}) = \frac{1}{\sigma_T \sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}\left(\frac{\varepsilon_\eta}{\sigma_T}\right)} \,, \tag{26.17}$$

where σ_T is the standard deviation of the estimate at a given value of $\hat{\eta}$.

Equation (26.16) can be integrated directly to convert to a return period representation using estimates of population parameters and standard deviation as shown by *Resio* et al. [26.44]

$$T(\eta) = \frac{1}{1 - F(\eta)} \,, \tag{26.18}$$

with

$$F(\eta) = \int_0^\infty \int_{-\infty}^\infty p[\hat{\eta}(T) + \varepsilon_\eta | \hat{\eta}] p(\hat{\eta}) p(\varepsilon_\eta) \times H(\hat{\eta} + \varepsilon_\eta - \eta)\, \mathrm{d}\varepsilon_\eta\, \mathrm{d}\hat{\eta}(T) \,, \tag{26.19}$$

where $H(g)$ is the heaviside function, equal to 1 if $(g) \geq 0$ and equal to 0 if $(g) < 0$.

To give an example of the impact of extrapolating many times the record length of the sample on the expected encounter probabilities, the *Resio* et al. [26.44]

study analyzed central pressures from a 70-year interval for the west coast of Florida. Since surges are related more to tropical cyclones which make a landfall than to exiting tropical cyclones, they stratified their storm sample to include only storms with a general west to east motion in a latitude-longitude box with boundaries at 81°W and 85°W longitude and 25°N and 30°N latitude. Additionally, to eliminate very weak storms which might behave differently to well-organized storms, they further limited the storms included within their sample to those with central pressures less than 990 mb within the selected geographic area. Using the most recent reanalysis data available from NCDC (National Climate Data Center), it was determined that most storms before about 1940 did not report central pressures; consequently, the analysis was limited to the interval 1940–2009.

Figure 26.7 shows the results of this integration for three Florida cases compared to the original deterministic estimate from [26.44]. Case 1 is the deterministic solution for the return period based on the best-fit Gumbel distribution. Case 2 is a test of the numerical algorithm generated by representing the probability of the deviations as a delta function (i. e., all of the probability exactly on the deterministic line), which was computed as a test case for the integration algorithm. As expected, this case exhibited no significant deviations from case 1, so the case 1 and case 2 lines are identical. Case 3 shows the results using standard deviations equal to the estimated distribution standard deviations divided by 2, simply to provide an indication of how nonlinear the dependence is on the standard deviation in the confidence band when compared to case 4; case 4 shows the results for the estimated standard deviation. These results indicate that in this region of the Gulf of Mexico, the encounter probabilities of central pressure (a surrogate for tropical cyclone intensity) are markedly affected by sample uncertainty.

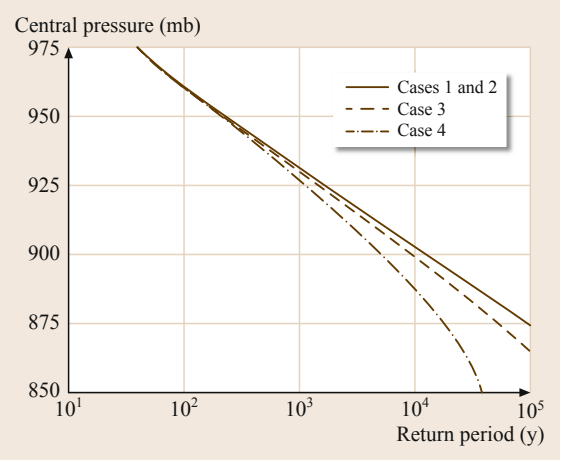


Fig. 26.7 Comparison of expected return periods for central pressures without uncertainty considered (cases 1 and 2) and with uncertainty considered (cases 3 and 4) for the West Florida region using data from 1940–2009

A curious aspect of these results is the increasing deviation between the lines as the return period increases. Whereas the 10 000-year value for the distribution which includes the full uncertainty (case 4) is slightly higher than the maximum possible intensity (MPI) value of 880 mb assumed here, the 40 000-year value has a much lower value (850 mb). This suggests that the role of uncertainty in very low probabilities is to act as a sort of filter for how long of a time period one can extrapolate before the uncertainty begins to become a dominant contributor to the distribution. Even at the 10 000-year value for the central pressure, the pressure differential has already been increased by about 20% over its value with uncertainty neglected. Since values for maximum possible surges in many areas along the west Florida coast are around 10 m, the impact of including this uncertainty would add about 2 m to the design surge levels over the deterministically estimated values.

26.3.3 The Development of the Historical Storm Method for Estimating Coastal Extremes

By the 1980s, it was widely recognized that record length was a significant factor in the confidence that could be placed on estimates of extremes. This led to a quandary of sorts. How could estimates of 100-year values be meaningfully estimated when observational data from measurements covered only relatively short time spans? In addition to the lack of long record lengths for waves and surges, careful review of the available measurements showed that data from large storms were often missing, further compromising the ability of measurements alone to provide a suitable base for the estimation of design values. The answer came via the development of improved physics-based models for predicting waves and surges from available meteorological information. Since meteorological information extended much farther back in time, it was hypothesized that reconstructions of winds and pressures for past events could be used to drive wave and surge models to obtain a surrogate for measurements during these storms. As noted previously, however, it is critical that the computational models used for this purpose are sufficiently accurate to provide accurate unbiased estimates of the extreme values.

This focus on storm hindcasts (the estimation waves and/or surges during past intervals of time) soon be-

came the primary method used in many offshore designs [26.60] as well as in many coastal designs [26.61–64]. Over time, it began to be evident that the selection of storms for hindcasts of waves and surges often exhibited different CDF characteristics in different magnitude ranges. This led to the development of the points over threshold (POT) method for estimating extremes. In this method, only the sample values above some threshold would be included within the sample to be fit by the selected parametric form.

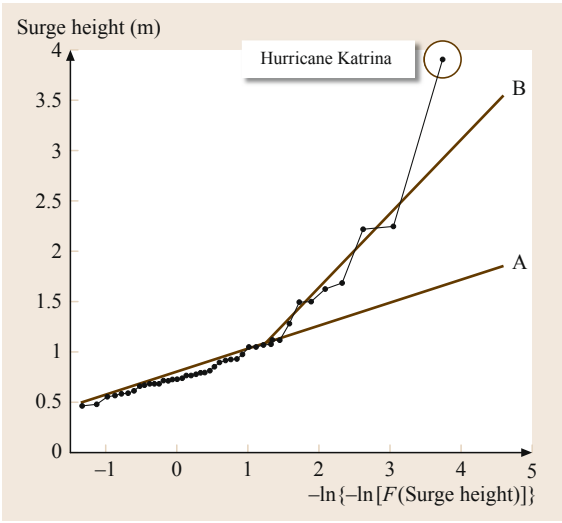


Fig. 26.8 Graph of a sample of hindcast surge heights as a function the double-negative logarithm of the CDF. The *line labeled A* represents storms which made landfall a significant distance away from the site, while the *line labeled B* represents hurricanes which made a landfall in the immediate vicinity of the site. The *single circled point* represents the surge at this site from Hurricane Katrina

Figure 26.8 is a good example of the situation faced by many planners and engineers when dealing with real-world data and helps to demonstrate the need for the POT approach in many situations. In this example, the data is actually more recent than the data that was used initially to justify the POT in earlier studies; however, the point is the same. The data shown here are simulated hindcast storm surges for a site in Lake Pontchartrain, Louisiana. A simplistic concept would be to use a three-parameter GEV (or GPD) distribution to fit the overall distribution. This would imply that Hurricane Katrina was only about a 65-year event in terms of the surges that it generated at this site; however, based on several independent analyses, it appeared that such a surge level would be expected considerably less frequently than this.

In Fig. 26.8, we have added two lines labeled A and B, representing a stratification of the surges at this site based on the proximity of landfall to the site. As can be readily seen and as expected from the physics of surge generation, there is a substantial difference between these two groups of events. This finding is consistent with the findings of researchers who had for years been advocating for the necessity of stratifying the sample before fitting the sample data to some parametric distribution. An interesting additional issue arises in Fig. 26.9. This relates to the issue of *outliers*. An outlier is an event that is believed to represent something

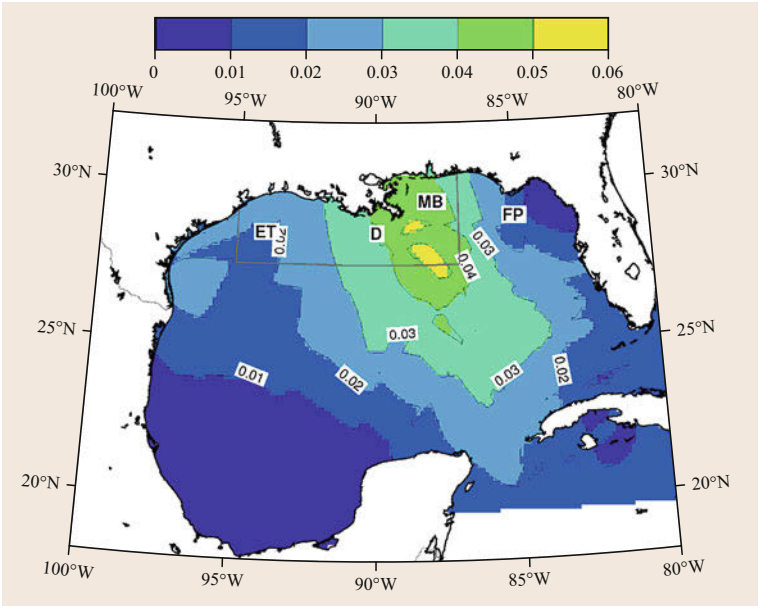


Fig. 26.9 Analysis of hurricane frequency within the Gulf of Mexico based on an analysis using an optimized spatial kernel (after [26.59]

that would only happen much less often than the time span of the sample record. It is intuitive that even the 1 000 000-year event must happen in some 10-year sample, so the historical treatment for these points was to consider them as not representative of the record length within which they were sampled. Such points were typically simply omitted from the distribution fitting process; however, it seems obvious that this omission for all points in a region neglects a potentially valuable piece of information from all of the analyses in an area.

26.3.4 The Development of Alternative Methods

The tragedy of Hurricane Katrina emphasized the point that reliable estimates of coastal hazards and risks should include the effects of all storms that can occur in an area and not just the events that occurred within a particular historical sample. However, as early as the 1970s, this had been recognized, and alternative methods started to be developed which afforded improved estimates of coastal hazards and risks. Probably the two most widely used of these methods have been the joint probability method (JPM), along with a derivative of the JPM termed the JPM-OS (joint probability method optimal sampling), and a second approach which, in this handbook, we will term the empirical track method (ETM). Material from a US Army Corps of Engineers White Paper [26.65] and a reduced version of the White Paper [26.66] will form a substantial portion of our discussions of these developments, and a considerable amount of details on the specifics can be found there.

The Development of the Joint Probability Method (JPM)

The JPM was developed for application to coastal surges in the 1970s [26.67, 68] and subsequently extended by a number of investigators [26.38, 69] in an attempt to circumvent problems related to limited historical records. In this approach, information characterizing a small set of storm parameters was analyzed from a relatively broad geographic area. In applications of this method in the 1970s and 1980s, the JPM assumed that storm characteristics were constant along the entire section of coast from which the sample was drawn. This assumption is inconsistent with more recent studies (Fig. 26.9) which show that storm frequencies vary substantially in this area.

The JPM used a set of parameters, including central pressure, radius of maximum wind speed, storm forward speed, storm landfall location, and the angle of the storm track relative to the coast, to generate parametric wind fields. Furthermore, initial applications of the JPM assumed that the values of these five parameters varied only slowly in storms approaching the coast; therefore, the values of these parameters at landfall could be used to estimate the surge at the coast, and these values could be treated as a constant during the approach of a storm to the coast. Recent data shows that this is not a good assumption [26.70, 71]; specifically, these studies suggest that tropical cyclones decay as they approach land. *Kimball* [26.72] has shown that such decay is consistent with the intrusion of dry air into a tropical cyclone during its approach to land, although other hypotheses for this consistent decay in intensity, such as the lack of energy production from parts of the tropical cyclone over land and increased drag in these areas have also been advanced. In any event, the evidence appears rather convincing that major tropical cyclones begin to decay substantially before they make landfall, rather than only after landfall as previously assumed.

The initial formulation of the JPM used computer simulations of straight-line tracks with constant parametric wind fields to define the maximum surge value for selected combinations of the basic five storm parameters. Each of these maximum values was associated with a probability

$$p(c_\mathrm{p}, R_\mathrm{max}, v_\mathrm{f}, \theta_\mathrm{l}, x)\,, \tag{26.20}$$

where c_p is the central pressure, R_max is the radius of maximum wind speed, v_f is the forward velocity of the storm, θ_l is the angle of the track relative to the coast at landfall, and x is the distance between the point of interest and the landfall location.

These probabilities were treated as discrete increments and the CDF was defined as

$$F(x) = \sum p_{ijklm} | x_{ijklm} < x\,, \tag{26.21}$$

where the subscripts denote the indices of the five parameters used to characterize the tropical cyclones.

The JPM represents a straightforward application of the *response surface* method for estimating extremes. In this method, the response of a particular variable (such as waves and surges) is derived as a function of several parameters, such as the set $(c_\mathrm{p}, R_\mathrm{max}, v_\mathrm{f}, \theta_\mathrm{l}, x)$ used in the above example. Given that the multivariate probability of the five-parameter space is known and a method for defining the response from this set of variables is known, the equation for the CDF for the response can be written for this example as

$$\begin{aligned} F(\eta) = \int\int\int\int\int & p(c_\mathrm{p}, R_\mathrm{max}, v_\mathrm{f}, \theta_\mathrm{l}, x) \\ & \times H[\eta - \Lambda(c_\mathrm{p}, R_\mathrm{max}, v_\mathrm{f}, \theta_\mathrm{l}, x)] \\ & \times \mathrm{d}c_\mathrm{p}\mathrm{d}R_\mathrm{max}\mathrm{d}v_\mathrm{f}\mathrm{d}\theta_\mathrm{l}\mathrm{d}x\,, \end{aligned} \tag{26.22}$$

where Λ represents the model used to convert the parameters to an estimate of maximum surge for each event as a function of $(c_p, R_{max}, v_f, \theta_l, x)$; however, it is easy to see that this approach can be generalized to any set of n parameters for which a response surface can be defined, along with the multivariate probability distribution. Variations on the methodology for evaluating the parameter probabilities and the storm set to be simulated have been presented by *Toro* et al. [26.59], *Niedoroda* et al. [26.73], *Resio* et al. [26.70], and *Irish* et al. [26.74]. Each has its advantages and disadvantages, and as long as the multivariate probability function is sufficiently resolved in the integration process, the results should be in reasonable agreement with each other.

Similar to the EST, this method is nonparametric with respect to the form of the CDF; however, the type of extrapolation used to extend the JPM result beyond the record length is very different to that used by the historical storm method. As shown by *Irish* et al. [26.75], the former of these is based on probability estimates which are much more smoothly varying than the local surge response inherent in the historical data, which means that when applied the JPM provides a significant decrease in sampling uncertainty, when compared to the historical storm method.

A potential advantage of the JPM over methods which depend heavily on historical storms is that the JPM attempts to consider all storms that might happen in an area; whereas, the EST considered only storms that did happen in that area. For example, the 100-year return period storm set to contain a full suite of $(c_p, R_{max}, v_f, \theta_l, x)$; whereas, the historical storm method will have just one or two storms available in this parameter space. Assuming that, for the purpose of surge generation, storm characteristics can be represented adequately by the set of parameters used, it is possible to construct a Katrina-like storm (high intensity combined with large size) even if one has not happened previously. Likewise, it is possible to interpolate between re-curved storms such as Hurricanes Opal and Wilma to understand probabilities of possible hurricane impacts in the Tampa, Florida area, even though neither of these storms produced significant surges in the Tampa area.

Perhaps the biggest controversy in JPM applications during the 1970s and 1980s centered on the sufficiency of this five-dimension set of parameters used in the joint-probability function to produce accurate wind fields. In addition to this concern, the lack of data on historical storms prior to 1950 made it very difficult to derive representative distributions, even for extended sections of coast. For example, information on storm size (R_{max}) was lacking for most historical storms; consequently, a statistical estimate of R_{max} (as a function of latitude and central pressure) was frequently substituted for actual values in the probability distribution. One wind field factor not considered in early JPM applications was the variable peakedness of tropical cyclone wind fields. This term is represented in terms of the Holland B parameter [26.76] in recent tropical cyclone wind models.

It is clear that the number of primary dimensions within the JPM must be capable of representing wind fields to sufficient accuracy so that they provide reasonable, relatively unbiased skill when used to drive coastal wave and surge models. For the case of extratropical storms, there is no known simple set of parameters that meets this criterion, and some extension of the EST or POT method may be the suitable choice for such applications, at least for recurrence intervals which are not too long (less than 100 years or so). For the case of tropical cyclones, dynamic models of tropical cyclone wind fields [26.77, 78] have been shown to capture a substantial portion of the wind field structure, when driven with the parameters listed above plus the Holland B parameter.

An argument against using the parametric wind fields in the JPM is that each historical tropical cyclone will tend to exhibit some degree of deviation from the theoretical parametric (planetary boundary layer) estimates. At any fixed time, such deviations could be produced by strong storm asymmetries, variations in R_{max} around the storm, enhanced spiral bands, etc. Hence, a *best-estimate* wind field crafted by experts to assimilate all the observations in a given tropical cyclone will typically represent the details of that particular storm much more faithfully than is possible via a parameterized theoretical model. Such wind fields today are produced primarily by groups like the National Oceanic and Atmospheric Administration (NOAA) Hurricane Research Division [26.79]. These wind fields are absolutely essential for advancing our understanding of tropical cyclone winds relative to wave and surge forcing in offshore and coastal areas. An excellent discussion of the reconstruction of historical wind fields and the impact of different approaches on model accuracies can be found in [26.80].

It is obvious that *best-estimate* wind fields contain an extremely large number of degrees of freedom in their formulation. Given the relatively small number of historical tropical cyclones, it is unlikely that we can understand/quantify the probabilistic nature of all the interrelated detailed factors that create these deviations. If these details were absolutely critical to coastal wave and surge estimation, we would be able to represent a past tropical cyclone very accurately but would know little about the probability of future tropical cyclones unless we retained the same number of degrees of free-

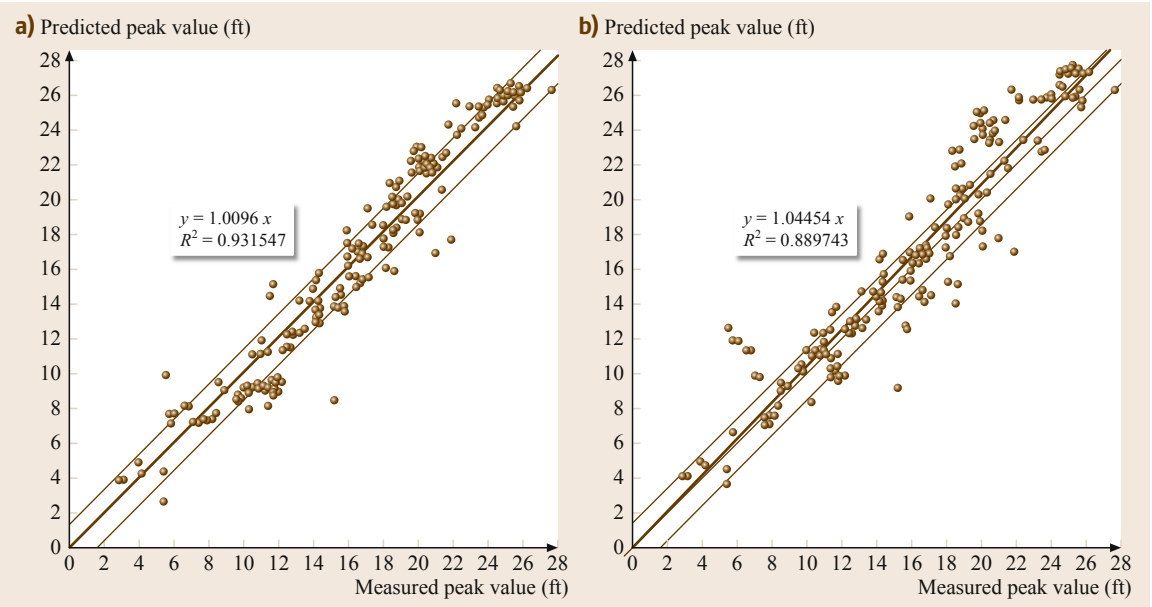


Fig. 26.10 **(a)** Comparison of observed high water marks (HWM, [26.81] for Hurricane Katrina and the computational simulation using *best-wind* wind fields. Points are the values at the recorded HWMs. Lines display a 1 : 1 correlation as well as 1.5 ft standard deviation on each side. **(b)** Comparison of observed HWM for Hurricane Katrina and the computational simulation using the best hand-analyzed post-event wind fields. *Points* are the values at the recorded HWMs. Lines display a 1 : 1 correlation as well as 1.5 ft standard deviation on each side

dom, including their expected variability in estimates of future storm surge and wave estimates. However, Figs. 26.10a and 26.10b show comparisons of hindcast storm surges using *best-estimate* wind and parametric wind fields in a hindcast of Hurricane Katrina. As can be seen in these figures, the differences in the surge model results are relatively small for this case.

Following Hurricane Katrina, improvements to the JPM were made to include additional physical factors affecting the storm and to improve the efficiency of the method [26.40, 66, 70, 74]. These newer methods have played a large role in the estimates of design levels for the new Hurricane Risk Reduction System for New Orleans. A more detailed discussion of these improvements can be found in [26.65].

Empirical Track Method (ETM)

Vickery et al. [26.78] presents a method for modeling hurricane risk in the United States. This method has been adopted for the development of design wind speed maps within the US (American National Standards Institute (ANSI), [26.82, 83]). The ETM uses a Monte Carlo approach to sample from empirically derived probability and joint probability distributions. The central pressure is modeled stochastically as a function of sea surface temperature along with storm heading, storm size, storm speed, and the Holland B parameter. This method has been validated for several regions along US coastlines and provides a rational means for examining hurricane wind risks associated with geographically distributed systems such as transmission lines and insurance portfolios.

A key requirement for the application of the ETM within its Monte Carlo framework is the ability to execute storms over many, many years (20 000 years in the application by *Vickery* et al. [26.78]). As such, a basin-scale study area with an average of three tropical cyclones occurring per year occur would require the simulation of 60 000 storms. Whereas this is not too demanding for the estimation of wind probabilities using an efficiently written planetary boundary layer (PBL) wind model, it is well beyond the range of current computer capacity for existing large, high-resolution ocean and coastal response models (wave models and surge models). For this reason, the ETM is typically not used for applications in which the computational burden for individual events is very high.

In its present form, the ETM is based on an autoregressive approach in which storm parameters at a subsequent time step are determined numerically from the same set of parameters at the current time step, utilizing a set of relationships derived from statistical and theoretical considerations. Although it is recognized that such predictions would contain very large statis-

tical errors for a specific storm, particularly in terms of predictions of storm intensity and size, the predictions appear to approximate certain statistical characteristics of the storm population reasonably well, such as the mean number of storms in an area. What is not so clear is whether they capture the complex multivariate structure that is inherent in interrelationships among storm characteristics, such as the interaction of rates of change of these characteristics and the effect of land proximity. The effect of the omission of these higher order interactions on the statistics of predicted extremes remains a topic for future research.

26.3.5 Probabilistic Analyses of Extratropical Storms

As noted earlier, extratropical storms tend to be more frequent and affect larger coast regions than tropical cyclones due to the relatively large size of these storms. Based on these two considerations, it has been assumed that, for most coastal areas, the use of the historical storm method is adequate for quantifying hazards/risks. However, there are four reasons why this may not be a justifiable assumption given the geometric complexity of many coastal areas, particularly inland areas such as major bays and lakes:

- In some areas, the use of in situ observations to characterize the extremes may contain mixed populations, with a few observations from storms in which strong winds blow along the optimal surge/wave generating direction mixed with many other observations with winds from other directions. In such situations, the former events often have the appearance of outliers. Since a record length of 30–40 years may only contain a single interval of intense storms in which winds line up with the optimal generation direction, this creates a difficult situation for the analysis of extremes and a difficult situation for planners and engineers to assess.
- Although the GPD is specifically derived to fit a distribution above a particular threshold, it is still subject to the statistical constraint that all the samples are drawn from a homogeneous population. Like the GEV, the GPD is still only a three-parameter distribution, and if too many small events are mixed into the analysis, the upper tail of the distribution is very likely to be misestimated. Thus, in applications where researchers choose the lower limit to be given by an arbitrary (pre-selected) number of events, this can lead to a very poor estimation of the actual hazards/risks.
- Earlier, we noted that storm characteristics and frequencies seem to be significantly influenced by multidecadal variations in atmospheric circulation patterns. In these situations, it is exceptionally difficult to quantify the effects of the climatic variability and relatively short-term sample durations may not be adequate for accurately quantifying expected extremes. This is an especially important point relating to the difference between storm events that can occur and storm events that have occurred.
- Due to all of the issues raised in items 1–3 and inherent effects of uncertainty on encounter probabilities, the use of historical storm methods for very low probabilities, as is needed for certain critical infrastructure vulnerability assessment and design, is very difficult to justify without the inclusion of some means to add substantial conservatism into the analysis. Additional research developing a variation of the JPM to fit extratropical storms would be very valuable in meeting this need.

26.3.6 Future Directions and Final Comments

Irish and *Resio* [26.40] show that the effects of storm size, storm forward speed, landfall location, and storm track angle at the coast on storm surges all have asymptotic upper limits in tropical cyclones. Although these upper limits are site dependent, the forms of these limiters can be written in terms of some relatively simple functions. This information can help simplify the estimation of very low probability extremes; since in this very extreme range of surges, it reduces the surge probabilities back to a univariate distribution, in this case a function of a single parameter, storm central pressure. However, as emphasized previously, sampling uncertainty must still be considered in the estimation process. It is likely that, in many areas, the inclusion of sampling uncertainty within relatively small sample sizes will potentially make the estimates of central pressure for AEP values of 10^{-6} unrealistically low. In such situations, it will likely be necessary to combine probabilistic estimates with theoretically derived upper limits to be able to estimate coastal surges. Presently, this theoretical upper limit, the MPI, is still developed via relatively simplistic theoretical formulations combined with empirical envelopes of lowest observed central pressures around the world [26.84–86]. In their present form, this limit is estimated primarily as a function of sea surface temperature; however, more theoretical work will likely show the importance of additional terms to the estimation of the MPI. It should certainly be recognized that any estimate of an MPI will directly contain epistemic errors and indirectly contain alleatory errors due to the sampling distribution used to derive and calibrate theories.

26.4 Summary

On the one hand, we can see that considerable progress has been made in the estimation of hazards and associated risks in coastal areas over the last 100 years. Like all decision-making processes which become codified, it is difficult to change a methodology used for such estimates very quickly; hence engineering practice has often significantly lagged the recognized state of the art of our understanding of coastal hazards and risks, tending to wait until the occurrence of a natural disaster, such as Hurricane Katrina and the tsunami affecting Fukushima, Japan, to foster a willingness to change. Prudent design concepts in areas with critical infrastructure and communities that can be affected by hazards/risks must continue to consider the effects of both alleatory and epistemic uncertainty.

While the discussion herein focuses primarily on the coastal storm problem, the key concepts translate to the wide array of hazards impacting the coast. While each of these hazards will have its own unique set of physical characteristics and statistical challenges, the general concepts discussed here are appropriate for evaluating their probability and risk, as well as for understanding the factors contributing to uncertainty in these extreme estimates.

26.5 Nomenclature

a	specified limit or coefficient
b	specified limit or coefficient
c	coefficient
c_p	storm central pressure
d	coefficient
k	coefficient
n, m	indices
$p(x)$	probability density function
q	coefficient
v_f	storm forward speed
x	variable
x_c	coefficient
AEP	annual exceedance probability
CDF	cumulative distribution function
EST	empirical simulation technique
ETM	empirical track method
$F(x)$	cumulative distribution function
GEV	generalized extreme value distribution
GPD	generalized Pareto distribution
JPM	joint probability method
MPI	maximum possible intensity
N	total number of items
P	probability of exceedance
PDF	probability density function
PMH	probable maximum hurricane
POT	peaks over threshold
$Pr(A)$	probability of event A occurring
$Pr(B\|A)$	probability of a consequence (B) occurring given that event A occurred
R	risk
R_max	storm radius to maximum winds
SPH	standard project hurricane
T	return period (or return interval)
T'	dimensionless return period (or return interval)
ε	coefficient
ε'	coefficient
ε_η	deviation from deterministic surge estimate
η	surge value
θ_l	storm track angle relative to coast at landfall
λ	sampling frequency
μ	coefficient
y	reduced Gumbel variate
σ	distribution standard deviation
σ_T	rms error at return period T
σ'_T	dimensionless root-mean-square error at dimensionless return period T'

26.A Appendix: Glossary of Probability and Risk Terms

- *Confidence intervals*: A confidence interval is a statistical representation of how certain one is that a given variable will lie within a given range. A confidence interval includes two components – the *interval* (the value will be between x and y) and the level of *certainty* (a 90% confidence interval indicates one can say that there is a 90% probability the value *will* fall between the cited interval. As the interval gets smaller, the certainty that the values will fall within the interval goes down. Confidence intervals are related to uncertainty in that they are a statistical method of measuring uncertainty of

a predicted value based on past data that provides the probability distribution and variance of the specific parameter of interest.

- *Error*: Error itself has multiple meanings. In the modeling/predictive sense, the concept of error requires that there be a *correct answer* against which to compare a predicted answer (modeling context) or sample result (statistical context). Error is then defined as the difference between the actual and predicted answer or measured sample. Therefore, in the sense of predictive risk assessment (looking forward in time), error is not a useful concept, as there is not yet any *actual* or *true* value against which to measure or compare. Error should not be confused with *mistake*, which is the result of an incorrect assumption, calculation or model formulation. The hallmark of a mistake is that it is avoidable.
- *Error* (*Types I and II*): Type I error and Type II error are technical terms used in statistics to describe particular types of erroneous results in a testing process. The terms relate to the acceptance or rejection of the hypothesis being tested (the null hypothesis). If the null hypothesis was rejected (found to be false) when it actually is true, then a Type I error occurred. Conversely, if the null hypothesis is not rejected (found to be true) when it actually is not true, a Type II error has occurred. These definitions correlate to the concepts of false positives (Type 1 error) and false negatives (Type II error) in testing. For a good technical tutorial on Type I and Type II errors [26.87].
- *Exposure*: The process of the receptor coming into contact with the hazard.
- *Exposure pathway*: The route by which the receptor is exposed. For animals, this is either by ingestion, inhalation, or dermal absorption.
- *Exposure assessment*: Quantitative analysis of how much (concentration, duration) of the hazard reaches the receptor.
- *Hazard/stressor*: An event (storm, accident), agent (chemical, radiation), situation or action with a potential for an undesirable consequence, such as harm to property, the environment, and human health or life. This term is often used synonymously with threat.
- *Probability*: In its most basic sense, probability is the chance or likelihood that something will happen. Qualitatively, the more likely an event is to happen, the more *probable* the event is. Quantitatively, probability is a value between 0 and 1, with 1 representing absolute certainty of the event occurring. The probability of an event is typically measured as the ratio of the number of times an event occurred over the total number of times the event could have occurred. For example, if we consider the event to be the occurrence of precipitation on any given day, we would collect information on whether it rained on a given day for a period of time (say 1 year). The probability of rain would then be

$$\text{probability of precipitation} = \frac{\text{number of days with precipitation}}{\text{total number of days}} .$$

- *Probability distribution*: The basic definition of probability works for discrete events, but when the event or parameter being evaluated is continuous in nature (the maximum temperature on any given day, the stage of a river on any given day), then a single probability is not sufficient. Instead, possible values are grouped into discrete ranges, and the number of occurrences in that range are counted then divided by the total number of measurements.
- *Receptor*: A receptor is the specific thing or entity being affected by the hazard/stressor. In a human health risk assessment – the receptor is a person.
- *Risk*: The potential for realization of unwanted, adverse consequences to human life, health, property, or the environment. The estimation of risk is usually quantified using the expected value of the conditional probability of the event occurring times the consequence of the event given that it has occurred. This definition is currently used by the Society of Risk Analysis. Mathematically, this is expressed as

$$R = P(A) * P(B|A) , \tag{26.23}$$

where R = risk (a probability from 0 to 1), $P(A)$ = probability of event (A) occurring, and $P(B|A)$ = probability of a consequence (B) occurring given that event A occurred.

This technical definition is what will be used through this handbook. However, there are other definitions in other fields of which the knowledgeable practitioner should be aware. Table 26.2 presents a summary of other definitions used in various fields and a list of citations to which one may refer for more information. In each case, the same mathematical representation can be used, but there are assumptions made that result in the definition being slightly different. This is also presented in Table 26.2.

Besides having multiple potential definitions, risk can also be differentiated by *types*. Different types of risk, even if they are of the same quantitative value, are often managed differently or even ig-

Table 26.2 Risk definitions in different applications

Source	Risk definition	Mathematical representation	References
Occupational Safety and Health Administration (OSHA)	The probability that an adverse effect will occur	In this approach, the event is assumed to happen as the hazard is the item being regulated. For example, the absence of the railing in a factory catwalk would be the hazard. For OSHA purposes, if the hazard exists, then the probability of the accident becomes the risk. Mathematically, the probability of the even $[P(A)]$ is assumed to be 1.0. Therefore, the risk equation becomes: $R = 1 * P(B\|A) = P(B)$	[26.88]
Ecological and Human Health Risk – US Environmental Protection Agency (EPA)	Risk is the overall the probability of injury, disease, or death to humans or damage to the environment resulting from exposure to a chemical, stressor, or occurrence of a hazard. Risk has three components: a hazard, a receptor, and a method for the hazard to reach the receptor (involves both transport and exposure pathway)	The modification in this formulation is that the risk of the effect is not only dependent on the occurrence of A, but on the ability of the hazard to reach the receptor to cause the consequence. Hence, if the occurrence is a release of a toxic chemical, and the outcome of the release on a person is cancer (B), there will be no risk if there is no pathway (P) to the receptor. Mathematically, this means that the probability of the event becomes a conditional probability on the occurrence of the pathway. Therefore, the risk equation becomes: $R = P(A\|P) * P(B\|A)$	[26.89–95]
Medicine, psychology, health and social services	The potential for realization of unwanted adverse consequences or events	This approach, much like the OSHA approach assumes the hazard to be a reality (in this instance, an illness or negative condition such as poverty or exposure to domestic violence). Therefore, there is no probability of a conditional event – the illness/condition *is* the adverse outcome (i. e., $P(A) = 1$). This is also referred to as being *at risk* for some illness or negative outcome. Therefore, the risk equation becomes: $R = 1 * P(B\|A) = P(B)$	[26.96–103]
Nuclear Regulatory Commission [26.104]	Risk is the combined answer to three questions that consider (1) what can go wrong, (2) how likely it is, and (3) what its consequences might be	While posed as three questions, the first two are the definition of the probability of a hazard (e.g., what could go wrong is a pump failure (A), and the probability of that failure is by definition $P(A)$. The third question is the conditional probability of the negative consequence given the occurrence of A. Therefore, mathematically: $R = P(A) * P(B\|A)$	NRC Glossary (Web) [26.105]

Table 26.2 (continued)

Source	Risk definition	Mathematical representation	References
Financial Professions	Risk is the chance of loss with respect to person, liability, or the property of the insured or the possibility that the actual return on an investment will be different from its expected return. This includes such things as the danger or probability of loss to an insurer, the amount that an insurance company stands to lose, the variability of returns from an investment, and the chance of non-payment of a debt. Also often defined as the standard deviation of the return on total investment or the degree of uncertainty of return on an asset	In this formulation, the B (outcome) is defined as the deviation from an expectation. Note – this can be either positive or negative (return can exceed expectation), but the scenarios examined typically are only those which will produce a negative impact. If we assume the expected return of an investment to be x, the actual return realized to be x_a, and the difference to be defined as $D = (x_a, -x)$, then the risk financially if some event A occurs is: $R = P(A) * P(D\|A)$	[26.106–113]
United Nations	Risk is defined as the uncertainty that surrounds future events and outcomes. It is an expression of the likelihood and impact of an event with the potential to influence the achievement of the organization's objectives and goals. Risks in the United Nations context are normally referred to as programmatic and operational areas that have the greatest exposure to inefficiencies, ineffectiveness, fraud, waste, abuse, and mismanagement	In this formulation, the only difference is that the specific *hazard* is pre-defined as inefficiencies, ineffectiveness, fraud, waste, abuse, and mismanagement. Therefore, the mathematical formulation remains the same: $R = P(A) * P(B\|A)$	[26.114]
The ISO 31000 (2009)/ ISO Guide 73:2002	Risk is the effect of uncertainty on objectives. In this definition, uncertainties include events (which may or may not happen) and uncertainties caused by ambiguity or a lack of information. It also includes both negative and positive impacts on objectives. This definition was developed by an international committee representing over 30 countries and is based on the input of several thousand subject matter experts	This approach is closest to the financial professions in that risk is measured with respect to a desired objective (an expected outcome in financial terms). Using the same approach as financial risk, define the expected outcome as $\bar{o}$; the actual as o_a and the difference $D = (o_a - \bar{o})$, then the risk if some event A occurs is: $R = P(A) * P(D\|A)$ The only difference between this and financial risk is that the event A being analyzed explicitly may cause o_a to increase (a positive result)	ISO31000:2009 Risk Management Standard [26.115]

nored. Some of the common risk types that greatly influence both how these risks are assessed, managed and communicated are defined below.

- *Actual risk*: A scientifically verifiable risk. For example, it is well researched and documented that smoking places you at-risk for cancer [26.3–6]. Actual risk is sometimes referred to as *objective risk*, but whether risk is subjective or objective is related more to its ability to be measured than it is to its actual verifiability.
- *Perceived risk*: Risk that is thought to exist by an individual or group that is non-existent or exaggerated. This often occurs in situations where the public is misinformed or in which media reports instill unnecessary panic. Food safety concerns often top the list of such events and lead to the significant public policy debates, see, for example, [26.7].
- *Assumed risk*: Risk that is taken by choice. Assumed risk can be quantifiably large or small, and actual or perceived. For example, individuals who choose to partake in risky activities (skydiving, mountain climbing) choose to assume the relatively large risks associated with these activities, but choosing to drive a car, take medicines or be involved in day-to-day activities all involve some assumed risk.
- *Comparative risk*: Risk placed in context through comparison with another, perhaps better known risk. For example, stating that one is more likely to be hit by a meteor than to be injured in a plane crash compares a risk that people perceive as low (being hit by a meteor) with one they *perceive* as high but which actually is not. This helps place the risk in a conceptual framework.
- *Imposed risk*: Risk that is forced upon an individual, either without the knowledge of the individual or if known, without consent. For example, second-hand smoke exposure is seen as an imposed risk [26.8–10]. Natural events such as earthquakes, hurricanes. and extreme weather events are, to a large extent, *imposed* risks, but to some extent individuals assume that risk based on where they choose to live. For an interesting approach to this, the reader is referred to the work of *Parsad* [26.11].
- *Relative risk*: Risk of a particular outcome compared between two different groups or conditions. The relative risk is calculated as

$$\text{relative risk} = \frac{\text{risk under condition 1}}{\text{risk under condition 2}} .$$

Thus relative risk proved a value of the risk for Condition 1 as a multiple of the risk for Condition 2. For example: suppose property bordering a coastline (Condition 1) has a 30% probability of being inundated during a storm surge, while properties over 100 m away from the coastline (Condition 2) have a 25% probability of inundation. The relative risk of inundation is, therefore, 1.2 times higher along the coastline (Condition 1).

It is imperative that the underlying information about the actual baseline risk be given (i. e., 30% and 25% probability of inundation). Relative risk statistics where no baseline information is given can be very misleading. For example, if there is a 1 in 1 000 000 (10^{-6}) chance of an event at location A, and a 1 in 10 000 000 chance of the same event at location B, saying that location A has a 10 times greater risk than B belies the fact that the risk is still exceedingly small at location A.

- *Percent increased risk*: Risk of a particular outcome compared between two different groups or conditions measured as a relative difference of the two risks related to a base risk. Percent increases risk is calculated as

$$\text{percent increased risk} = \frac{\text{risk under Condition 1} - \text{risk under Condition 2}}{\text{risk under Condition 2}} \times 100 .$$

Using our example from conditional risk, the percent risk increase for inundation by living on the coast would be

$$\text{percent increased risk} = \frac{30-25}{25} \times 100 = 20\% \text{ increase} .$$

Just as with relative risk, it is imperative that the underlying information about the actual baseline risk be given (i. e., 30% and 25% probability of inundation). Percentage increased risk can be even more misleading than relative risk, especially where small risks are involved. For example, if there is a 1 in 1 000 000 (10^{-6}) chance of an event at location A, and a 1 in 10 000 000 chance of the same event at location B, the percent increased risk at location A would be 900%, which clearly presents a very different picture than saying that location A has a *one in a million* chance of the event.

- *Risk analysis*: The overall name given to the application of risk concepts to decision making. It

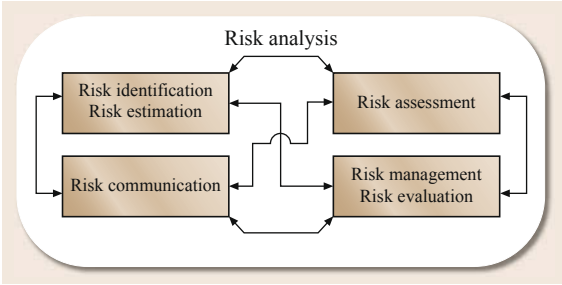


Fig. 26.11 Graphical representation of relationship of the elements of risk analysis

involves detailed examinations performed to understand the nature of unwanted, negative consequences to human life, health, property, or the environment. The process includes identification of potential events (scenarios), quantitative and/or qualitative assessment of risk, analysis of risk management alternatives, and communication of that risk to the necessary stakeholders; an analytical process to provide information regarding undesirable events; and the process of quantification of the probabilities and expected consequences for identified risks. Figure 26.11 gives a graphical representation of the iterative and highly interactive relationship of the various aspects of risk analysis.

- *Risk assessment*: The use of scientifically supported relationships to evaluate the magnitude and probability of adverse impacts on selected endpoints of specific actions, events or hazards/stressors. The assessment may be either qualitative or quantitative. An example of a qualitative risk assessment is the prediction of cancer based on decreased ozone in the atmosphere conducted by the World Meteorological Organization. The risk assessment predicted that if conditions did not change, there would be 50 million additional skin cancer cases due to sunburn by the year 2000 [26.116]. An example of a qualitative risk assessment concerning the same topic (cancer from sunburn) is that a person's risk for melanoma – the most serious form of skin cancer–doubles if he or she has had five or more sunburns [26.117].
- *Risk estimation*: The scientific determination of the characteristics of hazards/threats, usually in as quantitative a way as possible. This includes the magnitude, spatial scale, duration, and intensity of adverse consequences and their associated probabilities, as well as a description of the cause and effect links.
- *Risk evaluation*: A component of risk assessment in which judgments are made about the significance and acceptability of risk.
- *Risk identification*: Recognizing that a hazard exists and trying to define its characteristics. Often risks exist and are even measured for some time before their adverse consequences are recognized. In other cases, risk identification is a deliberate procedure to review, and it is hoped, to anticipate possible hazards.
- *Stochastic*: The property of having inherent random variation. The variation in a stochastic process, while random, is describable through probability theory (see *uncertainty* and *variability* below).
- *Threat*: See *hazard/stressor* above.
- *Uncertainty*: Uncertainty is a widely used term that is unfortunately often misused as a *catchall* term. In some instances, it is erroneously applied to the concept of confidence interval, which is actually a method of quantifying uncertainty. For the purpose of risk assessment, modeling, and/or prediction, uncertainty arises from three main components: error, variation, and lack of knowledge (see definitions herein and in the chapter text).
- *Variability* (*alleatory uncertainty*): Variability is a range of potential values for a given parameter. Variability is a natural characteristic of natural processes (also called *natural variation*). It is describable using probability distributions. The result of natural variation is sometimes called *alleatory uncertainty*. Variability in natural processes (and the resulting alleatory uncertainty) cannot be reduced, as it is an inherent property of the process itself.
- *Lack of knowledge* (*epistemic uncertainty*): The result of a lack of knowledge in risk assessments is also sometimes called epistemic uncertainty. A lack of knowledge can arise because the knowledge is not yet scientifically available – and as such, it can be reduced (along with the resulting epistemic uncertainty) going forward through additional data, experimentation, theoretical development, and scientific inquiry. However, lack of knowledge also includes things that we do not even know we do not know. This area of *lack of knowledge* is more difficult, because it is not easily identifiable and as such, becomes included in variability, or is called *error* when comparing model results to reality.

References

26.1 P.L. Bernstein: *Against the Gods: The Remarkable Story of Risk*, 3rd edn. (Wiley, New York 1996)

26.2 P.S. Laplace: *A Philosophical Essay on Probabilities* (Wiley, London 1902), translated by F.W. Truscott, F.L. Emory

26.3 S. Bišanović: The length of cigarette smoking is the principal risk factor for developing COPD, Int. J. Collab. Res. Intern. Med, Public Health **4**(1), 45–54 (2012)

26.4 H.G. Coleman, S. Bhat, B.T. Johnston, D. McManus, A.T. Gavin, L.J. Murray: Tobacco smoking increases the risk of high-grade dysplasia and cancer among patients with Barrett's Esophagus, Gastroenterology **142**(2), 233–240 (2011)

26.5 G.C. Kabat, M.Y. Kim, J. Wactawski-Wende, T.E. Rohan: Smoking and alcohol consumption in relation to risk of thyroid cancer in postmenopausal women, Cancer Epidemiol. **36**(4), 335–340 (2012)

26.6 G.C. Kabat, N. Shivappa, J.R. Hébert: Mentholated cigarettes and smoking-related cancers revisited: An ecologic examination, Regul. Toxicol. Pharmacol. **63**(1), 132–139 (2012)

26.7 O. Johansson-Stenman: Mad cows, terrorism and junk food: Should public policy reflect perceived or objective risks?, J. Health Econ. **27**(2), 234–248 (2007)

26.8 A. Besaratinia, G.P. Pfeifer: Second-hand smoke and human lung cancer, Lancet Oncol. **9**(7), 657–666 (2008)

26.9 L. Lazuras, A. Rodafinos, J.R. Eiser: Adolescents' support for smoke-free public settings: The roles of social norms and beliefs about exposure to secondhand smoke, J. Adolesc. Health **49**(1), 70–75 (2010)

26.10 J. Meadowcroft: Economic and political solutions to social problems: The case of second-hand smoke in enclosed public places, Rev. Political Econ. **23**(2), 233–248 (2011)

26.11 S. Parsad: *Planning for Human Settlements in Disaster Prone Areas* (Manglam Publ., Delhi 2004)

26.12 E.J. Gumbel: *Statistics of Extremes* (Columbia Univ. Press, New York 1958)

26.13 M.R. Leadbetter, S. Lindgren, H. Rootzen: Extremes and related properties of random sequences and processes, Z. Wahrsch. Geb. **65**, 291–306 (1983)

26.14 S. Resnick: *Extreme Values, Point Processes and Regular Variation* (Springer, New York 1987)

26.15 N. Cressie: *Statistics for Spatial Data* (Wiley, New York 1993)

26.16 S. Coles: *An Introduction to Statistical Modeling of Extreme Values*, Springer Series in Statistics (Springer, London 2001)

26.17 T. Aven: *Foundations of Risk Analysis: A Knowledge and Decision-Oriented Perspective* (Wiley, Chichester 2003)

26.18 M. Rausand: *Risk Assessment: Theory, Methods, and Applications* (Wiley, Hoboken 2011)

26.19 W.E. Fuller: Flood flows, Trans. Am. Soc. Civ. Eng. **77**, 564–617 (1914)

26.20 R.E. Horton: Frequency of recurrence of Hudson River Floods, US Weather Bur. Bull. **2**, 109–112 (1913)

26.21 A. Hazen: Discussion on flood Flows, Trans. Am. Soc. Civil Eng. **77**, 626–632 (1914)

26.22 V.T. Chow: *Handbook of Applied Hydrology* (McGraw-Hill, New York 1964)

26.23 Y. Goda: Random Waves and Spectra. In: *Handbook of Coastal and Ocean Engineering*, Vol. 1, ed. by J.B. Herbich (Gulf Publ. Company, Houston 1990) pp. 175–212

26.24 Y. Goda: Distribution of Sea State Parameters and Data Fitting. In: *Handbook of Coastal and Ocean Engineering*, Vol. 1, ed. by J.B. Herbich (Gulf Publ. Company, Houston 1990) pp. 371–404

26.25 N. Hogben, F.E. Lumb: *Ocean Wave Statistics* (HMSO, London 1967)

26.26 J.A. Battjes: Long-term wave height distribution at seven stations around the British Isles, Dtsch. Hydr. Z. **25**(4), 179–189 (1972)

26.27 M. Isaacson, N.G. Mackenzie: Long-term distributions of ocean waves, J. Waterw. Port Coast. Ocean Div. **107**(2), 93–109 (1981)

26.28 C. Graham: The parameterization and prediction of wave height and wind speed persistence statistics for oil industry operational planning purposes, Coast. Eng. **6**(4), 303–329 (1982)

26.29 S. Kuwashima, N. Hogben: The estimation of wave height and wind speed persistence statistics from cumulative probability distributions, Coast. Eng. **9**(6), 563–590 (1986)

26.30 B. Efron: Bootstrap methods: Another look at the jackknife, Ann. Stat. **7**(1), 1–26 (1979)

26.31 B. Efron: Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods, Biometrika **68**(3), 589–599 (1981)

26.32 B. Efron: *The Jackknife, the Bootstrap, and Other Resampling Plans*, CBMS-NSF Monographs, Vol. 38 (Society of Industrial and Applied Mathematics, Philadelphia 1982)

26.33 B. Efron: Better bootstrap confidence intervals, J. Am. Stat. Assoc. **82**(397), 171–185 (1987)

26.34 C.F.J. Wu: Jackknife, bootstrap and other resampling methods in regression analysis (with discussions). Ann, Stat. **14**, 1261–1350 (1986)

26.35 N. Scheffner, L. Borgman, D. Mark: Empirical simulation technique applications to a tropical storm surge frequency analysis of the coast of Delaware, Proc. 3rd Int. Conf. Estuar. Coast. Model. (1993)

26.36 V. Harcourt: *Harbours and Docks*, Vol. 1 (Clarendon Press, Oxford 1895)

26.37 H.E. Graham, D.E. Nunn: *Meteorological considerations pertinent to the Standard Project Hurricane, Atlantic and Gulf Coasts of the United States*, National Hurricane Res. Proj. Rep. No. 33 (Weather Bureau, Washington 1959)

26.38 R.W. Schwerdt, F.P. Ho, R.R. Watkins: *Meteorological criteria for Standard Project Hurricane and*

Probable Maximum Hurricane Windfields, Gulf and East Coasts of the United States, Tech. Rep. NOAA-TR-NWS-23 (National Oceanic and Atmospheric Administration, Washington 1979)

26.39 J.L. Irish, D.T. Resio, J.J. Ratcliff: The influence of storm size on hurricane surge, J. Phys. Oceanogr. **38**(9), 2003–2013 (2008)

26.40 J.L. Irish, D.T. Resio: A Hydrodynamics-based surge scale for hurricanes, Ocean Eng. **37**(1), 69–81 (2010)

26.41 D.T. Resio, J.J. Westerink: Modeling the physics of hurricane storm surges, Phys. Today **61**, 33–38 (2008)

26.42 S. Bunya, J.C. Dietrich, J.J. Westerink, B.A. Ebersole, J.M. Smith, J.H. Atkinson, R. Jensen, D.T. Resio, R.A. Luettich, C. Dawson, V.J. Cardone, A.T. Cox, M.D. Powell, H.J. Westerink, H.J. Roberts: A high-resolution coupled riverine flow, tide, wind, wind wave, and storm surge model for Southern Louisiana and Mississippi. Part I: Model Development and Validation, Mon. Weather Rev. **138**, 345–377 (2010)

26.43 J.C. Dietrich, M. Zijlema, J.J. Westerink, L.H. Holthuijsen, C. Dawson, R.A. Luettich Jr., R.E. Jensen, J.M. Smith, G.S. Stelling, G.W. Stone: Modeling hurricane waves and storm surge using integrally-coupled, scalable computations, Coast. Eng. **58**(1), 45–65 (2010)

26.44 D.T. Resio, J.L. Irish, J.J. Westerink, N. Powell: The effect of uncertainty on estimates of hurricane surge hazards, Nat. Hazards **66**(3), 1443–1459 (2013)

26.45 R.A. Fisher, L.H.C. Tippett: Limiting forms of the frequency distribution of the largest or smallest member of a sample, Math. Proc. Camb. Philos. Soc. **24**(02), 180–190 (1928)

26.46 B.V. Gnedenko: Sur la distribution limite du terme maximum d'une série aléatoire, Ann. Math. **44**, 423–453 (1943)

26.47 L.E. Borgman, D.T. Resio: Extremal statistics in wave climatology. In: *Topics in Ocean Physics*, ed. by A. Osborne, P.M. Rizzoli (North-Holland, Amsterdam 1982) pp. 387–417

26.48 P. Embrechts, C. Klüppelberg, T. Mikosch: *Modelling Extremal Events for Insurance and Finance* (Springer, Berlin, Heidelberg 1997)

26.49 A.F. Jenkinson: The frequency distribution of the annual maximum (or minimum) values of meteorological elements, Quart. J. R. Meteorol. Soc. **81**, 158–171 (1955)

26.50 J. Pickands: Statistical inference using extreme order statistics, Ann. Stat. **3**(1), 119–131 (1975)

26.51 A.C. Davison, R.L. Smith: Models for exceedances over high thresholds (with discussion), J. R. Stat. Soc. B **52**(3), 393–442 (1990)

26.52 E. Castillo, A.S. Hadi: Fitting the generalized Pareto distribution to data, J. Am. Stat. Assoc. **92**, 1609–1620 (1997)

26.53 D.T. Resio: Some aspects of extreme wave prediction related to climatic variations, Proc. 10th Annu. Offshore Technol. Conf. (1978), OTC-3278-MS

26.54 N.R. Mann, R.E. Schafer, N.D. Singpurwalla: *Methods for Statistical Analysis of Reliability and Life Data* (Wiley, New York 1974)

26.55 M. Engelhardt, L.J. Bain: On prediction limits for samples from a Weibull or extreme-value distribution, Technometrics **24**, 147–150 (1982)

26.56 J.K. Patel: Prediction intervals – A Review, Commun. Stat. Theory Methods **18**(7), 2393–2465 (1989)

26.57 I.I. Gringorten: A simplified method of estimating extreme values from data samples, J. Appl. Meteorol. **2**, 82–89 (1962)

26.58 I.I. Gringorten: *Extreme Value Statistics in Meteorology – A Method of Application*, Air Force Surveys in Geophys. N. 125 (Air Force Cambridge Research Center, Bedford 1963)

26.59 G.R. Toro, D.T. Resio, D. Divoky, A.W. Niedoroda, C. Reed: Efficient joint probability methods for hurricane surge frequency analysis, Ocean Eng. **37**(1), 125–134 (2010)

26.60 V.J. Cardone, W.J. Pierson, E.G. Ward: Hindcasting the directional spectra of hurricane generated waves, J. Petrol. Technol. **28**, 385–394 (1976)

26.61 S.G. Coles, J.A. Tawn: Statistics of coastal flood prevention, Philos. Trans. R. Soc. Lond. A **332**, 457–476 (1990)

26.62 N.W. Scheffner, L.E. Borgman, D.J. Mark: Empirical simulation technique based storm surge frequency analysis, J. Waterw. Port Coast. **122**, 93–101 (1996)

26.63 E.F. Thompson, N.W. Scheffner: *Typhoon-Induced Stage-Frequency and Overtopping Relationships for the Commercial Port Road, Territory of Guam* (Coastal and Hydraulics Laboratory, US Army Corps of Engineers Engineer Research and Development Center, Vicksburg 2002), ERD/CHL TR-02-1

26.64 H.W. van den Brink, G.O. Konnen, J.D. Opsteegh, G.J. van Oldenborgh, G. Burgers: Improving 104-year surge level estimates using data of the ECMWF seasonal prediction system, Geophys, Res. Lett. **31**, L17210 (2004)

26.65 US Army Corps of Engineers: *Performance Evaluation of the New Orleans and Southeast Louisiana Hurricane Protection System, Final Report of the Interagency Performance Evaluation Task Force. Vol. VII – The consequences* (US Army Corps of Engineers, Washington 2006)

26.66 US Army Corps of Engineers: *Performance Evaluation of the New Orleans and Southeast Louisiana Hurricane Protection System, Final Report of the Interagency Performance Evaluation Task Force. Vol. VIII – Engineering and Operational Risk and Reliability Analysis* (National Oceanic and Atmospheric Administration, Washington 2009)

26.67 V.A. Myers: *Storm Tide Frequencies on the South Carolina Coast*, NOAA Tech. Rep. NWS-16 (National Oceanic and Atmospheric Administration, Washington 1975)

26.68 F.P. Ho, V.A. Myers: *Joint Probability Method of Tide Frequency Analysis applied to Apalachicola Bay and St. George Sound, Florida*, NOAA Tech.

Rep. NWS, Vol. 18 (National Oceanic and Atmospheric Administration, Washington 1975)
26.69 F.P. Ho, J.C. Su, K.L. Hanevich, R.J. Smith, F.P. Richards: *Hurricane Climatology for the Atlantic and Gulf Coasts of the United States*, NOAA Tech. Rep. NWS, Vol. 38 (National Oceanic and Atmospheric Administration, Washington 1987), completed under agreement EMW-84-E-1589 for FEMA
26.70 D.T. Resio, J.L. Irish, M.C. Cialone: A surge response function approach to coastal hazard assessment: Part 1: Basic Concepts, Nat. Hazards **51**(1), 163–182 (2009)
26.71 D.H. Levinson, P.J. Vickery, D.T. Resio: A review of the climatological characteristics of landfalling Gulf hurricanes for wind, wave, and surge hazard estimation, Ocean Eng. **37**(1), 13–25 (2010)
26.72 S.K. Kimball: A modeling study of hurricane landfalls in a dry environment, Mon. Weather Rev. **134**, 1901–1918 (2006)
26.73 A.W. Niedoroda, D.T. Resio, G.R. Toro, D. Divoky, H.S. Das, C.W. Reed: Analysis of the coastal Mississippi storm surge hazard, Ocean Eng. **37**, 82–90 (2010)
26.74 J.L. Irish, D.T. Resio, M.C. Cialone: A surge response function approach to coastal hazard assessment: Part 2: Quantification of spatial attributes of response functions, J. Nat. Hazards **51**(1), 183–205 (2009)
26.75 J.L. Irish, D.T. Resio, D. Divoky: Statistical properties of hurricane surge along a coast, J. Geophys. Res. **116**, C10007 (2011)
26.76 G.J. Holland: An analytic model of the wind and pressure profiles in hurricanes, Mon. Weather Rev. **108**, 1212–1218 (1980)
26.77 E.F. Thompson, V.J. Cardone: Practical modeling of hurricane surface wind fields, J. Waterw. Port Coast. Ocean Eng. **122**(4), 195–205 (1996)
26.78 P.J. Vickery, P.F. Skerjl, L.A. Twisdale: Simulation of hurricane risk in the U.S. using empirical track model, J. Struct. Eng. **126**(10), 1222–1237 (2000)
26.79 M.D. Powell, P.J. Vickery, T.A. Reinhold: Reduced drag coefficient for high wind speeds in tropical cyclones, Nature **422**, 279–283 (2003)
26.80 V.J. Cardone, A.T. Cox: Tropical cyclone wind field forcing for surge models: Critical issues and sensitivities, Nat. Hazards **51**, 29–47 (2009)
26.81 J.J. Westerink, Personal communication with Dr. D.T. Resio. (2007)
26.82 Minimum Design Loads for Buildings and Other Structures, ANSI A58.1 (American National Standards Institute, New York 1982)
26.83 Standard Minimum Design Loads for Buildings and Other Structures, Report No. ANSI/ASCE 7-95 (American Society of Civil Engineers, New York 1996)
26.84 K.A. Emanuel: An air-sea interaction theory for tropical cyclones. Part I: Steady-state maintenance, J. Atmos. Sci. **43**, 585–604 (1986)
26.85 G.J. Holland: The maximum potential intensity of of tropical cyclones, J. Atmos. Sci. **54**, 2519–2541 (1997)
26.86 H. Tonkin, G.J. Holland, N. Holbrook, A. Henderson-Sellers: An evaluation of thermodynamic estimates of climatological maximum potential tropical cyclone intensity, Mon. Weather Rev. **128**, 746–762 (2000)
26.87 M.C. Ortiz, L.A. Sarabia, M.S. Sánchez: Tutorial on evaluation of type I and type II errors in chemical analyses: From the analytical detection to authentication of products and process control, Anal. Chim. Acta **674**, 123–142 (2010)
26.88 OSHA: Guidance For Hazard Determination For Compliance With The OSHA Hazard Communication Standard. (29 CFR 1910.1200) http://www.osha.gov/dsg/hazcom/ghd053107.html (2014)
26.89 USEPA: Community-based air pollution projects glossary, http://www.epa.gov/airquality/communitybase/glossary.html (2015)
26.90 USEPA: Waste and Cleanup Risk Assessment Glossary, http://www.epa.gov/oswer/riskassessment/glossary.htm (2015)
26.91 USEPA: *Exposure Factors Handbook: 2011 Edition*, EPA/600/R-09/052F (National Center for Environmental Assessment, USEPA Office of Research and Development, Washington 2011)
26.92 USEPA: Environmental Management System Glossary, http://www.epa.gov/region4/ems/glossary.htm (2015)
26.93 USEPA: *Terms of Environment: Glossary, Abbreviations and Acronyms*, EPA 175-8-92-001. (Office of Communications, Education, and Public Affairs, Washington 1992)
26.94 USEPA: Regional vulnerability assessment (ReVA) program glossary, http://www.epa.gov/glossary.html (2015)
26.95 USEPA: Integrated risk information system (IRIS) glossary, http://www.epa.gov/iris/help_gloss.htm (2015)
26.96 USEPA: Radiation protection radiation glossary, http://www.epa.gov/radiation/glossary/index.html (2015)
26.97 USEPA: RadNet Glossary (last updated on February 13, 2012), http://www.epa.gov/radnet/radnet-glossary.html
26.98 A. Barker, J. Kamar, M. Graco, V. Lawlor, K. Hill: Adding value to the stratify falls risk assessment in acute hospitals, J. Adv. Nursing **67**(2), 450–457 (2011)
26.99 J.C. Fowler: Suicide risk assessment in clinical practice: Pragmatic guidelines for imperfect assessments, Psychotherapy **49**(1), 81–90 (2012)
26.100 A.S. Prentiss: Early recognition of pediatric venous thromboembolism: A risk-assessment tool, Am. J. Crit. Care **21**(3), 178–184 (2012)
26.101 N. Scurich, R.S. John: Prescriptive approaches to communicating the risk of violence in actuarial risk assessment, Psychol. Public Policy Law **18**(1), 50–78 (2012)
26.102 G. Côté, A.G. Crocker, T.L. Nicholls, M.C. Seto: Risk assessment instruments in clinical practice, Can. J. Psychiatry **57**(4), 238–244 (2012)
26.103 US Department of Health and Human Services: *Guidance for Industry Development and Use of*

Risk Minimization Action Plans (FDA Center for Drug Evaluation and Research (CDER), FDA Center for Biologics Evaluation and Research (CBER), Washington 2005), online at http://www.fda.gov/downloads/RegulatoryInformation/Guidances/UCM126830.pdf

26.104 Nuclear Regulatory Commission: Glossary, http://www.nrc.gov/reading-rm/basic-ref/glossary.html#R (2015)

26.105 Nuclear Regulatory Commission: Glossary, www.nrc.gov/readng-rm/basic-ref/glossary/html

26.106 USEPA: Environmental insurance and risk management tools glossary of terms, http://www.epa.gov/brownfields/insurance/ei_glossary_06.pdf (2004)

26.107 Farlex, Inc: The Farlex financial dictionary, http://financial-dictionary.thefreedictionary.com/ (2012)

26.108 J. Guinan: *Investopedia: The (I)Investopedia Guide To Wall Speak* (McGraw-Hill, New York 2009)

26.109 A. Chernobai, P. Jorion, F. Yu: The determinants of operational risk in US financial institutions, J. Financ. Quant. Anal. **46**(6), 1683–1725 (2011)

26.110 G. van de Venter, D. Michayluk, G. Davey: A longitudinal study of financial risk tolerance, J. Econ. Psychol. **33**(4), 794–800 (2012)

26.111 K. Watson Hankins: How do financial firms manage risk? Unraveling the interaction of financial and operational hedging, Manag. Sci. **57**(12), 2197–2212 (2012)

26.112 M. Drehmann, K. Nikolaou: Funding liquidity risk: Definition and measurement, J. Bank. Finance **37**(7), 2173–2183 (2013)

26.113 E. Vasile, I. Croitoru, D. Mitran: Risk management in the financial and accounting activity, Anul VII **3**(27), 13–24 (2012)

26.114 United Nations: Risk management, Office of Internal Oversight Services http://www.un.org/Depts/oios/pages/risk_management.html (2010)

26.115 ISO 31000: Risk Management Principles and Guidelines (International Organization for Standardization, Geneva 2009)

26.116 W.M. Organization: Scientific assessment of ozone depletion. In: *WM0 Global Ozone Research and Monitoring Project*, Report No. 44, ed. by D.L. Albritton, P.J. Aucamp, G. Megie, R.T. Watson (World Meteorological Organization, Geneva 1998)

26.117 Skin Cancer Foundation: Facts about sunburn and cancer, http://www.skincancer.org/prevention/sunburn/factsabout-sunburn-and-skin-cancer (2013)

27. Modeling of Coastal Waves and Hydrodynamics

Patrick J. Lynett, James M. Kaihatu

This chapter presents an overview of the available methods for modeling coastal waves. First, an overview of the relevant coastal processes, from shoaling to turbulent mixing, is provided to establish a basis to compare the various modeling approaches. The bulk of the discussion centers on modeling wind waves and includes a brief overview of the linear and analytical theory available to quantify coastal transformation, and then follows with a summary of spectral and phase-resolving approaches. Modeling long waves is discussed next, with a focus on tsunami simulation. Finally, the chapter summarizes techniques to couple the various models and reviews recent advances in the topic.

27.1 **Wind Wave Modeling** 598
27.1.1 Linear, Analytical and Semi-Empirical Approaches . 598
27.1.2 Spectral Modeling: Phase-Averaged and Phase-Resolving Approaches 600
27.1.3 Depth-Integrated and Boussinesq-Type Approaches 602
27.1.4 Navier–Stokes Equation-Based Approaches 603
27.2 **Modeling Long Waves** 604
27.3 **Coupled and Nested Techniques** 605
27.4 **Summary of Model Properties** 606
27.5 **Conclusions** 608
27.6 **Nomenclature** 608
References 608

As an ocean surface wave enters the coastal zone, broadly defined here as the water landward of the continental shelf or some other depth transition, interaction with the seafloor and currents causes significant transformation. Waves will first start to feel the bottom when the depth (h) to wavelength (L) ratio is less than $1/2$, and waves are considered to be in shallow water, where the effects of the bottom are strong, when this ratio dips below approximately $1/25$. Two-dimensional (2-D) variations in the seafloor can lead to wave focusing due to the effects of refraction. Wave shoaling causes the wave amplitude to grow and the wavelength to shorten, steepening the wave front until the wave breaks.

Depending on the properties of the incoming wave, nonlinearities in the wave behavior are important for some distance before breaking and through the surf zone. Shallow water nonlinearity is governed by the wave height to water depth ratio; as the ratio gets larger, it is expected that nonlinearity becomes increasingly important. From this definition, it is clear that nonlinearity must be important near breaking, as this is the point when the wave height is similar to the depth. Nonlinearity in the coastal zone can change the speed at which a wave travels, where it breaks, and how it interacts with currents.

Nonlinear waves in the coastal zone also induce a net mass transport in the direction of the wave. This net mass transport is due to wave radiation stress, which is name given to the net positive momentum flux in the direction of wave propagation, and leads to wave-induced currents. Currents moving parallel to the shoreline are termed alongshore currents and are primarily responsible for alongshore sediment movement. These currents can generate instabilities, often driven by alongshore bathymetry variation, which can manifest as strong rip currents and shallow turbulent coherent structures. Such features induce high horizontal shear on the flow, and lead to significant horizontal and vertical mixing.

When designing coastal structures, engineers are often most concerned with extreme event loads. Here, extreme events may be storm surges and the accompanying storm waves or tsunamis, for example. It is a present challenge for coastal engineers to determine these extreme event loads, as there is currently no accepted design code for extreme wave loads. Furthermore, the sea level rise associated with climate change

has the potential to increase peak water levels during storms [27.1], also implying that wave energy may penetrate further inshore. The engineer is left with many different options to quantify the wave properties in the coastal zone, from overly simple methods to impractically complex numerical tools.

Relative to many engineering disciplines, the use of numerical coastal wave models for design is new. Many analyses still rely on the simple linear wave methods described in the following section. The usage of large-scale spectral wave models has become commonplace, as these models are efficient and well tested. Currently, it is less common to go to detailed phase-resolving models in engineering practice. These models provide a wave-by-wave resolution of the coastal dynamics, and can provide insight on a small spatial scale. Such insight is important when trying to, for example, accurately estimate wave overtopping of a coastal levee or wave impact loads on a structure. The following sections of this paper will outline the various approaches, providing a brief literature review and theoretical background.

27.1 Wind Wave Modeling

There is a wide range of useful techniques for estimating wind wave properties in the nearshore; the choice of which technique to use depends on limiting factors such as the resources available, the precision required in the result, and the physical properties of the waves themselves. The types of approaches discussed in this section have both highly variable physical approximation as well as computational demand; these two factors often have an inverse behavior. It is the purpose of the following presentation to outline the main types of coastal modeling options, providing theoretical background, numerical modeling schemes, and typical applications.

27.1.1 Linear, Analytical and Semi-Empirical Approaches

The simplest methods for estimating coastal wave properties utilize linear potential wave theory. While this theory is covered comprehensively in many textbooks [27.2], the most relevant equations are given here. Additionally, this presentation will serve to establish the basic variable quantities used in coastal hydrodynamic modeling. The free surface elevation η for a single frequency component is given by

$$\eta = \frac{H}{2}\cos\theta \,,$$

where H is the wave height equal to twice the wave amplitude a. For one horizontal dimension and constant depth, θ is the wave phase function given by

$$\theta = kx - \omega t \,,$$

where k is the wave number, x is a horizontal coordinate, ω is the angular frequency, and t is time. In general terms, k and x can be vector quantities when directional waves are considered. Wavenumber and angular frequency are related to other important physical quantities

$$k = 2\pi/L, \quad \omega = 2\pi/T, \quad f = 1/T \,,$$

where L is the wavelength, or horizontal distance between two successive crests or troughs, T is the wave period, and f is the wave frequency. The linear dispersion relation provides the connection between wave frequency and wavelength

$$\omega^2 = \mathrm{g}k\tanh(kh) \,,$$

where g is gravity, h is the local water depth, and the fluid is assumed inviscid. Expressions for fluid velocity and pressure under the wave can be found in textbooks, and will not be repeated here. Linear wave theory predicts closed, elliptical orbits of fluid particles under waves, and, therefore, no net mass transport.

In the limited case of very shallow water, where the ratio of water depth to wavelength approaches zero, simplifications can be made to the full linear potential theory discussed above. The dispersion relation reduces to

$$\omega^2 = \mathrm{g}hk^2 \,,$$

which yields a wave speed c of

$$c = \frac{\omega}{k} = \frac{L}{T} = \sqrt{\mathrm{g}h}$$

and a horizontal fluid velocity u of

$$u = \frac{H}{2}\sqrt{\mathrm{g}/h}\cos\theta \,.$$

These expressions should be applied only in the case of a very long wave, such as for tides, tsunamis, and swell

waves in very shallow water. However, even outside these cases, quick wave property calculations based on the above will still provide a useful preliminary reference point for shallow water wind waves.

Linear wave transformation in shallow water includes the effects of shoaling, refraction and diffraction, and can be expressed in the general form

$$\frac{H}{H_o} = K_S K_R K_D \,,$$

where H represents the wave height at some coastal location, H_o the height at some offshore location, K_S is the shoaling factor, K_R is the refraction factor, and K_D is the diffraction factor. Assuming no dissipation of energy during propagation, the conservation of wave power between any two locations in the propagation path gives the shoaling coefficient

$$K_s = \sqrt{\frac{n_o L_o}{nL}} \,,$$

where

$$n = \frac{1}{2}\left[1 + \frac{2kh}{\sinh(2kh)}\right] .$$

This n value varies from 0.5 for a deep water wave to 1.0 for a shallow water wave. It is remarked that the basic linear theory presented here does not include bottom reflection or bottom dissipation, and can thus be applied between any two points regardless of the bathymetry between these two points. However, the theory implicitly assumes inviscid propagation and a mildly sloping bottom, and this must be satisfied at a minimum for the shoaling estimation above to be physically appropriate.

Wave refraction is the process of wave crests bending due to propagation-normal variations in bathymetry. Wave crests will bend *into* shallow water regions as the propagation speed slows in such regions, with the refraction coefficient

$$K_R = \sqrt{\frac{\cos\alpha_o}{\cos\alpha}} \,,$$

where α is the angle of incidence relative to the shoreline or some prescribed depth contour, and can be found using Snell's Law

$$\sin\alpha = \frac{c}{c_o}\sin\alpha_o \,.$$

This relation may only be used in simple bathymetries with shore-parallel bottom contours. In more complex regions, modified forms of Snell's law can be used to create wave rays, which were commonly used in the twentieth century to show wave paths and outline possible areas of wave focusing and de-focusing. These techniques have been displaced by numerical methods, some of which are discussed in later sections of this chapter. There is no general equation for the diffraction coefficient. The engineer should reference the numerous diffraction diagrams that can be found in the Coastal Engineering Manual CEM) [27.3], which plot K_D surfaces for various combinations of breakwater configurations and incident wave directions.

Once the wave reaches the break point, the coastal transformation equation given above can no longer be used, and some other approach must be used to cap the wave height. A common engineering approach is the introduction of the breaker index, which is the ratio of the breaking wave height to the local water depth. For a given wave period and bathymetry slope, a breaker index can be estimated. Once the shoaling and refraction analysis estimates a wave height to depth ratio that exceeds the breaker index, it is assumed the wave starts to break; this is the break point. Breaking is then usually assumed to be *depth limited*, meaning that the breaker index remains constant as the water depth decreases onshore. The difficulty in this approach is determining a reliable estimation of the breaker index. There are numerous different equations for this index; the most common are found in the CEM. For waves shoaling over a very steep slope, the breaker index can be greater than 1.0, while for waves breaking over an essentially flat bottom, the breaker index may be lower than 0.4 [27.4]. A commonly assumed value for the breaker index found in engineering studies is 0.78; this is derived from the maximum theoretical solitary wave height, and its physical relevance to random coastal wind wave breaking is questionable.

For studies that seek to use straightforward analysis methods but also provide an estimate of the wave height envelope across a cross-shore beach profile, the energy flux method [27.5] is a useful tool. Here, it is assumed that the spatial rate of change of wave energy flux F is due to some dissipation function φ

$$\frac{\mathrm{d}F}{\mathrm{d}x} = -\varphi(x) \,.$$

The dissipation function is driven by breaking and is proportional to the difference between the local wave energy flux and some *stable* energy flux for the local depth. Determination of a stable energy flux is, similarly to the determination of a breaker index, difficult for a wide range of conditions. Commonly used models for the dissipation function include [27.6] and [27.7], which are both established for random wave break-

ing. Note that outside the breaker zone, where the left-hand side of the above equation is zero, the energy flux method will reduce to a linear shoaling model. This wave transformation model has been adopted by a number of large-scale morphology change simulators, which require a good estimate of the breaker location and resulting longshore currents for large horizontal areas. However, this approach has been largely displaced by more recent numerical tools, which can better include directionality, dissipation, and nonlinear effects. Some of these tools are discussed in the following sections.

27.1.2 Spectral Modeling: Phase-Averaged and Phase-Resolving Approaches

In the open ocean, wind wave generation and propagation are typically described using phase-averaged spectral models. These spectral models determine spatial and temporal changes in the wave spectrum, or the wave energy as a function of frequency and direction. A spectral energy balance is derived, accounting for wave growth, propagation, and dissipation based on wind energy input, whitecapping, and bottom friction. Examples of such models are WISWAVE (wave information study wave model) [27.8], WAVEWATCH III [27.9], and WAM (wave model) [27.10]. These models are highly developed for open ocean waves, but do not account properly for coastal effects such as shallow water wave–wave interactions and depth-induced breaking [27.11]. They output a directional spectrum, which can then be employed as boundary conditions in a coastal zone model to simulate nearshore propagation. For example, WAM could be coupled with SWAN (simulating waves nearshore) [27.12], a coastal spectral model, to estimate the spectral evolution from deep to shallow water [27.11]. However, even coastal spectral models only crudely approximate dynamic nearshore phenomena, due to the inherent approximations such as phase-averaging, weak nonlinear effects, and no diffraction.

In its most general form, the spectral wave action density balance equation is given by

$$\frac{D_w(N)}{Dt} = \frac{S}{\omega} ,$$

where N is the time-dependent wave action density, or energy density divided by frequency, defined for an individual frequency; the total derivative $D_w(N)/Dt$ is taken in the direction of wave motion, and S represents any external source or sink acting on the conservation volume for which the total derivative is taken. S by definition includes forcings such as those of wind and viscous dissipation, such as whitecapping, wave breaking, and bottom friction. If the nonlinear advection terms are neglected from the total derivative, then S must also include nonlinear energy transfers among the various frequencies. Expanding the above balance equation, and including the terms of coastal importance, we have

$$\begin{aligned}\frac{\partial N}{\partial t} &+ (\text{linear energy transfers}) \\ &= \frac{S_{\text{bottom friction}}}{\omega} + \frac{S_{\text{breaking}}}{\omega} + \frac{S_{\text{whitecap}}}{\omega} \\ &\quad - \frac{S_{\text{wind input}}}{\omega} - (\text{nonlinear energy transfers}) ,\end{aligned}$$

where the mechanism of each source/sink S term is given in its subscript, and the nonlinear energy transfer term represents energy transfer from one frequency to a different frequency. For example, the bottom friction sink is most commonly expressed in the Joint North Sea Wave Project (JONSWAP) form [27.13]

$$S_{\text{bottom friction}} = -C_b G(f) E ,$$

where C_b is a bottom friction coefficient dependent on local wave and bottom conditions, G is the *functional form* of the sink and is dependent on the frequency, and E is the local energy density. G has the physically reasonable form of being large at low frequencies (high dissipation in shallow water) and approaches zero (no dissipation) in the deep water region. The breaking dissipation source function is restricted by breaker-index uncertainties similarly as discussed with the linear modeling in the previous section. However, due to the ability of these coastal spectral models to include nonlinearity and approximate dissipation, they have been shown to be excellent predictors of wave height transformation for a wide range of physical configurations. It is reasonably stated that, if the user is not interested in detailed wave-by-wave breaking estimation, runup, infragravity waves, three-dimensional (3-D) flow profiles, and turbulent dynamics such as rip currents and coastal eddies, then coastal spectral models are the most efficient and accurate option available. For a further discussion and a nice review of state-of-the-art spectral modeling methods, the reader is directed to [27.14]. Figure 27.1 shows a result from the SWAN model [27.12], with an improved numerical scheme for the geographical propagation terms [27.15], applied to the area surrounding Santa Rosa and San Miguel Islands in the Southern California Bight. The extreme refraction around the islands served as a stringent test of the model; the single-peaked incident wave spectrum (Fig. 27.1c) has split into two discrete spectra at the location of the measurement station between the two islands.

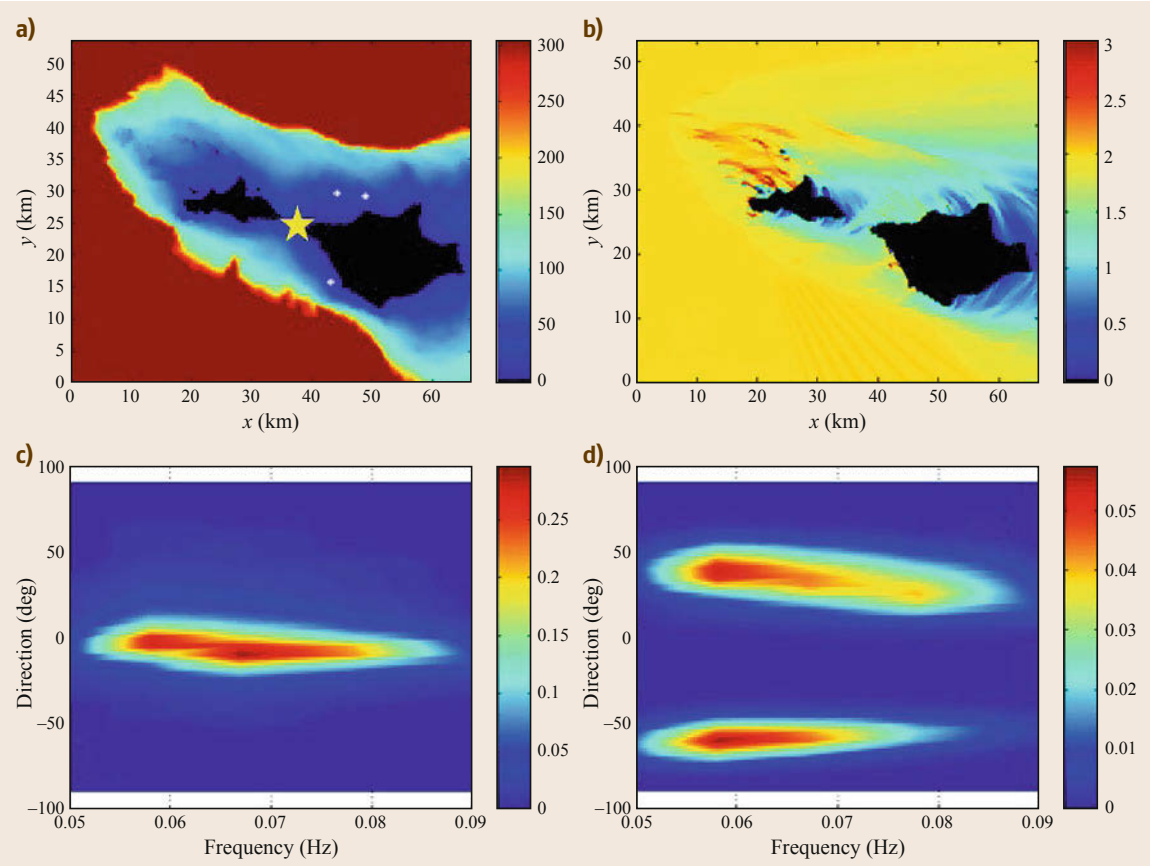


Fig. 27.1a–d SWAN simulation of wave refraction around San Miguel and Santa Rosa Islands in the Southern California Bight. **(a)** Bathymetry (in m) and pressure gage locations (*white dots*) near San Miguel (*left*) and Santa Rosa (*right*) Islands; location of pressure gage #10 denoted with a *star*. **(b)** Significant wave height (in m) on January 14, 1992 at 2300 Pacific Standard Time. **(c)** Incident swell wave spectrum measured offshore of Southern California Bight (January 14, 1992 at 2300 PST) and applied to the left and upper boundaries of local model domain shown in **(a)**; zero degrees denotes waves approaching from due west, while +90 degrees denotes waves approaching from due north. **(d)** Swell wave spectrum output at location of pressure gage #10

Modelers looking to perform phase-resolving simulations of waves from intermediate depths to the shoreline have few options. Well-established models such as SWAN and STWAVE (steady state spectral wave model) [27.16] are phase-averaged models and do not directly provide time histories of free surface and velocity fluctuations due to waves. Mild-slope equation models, such as REF/DIF (refraction/diffraction model) [27.17], are phase-resolving models and are computationally practical to run in most cases. However, these models have restrictions that limit their use, such as weak diffraction effects, a lack of wave reflection, and the limitation to narrow banded spectrums, and higher-order nonlinearity is generally not captured [27.18].

An alternative approach to spectral phase-averaged modeling, which retains both shallow water nonlinearity and the periodic nature of waves, is nonlinear frequency-domain wave modeling [27.19–22]. In essence, these models are nonlinear extensions of phase-resolving nearshore linear refraction-diffraction models such as REF/DIF [27.17]; a relatively comprehensive account of these models is given in [27.23]. These models retain second-order nonlinearity via near resonant interactions (both subharmonic and superharmonic interactions) between triads of frequencies. The strength of the interactions is dictated by both the interaction coefficients and the phase mismatches between the complex amplitudes of the spectral components. These models have been improved to include wide-

angle parabolic propagation [27.24] and interaction effects among triads of longshore wavenumbers [27.22, 25]. Several models also contain a dissipation mechanism representing spectral wave breaking, consisting of a lumped parameter breaking description [27.7] coupled with an adjustable distribution of dissipation over the frequency range. The distribution is an adjustable balance between a frequency-independent weighting of dissipation and one that is weighted as frequency squared. The latter distribution was shown by *Kaihatu* et al. [27.26] to be an inner surf zone asymptote of the dissipation dependence on frequency. Smoothed versions of these phase-resolving models that treat triad interactions in terms of averaged bispectral quantities [27.22, 25, 27] have also been developed; these models are closer akin to the phase-averaged spectral models discussed previously.

27.1.3 Depth-Integrated and Boussinesq-Type Approaches

While frequency domain models are advantageous due to their relatively small computational demand, a significant effort in the nearshore wave model community towards developing phase-resolving, time domain Boussinesq models has occurred in the past decade. Assuming that both nonlinearity and frequency dispersion are weak and are of the same order of magnitude, *Peregrine* [27.28] derived the *standard* Boussinesq equations for variable depth in terms of depth-averaged velocity and free surface displacement. Numerical results based on the standard Boussinesq equations or the equivalent formulations have been shown to give predictions that compared quite well with field data [27.29] and laboratory data [27.30, 31].

As mentioned above, the standard Boussinesq equations are derived based on an assumption between the nonlinearity of the wave, $\varepsilon = a/h$, and the frequency dispersion of the wave, $\mu = kh$. The precise relation of these parameters comes from a nondimensionalization of the full potential flow equations using a shallow water scaling and is given by

$$O(\epsilon) = O\left(\mu^2\right) \ll 1 \,,$$

which is the *true* Boussinesq assumption as related to nonlinear long surface waves. With this assumption, nonlinear two-dimensional-vertical potential flow can be reduced to the one-dimensional 1-D-horizontal equation set for a wave over constant water depth as

$$\text{continuity:} \quad \eta_t + (\eta u)_x + h u_x = 0 \,,$$

$$\text{momentum:} \quad u_t + g\eta_x + u u_x - \frac{1}{3} h u_{xxt} = 0 \,,$$

where the subscripts represent partial derivatives and u is the depth-averaged horizontal velocity. The nondispersive, shallow-water wave equations are easily extracted from this equation set by neglecting the last term ($\frac{1}{3} h u_{xxt}$) on the left-hand side of the momentum equation; it is this single term that adds dispersion to the shallow water base.

The mathematical effect of the additional dispersive term can be quantified by examining the linear form of the standard Boussinesq equations given above and substituting in the linear wave solution form

$$\eta = \eta_o e^{i\theta} \,, \quad u = u_o e^{i\theta} \,, \quad \theta = kx - \omega t \,.$$

After some algebra, the dispersion relation of this approximate equation set can be expressed as

$$\omega^2 = \frac{g h k^2}{1 + \frac{1}{3}(kh)^2} \,,$$

which is also the $[0, 2]$ Pade approximation of the full linear dispersion relation given in the Linear Wave Approaches above (Sect. 27.1.1). In the practical sense, this single additional term in the momentum equation allows for accurate linear propagation of waves up to $kh \approx 0.5$, which is approximately a fivefold increase in applicability compared to the shallow water model, using the same accuracy thresholds in phase and group velocity [27.32].

As it is required that both frequency dispersion and nonlinear effects are weak and of the same order, the standard Boussinesq equations are not applicable to very shallow water depth, where the nonlinearity becomes more important than the frequency dispersion, nor to the deep water depth, where the frequency dispersion is of order 1. The standard Boussinesq equations break down when the depth is greater than $\approx 1/10$ of the wavelength. For many engineering applications, where the incident wave energy spectrum consists of many frequency components, a lesser depth restriction is desirable. To extend the applications to shorter waves (or deeper water depth) many modified forms of Boussinesq-type equations have been introduced [27.22, 33, 34]. Although the methods of derivation are different, the resulting dispersion relations of the linear components of these modified Boussinesq equations are similar, and may be viewed as a slight modification of the $(2, 2)$ Pade approximation of the full dispersion relation for linear water waves [27.35]. It has been demonstrated that the *modified* Boussinesq equations are able to simulate wave propagation from intermediate water depth (the water depth to wavelength ratio is about 0.5) to shallow water including the wave–current interaction [27.36].

Despite the success of the modified Boussinesq equations in intermediate water depth, these equations are still restricted to weak nonlinearity. As waves approach the shore, the wave height increases due to shoaling, until they eventually break. The wave height to water depth ratios associated with this physical process violates the weakly nonlinear assumption. This restriction can be readily removed by eliminating the weak nonlinearity assumption [27.37, 38]. Numerical implementations of the highly-nonlinear, Boussinesq-type equations include FUNWAVE (fully nonlinear Boussinesq wave model) [27.38] and COULWAVE (Cornell University long and intermediate wave model) [27.39]. These models have been applied to a wide variety of topics, including rip and longshore currents [27.40, 41], wave runup [27.42], wave–current interaction [27.43], and wave generation by underwater landslides [27.42], among many others. Boussinesq models are steadily becoming a practical engineering tool. Directional, random spectra can readily be generated by the models, which capture nearshore evolution processes, such as shoaling, diffraction, refraction, and wave–wave interactions, with very high accuracy. The applicability of the Boussinesq models is limited by the fact that the models are fundamentally inviscid. The dissipation processes, such as breaking and bottom friction, must be parameterized in traditional Boussinesq models.

Recently, a number of *non-traditional* Boussinesq approaches have been developed, with the goal of including horizontal vorticity explicitly in the flow field. An attempt to include these dynamics under a breaking wave is found in [27.44], with further advances given in [27.45]. Integrated within a Boussinesq-type derivation, the stream function equation is used to determine the vertical variation of the velocity. This allows the inclusion of the vorticity generated by breaking. The breaking terms that appear as corrections to the momentum balance are a function of the amount of vorticity generated during the breaking process. This vorticity is obtained from the solution to the vorticity transport equation and has been shown to capture the flow field dynamics under a spilling breaker.

A similar attempt was made by *Kim* et al. [27.46] who included the viscous effects of a bottom shear, and the associated rotationality, directly in a Boussinesq-type derivation. While this leads to a far more complex equation model, it includes the physics necessary to simulate boundary shear and the complete coupling of these effects with a nonlinear, dispersive wave field. This model can predict the friction-induced changes to the vertical profile of velocity under weakly unsteady flow and can thereby provide good estimates of internal kinematics. It is also able to translate the bottom-created horizontal vorticity into a vertical vorticity field.

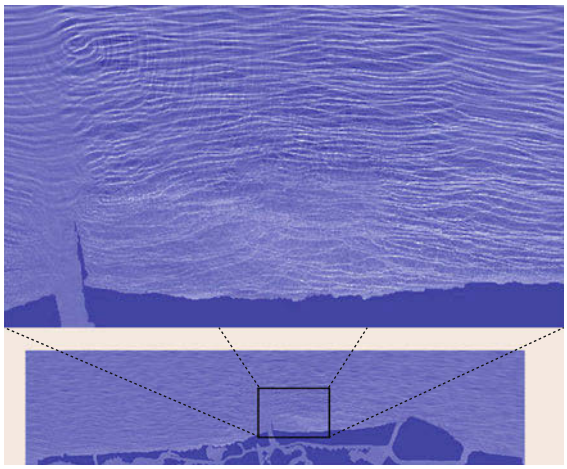


Fig. 27.2 Boussinesq simulation of waves near Freeport, TX. Resolution of the model is 5 m and the total spatial extent is $\approx 100\,\mathrm{km}^2$

Such a model is able to couple the dissipative and nonlinear effects of a bottom shear with a dispersive and nonlinear wave field. While this and similar recent advances allow the depth-integrated equations to model a wider range of physical processes, the computational cost for this inclusion is high [27.47]. In addition to the many individual calculations needed to quantify the various terms in these equations, since the model includes first to third-order spatial derivatives, it becomes necessary to resolve the leading order terms to fourth-order numerical accuracy. This leads to the requirement of a fourth-order spatial differencing and temporal integration scheme. Compared to an efficient shallow water solver, the highly nonlinear Boussinesq-type model can require 50−100 times more computational time for the same numerical grid and time step configuration. While this cost is relatively high, large domains can be tackled with parallel computing [27.48]. For an example of the possible scale of these simulations, see Fig. 27.2.

27.1.4 Navier–Stokes Equation-Based Approaches

Phase and depth-resolving surf zone hydrodynamic models, such as those that use the Reynolds-averaged Navier–Stokes(RANS) equations along with a turbulence closure and a robust free surface-tracking scheme, are an ideal alternative for simulation of complicated nearshore processes that involve breaking waves. In general, RANS-based models are capable of calculating turbulence energy and energy dissipation due to wave breaking and bottom friction. For example, in one of the RANS-based models, COBRAS (Cornell breaking wave and structure) [27.49–51], the two-

dimensional RANS equations are coupled with the k-ε turbulence closure and the volume of fluid (VOF) method for tracking free-surface position. This and similar models can adequately resolve the wave breaking process and its interactions with the seabed. An example of the types of output that this model provides is given in Fig. 27.3. RANS-based models have not yet been widely accepted into coastal science and engineering practice, as their computational cost is relatively high and the high-resolution physical output provided is often beyond what is needed by engineering standards of practice for design. Where these models are applied, it is usually for two dimensional vertical (2DV) transect studies, where, for example, details of wave breaking [27.52, 53], wave interaction with complex and/or porous structures [27.54], and fluid stresses and pressure distributions [27.55] are studied.

This general class of model is still not practically applicable, even in the academic sense, for simulating waves in 3-D coastal regions. 3-D solvers, using various turbulent closure schemes, are, however, becoming a common basis for large-basin numerical wave tanks. One of the more promising programs is OpenFOAM [27.56], which is an open-source and freely available computational fluid dynamics package. The flexibility in this package, and the ability to build off a significant body of previous computational fluid dynamics (CFD) work, makes it possible to apply the codes towards a variety of coastal modeling problems [27.57]. In the coming decade, it is anticipated that wind wave modeling will continually move towards the utilization of very high resolution, large eddy simulation (LES) tools, as this class of model is the main option for predicting vertical and turbulent structures which drive mixing and transport.

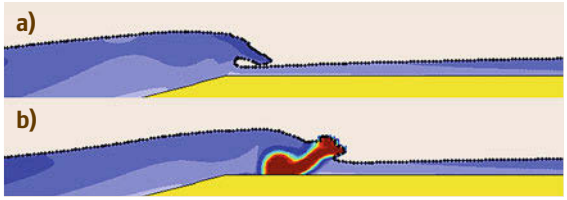


Fig. 27.3a,b Two snapshots from a RANS simulation of wind waves breaking on a shallow shelf. The *colors* indicate the intensity of turbulent kinetic energy. **(a)** shows the free surface and turbulent kinetic energy at moment of jet impingement on the free surface, and **(b)** shows the breaking wave a short time later.

27.2 Modeling Long Waves

A long wave is defined as having a wavelength much larger than the local water depth. Waves that commonly meet this requirement include tides, storm surges, and tsunamis. When the wave disturbance is considered to be *long*, it is reasonable to use the non-dispersive nonlinear shallow water (NSW) wave equations. The NSW equations can be derived in a number of different ways, but fundamentally all arise from an integration of the Euler or Navier–Stokes equations with an assumption of vertically-invariant horizontal velocity and hydrostatic pressure. Due to the simple and well-studied nature of the NSW equations, a wide variety of numerical schemes have been used to solve them.

Models that are used to examine tides and storm surges are often similar, if not the same model. Examples of this class of model include ADCIRC [27.58] and DELFT3D-FLOW [27.59], both of which have received widespread acceptance in the engineering community. These models need to include proper tidal forcing, wind stress for surge, and bottom friction in nearshore areas. Recently, circulation models have been coupled with nearshore short-wave spectral models, which provide a wave-induced water level and current forcing through radiation stress that is ingested by the circulation model [27.60]. This allows for a more complete nearshore simulation, as wave-induced currents become important near the surf zone. Of particular importance for accurate surge prediction is high resolution and accurate coastal bathymetry and topography, often down to 10 m or less. Therefore, as is a common challenge for simulation of long wave generation and evolution, a very wide range of scales must be simulated. Numerical approaches that can accommodate this issue, such as finite element meshing or grid nesting, must be adopted.

With the large number of deadly tsunamis in the past decade, tsunami simulation capabilities have increased relatively rapidly. Several tsunami computational models are currently used in the National Tsunami Hazard Mitigation Program, sponsored by the National Oceanic and Atmospheric Administration, to produce tsunami inundation and evacuation maps for the states of Alaska, California, Hawaii, Oregon, and Washington. The computational models include MOST (method of splitting tsunami), developed originally by researchers at the University of Southern California [27.61]; COMCOT (Cornell multi-grid coupled tsunami model), developed at Cornell University [27.62]; and TUNAMI-N2, developed at Tohoko University in Japan [27.63]. All three models solve the same depth-integrated and 2-D horizontal (2DH)

nonlinear shallow water equations with different finite-difference algorithms.

Successful simulation of tsunami propagation and accurate prediction of the arrival time and wave height at different locations rely on a correct estimate of the earthquake fault plane mechanism. Interplate faults in subduction zones are responsible for most of the large tsunamis in history. For interplate fault ruptures, the resulting seafloor displacement can be estimated approximately using linear elastic dislocation theory [27.64, 65]. For more sophisticated fault models, non-uniform stress-strength fields (faults with various kinds of barriers, etc.; [27.66]) are expected, so the actual seafloor displacement may be very irregular. Once the seafloor displacement has been determined, the initial ocean free surface profile is assumed to take the same configuration, based on the assumptions that the upward seafloor movement is impulsive and seawater is incompressible.

For a given source region condition specified by either the initial free surface elevations or a time history of sea floor displacement, existing models can accurately simulate propagation of a tsunami over a long distance, provided that bathymetry data exists. Figure 27.4 shows a snapshot of free surface elevation from a COMCOT simulation at 10 h and 44 min after the March 11, 2011 Japan tsunami. The shallow-water equation models by definition lack the capability of simulating dispersive waves, which could well be the dominating features in landslide-generated tsunamis [27.39] and for tsunamis traveling a long distance [27.67]. To address this issue of dispersivity, a different set of governing equations must be employed, or some manipulation of the numerical truncation error must be made.

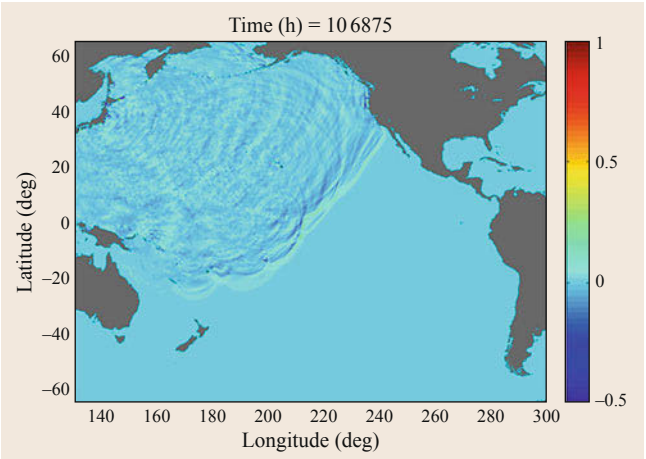


Fig. 27.4 Snapshot from tsunami simulation of the March 11, 2011 Japan tsunami

As a tsunami propagates into the nearshore region, the wave front undergoes a nonlinear transformation while it steepens through shoaling. If the tsunami is large enough, it can break at some offshore depth and approach land as a bore. Wave breaking in traditional NSW tsunami models has not been handled in a physically satisfactory manner. Numerical dissipation is commonly used to mimic breaking, and thus results become grid dependent. In Boussinesq models, this breaking is still handled in an approximate manner due to the fact that depth-integrated derivation does not allow for an overturning wave; however these breaking schemes have been validated for a wide range of nearshore conditions [27.68].

27.3 Coupled and Nested Techniques

As mentioned above, there have been recent successful attempts at coupling spectral wave models with circulation models, which allows for robust simulation of nearshore currents and the resulting transport. If one is interested in basin-scale simulation of wind waves with high coastal resolution, a natural coupling technique would be to use spectral wave information taken from a large-scale model, such as SWAN or STWAVE, to drive a detailed near-coast model, such as the Boussinesq model. Since these two models function in different dimensional space, the coupling between the two models is likely to be one way; meaning that information is given by the spectral model, but no information is received by the spectral model. The implementation of this coupling is straightforward, as it is with most one-way coupling methods, and has been used in a number of studies.

If examining transient wave phenomena or wave hydrodynamics on a fine scale, spectral models are of limited use. Of the two typical phase-resolving, depth-integrated models used in wave studies, the Boussinesq model can be considered a more physically complete (or at least a higher-order) approximation compared to the NSW. In coastal regions, where the water depth is very shallow and thus amplitude is large and wavelength becomes short, nonlinear and bathymetric interactions across a wide range of frequencies occur. These interactions can locally generate various shorter-crested, or dispersive wave components, even if the offshore forcing is considered a long wave. A well-known example is the transformation of a tsunami front into an undular bore. Thus, the nearshore is expected to be nonlinear and (possibly) dispersive, and a Boussinesq model is appropriate. However, the additional physics

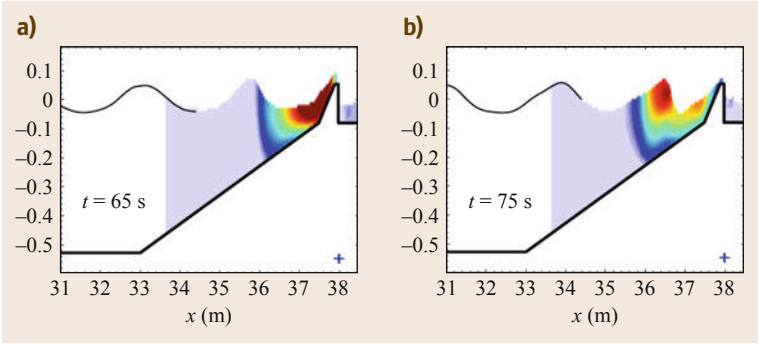


Fig. 27.5a,b Example hybrid Boussinesq-RANS setup; the Boussinesq domain stretches from $x = 0-34$ m and the RANS exists from $x = 34-39$ m. **(a)** shows an overtopping wave at the time of maximum speed on the front of the seawall, while **(b)** depicts the wave and turbulent kinetic energy during the downrush of water along the seawall

included in the Boussinesq approximation come with substantial computational cost, often making the model impractical for ocean basin scale simulations. If one wants to use the physical advantages of the Boussinesq model for a local region in the nearshore zone, it becomes necessary to couple that Boussinesq model with some other source of wave information for its boundary conditions. The obvious coupling choice would be the NSW, proven for both efficient and accurate basin scale tsunami prediction.

There are numerous challenges with this coupling, most notably that fact that the two approximations (NSW and Boussinesq) are different, and so there can be a physical mismatch across the coupling interface. Also, the NSW will typically have a low-order numerical solution approach (hence its computational efficiency), while high-order partial derivations in the Boussinesq model require a high-order scheme; matching these two schemes can also create numerical stability issues. Details regarding a coupling NSW-Boussinesq approach can be found in [27.69].

While Boussinesq models can provide a detailed description of the flow due to a tsunami, it can be necessary to use a model with even less restrictive physical approximations, and the ability to resolve very small-scale (sub meter) turbulent features, such as a 3-D Navier–Stokes model. A similar coupling argument as provided above for the Boussinesq can be made here for the fully 3-D model. Due to their high computational costs, full 3-D models would best be used in conjunction with a depth-integrated two-dimensional-horizontal model (2DH) (i. e., depth-integrated NSW or Boussinesq). While the 2DH model provides incident wave information, the 3-D model computes local wave–structure interactions. The results from 3-D models could also provide a better parameterization of small-scale features (3-D), which could then be embedded in a large-scale 2DH model. One-way coupling (using a NSW-generated time series to drive a 3-D model, but not permitting feedback from the 3-D model back into the NSW) is fairly straightforward to construct [27.70]. Two-way coupling, however, is difficult and requires consistent matching of physics and numerical schemes across model interfaces.

Sittanggang and *Lynett* [27.71] presented work on coupling of a Boussinesq model and 2-D Navier–Stokes model. The two models are two-way coupled, and so act as if they are a single model working on a continuous domain. In the coupling implementation, the Boussinesq model is applied in the non-breaking zone and the RANS model in the breaking/high-turbulence zone. The two models share a common domain interface for exchanging data, used as boundary conditions in the models. By coupling the two models, accurate large-scale wave simulation using a coarse grid and *simple* physics in the deep-to-intermediate water and fine grid and detailed physics in the nearshore area is computationally feasible. Figure 27.5 gives an example of this hybrid simulation approach for nonlinear wind waves overtopping a coastal structure. The wind waves propagate from offshore with the Boussinesq model. Approximately one wavelength offshore of the break point, the Boussinesq and RANS models are two-way coupled, giving the ability for high resolution of the wave overtopping and a relatively large domain.

27.4 Summary of Model Properties

Table 27.1 provides a descriptive listing of the various models described in this chapter. Included in the table are typical spatial and temporal extents used in simulations along with resolutions, the number of derivatives operators found in the governing equations, and common applications and outputs. The information in the derivative operators column, which also includes the highest order of differentiation, is

Table 27.1 Summary of model types, with basic information for comparison purposes

Theory	Typical spatial domain sizes and spatial resolution	Typical simulation time coverage and time step	Number of derivative operators on unknowns in governing equations and highest derivative order	Typical applications and output types
Spectral wave energy balance equations	From planet-wide domains at ≈ 30 arcmin resolution to $O(100)$ km^2 at $O(100)$m resolution	Hours of simulation time with time steps of $O(5)$min	Four derivatives per frequency, with ≈ 250 frequencies. All derivatives are first order	Global scale operational wave prediction, coastal wave prediction, and wave drivers for nearshore circulation models. Models provide frequency-directional spectra
Phase-resolving nonlinear frequency domain equations	$O(1)$km^2 domains at $O(1)$m resolution	Hours, though dependent on number of frequencies retained. Time periodic models – no time step	Two derivatives per frequency, with 150–300 frequencies. Highest derivative order is 2	Nearshore and surf zone wave prediction, wave drivers for nearshore circulation models and models of instantaneous sediment transport. Models provide complex (phase-dependent) amplitudes of free surface elevation
Nonlinear shallow water (NSW) equations	From planet-wide domains at ≈ 2 arcmin resolution to $O(100)$km^2 domains at $O(10)$m resolution	Hours to multiple days of simulation time with time steps of $O(1)$ s	Approximately 10 in the 2DH formulation, depending on the specific formulation. All derivatives are first order	Coastal circulation modeling, tides, river and overland flows, tsunamis. Model provides water surface elevation and depth-averaged velocity components
Boussinesq-type equations	$O(1)$km^2 domains at $O(1)$m resolution	Tens of minutes to a few hours, with time steps of $O(0.1)$ s	Approximately 15 in the 2DH weakly nonlinear, standard formulations, with highest derivative order of 3. In the highly nonlinear 2DH versions, approximately 50, with highest derivative order of 3	Coastal wave prediction, breaking waves, setup, runup, overtopping, and coastal flooding due to waves. Model provides water surface elevation and velocity components at some z-level, which can be post-processed to construct the vertical profile of velocity
Navier–Stokes (NS) equations	Coastal applications currently focus on transect (2DV) studies, covering $O(1)$km of horizontal length, with depths less than 10 m. Horizontal resolutions are $O(1)$m and vertical resolutions are $O(0.1)$m	Tens of minutes to a few hours, with time steps of $O(0.01)$ s	Approximately 40 in the 2DV RANS equations with standard k-e turbulence closure model. Highest derivative order is 2	Detailed coastal wave prediction, breaking wave dynamics, turbulence, pressure distributions, wave loads on structures, wave transmission through porous structures. Model provides velocity components and turbulence information throughout the water column

meant to provide a basis to compare the computational cost of the various models; this is a measure of the effort required to determine a solution at every model grid node at every time step. It is remarked that this table provides information relative to *typical* model usage, here meaning applications that might arise in engineering practice or on an operational level. Research-level applications push the domain sizes, resolutions, and applications well beyond the types given below.

27.5 Conclusions

In this chapter, we have provided a cursory overview of coastal wave modeling theory and methods. The information contained here provides a useful starting point for understanding the differences in the wide variety of approaches. Ultimately, the choice of model will be governed by the resources available for the study and the primary piece of physical information that requires estimation. If resources are light and a precise estimate of the wave conditions is not yet needed, as might be the case for a preliminary scoping analysis, linear wave theory may be an acceptable approach for coastal transformation. If accurate and relatively large-scale wave height and period information is the goal, spectral models will most likely be the best option, with the included level of nonlinearity and resolution controlled by the resources of the project. For studies requiring detailed, small-scale, and wave-resolved information, such as might be needed for critical infrastructure, a Boussinesq or Navier–Stokes approach is reasonable. It is reiterated that the background given here, in terms of the individual models discussed, does not represent the wealth of numerical modeling tools available to the international engineering community.

27.6 Nomenclature

η free surface elevation
H wave height
a wave amplitude
θ wave phase function
k wave number
ω angular frequency
L wavelength
T wave period
f wave frequency
g gravity
h water depth
c wave speed
u horizontal fluid particle velocity
H_o offshore wave height
K_S shoaling coefficient
K_R refraction coefficient
K_D diffraction coefficient
α wave angle of incidence
F wave energy flux
N time-dependent wave action density
S external wave forcing, source/sink term in spectral wave equations
C_b bottom friction coefficient
G functional form of wave energy sink
E local wave energy density
ε wave amplitude divided by water depth $= a/h$
μ wave number times water depth $= kh$

References

27.1 N. Lin, K. Emanuel, M. Oppenheimer, E. Vanmarcke: Physically based assessment of hurricane surge threat under climate change, Nat. Clim. Change **2**, 462–467 (2012)

27.2 R.G. Dean, R.A. Dalrymple: *Water Wave Mechanics for Engineers and Scientists* (Prentice Hall, Englewood Cliffs 1984)

27.3 US Army Corps of Engineers: Coastal Engineering Manual. Engineer Manual 1110-2-1100, http://chl.erdc.usace.army.mil/cem (2002), in 6 volumes

27.4 B. Raubenheimer, R.T. Guza, S. Elgar: Wave transformation across the inner surf zone, J. Geophys. Res. **101**(C11), 25589–25597 (1996)

27.5 W.R. Dally, R.G. Dean, R.A. Dalrymple: Wave height variation across beach of arbitrary profile, J. Geophys. Res. **90**(C6), 11917–11927 (1985)

27.6 J.A. Battjes, J.P.F.M. Janssen: Energy loss and set-up due to breaking of random waves, Proc. 16th Coast. Eng. Conf. (1978) pp. 569–587

27.7 E.B. Thornton, R.T. Guza: Transformation of wave height distribution, J. Geophys. Res. **88**, 5925–5938 (1983)

27.8 D.T. Resio: The estimation of wind-wave generation in a discrete spectral model, J. Phys. Oceanogr. **2**(4), 510–525 (1981)

27.9 H.L. Tolman: User manual and system documentation of WAVEWATCH-III. Version 1.15. NOAA/NWS/NCEP/OMB Technical Note 151, http://polar.ncep.noaa.gov/mmab/papers/tn151/OMB_151.pdf (1997)

27.10 G.J. Komen, L. Cavaleri, M. Donelan, K. Hasselmannn, S. Hasselmannn, P.A.E.M. Janssen: *Dynamics and Modeling of Ocean Waves* (Cambridge Univ. Press, Cambridge 1994)

27.11 S. Wornom, D.J.S. Welsh, K.W. Bedford: On coupling the SWAN and WAM wave models for accurate nearshore wave predictions, Coast. Eng. J. **43**(3), 161–201 (2001)

27.12 N. Booij, R.C. Ris, L.H. Holthuijsen: A third-generation wave model for coastal regions, Part I: Model description and validation, J. Geophys. Res. **104**(C4), 7649–7666 (1999)

27.13 K. Hasselmann, T.P. Barnett, E. Bouws, H. Carlson, D.E. Cartwright, K. Enke, J.A. Ewing, H. Gienapp, D.E. Hasselmann, P. Kruseman, A. Meerburg, P. Müller, D.J. Olbers, K. Richter, W. Sell, H. Walden: *Measurements of Wind–Wave Growth and Swell Decay During the Joint North Sea Wave Project (JONSWAP)* (Deutsches Hydrographisches Institut, Hamburg 1973)

27.14 L. Cavaleri, J.-H.G.M. Alves, F. Ardhuin, A. Babanin, M. Banner, K. Belibassakis, M. Benoit, M. Donelan, J. Groeneweg, T.H.C. Herbers, P. Hwang, P.A.E.M. Janssen, T. Janssen, I.V. Lavrenov, R. Magne, J. Monbaliu, M. Onorato, V. Polnikov, D. Resio, W.E. Rogers, A. Sheremet, J. McKee Smith, H.L. Tolman, G. van Vledder, J. Wolf, I. Young: Wave modelling – The state of the art, Prog. Oceanogr. **75**, 603–674 (2007)

27.15 W.E. Rogers, J.M. Kaihatu, H.A.H. Petit, N. Booij, L.H. Holthuijsen: Diffusion reduction in an arbitrary scale third generation wind wave model, Ocean Eng. **29**, 1357–1390 (2002)

27.16 J.M. Smith: Full-plane STWAVE with bottom friction: II. Model overview. CHETN-I-75, http://chl.erdc.usace.army.mil/chetn (2007)

27.17 J.T. Kirby, R.A. Dalrymple: A parabolic equation for the combined refraction-diffraction of Stokes waves by mildly varying topography, J. Fluid Mech. **136**, 453–466 (1983)

27.18 J.T. Kirby, R.A. Dalrymple: *Combined Refraction/Diffraction Model REF/DIF 1, Version 2.5. Documentation and User's Manual*, Vol. Res. Rep. No. CACR-94-22 (Center for Applied Coastal Research, Department of Civil Engineering, University of Delaware, Newark 1994)

27.19 M.H. Freilich, R.T. Guza: Nonlinear effects on shoaling surface gravity waves, Philos. Trans. R. Soc. A **311**, 1–41 (1984)

27.20 Y. Agnon, A. Sheremet, J. Gonsalves, M. Stiassnie: Nonlinear evolution of a unidirectional shoaling wave field, Coast. Eng. **20**, 29–58 (1993)

27.21 J.M. Kaihatu, J.T. Kirby: Nonlinear transformation of waves in finite water depth, Phys. Fluids **7**, 1903–1914 (1995)

27.22 T.T. Janssen, T.H.C. Herbers, J.A. Battjes: Generalized evolution equations for nonlinear surface gravity waves over two-dimensional topography, J. Fluid Mech. **552**, 393–418 (2006)

27.23 J.M. Kaihatu: Frequency domain models in the nearshore and surf zones. In: *Advances in Coastal Modeling*, ed. by V.C. Lakhan (Elsevier, Amsterdam 2003)

27.24 J.M. Kaihatu: Improvement of parabolic nonlinear dispersive wave model, J. Waterw. Port Coast. Ocean Eng. **127**, 113–121 (2001)

27.25 Y. Agnon, A. Sheremet: Stochastic nonlinear modeling of directional spectra, J. Fluid Mech. **345**, 79–99 (1997)

27.26 J.M. Kaihatu, J. Veeramony, K.L. Edwards, J.T. Kirby: Asymptotic behavior of frequency and wavenumber spectra of nearshore shoaling and breaking waves, J. Geophys. Res. **112**, C06016 (2007)

27.27 T.H.C. Herbers, M.C. Burton: Nonlinear shoaling of directionally spread waves on a beach, J. Geophys. Res. **102**, 21101–21114 (1997)

27.28 D. Peregrine: Long waves on a beach, J. Fluid Mech. **27**(4), 815–827 (1967)

27.29 S. Elgar, R.T. Guza: Shoaling gravity waves: A comparison between data, linear finite depth theory and a nonlinear model, J. Fluid Mech. **158**, 47–70 (1985)

27.30 D.G. Goring: Tsunamis – The Propagation of Long Waves Onto a Shelf, Ph.D. Thesis (California Institute of Technology, Pasadena 1978)

27.31 P.L.-F. Liu, S.B. Yoon, J.T. Kirby: Nonlinear refraction-diffraction of waves in shallow water, J. Fluid Mech. **153**, 184–201 (1985)

27.32 O. Nwogu: Alternative form of Boussinesq equations for nearshore wave propagation, J. Waterw. Port Coast. Ocean Eng. **119**(6), 618–638 (1993)

27.33 P.A. Madsen, R. Murray, O.R. Sørensen: A new form of the Boussinesq equations with improved linear dispersion characteristics (Part 1), Coast. Eng. **15**, 371–388 (1991)

27.34 Y. Chen, P.L.-F. Liu: The unified Kadomtsev-Petviashvili equation for interfacial waves, J. Fluid Mech. **288**, 383–408 (1995)

27.35 J.M. Witting: A unified model for evolution of nonlinear water waves, J. Comput. Phys. **56**, 203–236 (1984)

27.36 Q. Chen, P.A. Madsen, H.A. Schaffer, D.R. Basco: Wave-current interaction based on an enhanced Boussinesq approach, Coast. Eng. **33**, 11–39 (1998)

27.37 P.L.-F. Liu: Model equations for wave propagation from deep to shallow water. In: *Advances in Coastal and Ocean Engineering*, Vol. 1, ed. by P.L.-F. Liu (World Sciectific, Singapore 1994)

27.38 G. Wei, J.T. Kirby, S.T. Grilli, R. Subramanya: A fully nonlinear Boussinesq model for surface waves. I. Highly nonlinear, unsteady waves, J. Fluid Mech. **294**, 71–92 (1995)

27.39 P. Lynett, P.L.-F. Liu: A numerical study of submarine landslide generated waves and runup, Proc. R. Soc. Lond. A **458**, 2885–2910 (2002)

27.40 Q. Chen, R.A. Dalrymple, J.T. Kirby, A. Kennedy, M.C. Haller: Boussinesq modeling of a rip current system, J. Geophys. Res. **104**(C9), 20617–20637 (1999)

27.41 Q. Chen, J.T. Kirby, R.A. Dalrymple, F. Shi, E.B. Thornton: Boussinesq modeling of longshore currents, J. Geophys. Res. **108**(C11), 3362 (2001)
27.42 P. Lynett, T.-R. Wu, P.L.-F. Liu: Modeling wave runup with depth-integrated equations, Coast. Eng. **46**(2), 89–107 (2002)
27.43 S. Ryu, M.H. Kim, P. Lynett: Fully nonlinear wave-current interactions and kinematics by a bem-based numerical wave tank, Comput. Mech. **32**, 336–346 (2003)
27.44 J. Veeramony, I.A. Svendsen: The flow in surf zone waves, Coast. Eng. **39**, 93–122 (2000)
27.45 R.E. Musumeci, I.A. Svendsen, J. Veeramony: The flow in the surf zone: A fully nonlinear Boussinesq-type of approach, Coast. Eng. **52**(7), 565–598 (2005)
27.46 D.-H. Kim, P.J. Lynett, S.A. Socolofsky: A depth-integrated model for weakly dispersive, turbulent, and rotational fluid flows, Ocean Model. **27**(3/4), 198–214 (2009)
27.47 D.H. Kim, P.J. Lynett: Turbulent mixing and scalar transport in shallow and wavy flows, Phys. Fluids **23**, 016603 (2011)
27.48 K. Sitanggang, P. Lynett: Parallel computation of a highly nonlinear Boussinesq equation model through domain decomposition, Int. J. Numer. Methods Fluids **49**(1), 57–74 (2005)
27.49 P. Lin, P.L.-F. Liu: Turbulent transport, vorticity dynamics, and solute mixing under plunging breaking waves in surf zone, J. Geophys. Res. **103**(C8), 15677–15694 (1998)
27.50 P. Lin, P.L.-F. Liu: A numerical study of breaking waves in the surf zone, J. Fluid Mech. **359**, 239–264 (1998)
27.51 T.-J. Hsu, T. Sakakiyama, P.L.-F. Liu: A numerical model for wave motions and turbulence flows in front of a composite breakwater, Coast. Eng. **46**(1), 25–50 (2002)
27.52 Q. Zhao, S. Armfield, K. Tanimoto: Numerical simulation of breaking waves by a multi-scale turbulence model, Coast. Eng. **51**(1), 53–80 (2004)
27.53 J.L. Lara, I.J. Losada, M. Maza, R. Guanche: Breaking solitary wave evolution over a porous underwater step, Coast. Eng. **58**(9), 837–850 (2011)
27.54 N. Garcia, J.L. Lara, I.J. Losada: 2-D numerical analysis of near-field flow at low-crested permeable breakwaters, Coast. Eng. **51**(10), 991–1020 (2004)
27.55 A. Pedrozo-Acuña, A. Torres-Freyermuth, Q. Zou, T.-J. Hsu, D.E. Reeve: Diagnostic investigation of impulsive pressures induced by plunging breakers impinging on gravel beaches, Coast. Eng. **57**(3), 252–266 (2010)
27.56 OpenFoam: The OpenSource CFD toolbox, User guide, Version 1.4.1., http://www.openfoam.org/docs (2007)
27.57 P. Higuera, J.L. Lara, I.J. Losada: Realistic wave generation and active wave absorption for Navier–Stokes models: Application to OpenFOAM, Coast. Eng. **71**, 102–118 (2013)
27.58 C. Dawson, J.J. Westerink, J.C. Feyen, D. Pothina: Continuous, discontinuous and coupled discontinuous-continuous Galerkin finite element methods for the shallow water equations, Int. J. Numer. Methods Fluids **52**, 63–88 (2006)
27.59 WL|Delft Hydraulics: Delft3D-FLOW User Manual, Version 3.13., http://www.oss.deltares.nl/web/delft3d/manuals (2006), 638p.
27.60 J. Dietrich, S. Tanaka, J.J. Westerink, C.N. Dawson, R.A. Luettich Jr, M. Zijlema, L.H. Holthuijsen, J.M. Smith, L.G. Westerink, H.J. Westerink: Performance of the unstructured-mesh, SWAN+ADCIRC model in computing hurricane waves and surge, J. Sci. Comput. **52**(2), 468–497 (2012)
27.61 V.V. Titov, C.E. Synolakis: Numerical modeling of tidal wave runup, J. Waterw. Port Coast. Ocean Eng. **124**(4), 157–171 (1998)
27.62 P.L.-F. Liu, Y.-S. Cho, S.B. Yoon, S.N. Seo: Numerical simulations of the 1960 Chilean tsunami propagation and inundation at Hilo, Hawaii. In: *Recent Development in Tsunami Research*, ed. by M.I. El-Sabh (Kluwer, Dordrecht 1994) pp. 99–115
27.63 F. Imamura, N. Shuto, C. Goto: Numerical simulations of the transoceanic propagation of tsunamis, Proc. 6th Cong. Asian Pac. Reg. Div. (IAHR) (1988) pp. 265–272
27.64 L. Mansinha, D.E. Smylie: The displacement fields of inclined faults, Bull. Seismol. Soc. Am. **61**, 1433–1440 (1971)
27.65 Y. Okada: Surface deformation due to shear and tensile faults in a half-space, Bull. Seism. Soc. Am. **75**(4), 1135–1154 (1985)
27.66 H. Kanamori: The energy release in great earthquakes, J. Geophys. Res. **82**, 2981–2987 (1977)
27.67 S.T. Grilli, M. Ioualalen, J. Asavanant, F. Shi, J.T. Kirby, P. Watts: Source constraints and model simulation of the December 26, 2004 Indian Ocean tsunami, J. Waterw. Port Coast. Ocean Eng. **133**, 414–428 (2007)
27.68 P. Lynett: Nearshore modeling using high-order Boussinesq equations, J. Waterw. Port Coast. Ocean Eng. **132**(5), 348–357 (2006)
27.69 S. Son, P. Lynett, D.-H. Kim: Nested and multi-physics modeling of tsunami evolution from generation to inundation, Ocean Model. **38**(1/2), 96–113 (2011)
27.70 S. Guignard, S. Grilli, R. Marcer, V. Rey: Computation of shoaling and breaking waves in nearshore areas by the coupling of BEM and VOF methods, Proc. 9th Offshore Polar Eng. Conf., Vol. 3 (1999) pp. 304–309
27.71 K. Sitanggang, P. Lynett: Multi-scale simulation with a hybrid Boussinesq-RANS hydrodynamic model, Int. J. Numer. Methods Fluids **62**, 1013–1046 (2009)

28. Modeling of Coastal Morphological Processes

J.A. Dano Roelvink, Dirk-Jan R. Walstra, Mick van der Wegen, Roshanka Ranasinghe

Morphology is the study of forms, whether of animals, plants or words; coastal morphology is the study of coastal forms. Some of these forms are hard (like rock) and do not move much. Others, such as beaches, dunes, channels, sandbars, shoals, are made of mud, sand, or even gravel and move all the time. Understanding and predicting these motions are what coastal morphologists try to do.

The approach we take depends on the scale of the problem or process we are interested in. This scale ranges from small-scale ripples to entire tidal inlets or coast lines. The larger the length scale, the larger the timescale; the upper beach profile can adapt to a storm in hours; the coastline around a port can take decades to adapt, and the effects of closing a tidal inlet can be felt for many centuries.

In coastal engineering, morphology is a core discipline. It is needed to assess damage to beaches, dunes, and barrier islands during extreme events; it is needed to understand and mitigate coastal erosion; it is important in the design of ports, to assess impact on the coast and estimate dredging requirements for the navigation channel and basins; it is needed in the design and impact studies of land reclamations and finally, it provides scenarios to use in coastal zone management.

In this chapter, we will first discuss different types of coastal morphology models. We will then zoom in and focus on process-based modeling in more detail. First, the fundamentals of process-based modeling are described, which is followed by short descriptions of the commonly used types of process-based models. Subsequently, a number of modeling approaches considering different scales, from storm-event modeling to long-term modeling of estuaries are presented. We conclude by discussing some future directions.

28.1 **Types of Coastal Models** 611
28.1.1 Physical Scale Models 611
28.1.2 Analytical Models 612
28.1.3 (Semi-)Empirical Models 612
28.1.4 Process-Based Models 612
28.2 **Principles of Process-Based Morphodynamic Modeling** 613
28.2.1 What Is a Process-Based Morphodynamic Model? 613
28.2.2 Mass Balance Equation 613
28.2.3 Bed Celerity 613
28.2.4 Development Toward Equilibrium 614
28.2.5 Types of Process-Based Models 614
28.2.6 Upscaling Techniques 617
28.3 **Modeling Approaches** 619
28.3.1 Storm Event Modeling 619
28.3.2 Modeling Cross-Shore Breaker Bar Cycles 623
28.3.3 Long-Term Modeling of Estuaries 626
28.4 **Future Directions** 629
28.4.1 Ensemble Modeling 629
28.4.2 Biogeomorphology 629
28.4.3 Dune Modeling 630
28.4.4 Coral Reefs 630
28.4.5 Modeling Stratigraphy 630
28.4.6 Morphological Modeling of Tsunami Deposits 630
28.5 **Nomenclature** 631
References 631

28.1 Types of Coastal Models

28.1.1 Physical Scale Models

We generally approach these problems with models, of which we have a variety at our disposal or under development. To start with, there are the physical scale models, which appear to be starting a new life, after a period in which all the focus was on numerical models. The Dutch design rules for dune erosion

are almost entirely based on large-scale flume tests and just recently have been adapted for changed design wave conditions based on such tests [28.1]. Scour around structures is another example where physical models are still much in use, in this case because of the complexity of the three-dimensional (3-D) currents and turbulence around them. However, in many cases such a direct application of physical modeling to a problem is difficult, and such models are then used to look into the details of processes and to validate numerical models that can be used at larger scales.

Large-scale physical models of tidal systems, rivers, and estuaries have become too expensive in developed countries to compete with ever-improving numerical models, but are still widely used in countries like China, where labor cost is relatively cheap and there is considerable expertise in dealing with inevitable scale effects. Combining such models with numerical modeling can lead to interesting results, and is probably a direction worth exploring further.

When waves are important and we have to look at large systems at the same time, physical modeling is no longer an option. The modeled waves would be too small to be represented properly, so we will have to look at other methods.

28.1.2 Analytical Models

First, we turn to analytical models. They can be simple and elegant and provide almost instant insight. The coastline theory of *Pelnard-Considère* [28.2] has shaped our thinking about large-scale coastal evolution, and analytical and numerical successors of the theory are still abundantly used [28.3]. Other, more complex, analytical models explain point bars in rivers and estuaries [28.4], sand waves and tidal sandbanks [28.5], and though they use strong approximations of the physics, they tell us which processes are responsible for certain phenomena we see, so we can make sure that more complex numerical models reproduce these processes, which in theory they should, but in practice often fail to do.

28.1.3 (Semi-)Empirical Models

The next type of models is heavily based on data and therefore called *empirical models*. The equilibrium profile model for dune erosion, which has made it into Dutch law, is a good example [28.1, 6]. Based on the observation that beach profiles always have the same shape after a storm, the method consists of shifting that profile back and forth until the eroded area equals the accreted area. A more advanced approach for beach evolution and dune erosion is used in SBEACH [28.7]. SBEACH is a one-dimensional (1-D) (profile) beach erosion model for storm where net sediment transport occurs when the local wave dissipation rate is different from the equilibrium rate. For tidal inlets and estuaries, there is a host of equilibrium equations [28.8], relating the tidal volume to channel cross-section, to total channel volume, to ebb delta volume, etcetera. If you divide a tidal inlet or estuary up into a number of units (ebb delta, channels, tidal flats) you can define equilibrium situations for each of them. However, that does not tell you how they get to their equilibrium. To solve this, *semiempirical* models were developed, where the units all try to get to their equilibrium but sediment has to be transported between them. The rate of sediment exchange depends on how far each unit is from equilibrium. *Stive* and *Wang* use this concept to model long-term, large-scale effects of sea-level rise and gas extraction, inlet closure, dredging, and other scenarios [28.9].

Such models have shown a lot of value for specific situations, where there is a lot of historical data and nice test cases from the recent past (closure of inlets, channel deepening for navigation) to calibrate the coefficients in such models. In many places around the world this is not the case; also, more complex situations or more detailed questions need to be addressed than can be represented by the semiempirical models we just discussed. In such cases, we turn to process-based morphodynamic modeling.

28.1.4 Process-Based Models

In process-based models, we start from a detailed description of processes and let sediment transports, bottom changes, and the evolving bathymetry follow from that. It is a bottom-up approach, where we do not specify any equilibrium situation beforehand but let the model evolve freely to wherever it wants to go. Obviously, this raises the question of whether such an approach can ever lead to sensible solutions in the longer term. As we will see in the following sections, there is ample scope for reasonable behavior using this approach and in the rest of this chapter we will focus on the process-based approach.

28.2 Principles of Process-Based Morphodynamic Modeling

The term *morphodynamic model* is short for *dynamic morphological model*, where *morphological* means the study of shapes, in this case of the sea bed, and *dynamic* indicates that we consider how sea bed changes as a result of processes acting on it. The basic flowchart given in Fig. 28.1 is valid for most types of morphodynamic models. We call it process-based modeling since we take a detailed description of the physical processes leading to sediment transport and bottom changes as a starting point.

28.2.1 What Is a Process-Based Morphodynamic Model?

We start from a bathymetry, given in two dimensions (in the case of area models) or one dimension (in the case of coastline or coastal profile models). Given boundary conditions for waves and currents, we predict the wave and current fields, which usually interact; together, these processes determine the sediment transport. As we will discuss next, the sediment transport gradients lead to bottom changes which then feed back into the bathymetry, the currents and waves and the sediment transports, and so on. Different models may differ greatly in the way they compute currents, waves, and transports and in the way the bottom changes are computed and fed back into the bathymetry. In the following, we will first discuss the sediment balance and the importance of the feedback between the bed evolution and the evolution of the sediment transport.

28.2.2 Mass Balance Equation

The dominant process acting to change the bottom is the sediment transport, and for coasts as in rivers, the most fundamental equation describing erosion and sedimentation is the mass balance equation or the Exner equation

$$(1-\varepsilon)\,\frac{\partial z_\mathrm{b}}{\partial t}+\frac{\partial q_x}{\partial x}+\frac{\partial q_y}{\partial y}=0\,. \tag{28.1}$$

Here ε is the porosity of the bed, z_b is the vertical position of the bottom (positive upwards) and q_x and q_y the components of sediment transport in x and y direction, respectively, expressed as volume transport per second crossing a unit width ($\mathrm{m}^3/\mathrm{m}/\mathrm{s}$).

The first thing we learn from this equation is that the rate of bed-level change depends on the sediment transport *gradients* rather than on the transport rate itself; an increasing transport (positive gradient) leads to a decrease in bed level (erosion), since more sediment leaves a control volume than enters it; likewise, accretion (sedimentation) occurs when the sediment transport decreases (negative gradient).

If the transport rates would not change with time, the erosion and sedimentation rates would be constant and the bed would move up or down with a constant speed. However, usually the sediment transport changes when the bottom changes, and this interaction creates a dynamic behavior where bed forms move around.

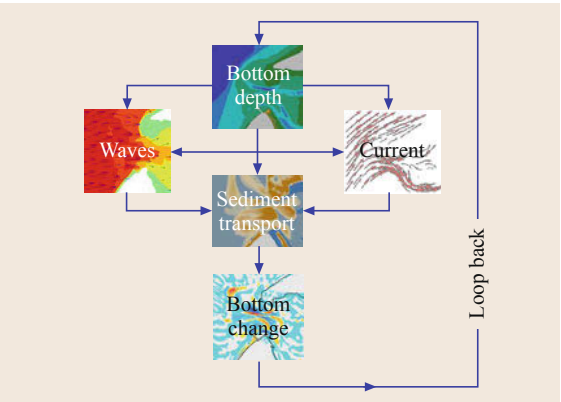


Fig. 28.1 Basic flow chart of a morphodynamic model

28.2.3 Bed Celerity

For one-dimensional, stationary flow through a channel, under the assumptions that transport is a function of velocity to a certain power b and that the current velocity is given by a constant discharge divided by depth h, it follows (see [28.10] for a derivation) that the bed evolution is described by the *simple wave* equation

$$\frac{\partial z_\mathrm{b}}{\partial t}+c_\mathrm{bed}\frac{\partial z_\mathrm{b}}{\partial x}=0\,. \tag{28.2}$$

Here

$$c_\mathrm{bed}=\frac{bq_x}{(1-\varepsilon)\,h}\,. \tag{28.3}$$

These simple equations explain how a hump on the bottom propagates at a speed proportional to the transport rate and to the power b in the transport relation, but inversely proportional to the depth; in other words, in shallow water bed forms move faster. It also explains why humps turn into shock-like dunes (because the top moves faster than the bottom) and for trenches the upstream side steepens, while the downstream side tends

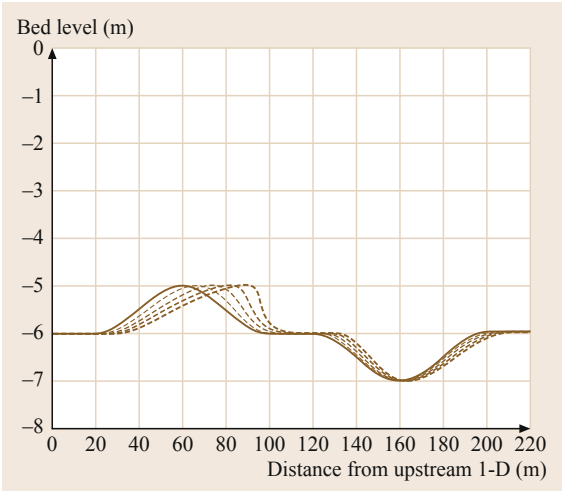


Fig. 28.2 Behavior of a hump and a trench in a channel

to smooth out. This behavior is illustrated in Fig. 28.2 for a 1 m high hump and 1 m deep trench across a channel with uniform depth of 6 m. In coastal areas, the development of such fronts is often seen, for instance, in the development of ebb shoals; likewise the behavior of a dredged channel perpendicular to a longshore current is often similar to what we see in the figure: a downstream migration and steepening of the upstream slope.

The so-called bed celerity c_{bed} also plays an important role in determining the time step of morphological updating in a numerical model, as we will see further on.

Through the simple example of the moving hump, we see that the shape of the transport relation is important in determining the morphological response. If the power b in the sediment transport relation changes, so does the bed celerity. Likewise, adding complexity to the sediment transport relation will lead to a more complex morphological response.

28.2.4 Development Toward Equilibrium

Another important feature of the mass balance equation is that when the divergence of the transport field goes to zero, the bottom change rate also goes to zero and an equilibrium or near-equilibrium is possible even though sediment transport circulations are present. Typical examples are a straight coastline where longshore transport is constant and therefore gradients are zero, or even a tidal inlet bypassing longshore sediment transport in a complex pattern, while experiencing little change. To illustrate this process, let us look at the bed evolution around a newly constructed, 300 m long groyne in an otherwise uniform, 1 km wide channel. The initial water depth is 5 m and the undisturbed current speed is around 1 m/s. Figure 28.3b shows the sediment transport magnitude just before we start updating the bed; the initial rate of erosion and sedimentation (Fig. 28.3a) for this example is of the order of 1.5 m/day. If the same transport pattern would persist, we would have around 500 m of erosion per year; however, since a scour hole develops the sediment transport is reduced by some orders of magnitude, and after 3 years the scour hole is *just* around 10 m deep. Upstream and downstream of the groyne shallow shoals develop; the overall morphological pattern leads to a much smoothed out and reduced sediment transport pattern, which is very close to equilibrium, as we can see from the time series of transport through a number of cross-sections (Fig. 28.3d), which all converge to the same value, almost two orders of magnitude smaller than the initial transport rates.

From these simple examples, we can see that the essence of morphodynamic modeling is the interaction between the sediment transport and the evolving bottom, which leads to moving bed features during transient states and under certain conditions will lead to equilibrium.

The way the sediment transport is a function of the evolving bottom is governed by the currents and waves acting on it and on the sediment properties both in the bed and in suspension. The level of detail at which these processes are described depends on the type of model; modeling approaches that are commonly adopted at present can be broadly divided into coastline models, coastal profile models, and coastal area models. These different approaches are described below.

28.2.5 Types of Process-Based Models

Coastline Models

The oldest type of morphodynamic model was first developed by *Pelnard-Considère* [28.2], who developed an analytical solution for the horizontal movement of the coastline, assuming that the cross-shore profile is constant in shape and that the part between the dune crest and the so-called depth of closure moves seaward or landward depending on the longshore transport gradient

$$h\frac{\partial y}{\partial t}+\frac{\partial S}{\partial x}=0\,. \tag{28.4}$$

Here, h is the active profile height, y is the cross-shore position of the coastline, x is the longshore position along the coastline, and S is the total longshore transport. When, for small deviations of the coast angle, we assume that transport varies linearly with the coast an-

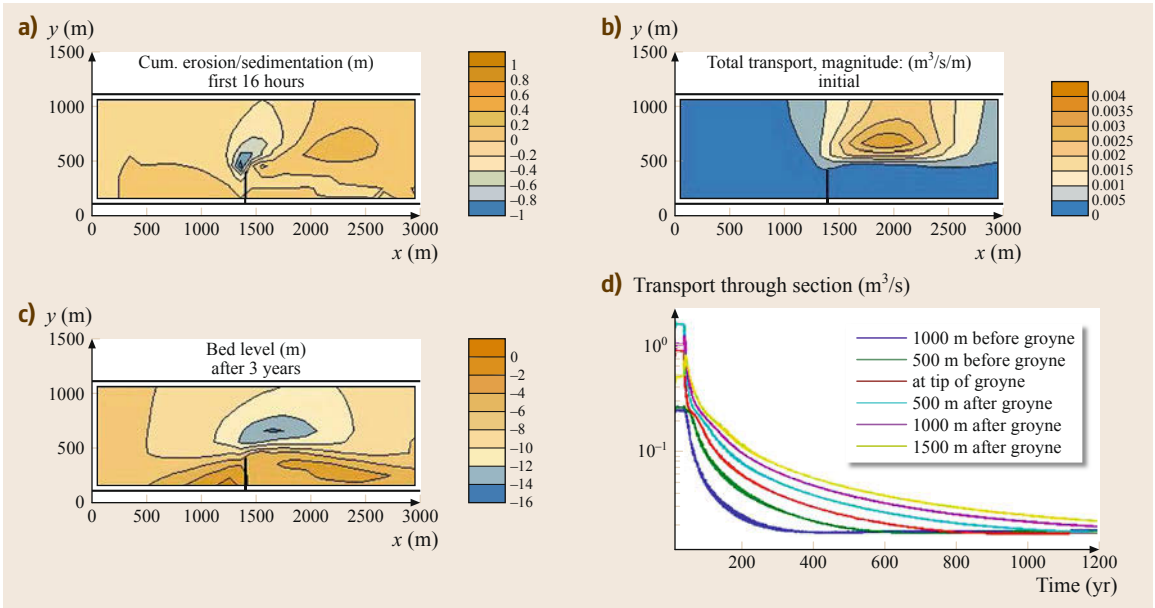


Fig. 28.3a–d Initial transport **(b)** and sedimentation/erosion **(a)**, final bathymetry **(c)** and time series of transport through selected cross-sections **(d)**, for the development of a scour hole near a groyne; sample simulation using Delft3D in 2-D horizontal (2DH) mode

gle and therefore with $\partial y/\partial x$, the coastline evolution is given by a diffusion equation as given by Pelnard-Considère

$$\frac{\partial y}{\partial t} - \frac{s}{h}\frac{\partial^2 y}{\partial x^2} = 0\,. \qquad (28.5)$$

Here s is a coastal constant depending mostly on the wave climate. Analytical solutions exist for this equation, describing for instance the accretion due to a groyne or the longshore dispersion of a nourishment. However, for more practical situations and a wider range of angles and processes the transport and resulting coastline evolution can be numerically evaluated using a bulk formula such as the CERC or Kamphuis formulas, in which case it is mostly a function of the incident wave energy flux and angle, or it may be evaluated using a coastal profile model. Coastline models such as GENESIS, UNIBEST LT/CL, BEACHPLAN, and LITCROSS (for a summary see for instance [28.10]) are widely used in engineering studies, but a word of caution is in order.

Coastline models assume gradually varying flow conditions and approximately parallel depth contours. Under such circumstances, the longshore transport can locally be treated as if it is fully adapted to the local incident wave conditions relative to the coast orientation. Since it typically takes hundreds of meters for the longshore current to spin up, coastline models should in principle be applied for large-scale applications, over alongshore distances of many kilometers. Good examples are the accretion of beaches updrift of a harbor [28.11], the large-scale evolution of river deltas, longshore spreading of large beach nourishments [28.12], evolution of headland embayments, and realignment of coastlines due to changes in wave climate [28.13].

However, with the help of various additional techniques coastline models are also frequently applied at relatively small scales, such as behind emerged offshore breakwaters, around T-groynes and near tidal inlets. This is achieved by combining a coastline model with a two-dimensional (2-D) wave model to predict nearshore wave climates, by extra terms in the longshore transport equations to account for set up differences, and sometimes by schematizing the coast into two interconnected coastlines. Such applications meet with increasing criticism, since a lot of the processes acting on such small scales are not represented in this approach.

Sometimes small-scale structures are somehow represented in a coastline model that is aimed at predicting large-scale trends. This in our opinion is fine, as long as one realizes that the details of the evolution in the vicinity of the structures cannot be trusted.

As explained above, the behavior of coastline models is mostly diffusive. However, when dominant waves approach the coast at an angle of more than approx-

imately 45°, the coastline evolution may become unstable: the maximum longshore transport rate occurs at an angle of around 45°, so if the angle at a given point is higher than that of its upstream neighbor, sand will accumulate there until the maximum transport is reached again. *Ashton* and *Murray* [28.14] elegantly show through coastline modeling that this can explain spectacular coastal shapes such as sand waves, flying spits, and capes.

It is shown in [28.10] that the behavior of coastline models can also turn from diffusive to propagating if the sediment supply is limited, for instance because of the presence of sea walls or groynes. As an example, if the coastline behind a groyne cannot retreat, the erosion shadow propagates alongshore in the downdrift direction. The mechanism for this is that the longshore transport does not only depend on the coast orientation, but also on the absolute coastline position. In Fig. 28.4, we show an example for a nourishment on a sandy versus a sand-starved coast. It must be noted that the latter situation often occurs as population encroaches on the sea and rivers cease to supply sand to the coast.

Coastal Profile Models

In profile models, the main assumption is that the coast is almost uniform in the longshore direction and therefore profile changes are dominated by cross-shore processes. A review of some possible approaches was given in [28.15]; an extensive discussion of physical processes relevant to cross-shore behavior can be found in [28.10]. Profile models typically consist of:

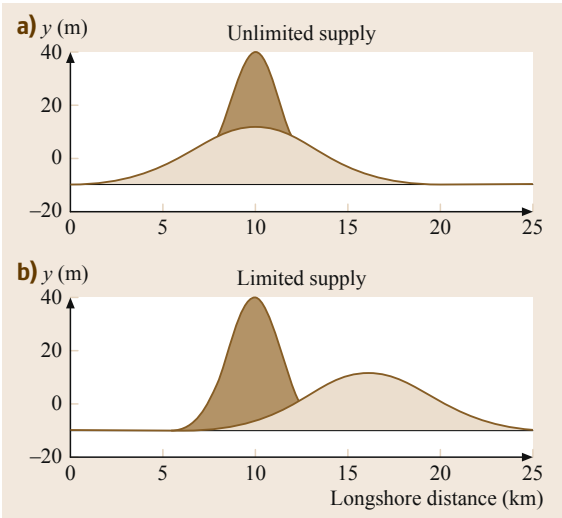


Fig. 28.4a,b Behavior of nourishment on **(a)** sandy coast versus **(b)** sand-starved coast; initial distribution (*dark brown*) and distribution after some time with obliquely incident waves (*light brown*)

- A wave model based on some form of the wave energy balance; this provides local wave conditions at each point along the profile.
- A 1DV (vertical profile) or 2DV model for the undertow and other near-bottom flow processes.
- A sediment transport model, which includes contributions due to undertow, wave skewness, and asymmetry, bed slope, etc.
- A sediment balance equation from which bottom changes are computed.

Since such models are not too computationally heavy, they can be run using realistic variations of wave, wind, and tide boundary conditions (brute force simulation), over periods of up to many years (Sect. 28.3.2).

Coastal profile models are typically applied to address two kinds of problems: evaluation of storm impacts on a coastal profile, with or without hard structures, and evaluation of longer-term behavior of sandbars and nourishment schemes, both on the beach and on the shoreface. The first type is the oldest, and in a way the simplest, since we try to model a situation with a very dominant seaward-directed transport, where the main difficulties lie in predicting the right magnitude and speed of erosion and in achieving the correct profile shape after the storm. Most of the profile changes take place in a few hundreds of meters seaward of the dune front or sea wall, and the duration of such events is at most a few of days. The second type involves the whole active profile, typically of the order of a kilometer cross-shore, down to a depth of some 6–15 m depending on the wave climate and on the simulation duration. Changes in the pattern of bars and troughs take place over years, where in some systems a single bar moves back and forth, and in others bars are created near the shoreline, move seaward initially while growing, and then damping out, allowing a new cycle to begin. This process may take from a couple of years to more than a decade. Such behavior is the result of a subtle balance between various counteracting processes, none of which are modeled very accurately. *Roelvink* et al. [28.16] obtained somewhat realistic, almost periodic bar behavior on a timescale of approx. a decade; *Ruessink* et al. [28.17] after rigorous model sensitivity study and extensive calibration reproduced bar behavior on a timescale of months for three different sites. Systems like this are affected by beach or foreshore nourishments that are of similar scale and clearly interact with the bar pattern. This makes modeling the behavior and longevity of nourishments very difficult: only when the natural bar behavior is modeled somewhat realistically can we have some confidence in the model's capability to predict the relative impact of the nourishment. Similar arguments are valid for hard

structures, which also interact with the profile on the same scale as the longshore bars.

Coastal Area Models

Coastal area models are applied where a separation between longshore and cross-shore scales is not possible, for instance in the vicinity of a tidal inlet, with a complex bathymetry with channels and shoals at varying angles with respect to the undisturbed coast orientation.

The typical composition of an area model is as follows (see also Fig. 28.1):

- A 2-D wave propagation model; by far the most used are spectral wave models, e.g., SWAN [28.18, 19] or Mike21SW; such models solve for wave shoaling and refraction, wind growth, whitecapping, and wave breaking and provide spatially varying field of wave properties and radiation stresses used in the flow and sediment transport models.
- A 2DH, Q3D or 3-D flow model, usually based on the shallow water equations, solving the nonstationary flow fields and water level variations due to tide, wind, waves, and density gradients.
- A sediment transport model, usually combining bed load transport with suspended load transport based on the 2DH or 3-D advection–diffusion equation or on an energetics approach, with specifically for coastal areas important effects of wave stirring included.
- A sediment balance and bottom composition bookkeeping system, which updates the bathymetry and composition of the bottom (in case of multiple sediment fractions).

Area models are applicable at a range of scales, from small-scale coastal engineering problems to macro-scale evolution of tidal basins.

At the smallest scales, we are talking about the details of the waves and wave-driven flow patterns and resulting morphology changes around small structures such as groynes and detached breakwaters, or the erosion of dunes where the coast is not alongshore uniform. The small grid cell sizes required to resolve the main processes and associated small time steps make such problems computationally intensive, especially when predictions over longer time than some storms are required. An important choice is whether to treat the problem as two-dimensional, where the focus is on the effect of the horizontal circulation patterns, or as (quasi-) three-dimensional, where vertical circulations (undertow) are accounted for as well as additional wave-related effects providing an onshore component [28.20, 21]. In the first case, beach profiles may evolve that are far from an equilibrium shape; in the second case, there will be a tendency toward restoring a certain profile shape, but much care must be given to tuning the cross-shore processes in such a way that reasonable profiles develop or are maintained. At these scales, sandbars may also be resolved, but as we discussed under profile models, obtaining the right bar behavior is a tight balancing act. An especially difficult challenge is to model the natural evolution of complex three-dimensional beach topography, where it is a key to get the transitions between different beach states right [28.22–24].

At a larger scale, for instance, in studies of harbor extension, inlet stabilization or large-scale nourishments or land reclamations, details of processes within the surfzone usually cannot be resolved, but the surfzone must be resolved at least enough to generate a decent longshore current and transport; this means a minimum of 5−10 grid cells across the surfzone. As these studies are often aimed at predicting the evolution over months to years, it is important that some cross-shore transport processes are accounted for, so cross-shore profiles stay in a reasonable shape even though bars cannot be represented and must perhaps even be prevented from occurring. Part of the evolution of coastal profiles depends on the interaction with the dunes, in terms of dune erosion and regeneration by wind-blown transport. Such processes are typically ill-represented and some heuristic approaches to exchange sand between the wet model domain and the dry beach and dunes are called for.

In larger scale, longer term situations, for example the long-term evolution of tidal basins, simulations over time spans of centuries are now possible with the help of new acceleration techniques, allowing us to study the equilibrium behavior of tidal basins and the potential impacts of the sea-level rise on such basins [28.25–29]. Surprisingly, realistic results are obtained using relatively simple physics, though a general tendency in such simulations is to generate too deep and narrow channels. Recent research, however, shows that much better channel profiles are obtained when taking into account the spatial variation in grain sizes and especially the tendency of sediment to coarsen with increasing mean shear stress. When the adjacent coasts and ebb deltas are to be properly represented wave effects must also be included. This poses severe restrictions on the nearshore grid resolution and drastically increases the computational cost, typically, by at least an order of magnitude. The applications and scales discussed in the previous paragraphs are summarized in Table 28.1.

28.2.6 Upscaling Techniques

Morphodynamic model components each have their typical timescales and numerical time steps. For flow

Table 28.1 Overview of model applications and scales (after [28.10])

Time scale	Space scales		
	1 m–1 km	10 m–10 km	100 m–100 km
1 hour–10 days	Dune erosion (1-D)[a] Reset events[c] Dune erosion, overwashing and breaching (2-D)[c]		
1 day–10 years	Cyclic bar behavior[a] Effect of shoreface nourishments[a] Evolution of beach states[c] Effects of small-scale coastal structures[c]	Impact of port extensions, land reclamations[c] Mega nourishments[c]	Longshore spreading of nourishments[b] Coastal realignment in response to climate variability[b]
1−1000 years	Climate change impact on profile behavior[a]	Evolution of tidal inlets[c]	Evolution of tidal basins, including climate change impacts[c] Large-scale coastline evolution[b]

Scales of application of [a] profile models, [b] coastline models, [c] area models

models that resolve tide and wind-related variations (hours–days) typical time steps are from seconds to minutes; wave models used are often stationary or slowly-varying with 5−10 min time steps; transport models usually have time steps similar to the flow model. A straightforward way to update bathymetry in a morphodynamic model is to update the bathymetry every flow time step, and to have the morphological time step equal to the flow time step. However, since morphological timescales are often several orders larger than hydrodynamic timescales (compare bed celerity with wave celerity), this is very inefficient, except during extreme conditions when morphology changes rapidly.

To overcome this gap in timescales, several methods have been proposed, see [28.30, 31] or [28.32]. The most common methods used today are:

- Tide-averaging approach: Here, sediment transports are computed through a tidal cycle, while keeping the bottom constant; the transports are then averaged and from the averaged transports the bottom change rate is computed; bottom changes are computed using a morphological time step that may be larger than the tidal period. Often this method is combined with simplified updating of the transports using so-called *continuity correction* or *rapid assessment of morphology* (*RAM*) techniques, to avoid having to re-evaluate the full hydrodynamics too often. Drawbacks of the method are that morphological time steps are limited by bed celerity Courant number and in dynamic systems can become less than the tidal period, which makes the method inefficient; the continuity correction or RAM steps reduce the accuracy, especially in wave-dominated cases.
- *Online* or *morfac* approach: Here the bottom is updated every flow time step but bottom changes are multiplied by a morphological factor or *morfac*. This way, after 1 tidal cycle one obtains the morphological changes of *morfac* tidal cycles. The results within the tidal cycle cannot be trusted because the amplitude of bed changes is exaggerated *morfac* times, but since these intratidal changes are typically very small, the end result is surprisingly accurate. Since updating takes place so often, the method is numerically very smooth; the implementation is also very straightforward, so nowadays in most long-term simulations the *morfac* approach is used. Typical *morfac* values range from 10 for very dynamic, small-scale problems to 200 or more for large-scale tidal basin simulations. With some caution, it is possible to use a time-varying *morfac*, for instance when simulating a series of wave or river discharge conditions, where short energetic conditions are simulated with a lower *morfac* than long moderate conditions. In [28.33], based on the results of a numerical modeling exercise using the morphodynamic model Delft3D, the main dependencies and sensitivities of the MORFAC approach were investigated. A criterion (based on the Courant–Friedrichs–Lévy (CFL) condition for bed form migration) for the a priori determination of the critical MORFAC was also proposed, but further work is needed to extend this criterion to tidal situations.
- *Parallel online* approach, where in the case of a rapidly varying wind and wave climate the chronology of different conditions is not important, simulations for different conditions are carried out in parallel, while the bottom changes due to all conditions are exchanged, averaged, and passed back after every time step. This has the advantage that the simulations for different conditions can be carried out on different processors, and since the averaged result is rather smooth, a relatively high *morfac*

can be used. A drawback of the method is that it is technically somewhat involved, but once implemented correctly it makes long-term simulations in complex areas including wave effects feasible; *Dastgheib* [28.34] used it to make a very detailed 50-year hindcast of the development of the Marsdiep inlet in the Netherlands after closure of the Zuiderzee.

An important aspect of the upscaling approach is the necessary input reduction in the forcing; in order to simulate with high *morfac* or using the tide-averaging approach, the wind, discharge, and wave climate must be schematized to a smaller number of representative conditions. A number of approaches have been developed to do this; for extensive discussions, see [28.10, 35, 36].

28.3 Modeling Approaches

In the following sections, we will discuss a sampling of process-based coastal morphodynamic modeling approaches for cases where there has been much recent development.

In the first approach, we discuss modeling of storm or swell events impacting complex coasts, where longshore uniformity cannot be assumed, impacts may go from collision to full inundation regimes [28.37], and where wave-group-related long waves can be dominant.

In the second study, we describe recent developments and insights into modeling the mechanisms of cross-shore breaker bar cycles, the understanding of which is necessary to design local engineering works or nourishments, which often interact with the bar behavior and are studied on timescales of months to years. In the discussion, we stick to alongshore uniform conditions but the model concepts are readily extendable to coastal area models.

The last case describes the long-term modeling of estuaries, using a schematized 2DH process-based modeling approach. This type of modeling has important implications for understanding the response of tidal systems to sea-level rise, but can also support geological research as it can actually generate realistic morphologies, not just predict bottom changes.

28.3.1 Storm Event Modeling

Background

In this approach, we describe the modeling of fast and complex processes during extreme events, where high temporal and spatial resolution are required. We do this based on a recent model development which has allowed a paradigm shift from simple, quasi-stationary, largely empirical approaches to a physics-based approach where the lessons learned from decades of model development have led to a robust modeling tool where the important processes during such events are resolved. In the following section, we will focus on storm events rather than tsunamis, since the vast majority of modeling studies related to tsunamis, their impact and design consequences deal with hydrodynamics only, as the morphology changes during a tsunami do not affect the flow speeds and runup heights much. However, morphological studies are increasingly carried out in order to interpret tsunami deposits, which will be discussed in Sect. 28.4.6.

The devastating effects of hurricanes on low-lying sandy coasts, especially during the 2004 and 2005 seasons in the USA (Fig. 28.5), have pointed toward an urgent need not only to be able to assess the vulnerability of coastal areas and (re-)design coastal protection for future events, but also to evaluate the performance of existing coastal protection projects compared to *do-nothing* scenarios. Existing dune erosion models such as DUROS [28.38], DUROSTA [28.39], and SBEACH [28.7] had proven their usefulness but could not cope with longshore varying topography, overwashing and inundation, hard structures, and other complexities.

The XBeach model [28.40] was developed to overcome these limitations. It can not only be used as a stand-alone model for small-scale (project-scale) coastal applications, but can also be used within the operational model systems, where it is driven by boundary

Fig. 28.5 Post-Katrina aerial photograph of Dauphin Island, Alabama (USA). Source: USGS/NASA

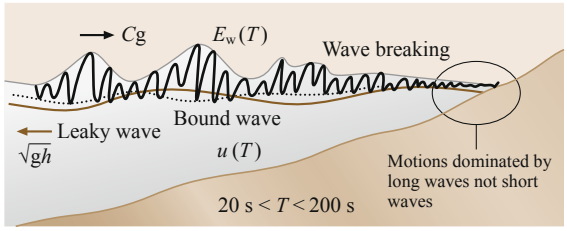


Fig. 28.6 Principle sketch physics of wave groups and surf beat

conditions provided by the wind, wave, and surge models and its main output to be transferred back will be the time-varying bathymetry and possibly discharges over breached barrier island sections.

Physical Processes Included in XBeach

The model solves coupled 2-D horizontal equations for wave propagation, flow, sediment transport, and bottom changes, for varying (spectral) wave and flow boundary conditions.

Because the model takes into account the variation in wave height in time (long known to surfers) it resolves the long wave (infragravity) motions created by this variation. This so-called *surf beat* is responsible for most of the swash waves that actually hit the dune front or overtop it. Because of this innovation, the XBeach model is better able to model the development of the dune erosion profile and to predict when a dune or barrier island will start overwashing and breaching. The physics behind the surf beat formulations in XBeach is illustrated in Fig. 28.6. The individual waves travel in groups and their envelope slowly varies in time and space. The short-wave energy and its modulation propagate at the group speed Cg; associated with these variations are bound long waves; as they propagate into the surf zone, the short-waves break but the bound waves are released and are reflected in offshore direction as *leaky waves*. In total, the result is that the short wave energy at the shoreline is small and the motions there are dominated by the infragravity motions. In XBeach, the individual wave motions are not resolved but the wave energy variation on the wave group scale is, as well as the resulting long-wave motions.

Furthermore, in contrast to most existing dune erosion models, XBeach is a fully two-dimensional (area) model and can thereby take into account longshore variations in topography, effects of the longshore current and gradients in longshore transport.

1-D Validation Tests

The model has already been validated against extensive large-scale flume data sets including short- and long-wave distributions, return flow, orbital velocities, concentrations, and profile change during dune erosion events. An essential part is an avalanching mechanism which allows a surprisingly accurate description of the evolution of the upper profile and dune face. Figure 28.7 shows a comparison between measured and computed dune profile development for large-scale tests carried out at Delft Hydraulics' Delta Flume in 1993 [28.41]. In the lower panels, the importance of both the

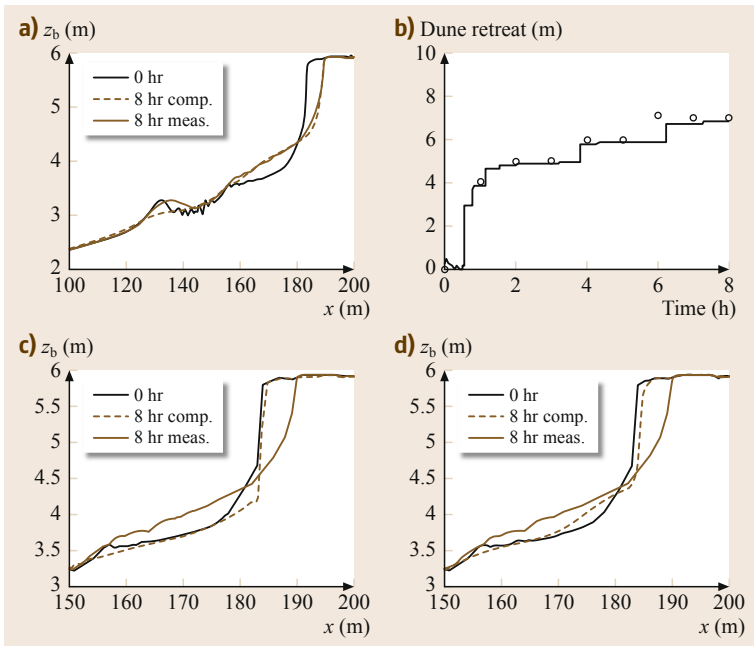


Fig. 28.7a–d Measured and simulated profile development for Deltaflume test (after [28.41]), test 2E; **(a)** all processes; **(b)** retreat of dune front in time, measured (*circles*) versus computed (*drawn line*); **(c)** profile evolution without avalanching; **(d)** without wave groups (after [28.40])

long-wave motions and the avalanching mechanism is illustrated.

In [28.40], extensive validation is shown for hydrodynamics, sediment transport, and profile evolution for a number of test cases. This number has since been extended to practically all large-scale tests used in the past to develop empirical dune erosion models, and results can be found in the regularly updated skillbed report on www.xbeach.org.

2DH Validation Tests

In [28.42], the first 2DH field validation test was carried out, in a hindcast of the impact of hurricane Ivan on Santa Rosa Island, Fl, in the Gulf of Mexico. This simulation has shown that XBeach can simulate complex runup and inundation overwash over longshore varying terrain. The level of detail of these simulations is illustrated in Fig. 28.8, which shows a snapshot of an animation during the onset of overwashing. The model is capable of producing morphological features common to overwash, such as foredune erosion, back barrier deposition, and washover fans. The uncalibrated result showed an overestimation of the erosion during the inundation phase and accompanying extreme current speeds, as the Soulsby–van Rijn transport model used at the time was not developed for such conditions, but after a shear stress limiter was included the behavior was much improved.

In Fig. 28.9, the erosion and sedimentation pattern as simulated and observed from LIDAR (light radar) data is shown; there is a striking agreement both in the erosion zone and in the deposition areas. The quantitative skill of the calibrated model was reasonably good.

Erosion of Curved Coastlines

The original implementation of XBeach uses a rectilinear, nonequidistant finite-difference grid, which is adequate for applications with a relatively short longshore extent or with a rather straight coast, such as the Santa Rosa Island case. However, for longer stretches of a curved coast (typical applications where you would want to use it) it becomes cumbersome to maintain the required fine resolution to resolve dune erosion processes. A curvilinear approach can overcome this problem and allow including the effects of longshore gradients on coastal evolution in a smooth manner.

This first field test case for the curvilinear version of XBeach, described in [28.43], concerns the response of the sandy sea defense of the new Rotterdam port extension Maasvlakte-2 during an early construction phase when it had a banana-like shape (Fig. 28.10). During a storm event on September 3–5 2009 considerable scarping of the steep profiles took place, where

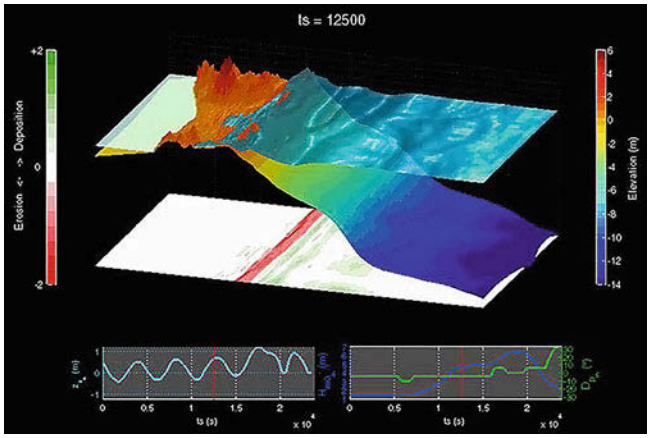


Fig. 28.8 Snapshot of water level, topography and erosion/sedimentation pattern during the overwash phase, Santa Rosa island, hurricane Ivan (after [28.42], courtesy of Dave Thompson, USGS)

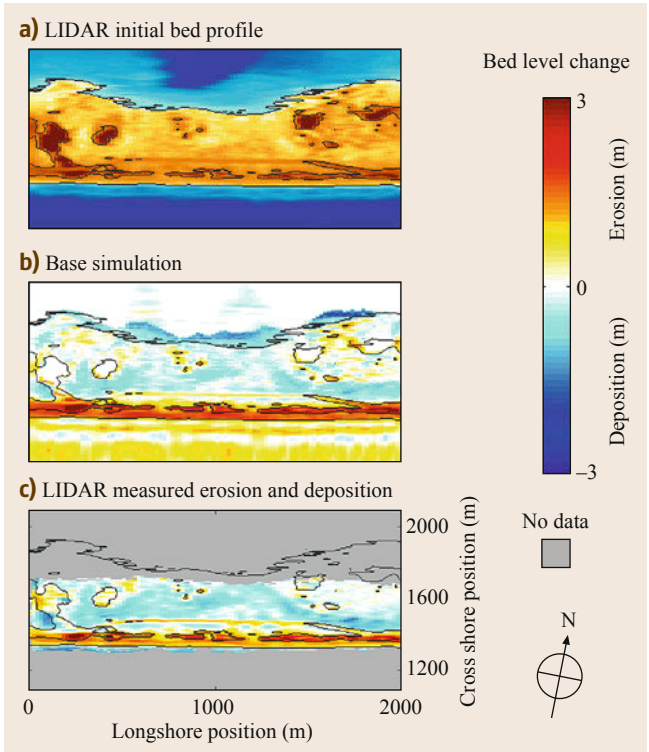


Fig. 28.9 **(a)** Measured LIDAR topography, **(b)** simulated and **(c)** observed sedimentation and erosion, Santa Rosa Island

the eroded sediment was not just deposited seaward but to a large extent was moved alongshore, as can be seen in the third panel. A curvilinear grid was set up with the resolution of 6–20 m in cross-shore and 17–50 m in the longshore direction. Observed time

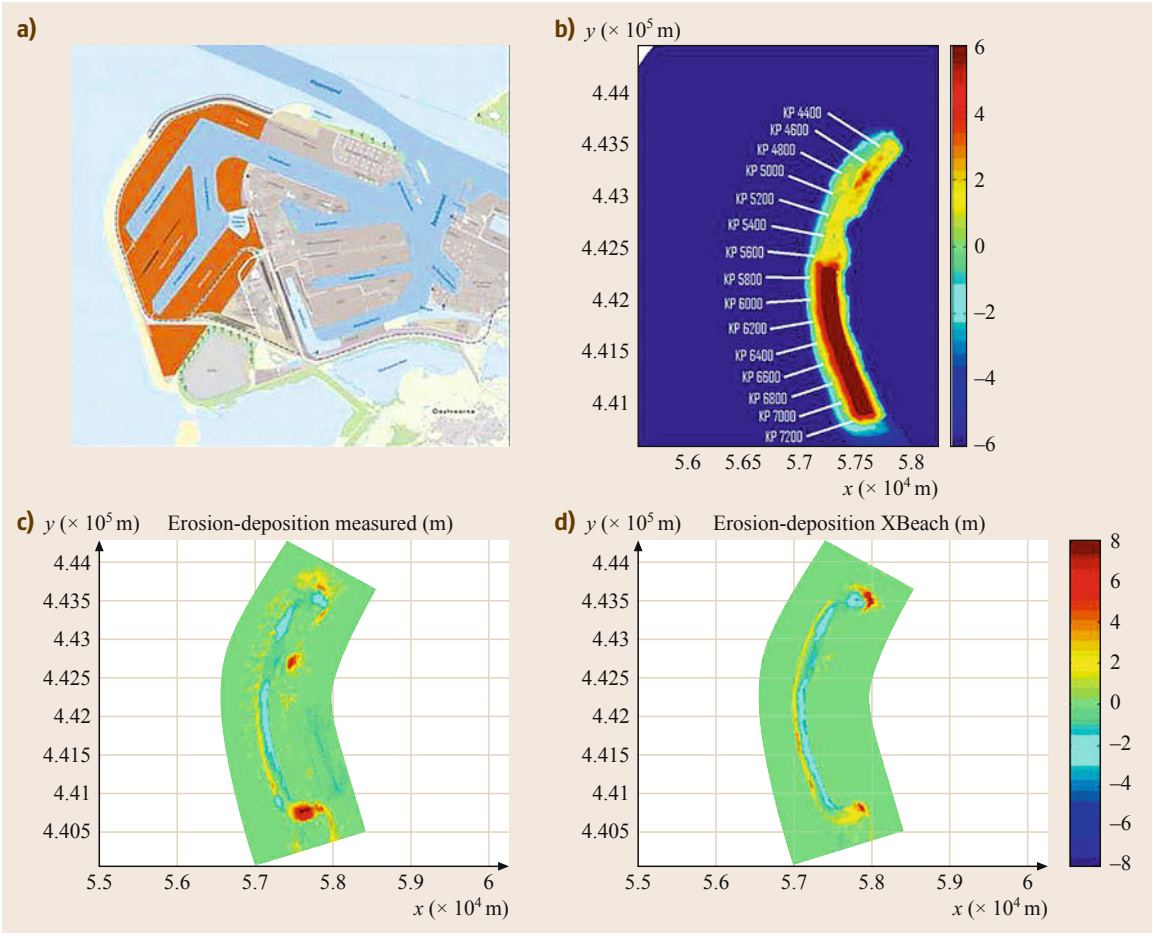


Fig. 28.10 **(a)** Layout of new port extension; **(b)** topography during building phase; **(c)** measured and **(d)** computed sedimentation/erosion

series of water levels and wave conditions were imposed and a simulation was carried out over a period of three days, using default sediment transport settings. Figure 28.10b,d shows the computed erosion and sedimentation. Both in observations and simulation, considerable scarping took place with only a small part being deposited offshore and a larger part moved alongshore and deposited at the ends. This is illustrated further in a cross-sectional view of KP4800 in Fig. 28.11, where the observed profile change is compared against 2DH and 1-D simulations.

This case clearly shows the importance of including the longshore dimension in the modeling of storm erosion.

The model was tested for nine field sites in the framework of the EU-MICORE project (Fig. 28.12). Each field site had a unique bathymetry/topography and/or wave/tidal climate and contributed to testing the

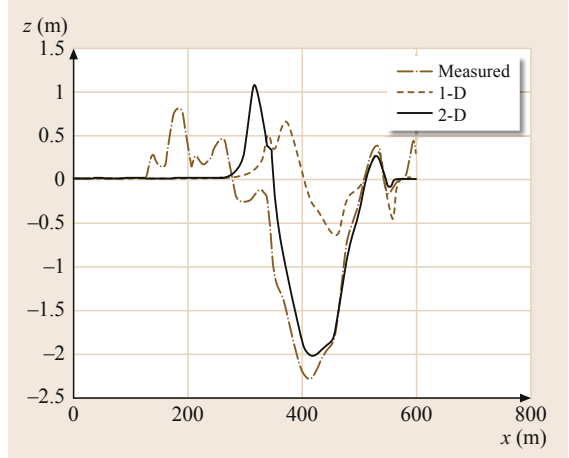


Fig. 28.11 Comparison of profile change between observed, 1-D and 2-D simulations

modeled physics under a wide range of environmental conditions. The full results are presented in [28.44].

In the previous section, we have shown that modeling dune erosion and overwash processes do no longer have to rely on empirical or semiempirical approaches, but that a process-based approach is feasible. The big advantage of such an approach is that it is much more generic and the same model concepts may be applied, after thorough validation, in widely varying environments, from the sandy North Sea coasts to open-ocean coral reef environments.

28.3.2 Modeling Cross-Shore Breaker Bar Cycles

Most sandy coasts have sand-bar features [28.45, 46], which exhibit a range of behavior, on timescales of days to decades. Such features are at similar length and timescales as engineering measures such as groynes, offshore breakwaters, and nourishments, and their state at any given time affects the safety of dikes and dunes behind them. It is therefore very important to unravel the dynamics behind their generation, migration, and fading out. In a recent paper [28.47], important steps were taken to do just that; a summary is given below.

Multiple bar systems often show a cyclic net offshore directed migration with return periods of the order of years. Typically, a bar is generated near the shoreline, grows in height and width, while migrating offshore before finally decaying at the seaward limit of the surf zone [28.48–51]. The observed bar behavior reflects cross-shore oriented bar variability and is not the result of alongshore propagating shore oblique bars [28.52]. This implies cross-shore hydrodynamic and sand transport processes are dominant over longshore processes. Cycle periods vary worldwide between approximately 1 and 15 years [28.49]. The net offshore bar migration is the result of gradual onshore movement during calm periods and episodic strong offshore movement during storms. However, the longshore coherence does not imply that bars along the coast are in the same phase of the cycle. At certain locations, distinct shifts can be observed in which for example the outer is attached to an inner bar, this is often referred to as bar switching. Furthermore, small alongshore variations in bar position and crest height may result in a different morphological response during parts of the cycle period.

Walstra and *Ruessink* [28.53] explored the suitability of a process-based cross-shore profile model (Unibest-TC, [28.17]) to hindcast one cycle at Noordwijk, the Netherlands from 1984 to 1987 using the standardized calibration approach of [28.17]. The multiple bar system at Noordwijk in the Netherlands was selected because it has a relative short bar cycle of about 3–4 years.

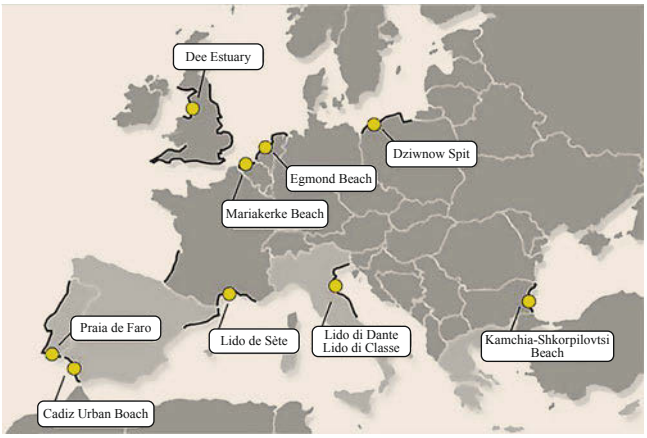


Fig. 28.12 MICORE field test sites

Quantitative descriptions of cyclic large-scale morphological developments have been provided, but the physical reasons underlying the intersite difference in the bar cycle period are not well understood. Among others [28.49] and [28.52] have found various correlations between bar response and environmental parameters. However, these correlations only provide a phenomenological explanation for observed intersite differences. In the following, we summarize [28.47] which investigated the bar amplitude growth and decay within the bar cycle using the [28.17] wave-averaged cross-shore process model to identify the dominant mechanisms driving this transient bar amplitude response.

Mechanisms Underlying Bar Amplitude Change

A detailed analysis, in which the Noordwijk hindcast is combined with a range of schematic cases, is undertaken with the aim of determining the dependency of bar amplitude growth and decay on the offshore wave conditions (height, period, and angle) and cross-shore bar location. Finally, the identified dominant processes that govern bar amplitude change are related to the observations to explain the transient bar amplitude change during the interannual net offshore bar migration.

As a first step to determine when bars grow and decay, dA_b/dt, extracted from the 3.4 year simulation (Fig. 28.13), was correlated with offshore and local (i. e., at the bar crest) wave parameters. Figure 28.14 demonstrates the correlation coefficients significant but lower than 0.5. The offshore wave height and the wave height at the bar crest result in similar correlations. Interestingly, the sign of the correlation changes from bar 1 (most landward) toward bar 4 (most seaward). This indicates the opposite response to similar wave conditions, also observed in Fig. 28.13. Especially the large

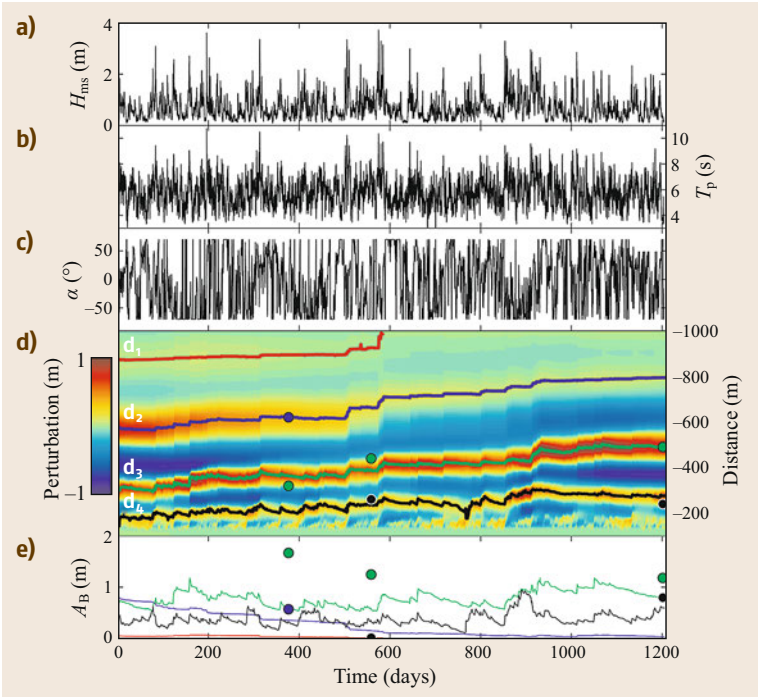


Fig. 28.13a–e Time series of **(a)** offshore root-mean-square wave height $H_{\widehat{rms}}$, **(b)** offshore peak wave period $T_{\hat{p}}$, **(c)** offshore incident wave angle α, **(d)** time stack of predicted profile perturbation and indication of bar crests (after [28.1–4]), and **(e)** predicted bar amplitudes $A_{\hat{b}}$

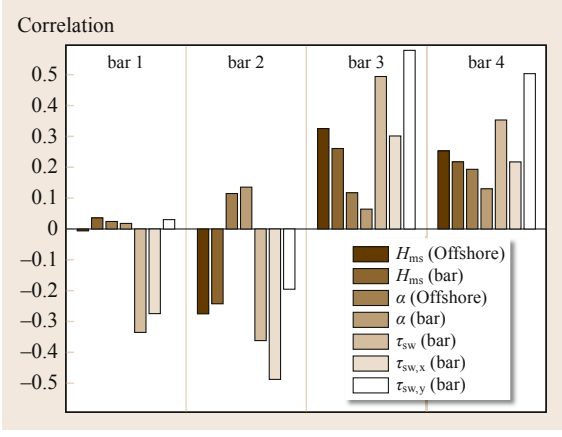


Fig. 28.14 Correlation coefficients between wave parameters and dA_b/dt

correlation with the longshore surface stress $\tau_{sw,y}$ for bars 3 and 4 indicates that longshore forcing may play a significant role in the bar growth in the inner surf zone.

Application of a range of conditions systematically varying H_{rms}, T_p, α, and η revealed that bars only grow for specific wave conditions and that the angle of wave incidence controls whether a specific (H_{rms}, T_p, η) condition results in bar growth or decay (the same wave height and period can result in either bar decay or growth depending on the wave angle). However, the growth rates ($dA_b/dt > 0$) are more frequent and significantly higher for the more shoreward inner bars. This also explains the transient bar response in the hindcast simulation (net decay of bars 1 and 2 and net growth of bars 3 and 4).

Detailed evaluation of model results showed that obliquely incident waves induce longshore wave-driven currents which influence the magnitude of bed shear stresses and subsequently the cross-shore distribution of the sediment transport via enhanced sediment stirring. As waves break, they do not instantaneously dissipate energy, it is first transferred to the surface rollers in which the energy is dissipated: this results in a landward shift of τ_{sw}. However, the local distribution of the cross-shore currents over the bar is particularly sensitive to variations in the local water depth due to the wave-induced undertow which causes the location of the maximum cross-shore currents to coincide with the bar crest location. In contrast, the wave-driven longshore current originates directly from $\tau_{sw,y}$, and consequently has a very similar distribution resulting in a concomitant landward shift of the longshore current [28.54, 55]. Since the longshore current is typically much larger than the cross-shore current for oblique incident waves it dominates the cross-shore distribution of the sediment concentration. If the longshore current is strong enough, the offshore transport peak is shifted landward of the bar crest promoting bar amplitude growth when bars migrate offshore. During onshore bar migration the

presence of longshore current results in enhanced offshore transports in the trough and on the landward bar slope. Thus, for onshore bar migration also, the incident wave angle and the associated longshore currents are very important as sand is eroded from the troughs and deposited on the bar enhancing bar growth and onshore migration.

The implications for the bar response are further illustrated using schematic profiles with bars at various cross-shore positions, thus varying the water depth at the bar crest h_{Xb} (Fig. 28.15). The initial bar amplitude (dA_b/dt) and migration (dX_b/dt) responses were analyzed using a single wave condition ($H_{rms} = 1.7\,m, T_p = 8\,s$) combined with a range of wave angles ($\alpha = 0°–70°$, $\Delta\alpha = 2.5°$). By combining the transitions from growth ($dA_b/dt > 0$) to decay ($dA_b/dt < 0$) (indicated by gray line in Fig. 28.16) with the transition from landward ($dX_b/dt > 0$) to seaward ($dX_b/dt < 0$) (black line Fig. 28.16) it is clear that landward or seaward migration is uncorrelated with amplitude change as all combinations of bar response: landward, seaward migration (denoted as L, S) with bar growth and decay (denoted as: D, G; Fig. 28.16) can occur for the same wave condition. It is noteworthy that in deeper water ($h_{Xb} > 6\,m$) bars decay but do not migrate anymore. Furthermore, onshore migration rates are usually lower than offshore migration rates and are associated with lower growth rates. The largest bar amplitude change (growth and decay) takes place during offshore migration.

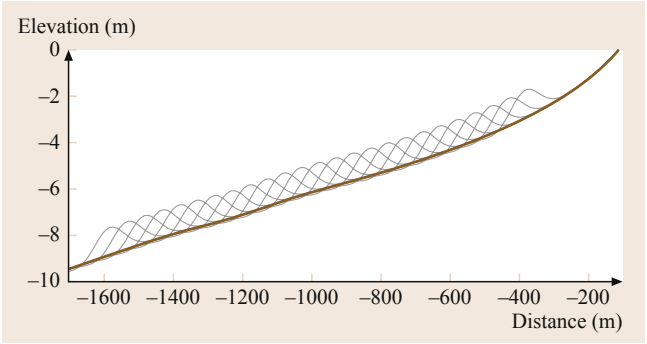


Fig. 28.15 Averaged profile (*black*) and bottom profiles with the schematic bars (*gray*)

Cyclic Bar Behavior Explained

The water depth at the bar crest h_{Xb} and the angle of wave incidence α control the generation of the longshore current. For bars in shallow water, the bar amplitude response and α are strongly related, due to the relatively strong longshore current that waves breaking under an angle generate. The absence of breaking waves and the associated longshore current at bars in deeper water result in the dominance of cross-shore currents on sediment transport. This results in transport peaks that coincide with the bar crest location forcing the bar amplitude decay at larger water depths. The strong dependency on the longshore current also explains the observed transient bar amplitude response during the net interannual offshore migration. For bars in relatively shallow water wave breaking is more frequent, promoting net bar amplitude growth in case of oblique wave incidence, whereas in deeper water wave breaking on the bars is limited, leading to net bar amplitude decay. The latter usually is the onset of a generation of a new bar being generated near the shoreline perpetuating the cycle.

The studies described here indicate that even assuming longshore uniformity, much of the cross-shore behavior of sandbars can be described with relatively simple physics, where undertow, wave asymmetry and skewness, longshore current, and bed slope effects are the dominant mechanisms. Since these processes can be readily implemented in 2DH or 3-D models, there is a reason for optimism that with the tuning provided in this study, such models will be able to describe the full rich behavior in the case of nonlongshore uniform morphologies. However, this still remains to be proven in practice.

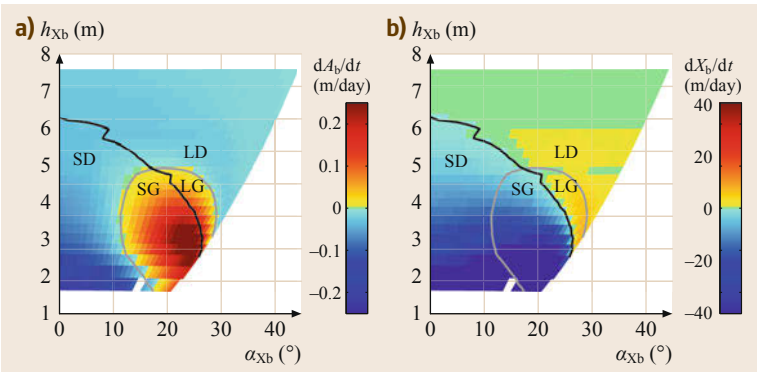


Fig. 28.16a,b Predicted bar amplitude change **(a)** and bar migration **(b)** as a function of the water depth and the wave angle at the bar crest. *Black line* indicates $dX_b/dt = 0$ contour line, *gray line* indicates $dA_b/dt = 0$ contour line. Identified bar response (SD=Seaward, Decay; SG=Seaward, Growth; LD=Landward, Decay; LG=Landward, Growth)

28.3.3 Long-Term Modeling of Estuaries

The morphodynamic modeling of estuaries is an area where much progress has been made in recent years, in physical modeling, analytical, semiempirical, or aggregated methods and in process-based numerical modeling. In the following section, we aim to present an overview of recent advances and case studies of estuarine morphodynamic modeling efforts with process-based models. We distinguish studies with a *fundamental* approach from studies with a more *realistic* approach. The fundamental approach is more schematized and aims to develop understanding on the long-term evolution or equilibrium conditions of morphodynamic systems. The *realistic* approach aims at morphodynamic predictions in a realistic environment on an engineering timescale up to decades.

System Understanding – Schematized Approach

Process-based numerical models can be used to explore equilibrium conditions of the longitudinal estuarine bed profile by tidal forcing under idealized 1-D forcing conditions [28.27, 56–60]. Idealized 2-D models including the evolution of channel-shoal patterns [28.4, 26, 27, 61–64] show that stable channel-shoal patterns develop within decades by starting from a flat sandy bed in a rectangular basin and applying highly schematized tidal forcing. The development of the longitudinal bathymetric profile is subject to a second, much longer timescale and is caused by tide-residual sediment transports resulting from tidal asymetry along the basin. In a similar modeling exercise [28.65, 66] show that morphodynamic evolution leads to less morphodynamic activity due to continuously smaller spatial shear stress gradients. The model results compare well the empirical relationship between the cross-sectional area (A) and tidal prism (P) [28.67] (Fig. 28.17).

The importance of a flow-confining estuarine geometry becomes clear once a realistic plan form is applied within the schematized modeling approach. 2-D process-based models are able to reproduce realistic morphological patterns observed in real estuaries [28.28, 29, 68, 69]. Van der Wegen et al. [28.70] show that modeled channel-shoal patterns significantly resemble the measured bathymetry by just imposing the Western Scheldt geometry and major tidal forcing. A systematic sensitivity analysis shows that including external overtides, varying grain size, different sediment transport formulations, and dredging and dumping activities only have a secondary effect on the evolving channel-shoal patterns. This suggests that the estuarine plan form is a governing factor in the estuaries' morphology and morphodynamic development.

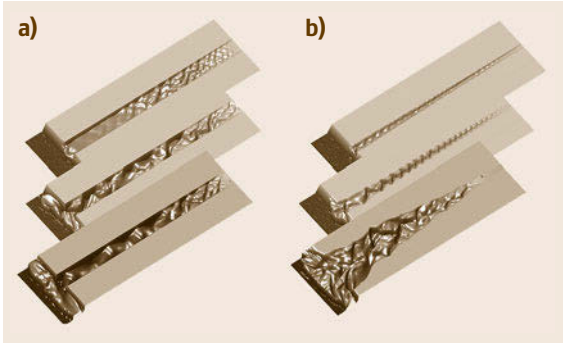


Fig. 28.17a,b Impression of bathymetry after 25, 400, and 3200 years **(a)** with fixed banks (2.5 km wide and 80 km long) and **(b)** with erodible banks (initially 0.5 km wide and 80 km long). Scales are distorted (after [28.67])

Given that a highly schematized model is able to reproduce a realistic bathymetry, long-term varying forcing conditions may be adapted. An example is the study by the authors of [28.71, 72] who explore the impact of the sea-level rise on the morphology previously generated by the same model under constant forcing conditions. They show that intertidal area slowly disappears when realistic sea-level rise scenarios are imposed on a stable channel-shoal pattern earlier generated by the model itself. The advantage of this methodology is that model parameter settings are kept constant. In such a case, the model results will not suffer from morphological adaptations by poor model parameter settings or limited process description, which would probably occur when the sea-level rise is imposed on a measured bathymetry in a more realistic setting.

Realistic Conditions – Dealing with Complexity and Uncertainty

In contrast to highly schematized model settings, more *realistic* models take an existing and often complex environment as a starting point. Policy makers, decision makers and engineers have an interest in short-term high-resolution predictions, since there is an urge to anticipate on likely morphodynamic evolution to mitigate unwanted developments, or else, take action. The morphodynamic effect of man-made works in an estuary may be relatively large within a relatively short (decadal) period compared to system adaptations due to slowly varying forcing conditions such as sea-level rise. As an example, *Apostos* et al. [28.73] and *Van der Wegen* and *Jaffe* [28.74] describe process-based modeling efforts to reproduce morphodynamic developments in San Pablo Bay, a subembayment of San Francisco Estuary. A unique data set describes a 150-year period measured on a 30-year interval. It covers

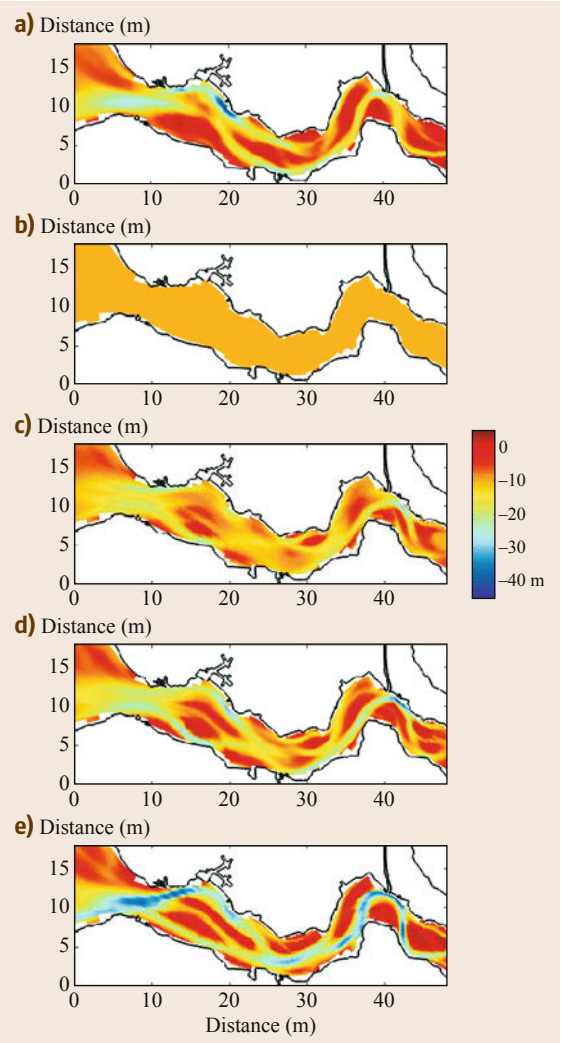


Fig. 28.18 **(a)** Measured 1998 bathymetry; bed development as the result of a 3-D hydrodynamic model with the Engelund–Hansen sediment transport formulation, **(b)** starting from an initially flat, **(c)** the bathymetry after 15 years, **(d)** the bathymetry after 30 years, **(e)** the bathymetry after 200 years (after [28.70], courtesy of Elsevier)

periods of net deposition and erosion as the result of the excessive sediment supply by upstream hydraulic mining (between 1849 and 1884) and the subsequent period of decaying sediment supply also due to the construction of reservoirs in the catchment. The environment required 3-D modeling of salt and fresh water interactions, sand, and mud sediment fractions, wind waves and varying boundary conditions by river flow leading to a highly complex model setup. Despite the large amount of processes and forcing conditions, Fig. 28.19 shows that the model results resemble measured morphodynamic development remarkably well. Furthermore, varying the model input parameters (i. e., uncertain sediment characteristics and forcing condition) within reasonable limits showed only limited and quantitative impact on the model results. The San Pablo Bay plan form and bathymetry seem to determine the evolving erosion and deposition patterns to a high degree.

Process-based models have evolved into powerful models that approach the complexity of reality. The models generate detailed and high-resolution output that allows for a close analysis of processes and forcing mechanisms. However, two critical remarks are made:

- The large number of processes included also requires high-quality model input, which cannot always be guaranteed by measurements. Collecting detailed knowledge on bed composition and SSC (suspended sediment concentration) boundary conditions, for example, is costly, especially when the whole model domain is involved. Other model input such as the diffusion coefficient (varying in space and time) is almost impossible to measure. *Van der Wegen* and *Roelvink* [28.70] present a promising message by showing that model results do not change fundamentally when sediment characteristics are varied within a reasonable range.
- Uncertain model input leads to uncertain model outcomes. Process-based models may describe reality well enough to justify a certain level of forcing schematization or best-guess model input. Still, model results should be presented in terms of probable developments and not as the single outcome of an optimized model. *Pinto* et al. [28.75] present examples on the uncertainty associated with different sediment transport formulations. *Fortunato* et al. [28.76] and *Van der Wegen* and *Jaffe* [28.74] are among the first to explore the possibilities of evaluating ensemble-averaged morphodynamic predictions in an estuarine environment.
- Starting from a measured bathymetry, model result may initially show developments due to limited model definition rather than reproducing realistic developments. Only after decades the bathymetry may have adjusted to the model parameter settings enough to lead to reliable forecasts. It is a major task to discriminate between actual developments and the effect of a period of such a *morphodynamic spin-up*. *Van der Wegen* et al. [28.77] describe a methodology to limit *morphodynamic spin-up*. They let the model itself generate a distribution of sediment classes over the model domain before

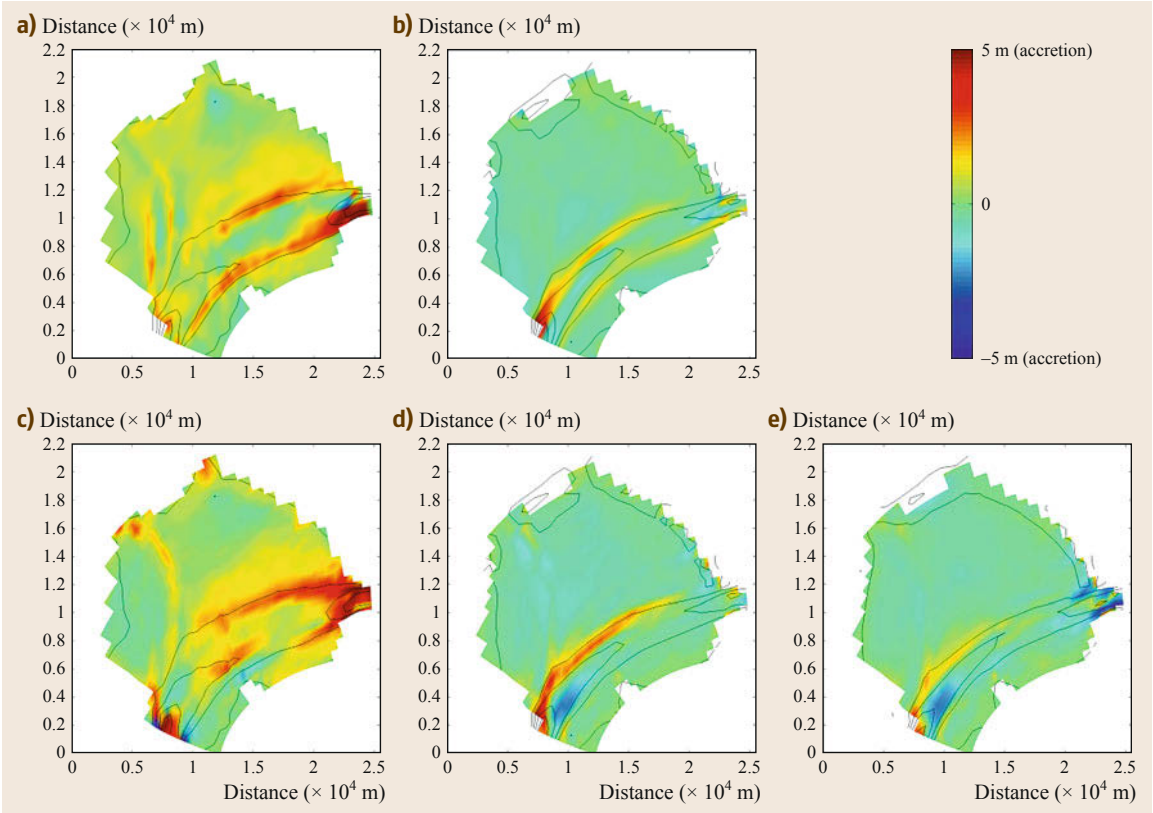


Fig. 28.19a–e Erosion and sedimentation patterns (in meter) measured **(a,b)** and modeled **(c,d,e)** for the 1856–1887 net depositional period **(a,c)** the 1951–1983 net erosional period **(b,d)** and the 1983–2013 forecast period **(e)**. The difference in modeled patterns between the depositional and erosional periods is mainly from a decreasing sediment supply at the landward boundary (after [28.74], courtesy of Elsevier)

allowing for bed level updates. In this way, they prevent an unrealistic peak in morphodynamic development as the result of an ill-defined initial distribution of sediment classes over the model domain.

From the preceding section, we may conclude that process-based morphodynamic modeling has grown up in the sense that stable, long-term simulations can be carried out that tend toward sensible near-equilibrium states; that the models are robust enough to create morphologies out of flat-bed initial conditions and that external forcing and geological or man-made constraints define the morphology to a large extent.

The classical assumption that morphodynamic models get derailed after a certain time and are useless in the long term has been proven wrong in many cases. On the contrary, recent studies have shown that the estuaries' outline determines the morphodynamics to a high degree and that morphodynamic modeling results are sensitive to model input parameter settings only to a limited extent.

Although it is formally difficult to prove and scientific discussion continues, experience suggests that process-based models will perform better once:

- The estuarine geometry confines morphodynamic development (the geometry is of similar length scale as the morphological features of interest [28.70]).
- A perturbation dominates other secondary processes (the scour hole as the result of breakwater construction, a large supply of sediments [28.78, 79] or the presence of major tidal forcing compared to river flow [28.70]).
- Longer timescales (> decades) are considered [28.78, 80, 81]; however, this requires further research; *Fortunato* et al. [28.76] suggest that uncertainty increases with time for a wave-dominated system, whereas *Ruessink* and *Kuriyama* [28.82] conclude that in a wave-dominated environment storms dominate the predictability.

Typically, estuarine morphodynamic development is slower than beach morphodynamics described in

the previous sections. The major reasons may be that wave forcing is relatively limited due to short fetch length and a shallow and wind-sheltered environment. Also, estuarine channel-shoal patterns develop by tide-residual sediment transports that are about an order of magnitude smaller than the maximum ebb, flood, or river-induced sediment transports. As a result, data sets for calibration and validation of these models should cover decades and (preferably) longer time spans. This makes case studies rare and often limited to estuaries with considerable human impact such as environments of ports.

28.4 Future Directions

28.4.1 Ensemble Modeling

Morphodynamic forecasts are never the result of a single run. Modeling practice shows the need for considerable testing of model parameter settings and model behavior until the results can be presented with confidence. The testing is sometimes caused by model shortcomings or bugs, but another important reason is that model input parameters are unknown beforehand.

When we take process-based modeling as an example, uncertainty exists with respect to:

1. Process definitions in the model itself: Are sediment transport formulae described adequately? Is there a need for including more processes? Can we apply a morphological factor?
2. Input parameter values: What should be the values for the viscosity/diffusivity coefficient, and sediment characteristics such as the critical bed shear stress or fall velocity? How many sediment classes should be defined?
3. Initial conditions: What is the initial sediment composition in the bed? What should be the initial SSC or salinity field in the water column?
4. Forcing conditions: Is it possible to schematize the wind climate? Can we apply a morphological tide? What is the impact of extreme events? What are expected rates in the sea-level rise?
5. The modeled system itself: Does the modeled system itself reflect (deterministic) chaos or is it insensitive to slightly different model input?

Until now, the literature shows only limited attention for a systematic analysis on morphodynamic model uncertainties. *Baart* et al. [28.83] point to confidence intervals around morphological forecasts as the result of the propagation of uncertainty through a chain of models from weather (wind and pressure) to hydrodynamic (waves and tides) and finally morphodynamic predictions. *Vreugdenhil* [28.84] and *Pinto* et al. [28.75] present examples on the uncertainty associated with different sediment transport formulations, whereas *Fortunato* et al. [28.76] and *Ruessink* and *Kuriyama* [28.82] explore the development of uncertainty over time. Further, *Ruessink* and *Kuriyama* [28.82] and *Van der Wegen* et al. [28.79] show that ensemble runs in morphodynamic predictions can be used to discriminate dominant forcing from secondary processes.

There lies a challenge in future research to further explore the conditions in which morphodynamic models function well, what the quality of model input parameters should be to come to trustworthy forecasts and what the quality of model output is in the light of uncertain model input parameters. Ensemble runs that include a systematic variation of (unknown) input parameters as well as model output presented in terms of a probability range may be a promising way to explore the value of models under uncertain conditions.

28.4.2 Biogeomorphology

In the last two decades, recognition has grown on the importance of the interaction between geomorphological and biological processes, and a new field of biogeomorphology has sprung up. In the coastal context, important areas where such interactions are dominant are mudflats, salt marshes, mangroves, coral reefs, and floodplains. Erosion and sedimentation of the estuarine bed may be subject to biological processes. Bed developments are influenced by benthic flora and fauna and can be a function of biological processes like the development of bio-films or the presence of bio-engineers [28.85–87]. Salt marshes eventually may be the most dramatic example where vegetation cover governs flow patterns and sedimentation trapping rates. *D'Alpaos* et al. [28.88] and *Temmerman* et al. [28.89] present process-based modeling efforts including the feedback process of vegetation on morphodynamics. In this way, salt marsh dynamics and its feedback on channel shoal morphodynamics can be included in the modeling effort.

Bed stabilization may occur due to bio-films or vegetation roots, whereas surface deposit feeders destabilize the upper bed layer during feeding activities. All these processes may be a function of time so that they change considerably over months, seasons, or even over decades.

Models such as Delft3D and SWAN nowadays have many algorithms describing effects of vegetation on wave dissipation, flow resistance, and trapping of sediments by vegetation and recently [28.90], generic vegetation succession formulations were implemented, which provides a toolbox to explore these complex interactions.

28.4.3 Dune Modeling

Aeolian transport models can be a valuable addition to hydrodynamic models to describe the long-term interaction between land and water. In a recent paper [28.91] present the application of an aeolian transport model [28.92, 93] as a tool for the evolution of coastal nourishments and dunes. The model was originally developed to describe important features and dynamics of typical desert dunes and was modified to apply to sandy beaches and coastal nourishments. This is expected to be the start of an important line of research as with the increase of large-scale nourishments natural dune-growth processes are being restored and need to be considered when trying to predict the long-term evolution of such nourishments.

28.4.4 Coral Reefs

A subject receiving increasing attention is the role coral reefs play in protecting and building up coastal environments. Coral reefs are different from other hard-protection structures in that they are incredibly rough and thereby dissipate wave energy very effectively; they also generate sediment and thereby, in a healthy situation, can keep up with sea-level rise. Modeling of the hydrodynamics of coral reef systems has got increasing attention in the last decade; a thorough review and new insights into the importance and modeling of infragravity waves on such systems are provided in [28.94]. The important problem on how island and low-lying coastal communities protected by reefs can adapt to climate change and sea-level rise will need much attention in years to come; see for instance [28.95].

28.4.5 Modeling Stratigraphy

Morphodynamic models are able to reproduce realistic morphological patterns. By including different classes of sediment and adding a model for bed composition administration, they are also able to reproduce stratification in the bed. Depositional layers of fine sediments can alternate with coarse sediment layers depending on hydrodynamic conditions in the water column and variations in sediment supply. Long-term ($\approx$ centuries to millennia) simulations may thus reproduce stratification characteristics, for instance in a reservoir or river delta [28.96, 97].

This information is of particular interest to geologists. Morphodynamic models offer a tool to explain what processes could have caused certain observed stratification features. Stratification models could also be used the other way round. Once it is known what prevailing hydrodynamic conditions in a certain domain were, the stratification model may reproduce stratification characteristics and permeability parameters. These data are of particular interest to investigating groundwater flows or to determining in a better way the placing of wells or bore holes for the oil industry.

28.4.6 Morphological Modeling of Tsunami Deposits

There is a wide body of research related to the analysis of historical tsunami deposits in order to obtain information on the frequency and strength of tsunamis. In [28.98], a simple model is proposed to link the tsunami flow speed to tsunami deposits; *Goto* and *Imamura* [28.99] consider the requirements of numerical models capable of modeling transport of sediment and large boulders during tsunamis; in [28.100], detailed morphodynamic modeling of the 2004 Indian Ocean Tsunami is carried out and the damping effect of vegetation is considered, using the Delft3D model; the same system is used in an extensive 2DV analysis of flow patterns, turbulence, sediment concentration, and erosion/deposition for different cross-shore profiles [28.73]. In an extensive hindcast of the tsunami at Bandar Aceh, *Li* et al. [28.101] apply a two-way coupled system of nested models to simulate both the generation of the tsunami waves and the propagation toward the area of interest; for the finest grid where morphology changes are important they apply the XBeach model incorporated into their COMCOT–SED system. They found that the model presented in the paper was capable of simulating extreme tsunami events (tsunami wave height in the order of 30 m) over a large domain, that the thickness of tsunami deposits could be quantitatively simulated and that therefore forward models of tsunami sediment transport can be a promising tool to help tsunami geologists understand tsunami deposits.

28.5 Nomenclature

α	wave angle [rad]
b	power in transport relation [–]
c_{bed}	bed celerity [m/s]
$\mathrm{d}A_{\mathrm{b}}/\mathrm{d}t$	initial bar amplitude [m/s]
$\mathrm{d}Xb/\mathrm{d}t$	bar migration [m/s]
ϵ	porosity of seabed [–]
h	depth [m]
$h_{\mathrm{X}b}$	water depth at bar crest [m]
q_x, q_y	horizontal components of sediment transport [$\mathrm{m}^3/\mathrm{m/s}$]
S	total longshore transport [m^3/s]
s	longshore transport coefficient [$\mathrm{m}^3/\mathrm{s/rad}$]
z_{b}	vertical position of seabottom [m]

References

28.1 J. van de Graaff: Dune erosion during a storm surge, Coast. Eng. **1**(0), 99–134 (1977)

28.2 R. Pelnard-Considere: *Essai de theorie de l'Evolution des Formes de Rivages en Plage de Sable et de Galets*, Les Energies de la Mer: Compte Rendu Des Quatriemes Journees de L'hydraulique, Paris Question III, Rapport No. 1, 1956) pp. 289–298, french

28.3 A. Ashton, A.B. Murray, O. Arnoult: Formation of coastline features by large-scale instabilities induced by high-angle waves, Nature **414**(6861), 296–300 (2001)

28.4 H.M. Schuttelaars, H.E. de Swart: Initial formation of channels and shoals in a short tidal embayment, J. Fluid Mech. **386**, 15–42 (1999)

28.5 S.J.M. Hulscher, G.M. van den Brink: Comparison between predicted and observed sand waves and sand banks in the North Sea, J. Geophys. Res. **106**(C5), 9327–9338 (2001)

28.6 P. Vellinga: Beach and dune erosion during storm surges, Coast. Eng. **6**(4), 361–387 (1982)

28.7 M. Larson, N.C. Kraus, M.R. Bymes: *SBEACH: Numerical Model for Simulating Storm-Induced Beach Change, Report 2: Numerical Formulation and Model Tests*, Tech. Rep. CERC-89-9 (US Army Engineer Waterways Experiment Station, Coastal Engineering Research Center, Vicksburg 1990)

28.8 W.D. Eysink, E.J. Biegel: *ISOS*2 Project: Impact of Sea Level Rise on the Morphology of the Wadden Sea in the Scope of its Ecological Function, Phase 2* (Delft Hydraulics, Delft 1992)

28.9 M.J.F. Stive, Z.B. Wang: Morphodynamic modeling of tidal basins and coastal inlets. In: *Advances in Coastal Modeling*, ed. by V.C. Lakhan (Elsevier, Amsterdam 2003) pp. 367–392

28.10 D. Roelvink, A. Reniers: *A Guide to Modeling Coastal Morphology*, Advances in Coastal and Ocean Engineering, Vol. 12 (World Scientific, Singapore 2011) p. 292

28.11 M. Szmytkiewicz, J. Biegowski, L.M. Kaczmarek, T. Okrój, R. Ostrowski, Z. Pruszak, G. Różyński, M. Skaja: Coastline changes nearby harbour structures: Comparative analysis of one-line models versus field data, Coast. Eng. **40**(2), 119–139 (2000)

28.12 R.G. Dean: Beach nourishment: Design principles, Proc. 23rd Int. Conf. Coast. Eng. (1992) pp. 301–349

28.13 M.C. Buijsman, P. Ruggiero, G.M. Kaminsky: Sensitivity of shoreline change predictions to wave climate variability along the southwest Washington coast, USA, Proc. Conf. Coast. Dyn. (2001) pp. 617–626

28.14 A. Ashton, B. Murray: High-angle wave instability and emergent shoreline shapes: 1. Modeling of sand waves, flying spits, and capes, J. Geophys. Res. **111**, F04011 (2006)

28.15 J.A. Roelvink, I. Broker: Coastal profile models, Coast. Eng. **21**, 163–191 (1993)

28.16 J.A. Roelvink, T.J.G.P. Meijer, K. Houwman, R. Bakker, R. Spanhoff: Field validation and application of a coastal profile model, Proc. Coast. Dyn. (1995) pp. 818–828

28.17 B.G. Ruessink, Y. Kuriyama, A.J.H.M. Reniers, J.A. Roelvink, D.J.R. Walstra: Modeling cross-shore sandbar behavior on the timescale of weeks, J. Geophys. Res. **112**(F3), F03010 (2007)

28.18 N. Booij, R.C. Ris, L.H. Holthuijsen: A third-generation wave model for coastal regions – 1. Model description and validation, J. Geophys. Res. **104**(C4), 7649–7666 (1999)

28.19 R.C. Ris, L.H. Holthuijsen, N. Booij: A third-generation wave model for coastal regions – 2. Verification, J. Geophys. Res. **104**(C4), 7667–7681 (1999)

28.20 G.R. Lesser, J.A. Roelvink, J.A.T.M. van Kester, G.S. Stelling: Development and validation of a three-dimensional morphological model, Coast. Eng. **51**(8–9), 883–915 (2004)

28.21 R. Ranasinghe, C. Pattiaratchi, G. Masselink: A morphodynamic model to simulate the seasonal closure of tidal inlets, Coast. Eng. **37**(1), 1–36 (1999)

28.22 A.J.H.M. Reniers, J.A. Roelvink, E.B. Thornton: Morphodynamic modeling of an embayed beach under wave group forcing, J. Geophys. Res. **109**(C1), 10.1029/2002JC001586 (2004)

28.23 M.W.J. Smit, A.J.H.M. Reniers, B.G. Ruessink, J.A. Roelvink: The morphological response of a nearshore double sandbar system to constant wave forcing, Coast. Eng. **55**(10), 761–770 (2008)

28.24 R. Ranasinghe, G. Symonds, K. Black, R. Holman: Morphodynamics of intermediate beaches: A video imaging and numerical modelling study, Coast. Eng. **51**, 629–655 (2004)

28.25 Z.B. Wang, C.T. Louters, H.J. De Vriend: Morphodynamic modelling for a tidal inlet in the Wadden Sea, Mar. Geol. **126**, 289–300 (1995)

28.26 A. Hibma, H.J. de Vriend, M.J.F. Stive: Numerical modelling of shoal pattern formation in well-mixed elongated estuaries, Estuar. Coast. Shelf Sci. **57**(5/6), 981–991 (2003)

28.27 M. van der Wegen, J.A. Roelvink: Long-term morphodynamic evolution of a tidal embayment using a two-dimensional, process-based model, J. Geophys. Res. **113**(C3), C03016 (2008)

28.28 A. Dastgheib, J.A. Roelvink, Z.B. Wang: Long-term process-based morphological modeling of the Marsdiep Tidal Basin, Mar. Geol. **256**(1–4), 90–100 (2008)

28.29 D.M.P.K. Dissanayake, J.A. Roelvink, M. van der Wegen: Modelled channel patterns in a schematized tidal inlet, Coast. Eng. **56**(11–12), 1069–1083 (2009)

28.30 H.J. de Vriend, M. Capobianco, T. Chesher, H.E. de Swart, B. Latteux, M.J.F. Stive: Approaches to long-term modelling of coastal morphology: A review, Coast. Eng. **21**(1–3), 225–269 (1993)

28.31 B. Latteux: Techniques for long-term morphological simulation under tidal action, Mar. Geol. **126**(1–4), 129–141 (1995)

28.32 J.A. Roelvink: Coastal morphodynamic evolution techniques, Coast. Eng. **53**(2–3), 277–287 (2006)

28.33 R. Ranasinghe, C. Swinkels, A. Luijendijk, D. Roelvink, J. Bosboom, M. Stive, D.J.R. Walstra: Morphodynamic upscaling with the MORFAC approach: Dependencies and sensitivities, Coast. Eng. **58**(8), 806–811 (2011)

28.34 A. Dastgheib: Long-Term Process-Based Morphological Modelling of Large Tidal Basins, Ph.D. Thesis (UNESCO-IHE/Delft University of technology, Delft 2012)

28.35 D.J.R. Walstra, R. Hoekstra, P.K. Tonnon, B.G. Ruessink: Input reduction for long-term morphodynamic simulations in wave-dominated coastal settings, Coast. Eng. **77**, 57–70 (2013)

28.36 G.R. Lesser: An Approach to Medium-Term Coastal Morphological Modelling, Ph.D. Thesis (Delft Univ. of Technology, Delft 2009)

28.37 A. Sallenger: Storm impact scale for barrier islands, J. Coast. Res. **16**(3), 890–895 (2000)

28.38 P. Vellinga: Beach and Dune Erosion During Storm Surges, Ph.D. Thesis (Delft University of Technology, Delft 1986)

28.39 H.J. Steetzel: Cross-shore transport during storm surges, Proc. 22nd Int. Conf. Coastal Eng. (1990) pp. 1922–1934

28.40 D. Roelvink, A. Reniers, A. van Dongeren, J. van Thiel de Vries, R. McCall, J. Lescinski: Modelling storm impacts on beaches, dunes and barrier islands, Coast. Eng. **56**(11-12), 1133–1152 (2009)

28.41 A.S. Arcilla, J.A. Roelvink, B.A. O'Connor, A. Reniers, J.A. Jimenez: The delta flume '93 experiment, Proc. Coast. Dyn. (1994) (1994), pp. 488-502–1061

28.42 R.T. McCall, J.S.M.V. de Vries, N.G. Plant, A.R. Van Dongeren, J.A. Roelvink, D.M. Thompson, A.J.H.M. Reniers: Two-dimensional time dependent hurricane overwash and erosion modeling at Santa Rosa Island, Coast. Eng. **57**(7), 668–683 (2010)

28.43 D. Roelvink, G. Stelling, B. Hoonhout, J. Risandi, W. Jacobs, D. Merli: Development and field validation of a 2dh curvilinear storm impact model, 33rd Int. Conf. Coast. Eng. (ICCE) (2012)

28.44 J. van Thiel de Vries, A. van Dongeren: *Validation of Dune Impact Models Using European Field Data* (Deltares, Delft 2011)

28.45 L.D. Wright, A.D. Short: Morphodynamic variability of surf zones and beaches: A synthesis, Mar. Geol. **56**(1-4), 93–118 (1984)

28.46 T.C. Lippmann, R.A. Holman: Quantification of sand bar morphology: A video technique based on wave dispersion, J. Geophys. Res. **94**, 995–1011 (1989)

28.47 D.J.R. Walstra, A.J.H.M. Reniers, R. Ranasinghe, J.A. Roelvink, B.G. Ruessink: On bar growth and decay during interannual net offshore migration, Coast. Eng. **60**, 190–200 (2012)

28.48 K.M. Wijnberg, J.H.J. Terwindt: Extracting decadal morphological behaviour from high-resolution, long-term bathymetric surveys along the Holland coast using eigenfunction analysis, Mar. Geol. **126**(1–4), 301–330 (1995)

28.49 R.D. Shand, D.G. Bailey, M.J. Shepherd: An inter-site comparison of net offshore bar migration characteristics and environmental conditions, J. Coast. Res. **15**(3), 750–765 (1999)

28.50 Y. Kuriyama: Medium-term bar behavior and associated sediment transport at Hasaki, Japan, J. Geophys. Res. **107**(9), 15-1–15-12 (2002)

28.51 B.G. Ruessink, A. Kroon: The behaviour of a multiple bar system in the nearshore zone of Terschelling, the Netherlands: 1965–1993, Mar. Geol. **121**(3-4), 187–197 (1994)

28.52 B.G. Ruessink, K.M. Wijnberg, R.A. Holman, Y. Kuriyama, I.M.J. van Enckevort: Intersite comparison of interannual nearshore bar behavior, J. Geophys. Res. **108**(8), 5-1–5-12 (2003)

28.53 D.J. Walstra, B.G. Ruessink: Process-based modeling of cyclic bar behavior on yearly scales, Proc. Coast. Dyn. (2009)

28.54 A.J.H.M. Reniers, J.A. Battjes: A laboratory study of longshore currents over barred and non-barred beaches, Coast. Eng. **30**(1-2), 1–21 (1997)

28.55 B.G. Ruessink, J. Miles, F. Feddersen, R. Guza, S. Elgar: Modeling the alongshore current on barred beaches, J. Geophys. Res. C: Oceans **106**(C10), 22451–22463 (2001)

28.56 A.R. van Dongeren, H.J. de Vriend: A model of morphological behaviour of tidal basins, Coast. Eng. **22**(3–4), 287–310 (1994)

28.57 H.M. Schuttelaars, H.E. de Swart: An idealized long term morphodynamic model of a tidal embayment, Eur. J. Mech. B/Fluids **15**(1), 55–80 (1996)

28.58 H.M. Schuttelaars, H.E. de Swart: Multiple morphodynamic equilibria in tidal embayments, J. Geophys. Res. **105**(C10), 24105–24118 (2000)

28.59 S. Lanzoni, G. Seminara: Long-term evolution and morphodynamic equilibrium of tidal channels, J. Geophys. Res. **107**(C1), 3001 (2002)

28.60 I. Todeschini, M. Toffolon, M. Tubino: Long-term morphological evolution of funnel-shape tide-dominated estuaries, J. Geophys. Res. **113**(C5), C05005 (2008)

28.61 G. Seminara, M. Tubino: Sand bars in tidal channels. Part 1. Free bars, J. Fluid Mech. **440**, 49–74 (2001)

28.62 G.P. Schramkowski, H.M. Schuttelaars, H.E. de Swart: The effect of geometry and bottom friction on local bed forms in a tidal embayment, Cont. Shelf Res. **22**(11–13), 1821–1833 (2002)

28.63 G.P. Schramkowski, H.M. Schuttelaars, H.E. de Swart: Non-linear channel-shoal dynamics in long tidal embayments, Ocean Dyn. **54**(3/4), 399–407 (2004)

28.64 M. ter Brake, H.M. Schuttelaars: Channel and shoal development in a short tidal embayment: An idealized model study, J. Fluid Mech. **677**, 503–529 (2011)

28.65 M. Van der Wegen, Z.B. Wang, H.H.G. Savenije, J.A. Roelvink: Long-term morphodynamic evolution and energy dissipation in a coastal plain, tidal embayment, J. Geophys. Res. **113**, F03001 (2008)

28.66 M. van der Wegen, Z.B. Wang, I.H. Townend, H.H.G. Savenije, J.A. Roelvink: Long-term, morphodynamic modeling of equilibrium in an alluvial tidal basin using a process-based approach, River Coast. Estuar. Morphodynamics, RCEM 2009 (2010)

28.67 M. van der Wegen, A. Dastgheib, J.A. Roelvink: Morphodynamic modeling of tidal channel evolution in comparison to empirical PA relationship, Coast. Eng. **57**(9), 827–837 (2010)

28.68 F. Cayocca: Long-term morphological modeling of a tidal inlet: the Arcachon Basin, France, Coast. Eng. **42**(2), 115–142 (2001)

28.69 R. Marciano, Z.B. Wang, A. Hibma, H.J. de Vriend, A. Defina: Modeling of channel patterns in short tidal basins, J. Geophys. Res. **110**(F1), F01001 (2005)

28.70 M. van der Wegen, J.A. Roelvink: Reproduction of estuarine bathymetry by means of a process-based model: Western Scheldt case study, the Netherlands, Geomorphology **179**, 152–167 (2012)

28.71 D.M.P.K. Dissanayake, R. Ranasinghe, J.A. Roelvink: The morphological response of large tidal inlet/basin systems to relative sea level rise, Clim. Change **113**(2), 253–276 (2012)

28.72 M. van der Wegen: Numerical modeling of the impact of sea level rise on tidal basin morphodynamics, J. Geophys. Res. Earth Surf. **118**(2), 447–460 (2013)

28.73 A.G. Apostos, G. Gelfenbaum, B. Jaffe: Process-based modeling of tsunami inundation and sediment transport, J. Geophys. Res. **116**(F1), 10.1029/2010JF001797 (2011)

28.74 M. Van der Wegen, B.E. Jaffe: Towards a probabilistic assessment of process-based, morphodynamic models, Coast. Eng. **75**, 52–63 (2013)

28.75 L. Pinto, A.B. Fortunato, P. Freire: Sensitivity analysis of non-cohesive sediment transport formulae, Cont. Shelf Res. **26**(15), 1826–1839 (2006)

28.76 A.B. Fortunato, X. Bertin, A. Oliveira: Space and time variability of uncertainty in morphodynamic simulations, Coast. Eng. **56**(8), 886–894 (2009)

28.77 M. van der Wegen, A. Dastgheib, B. Jaffe, D. Roelvink: Bed composition generation for morphodynamic modeling: Case study of San Pablo Bay in California, USA, Ocean Dyn. **61**(2–3), 173–186 (2011)

28.78 N.K. Ganju, D.H. Schoellhamer, B.E. Jaffe: Hindcasting of decadal-timescale estuarine bathymetric change with a tidal-timescale model, J. Geophys. Res. **114**(F4), F04019 (2009)

28.79 M. Van der Wegen, B.E. Jaffe, J.A. Roelvink: Process-based, morphodynamic hindcast of decadal deposition patterns in San Pablo Bay, California, 1856–1887, J. Geophys. Res. **116**(F2), 10.1029/2009JF001614 (2011)

28.80 N. Ganju, D. Schoellhamer: Decadal-timescale estuarine geomorphic change under future scenarios of climate and sediment supply, Estuar. Coasts **33**(1), 15–29 (2010)

28.81 M. Van der Wegen: Modeling Morphodynamic Evolution in Alluvial Estuaries, Ph.D. Thesis (Delft Univ. of Technology, Delft 2010)

28.82 B.G. Ruessink, Y. Kuriyama: Numerical predictability experiments of cross-shore sandbar migration, Geophys. Res. Lett. **35**(1), L01603 (2008)

28.83 F. Baart, P.H.A.J.M. Van Gelder, M. Van Koningsveld: Confidence in real-time forecasting of morphological storm impacts, J. Coast. Res. **64**, 1835–1839 (2011)

28.84 C.B. Vreugdenhil: Appropriate models and uncertainties, Coast. Eng. **53**(2–3), 303–310 (2006)

28.85 A.R.M. Nowell, P.A. Jumars, J.E. Eckman: Effects of biological activity on the entrainment of marine sediments, Mar. Geol. **42**, 133–153 (1981)

28.86 S. Temmerman, T.J. Bouma, J. Van de Koppel, D. Van der Wal, M.B. De Vries, P.M.J. Herman: Vegetation causes channel erosion in a tidal landscape, Geology **35**(7), 631–634 (2007)

28.87 B.W. Borsje, M.B. de Vries, S.J.M.H. Hulscher, G.J. de Boer: Modeling large-scale cohesive sediment transport affected by small-scale biological activity, Estuar. Coast. Shelf Sci. **78**(3), 468–480 (2008)

28.88 A. D'Alpaos, S. Lanzoni, M. Marani, A. Rinaldo: Landscape evolution in tidal embayments: Modeling the interplay of erosion, sedimentation, and vegetation dynamics, J. Geophys. Res. **112**(F1), F01008 (2007)

28.89 S. Temmerman, T.J. Bouma, J. Van de Koppel, D. Van der Wal, M.B. De Vries, P.M.J. Herman: Impact of vegetation on flow routing and sedimentation patterns: Three-dimensional modeling for a tidal marsh, J. Geophys. Res. **110**(F4), F04019 (2005)

28.90 Q. Ye: An Approach Towards Generic Coastal Geomorphological Modelling with Applications, Ph.D. Thesis (UNESCO-IHE and TU Delft, Delft 2012)

28.91 M. Muller, D. Roelvink, S. de Vries, A. Luijendijk, J. van Thiel de Vries: Process-based modeling of coastal dune development, ASCE Int. Conf. Coast. Eng., Santander (2012)
28.92 G. Sauermann, K. Kroy, H.J. Herrmann: Continuum saltation model for sand dunes, Phys. Rev. E **64**(3), 031305 (2001)
28.93 K. Kroy, G. Sauermann, H.J. Herrmann: Minimal model for aeolian sand dunes, Phys. Rev. E **66**(3), 031302 (2002)
28.94 A. Van Dongeren, R. Lowe, A. Pomeroy, D.M. Trang, D. Roelvink, G. Symonds, R. Ranasinghe: Numerical modeling of low-frequency wave dynamics over a fringing coral reef, Coast. Eng. **73**(0), 178–190 (2013)
28.95 C.T. Perry, P.S. Kench, S.G. Smithers, B. Riegl, H. Yamano, M.J. O'Leary: Implications of reef ecosystem change for the stability and maintenance of coral reef islands, Glob. Change Biol. **17**(12), 3679–3696 (2011)
28.96 N. Geleynse, J.E.A. Storms, M.J.F. Stive, H.R.A. Jagers, D.J.R. Walstra: Modeling of a mixed-load fluvio-deltaic system, Geophys. Res. Lett. **37**(5), L05402 (2010)
28.97 N. Geleynse, J.E.A. Storms, D.R. Walstra, H.R.A. Jagers, Z.B. Wang, M.J.F. Stive: Controls on river delta formation; insights from numerical modelling, Earth Planet. Sci. Lett. **302**(1–2), 217–226 (2011)
28.98 B.E. Jaffe, G. Gelfenbuam: A simple model for calculating tsunami flow speed from tsunami deposits, Sediment. Geol. **200**(3–4), 347–361 (2007)
28.99 K. Goto, F. Imamura: Numerical models for sediment transport by tsunamis, Quat. Res. (Daiyonki-Kenkyo) **46**, 463–475 (2007)
28.100 G. Gelfenbaum, D. Vatvani, B. Jaffe, F. Dekker: Tsunami inundation and sediment transport in vicinity of coastal mangrove forest, Coastal Sediments (2007) (2007) pp. 1117–1128
28.101 L. Li, Q. Qiu, Z. Huang: Numerical modeling of the morphological change in Lhok Nga, west Banda Aceh, during the 2004 Indian Ocean tsunami: Understanding tsunami deposits using a forward modeling method, Nat. Hazards **64**(2), 1549–1574 (2012)

29. Beach Nourishment

Robert G. Dean, Thomas J. Campbell

Although beach nourishment is a relatively young technology, some major projects have been in place for 80 years and monitored sufficiently to provide a basis for design and performance prediction with a reasonable degree of confidence. Of the various methods available for shoreline stabilization, beach nourishment is unique in addressing the sand deficit and with the capability to restore the recreational, storm protection, and ecological functions of a natural beach. The quality and quantity of sand placed in beach nourishment are shown to be essential to effective performance. Design and performance prediction methods are reviewed including the analytical method of Pelnard Considère, which provides interrelationships between various design parameters and is an excellent pedagogical tool for both understanding these relationships and guiding the more detailed methods which are evolving toward greater and greater complexity. Legacy projects are defined as those that were nourished many years ago, require very little maintenance, have been documented to the degree that their performance is well established and are generally not recognized as constructed beaches by the typical beach user. Two of these legacy projects are reviewed in some detail as well as several other projects for which the performance is well established. The challenges facing beach nourishment in the future include: rising energy costs, environmental issues, public perception and limited sand resources.

29.1 **Advantages of Beach Nourishment Over Other Approaches** 636
29.2 **Methods of Delivery of Sand for Beach Nourishment** 636
29.2.1 Characteristics and Considerations in Design of Beach Nourishment Projects 637
29.2.2 Significance of Good Quality Sand 638
29.2.3 Various Settings for Beach Nourishment 638
29.3 **Role of Structures in Beach Nourishment** 639
29.3.1 Terminal Structures 639
29.3.2 Groins 640
29.3.3 Detached Breakwaters 640
29.4 **Design and Prediction Approaches and Methods** 640
29.4.1 Pelnard Considère Methodology 641
29.4.2 One-Line Numerical Models 643
29.4.3 N-Line and more Detailed Models 643
29.5 **Additional Design Considerations** 643
29.5.1 Erosional Hot Spots 643
29.5.2 Nearshore Placement 644
29.5.3 Ad-Hoc Transformation for Modeling 645
29.6 **Legacy Beach Nourishment Projects** 645
29.6.1 Examples and Discussions of Legacy Projects 645
29.7 **Other Beach Nourishment Projects** 648
29.7.1 Seabright to Manasquan Inlet, NJ 648
29.7.2 Captiva Island, FL 648
29.7.3 Treasure Island, FL 649
29.8 **Summary and Conclusions** 649
29.9 **Nomenclature** 650
References 650

Beach nourishment, the placement of large volumes of good quality sand along the shoreline to advance it seaward, is a comparatively young technology. Apart from some early *projects of opportunity* with sand available from construction projects, the first large-scale designed nourishment projects in the United States commenced in the mid-1970s. The earliest projects were placed with comparatively little design background, regulations, or concerns for environmental effects. Yet, some of these projects continue to function

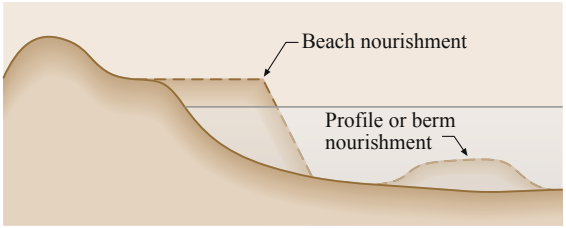


Fig. 29.1 Beach and profile nourishments

today with relatively little maintenance and generally without recognition of their nourishment history by the beach-user community. Two of these so-called *legacy* projects will be described in some detail later in this chapter.

Beach nourishment may be placed directly on the beach or as an underwater berm with the intention that this subaqueous feature will move ashore and widen the dry beach. An alternate design objective is that this feature will damp waves and stabilize the beach (Fig. 29.1).

Alternatives to beach nourishment generally involve hard structures including groins, seawalls, and detached breakwaters. Planting and stabilization of dune vegetation is effective against wind-induced erosion and in trapping wind-blown sand, but is less so against water-induced erosion. Although structural approaches can be effective in shoreline stabilization, and can accumulate sand, this is usually at the expense of the adjacent beach. In some cases, structures can be used in combination with beach nourishment projects, thereby improving project longevity. A number of lower cost so-called *innovative methods* have been proposed and some have been installed and monitored on an experimental basis; however, in our experience, none has been proven successful. These include artificial seaweed, beach drains, nets placed perpendicular to the shoreline, etc. These will not be discussed further in this chapter.

For information beyond that presented in this limited length chapter, the reader is referred to the *US Army Corps of Engineers Coastal Engineering Manual* and various journals and proceedings including those of the *International Conference on Coastal Engineering* series.

29.1 Advantages of Beach Nourishment Over Other Approaches

Nourishment is an attractive alternative for shoreline stabilization for a number of reasons. With the placement of good quality sand, after a relatively short period of adjustment, the beach appears and functions as the natural beach including the restoration of the natural recreational and environmental functions. Finally, the wider beach serves as an effective buffer against storm surge and waves and usually results in a substantial increase in upland property values.

Beach nourishment is usually conducted to restore an eroded beach to an earlier state. The causes of an eroding beach can be either natural or artificial or a combination. The most prevalent and persistent natural cause is rising sea level, which according to the *Bruun* Rule [29.1] causes the shoreline to retreat at a rate which is 50–100 times the rate of sea level rise. With the average 20th century rate of global sea level rise of 1.7 mm/year, this would translate into a beach retreat of 8.5 cm to 17 cm/year, not a large amount that certainly could be offset economically by beach nourishment in many areas. Often, areas in need of nourishment have experienced much larger erosion rates, frequently due to human activity affecting the nearshore system, in particular interference with the longshore sediment transport. Examples are provided later.

The general characteristics of an eroded beach that are well suited for beach nourishment include a relatively small erosion rate, valuable upland property, and large quantities of suitable sand located offshore in proximity to the candidate nourishment project. Small erosion rates ensure that the system can be restored to a previous desired state with a reasonable amount of sand and funding.

29.2 Methods of Delivery of Sand for Beach Nourishment

The delivery of sand to the beach is usually done by dredge; however, a comparatively small number of reasonably large projects (on the order of hundreds of thousands of cubic meters) have been conducted by truck haul. Factors favoring dredge placement include minimal interference with normal traffic and beach activities. Due to the large volumes of sand required for some projects, the truck haul approach can disrupt normal traffic activities and degrade road surfaces. Considering a beach nourishment project of 1 million

cubic meters, large dump trucks delivering approximately 20 m^3 per trip would require 50 000 trips. With favorable weather conditions, a large dredge can deliver up to 40 000 m^3 per day resulting in a project duration of 1–2 months or so allowing for maintenance, weather down time, etc. The area of the seafloor from which sand is removed by a dredge is termed a *borrow area* and must be investigated thoroughly for sand size suitability and other potential environmental effects.

Dredges usually employed for beach nourishment include pipeline dredges and hopper dredges. Pipeline dredges remove the sand from the seafloor with an intake pipe and convey the sand through a pipeline directly to the nourishment area on the dry beach. The diameter of the pipeline can vary from 0.5 to 1 m. Hopper dredges operate in more of a *batch mode* process. The dredges contain a large storage area in their hull and have *dragarms* that trail over the sides of the dredge through which sand is extracted from the sea floor and deposited in the hull while the dredge is underway at a speed of a few knots. In contrast to a pipeline dredge which operates best while excavating at least a meter *cut* depth, a hopper dredge skims a relatively thin layer on the order of 30 cm from the surface of the sea floor. Hopper dredges are classified by the volumes of sediment that they hold which can range up to several thousand cubic meters. Once full, the dredge moves to the nourishment area where various means are used to place the sand as beach nourishment. Direct pump out capacity from the dredge to the beach is becoming more common although the sand can be dropped to the sea floor and a pipeline dredge used to transport the sand to the beach while the hopper dredge returns to the borrow site for another load.

29.2.1 Characteristics and Considerations in Design of Beach Nourishment Projects

Beach nourishment projects are usually constructed with profiles steeper than equilibrium, in part, to facilitate quantification of sand volumes placed and also to minimize the immediate introduction of silts and clays into the nearshore waters during nourishment. Additionally, the project represents a *planform* which is out-of-equilibrium. Immediately after placement, wave action tends to cause the profile and planform to evolve toward equilibrium (Fig. 29.2). For most projects, the time scales for profile and planform equilibration are quite disparate with the profile equilibration occurring on a considerable shorter time scale than that for planform evolution. The profile adjusts such that approximately one-half the placed beach width is lost in the equilibration process and this adjustment is about

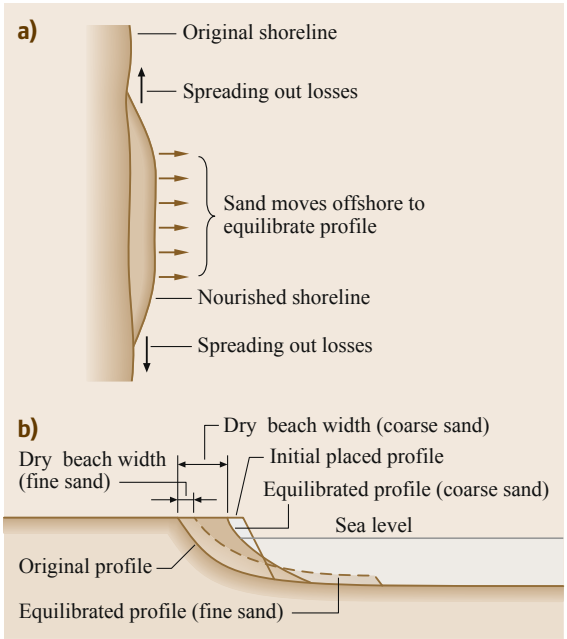


Fig. 29.2a,b The two modes of design consideration. The planform **(a)** evolves on a considerably longer time scale than the profile **(b)**

one-half complete in several years in moderate wave climates [29.2]. As a result of the difference in profile and planform equilibration time scales, design and prediction of the performance of beach nourishment projects usually consider the profile equilibration to occur instantaneously, leaving only the requirement for considerations of planform evolution which is a measure of project performance.

For compatible nourishment sediment, the volume density placed per unit length of beach v is translated to equilibrated additional dry beach width, Δy, by the relationship

$$\Delta y = \frac{v}{(h_* + B)} , \tag{29.1}$$

in which h_* is the so-called *depth of closure* considered to be the seaward limit of nourishment sand distribution and B is the berm height (Fig. 29.3). For example, with a value of $h_* + B = 10$ m, a volume density of 200 m^3/m will result in an equilibrated beach width of 20 m. We will discuss sand that is finer or coarser than the native sand later. The quantity, v is the volume per unit length of beach, is referred to as the *volume density* and is a useful parameter in comparing beach nourishment projects. Along Florida's east coast, an effective volume density has been found to be on the order of 250 m^3/m.

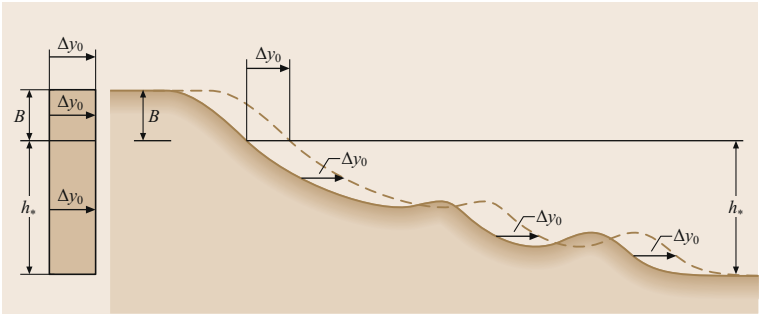


Fig. 29.3 Seaward profile advancement due to beach nourishment

29.2.2 Significance of Good Quality Sand

Understanding the significance of nourishing with good quality sand is essential to design and prediction of a beach nourishment project. *Good quality* infers that the nourishment sand hydraulic characteristics are similar to the native, a feature usually referred to as *compatibility*. In most cases, sand hydraulic characteristics equate to sand size characteristics; however, in those cases where the sand particle shape differs, such as for shell fragments, the fall velocity is a better indicator of sediment (and hydraulic) compatibility. Thus, our main concern here in the performance of nourishment projects is with the physical and hydraulic properties of the sediment although color may be of significant interest in some projects.

The main issue with sediment compatibility is that finer sands will equilibrate to a milder slope than for compatible sand. Thus for a given nourishment volume density, the additional dry beach width, an important parameter for the beach user and funding agencies, will be less for the finer sediment than for compatible sediments. A second effect is that the planform centroids of projects constructed of finer sand will migrate in the direction of net longshore sediment transport whereas those constructed with compatible sand will be stationary (discussed later). The planforms of projects constructed with finer sands will also evolve faster. Nourishment sand which is coarser than the native will yield a wider beach than with compatible sand although normally sand from offshore tends to be finer than the native; however, there are exceptions. Figure 29.4 shows calculated relationships between volume density and equilibrated shoreline advancement for various sediment size characteristics [29.2]. The term σ represents the so-called *sorting* or range of sediment sizes and a value of 0.5 is more representative than the more idealized $\sigma = 0.0$ (all of a single size). (Note that the line shown as $d_F = 0.2$ mm applies for $\sigma = 0.0$ and $\sigma = 0.5$. See *Dean* [29.2] for details). In this plot, $h_* + B = 7.8$ m. Considering only those cases with $\sigma = 0.5$, for a volume density, $v = 250\,\mathrm{m^3/m}$, it is seen that for a native sediment size of 0.2 mm, nourished (fill, subscript F) sediment sizes of 0.14, 0.2, and 0.275 mm yield equilibrated dry beach widths of 14, 32 and 48 m, respectively. Thus, the benefits of compatible sediment and sediment coarser than the native sand are substantial.

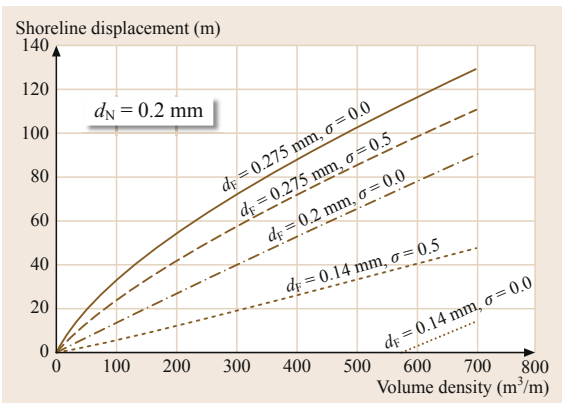


Fig. 29.4 Calculated variation of shoreline displacement vs volume density added and sediment characteristics (after [29.2])

29.2.3 Various Settings for Beach Nourishment

The geomorphic settings under which beach nourishment projects are constructed include:

1. Long beaches uninterrupted by artificial or natural barriers
2. Downdrift of a jettied or unjettied inlet
3. Pocket beach.

The first two of these settings are discussed briefly below.

Long Uninterrupted Beach

From design and prediction considerations, this is one of the simplest settings. As discussed later, for nourishment with compatible sediment, the project evolution is

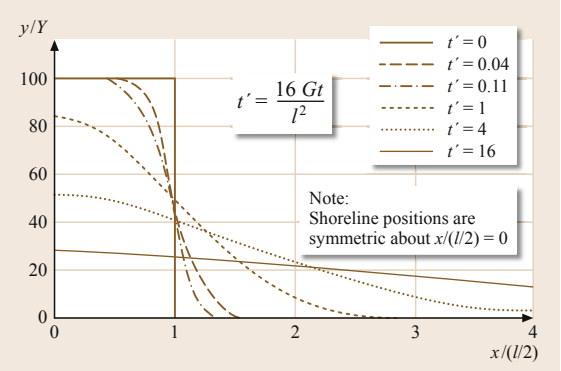


Fig. 29.5 Nondimensional planform evolution of an initially rectangular beach nourishment project of length ℓ and width Y. Note that only one-half of the project is shown and the evolution is symmetric about $x/(\ell/2) = 0$

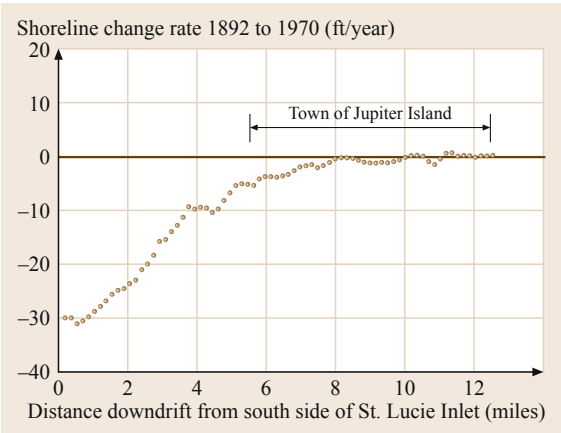


Fig. 29.6 Shoreline change rates downdrift of St. Lucie Inlet along the Southeast Florida coast

essentially independent of wave direction. Thus these projects act as *feeder beaches*. That is, as they evolve, they benefit both updrift and downdrift beaches. Figure 29.5 presents the calculated evolution of an initially rectangular planform for the case in which there is no background erosion. In the more usual case, background erosion (discussed later) must be taken into account in the design and prediction phases. It is seen that initially, the planform adjusts rapidly followed by a slower evolution as the planform becomes smoother. This effect will be quantified later.

Downdrift of a Littoral Barrier

Littoral barriers such as jetties and groins can cause substantial downdrift erosion which must be considered during the design and prediction phases. Two approaches can be applied to account for this background erosion and are discussed later. The most realistic approach of the two requires specification of the *effective* wave direction. Figure 29.6 presents the downdrift recession rate from 1883 to 1970 due to the cutting in 1892 of St. Lucie Inlet located along the southeast coast of Florida. The shoreline recession rate immediately downdrift (south) of the inlet was approximately 9 m/year, certainly among the highest in the State of Florida. With increasing time, the erosion within the Town of Jupiter Island increased and was addressed by a large-scale beach nourishment program. To date, the general area has been nourished with more than 12 million m^3 of sand.

29.3 Role of Structures in Beach Nourishment

Coastal structures can be used to slow or stop erosion of a specific section of beach by slowing littoral sediment transport. The main concern regarding coastal structures is that they can have unintended consequences such that by reducing beach erosion in one area, they may transfer the erosion to adjacent beaches. This is illustrated by the example of many groins constructed prior to understanding the interaction with adjacent beaches. In recognition of the possible adverse effects of coastal structures, they are essentially not permittable in some US coastal states. We regard this stance as unfortunate because, applied carefully, coastal structures can have a beneficial effect on the entire system. For example, using structures to reduce erosion of erosional hot spots (Sect. 29.5.1) within a nourishment project can reduce sand needs and long term costs. *Campbell* et al. [29.3] demonstrated that on an open coastline structural fields need to extend beyond the hot spot erosion areas to manage down-drift effects.

29.3.1 Terminal Structures

The term *terminal structures* indicates that the structures are placed at the end of a project or at the end of a littoral system. Although, as noted, structures are not considered favorably in many regulatory climates, terminal structures can provide a substantial benefit to beach stability. This is especially the case adjacent to inlets that have been dredged for navigation purposes. The deepened inlet can act as a *sink* on both the updrift and downdrift beaches causing sand to flow into the inlet. Thus, a short structure on the updrift or downdrift

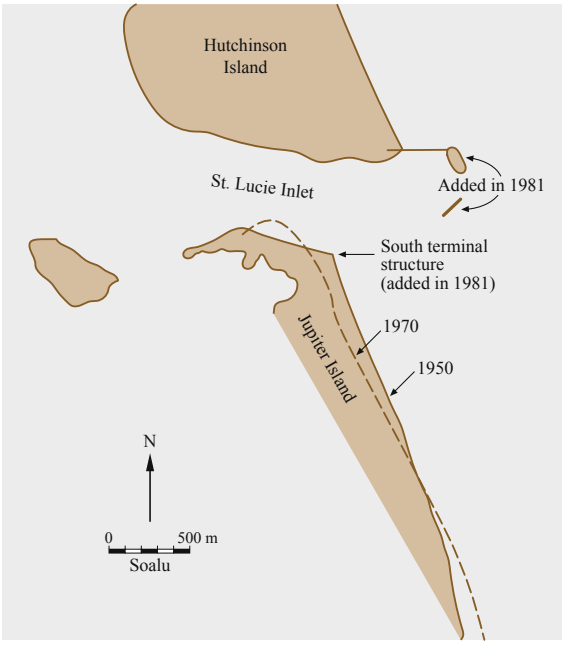


Fig. 29.7 Beneficial effect of downdrift jetty in reducing sand drawn into St. Lucie Inlet from Jupiter Island (after [29.4])

side of the inlet has been shown to provide a substantial stabilizing influence on the adjacent beach [29.4]. This can even be the case on a coastline with a substantial net longshore sediment transport.

Figure 29.7 shows the beneficial effect of a short south (downdrift) jetty constructed at St. Lucie Inlet. Prior to the construction of this jetty, sand was mobilized and drawn northward from Jupiter Island into the inlet, primarily on flood tidal currents at a rate estimated to be 170 000 m^3/year which is almost equal to the net southerly net transport of 180 000 m^3/year.

29.3.2 Groins

Groins have been discussed briefly above. They are best designed as a *template structure* which governs the profile of the affected beach. In designing a single groin or a field of groins, the potential effect on adjacent beaches should be examined carefully and understood along with a monitoring and contingency plan if the groins cause adverse effects. Adjustable groins have been applied with some success and provide a means of *fine-tuning* the structures in response to the monitoring results.

29.3.3 Detached Breakwaters

Detached breakwaters can be emergent or submerged and act to reduce wave energy in their lee and thus reduce longshore sediment transport in the case of oblique waves and cause sand to flow behind and be sequestered in the lee of the breakwater in the case of normally incident or oblique waves. Detached breakwaters have the same potential to interact with the nearshore sediment transport processes as groins and the same cautions apply for these structures especially if considered in areas with substantial net longshore sediment transport. An advantage of groins over detached breakwaters is their direct control on the beach profile. Additionally during elevated water levels which usually accompany storms, the effectiveness of detached breakwaters is reduced if the crest elevation is exceeded.

Submerged breakwaters have been proposed and implemented as an approach that is less visually intrusive than emergent detached breakwaters. However, it has been found that for effective performance, the combination of the breakwater crest width and elevation must be such that considerable energy dissipation occurs. The effect of elevated water levels during storms as discussed above is even more of a factor for submerged breakwaters in reducing their wave energy dissipation effectiveness. Also, with submerged breakwaters, there is the possibility that water transported over the breakwater by the waves will *pond* and flow along the coast removing sand from the area intended for stabilization, see *Dean* et al. [29.5] for monitoring results of a full scale submerged breakwater where this occurred.

29.4 Design and Prediction Approaches and Methods

The brevity of this chapter precludes a detailed presentation of design and prediction methodologies. For additional information, the reader is referred to *Dean* [29.2] and other references in this chapter. The sections below describe some of the key considerations in design and performance prediction of beach nourishment projects. The stakeholders, those funding the project, will be vitally interested in the future performance of the project especially when and how much future renourishment will be required. It is therefore incumbent on the engineer to be able to provide as accurate an assessment of future performance as possible along with a quantification of the uncertainties. It is our position that the *evolution of*

shorelines can only be predicted with some success if the initial shorelines are significantly out of equilibrium such as due to a nourishment project or a feature (such as a groin) is present that interrupts the longshore sediment transport. In the usual case in which shoreline erosion was occurring prior to the nourishment, this must be accounted for in the performance predictions.

Prior to designing a beach nourishment project, it is very helpful to try to understand the cause and magnitude of the erosion leading to the need. The cause may or may not be apparent. In most cases, nourishment will not correct the cause, so best estimates are that the project will continue to erode at a more-or-less predictable rate equal to the sum of the project-related evolution rate and the historic rate.

There are several levels of methodologies or *models* that can be applied to performance prediction. These range from an analytical model to one-line numerical models to advanced three-dimensional models. One-line models track only one contour of the profile, usually mean sea level, or some other contour representing the shoreline. Each of these model types can play a useful role in design and prediction.

29.4.1 Pelnard Considère Methodology

The simplest approach, that of Pelnard Considère (P-C) [29.6], is analytical and is useful for developing a basic understanding of the interrelationships between the design variables including project length, wave height, and sand transportability. Relatively simple one-line numerical models provide flexibility not available in the analytical models. The more advanced models provide greater detail than the other two levels. Due to the simplicity of the P-C approach, it is suggested that, if applicable, it be applied in conjunction with one of the other two more advanced level models. All models include a transport equation and a conservation of sediment equation.

A brief discussion of the P-C model and some results obtained will illustrate its utility. This one-line model combines a linearized version of the longshore sediment transport equation with the sediment conservation equation to yield the classic heat conduction equation

$$\frac{\partial y}{\partial t} = G\frac{\partial^2 y}{\partial x^2}\,, \tag{29.2}$$

in which y is the shoreline displacement, t is time, x is the longshore coordinate (the sign convention is positive to the right of an observer looking seaward), and G is the so-called *longshore diffusivity*, defined in terms of breaking wave conditions as

$$G = \frac{KH_b^{5/2}\sqrt{g/\kappa}}{8(S-1)(1-p)(h_* + B)}\,. \tag{29.3}$$

The quantity K is a sediment transport coefficient (of order 1) related to sediment size, H_b is the *effective* breaking wave height (similar to the root-mean-square wave height), g is gravity, κ is the ratio of breaking wave height to breaking depth (on the order of unity), S is the ratio of the sand specific weight to that of the water in which it is immersed (usually 2.57), p is the in-place porosity (0.35–0.4), h_* is the so-called depth of closure, and B is the beach berm height. The latter two variables are defined in Fig. 29.3.

There are several significant advantages to the P-C methodology. The governing equation is a classic linear equation with a number of known solutions that are relevant to beach erosion and beach nourishment. Due to its linearity, it is possible to superpose solutions, for example, the shoreline recession due to spreading out of a nourishment project and the historic background erosion rate. Finally, although the equation is linearized, it provides surprisingly valid results for many cases [29.7]. In our view, perhaps the greatest value provided by this methodology is the insight into the interrelationships between the various design and project variables, several of which are discussed below.

The P-C Solution for Nourishment of an Initially Rectangular Planform on a Long Straight Beach

The most valuable results apply for nourishment with compatible sediment on a long straight beach and, unless noted otherwise in this subsection, the following results apply to this case. The evolution of a beach nourishment project of an initially rectangular planform in the absence of background erosion (Fig. 29.5) is given by

$$y(x,t) = \frac{Y}{2}\left\{\operatorname{erf}\left[\frac{\ell}{4\sqrt{Gt}}\left(\frac{2(x-x_o)}{\ell}+1\right)\right] - \operatorname{erf}\left[\frac{\ell}{4\sqrt{Gt}}\left(\frac{2(x-x_o)}{\ell}-1\right)\right]\right\}\,, \tag{29.4}$$

where erf is the so-called error function, x_0 is the longshore location of the project center, ℓ is the project length, and Y is the project width.

A surprising result from the P-C methodology is that for nourishment with compatible sand, the centroid of a nourishment project remains fixed and is independent of wave direction (note that wave direction does not appear in the above equation). This is a valuable

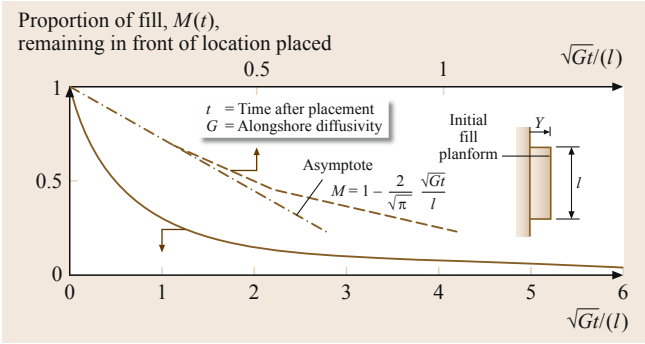


Fig. 29.8 $M(t)$ for case of beach nourishment of initially rectangular planform and no background erosion

result in design because usually the uncertainties of estimated wave heights and periods for design are much less than wave directions. Additionally, in the absence of background erosion, the longevity of a project is proportional to the square of the project length and inversely proportional to the effective breaking wave height raised to the 2.5 power, thus explaining why projects constructed in areas of reasonably mild wave climates are much more long lasting than those in more energetic wave environments and why longer projects last longer than shorter projects. In all settings, the evolution of a project depends only on the cumulative wave energy flux affecting the project. Thus, the evolution at a particular time is independent of the sequence of prior wave conditions that have resulted in the project evolution and depends only on the cumulative wave *loading*. The presence of a background erosion adds nuances to the above solution, but these can be incorporated by fairly direct approaches for simple background erosion distributions in the longshore direction.

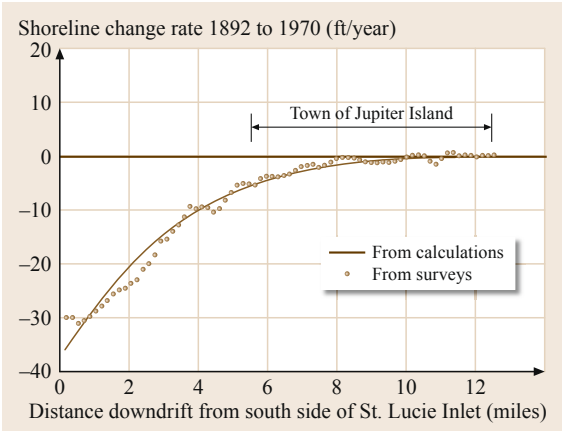


Fig. 29.9 Comparison of P-C theory with survey data south of St. Lucie Inlet, FL. Survey data: 1883 to 1970. For the theory, $G = 0.0111\,\mathrm{m}^2/\mathrm{s}$, $\theta - \alpha = 8.65°$

An example of the evolution of a nourishment project of initial rectangular planform as a function of time and distance has been presented earlier as Fig. 29.5. The evolution shown is symmetric about the centerline and the project center in this example is taken as $x = 0$. It is seen that the project first evolves rapidly, then at a slower rate. It is possible to integrate (29.4) to determine the proportion of sediment remaining at any time within the project limits, here defined as $M(t)$. The result is

$$M(t) = \frac{2\sqrt{Gt}}{\ell\sqrt{\pi}}\left(e^{-(\ell/2\sqrt{Gt})^2} - 1\right) + \mathrm{erf}\left(\frac{\ell}{2\sqrt{Gt}}\right), \qquad (29.5)$$

which is graphically shown in Fig. 29.8 where it is seen that initially the rate of loss is more rapid than after the project planform has become longer and with less abrupt changes (also, Fig. 29.5).

The P-C Solution for Shoreline Change Updrift or Downdrift of a Littoral Barrier

A second solution for shoreline change is that for downdrift (or updrift) of a littoral barrier

$$y(x,t) = \pm\left[\sqrt{\frac{4Gt}{\pi}}e^{-x^2/4Gt} - x\mathrm{erfc}\left(\frac{x}{\sqrt{4Gt}}\right)\right] \times \tan(\theta - \alpha_0), \qquad (29.6)$$

where the plus and minus signs apply for updrift and downdrift, respectively, of the barrier, θ and α_b are the azimuth of the outward normal from the unaffected shoreline and the azimuth of the direction from which the breaking wave originates, respectively and $\mathrm{erf}c = 1 - \mathrm{erf}$. The above solution applies to the case before bypassing commences around the barrier, after which a different solution applies (see *Dean* and *Dalrymple* [29.8] for additional information). To apply (29.6), an estimate of breaking wave direction relative to the shoreline, α_b, is required. An alternative if shoreline

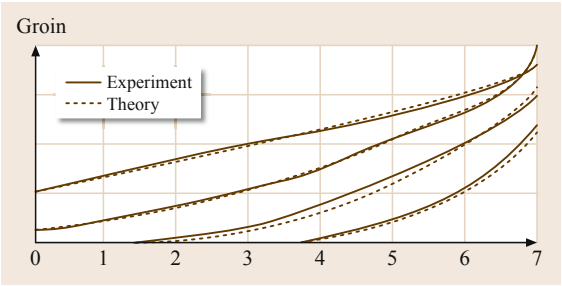


Fig. 29.10 Comparison of P-C theory with laboratory data updrift of a groin (after [29.9])

change data are available is to determine the wave direction corresponding to the best fit of the solution to the measured data. Figure 29.9 presents the *best fit* of the P-C methodology to the data presented earlier in Fig. 29.6. In this case, the associated value of the longshore diffusivity G, and angle $\theta - \alpha$ are $0.0111\,m^2/s$ and 8.65°, respectively.

Figure 29.10 presents results from a laboratory experiment in which measurements on the updrift side of a groin compare favorably with the P-C theory.

29.4.2 One-Line Numerical Models

One-line numerical models have a considerable advantage over the P-C method in terms of the flexibility they offer allowing groins or other coastal features such as inlets with or without bypassing to be introduced into the design. Many such models have been developed, some to an advanced level of detail. The US Army Corps of Engineers has developed the GENESIS model to a level that allows application to a variety of useful design conditions including groins and detached breakwaters [29.7, 10, 11].

29.4.3 N-Line and more Detailed Models

More complex models allow the third (vertical) dimension to be represented and thus, in principle, provide much more valuable information than the less complex models. Some of these models are in the research rather than the application arena. Application of all models requires some calibration with the more complex models usually requiring the greatest amount of calibration to achieve the level of detail desired.

Delft3D is a comprehensive, publicly available, process-based modeling suite that simulates two-dimensional (2-D) and three-dimensional (3-D) flows, sediment transport and morphology, waves, water quality and ecology and is capable of representing the interactions between these processes. The suite is designed for use by experts and nonexperts, which may range from consultants and engineers or contractors, to regulators and government officials, all of whom are active in one or more of the stages of the design, implementation, and management process. Delft3D is suitable for simulating erosion and deposition over storm-length (7–14 day) or project-length (5–10 year) time scales and handles a variety of coastal structures. Morphology type models like Delft3D offer distinct advantages over N-Line models because they are capable of updating three-dimensional bathymetric features at each time-step, thus providing feedback to the wave and flow models for a more complete simulation of the coastal environment. Delft3D requires an extensive amount of data and time to set up and run, and therefore may not be the best option for many applications.

29.5 Additional Design Considerations

29.5.1 Erosional Hot Spots

In beach nourishment projects, erosional hot spots (EHSs) are defined as areas that erode more rapidly than the adjacent areas or possibly areas that erode more rapidly than anticipated in the design. EHSs occur in nearly all beach nourishment projects and may or may not be predictable. In a celebrated case at Grand Isle, LA, USA [29.12], the nourished beach advanced significantly landward of the two seaward borrow areas. These borrow areas which may reported to have filled with very fine sediment and caused significant damping of the waves (personal communication, A. J. Combe). These borrow pits acted as detached breakwaters in reducing the shoreward wave heights causing shoreline advancement landward of the dredge pits and erosional hot spots adjacent to these advancement areas. *Benedet* and *List* [29.13] applying Delft3D, have correlated EHSs on the Delray Beach, FL, nourishment project with the seaward borrow areas.

The EHSs experienced at Grand Isle, LA, are exceptional; however, EHSs that occur in typical nourishment projects are not inconsequential. Figures 29.11 and 29.12 present the performance of the 1992 Anna Maria Key, FL, beach nourishment project. This project was 6.7 km in length with a nourishment volume of 1.78 million m^3. The initial additional beach width varied from 55 to 107 m and the nourishment volume density varied from 200 to $350\,m^3/m$. The most recent survey (February 1999) in these figures is for 6 years after nourishment and as of that time, the change in immediate postnourishment volumes ranged between a loss of 320 to a gain of $160\,m^3/m$. The corresponding shoreline width reductions ranged from 37 to 69 m. The borrow area for this project was located between monuments 25–34 and was fairly close to shore. Thus, some of the postnourishment changes in these figures could be due to the proximity of the borrow pit; however, some changes for other areas within this project are significant and generally cannot be predicted by any numerical models. The shoreline recession includes both volume loss and profile adjustment toward equilibrium.

Shoreline irregularities also exist along natural beaches. Often this occurs during mild wave conditions

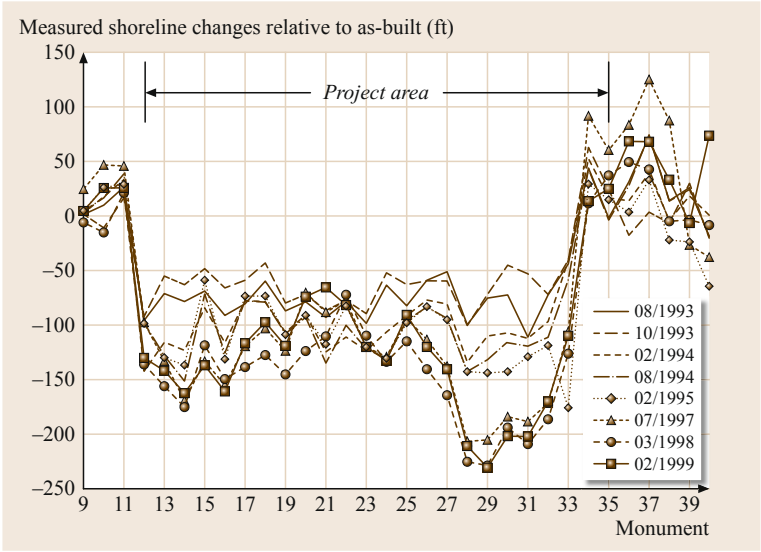


Fig. 29.11 Postnourishment shoreline changes in Anna Maria Key beach nourishment project. Borrow area is quite close to shore and between monuments 25 and 34. Monuments are spaced at approximately 300 m

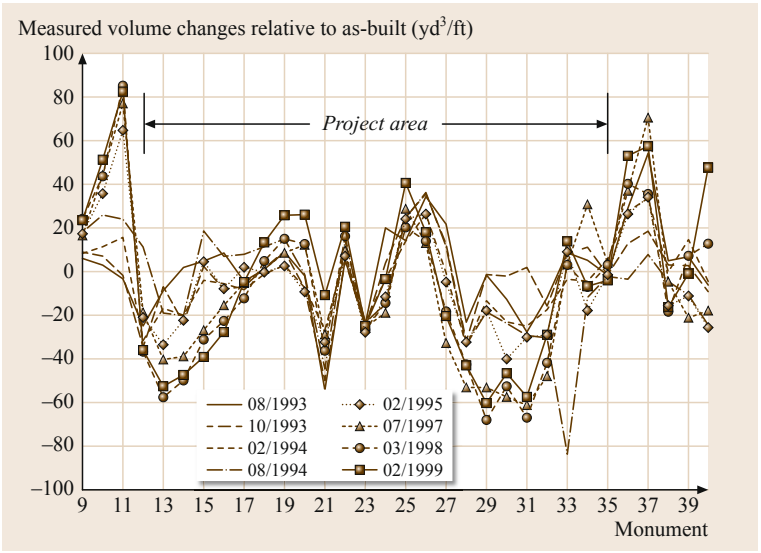


Fig. 29.12 Postnourishment volume density changes in Anna Maria Key beach nourishment project. Borrow area is quite close to shore and between monuments 25 and 34. Monuments are spaced at approximately 300 m

after an erosional event that has deposited a more or less linear offshore bar. The milder wave conditions tend to result in portions of the bar moving shoreward and landward of these portions, the shoreline tends to advance. This stage of recovery is often associated with the presence of rip currents.

29.5.2 Nearshore Placement

Figure 29.1 showed the case of sediment placed in the nearshore with the expectation that either the sand would move landward and widen the beach or dampen the incident waves thus sheltering and stabilizing the beach. The following discussion only addresses the conditions under which the nearshore placement would move landward. However, it is noted that during storms with elevated water levels, the effectiveness of the nearshore mound in wave damping would be reduced.

A number of studies have been conducted to address conditions under which nearshore deposits would be stable or move landward or seaward. Some field observations are inconclusive partly due to the lack of detailed wave measurements and surveys of the deposit. An analytical model has been developed by *Douglass* [29.14, 15] and *Hands* and *Allison* [29.16] have developed criteria based on the wave-induced bottom water particle velocities in the vicinity of the placement. *Otay* [29.17] has developed an extensive summary of

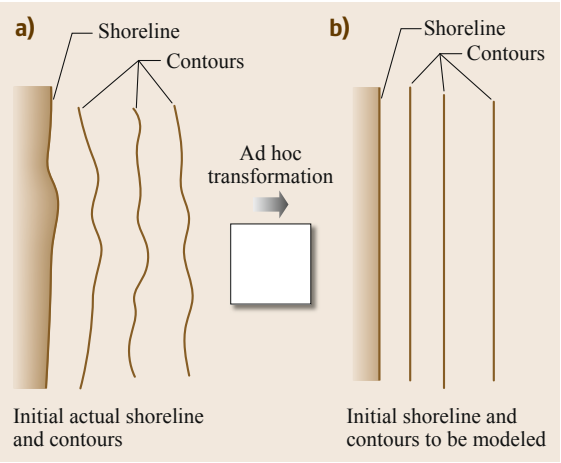


Fig. 29.13 Recommended ad hoc transformation for modeling systems in which a large perturbation (nourishment and/or a littoral barrier) is to be introduced (after [29.20]) (courtesy of the American Society of Civil Engineers)

the field tests of nearshore deposits and information regarding their movement.

Andrassy [29.18] described a fairly small (113 500 m^3, 365 m long) nearshore placement off southern California. The deposit was placed in a water depth of approximately 7 m with the crest in 4 m depth. It was found that the deposit moved landward over a period of approximately 2 year. The sediment size was not stated and the wave heights and periods encountered were up to 1.5 m and 15 s, respectively.

The best instrumented field test in the United States has been described by *Browder* and *Dean* [29.19] in which 3 million m^3 was placed in water depths of 6 m over a 7 −km length with the deposit crest 4.5 m below water level. This deposit was monitored for approximately 9 years during which four hurricanes passed over the site. Directional wave gages were deployed for 4.5 year and measured significant wave heights up to 2.7 m with a period of 13 s during Hurricane Andrew in 1992. These conditions certainly would have caused onshore transport of the placement based on the Douglass or Hands and Allison predictive methodologies. During the 9 year monitoring period, it was found that the centroid of the deposit moved landward a distance of approximately 20 m compared to the centroid location some 700 m seaward of the shoreline.

In summary of this issue, it is concluded that sediment size, while not included in most predictive methods, may be important. Second, to be sure that the placement will move landward within a reasonable time frame, it is recommended that beach profiles be examined to determine the seasonal limit of the offshore bar and place the sand landward of this limit.

29.5.3 Ad-Hoc Transformation for Modeling

Most shorelines are somewhat irregular. The question arises as to the best approach to represent these shorelines for numerical modeling, It has been noted that shoreline evolution can only be effectively modeled if the system is significantly out of equilibrium. Also, generally prenourishment shoreline change rates must be included in postnourishment shoreline change predictions. Our recommendation, sketched in Fig. 29.13, is that an ad-hoc transformation be applied in which the modeled shoreline and offshore contours are straight and parallel and the prenourishment shoreline change rates be applied empirically. In this way, the system perturbation (nourishment) is modeled in a geometry which is most appropriate for that purpose and the subtleties of the offshore contours that cause the background erosion are applied directly through application of the prenourishment shoreline changes. The question arises of whether the prenourishment conditions should be applied directly as shoreline changes or as gradients in sediment transport. The best approach requires judgment by the modeler of the cause of the changes. If they are due to cross-shore transport, then obviously, they should be applied directly. If no coastal structures such as groins are present, the results will not be affected by the choice of the application method.

29.6 Legacy Beach Nourishment Projects

Legacy projects as defined here are those that have been nourished and in place for many years, have performed well, have reasonable documentation, and in many cases, the average beach user is unaware of their nourishment history. Although there are undoubtedly many such projects that qualify, we illustrate two US projects in some detail and provide references for several other legacy projects.

29.6.1 Examples and Discussions of Legacy Projects

Santa Monica Bay Beach Nourishment

Santa Monica Bay in southern California is approximately 64 km in length. The central and southern portions of this bay have been nourished primarily with *sand of opportunity* generated through construction un-

Fig. 29.14 Santa Monica Bay. General Area of Interest (after [29.21])

related to a primary need for beach nourishment. The nourishment commenced in 1939 with the two main nourishments occurring in 1947 (13 million m^3, construction of the Hyperion Sewage Treatment Facility) and in 1963 (7.7 million m^3, construction of Marina del Ray Harbor). These two projects accounted for 93% of the total of 22.2 million m^3 placed between 1939 and 1989. Of the total sand volume placed, less than 4% was for shoreline stabilization purposes. Several structures consisting of groins and detached breakwaters, two harbors and a submarine canyon are located within the 35 km segment of the central and southern portions of Santa Monica Bay. There are no known large natural sources of sediment input to this system and the submarine canyon is believed to represent a substantial sediment sink. Some losses into the submarine canyon have been controlled through groin construction. This overall project qualifies as a Legacy Project due to its longevity, monitoring data, and general lack of recognition as a nourished beach by the average beach visitor. Most of the description of this project presented herein is from *Leidersdorf* et al. [29.21] and the reader is referred to that source for a more complete description.

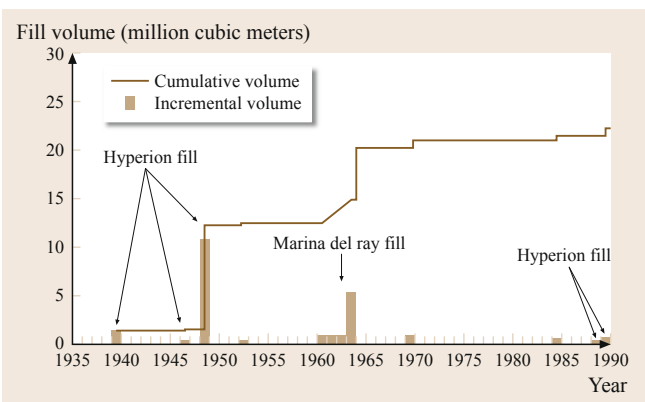


Fig. 29.15 History of nourishment along Santa Monica Bay (after [29.21])

The beaches remain substantially advanced relative to their pre-1939 positions. On average, it is estimated that the beaches are approximately 100 m wider than their pre-1939 conditions. Figure 29.14 shows the general area of interest and Fig. 29.15 presents the history of beach nourishment. Figure 29.16 shows the shoreline changes at two locations from 1935 to 1990, a period of 55 years.

In summary, the beaches of Santa Monica Bay have performed extremely well, have had a profound impact on the economy and recreational activities of the area, are recognized internationally, and are rich in documented results thereby providing learning opportunities for students of coastal engineering.

Delray Beach, FL

The first City of Delray Beach nourishment project was constructed in July of 1973 along 4.3 km of shoreline. The immediate postconstruction surveys showed an average mean high water beach widening of 80 m but the beach equilibrated to about half that width within the first few years. The 1973 project placed approximately 1 250 000 m^3 of material, of which 380 000 m^3 had eroded by 1977. In 1974, dune vegetation was planted to augment the beach restoration project and assist in reducing the losses due to wind blowing sand across the coastal road and covering the adjacent lawns.

Fig. 29.16a,b Variation with time of two profiles in Santa Monica Bay (after [29.21]) ►

As of 2012, a total of over 4.78 million m^3 of sand has been placed on Delray Beach over a period of 39 years as a result of four beach nourishments (1978, 1984, 1992, 2002). A storm damage repair was also constructed in 2005 with 190 000 m^3 following a series of hurricanes that impacted Florida during the 2004/5 hurricane seasons. The Delray Beach Fifth Periodic Beach Renourishment Project has been designed and permitted for construction in 2012. That project will place approximately 920 000 m^3 of fill on the beach along a distance of about 3.1 km.

The time between nourishment events has increased from 5 to 10 year (except for the hurricane repair in 2005). Additionally, as is evident in Fig. 29.17, the additional sand volume within the project area has increased from 1.25 million m^3 in 1973 to 2.9 million m^3 in 2009. The increase in the nourishment interval with increasing time can be attributed to reduced spreading losses as the earlier projects perform as longer projects and thus lose sand at slower rates.

Figure 29.17 also presents a comparison of the measured and calculated volume changes remaining within the project area. The calculations are based on the Pelnard Considère method (29.4) and present results for two G values, $G = 0.3\,ft^2/s$ ($0.0028\,m^2/s$) and $G = 0.6\,ft^2/s$ ($0.0056\,m^2/s$). Of interest in this comparison is that a factor of 2 in the G values has a relative small effect on the calculated performance; however, of course, this effect increases with time. Second, referring to the calculated performance for the smaller G value, it is seen that at some times the calculated volumes are greater than the measurements and at other times less. This is due to the calculations based on a single *representative* wave height whereas in nature, some years are more *stormy* than others (the G values vary with time as shown the dependency on wave height in (29.3)). Comparisons such as this provide an effective basis for calculating performance of future projects.

Other Legacy Projects

Here we list three other legacy projects with a very brief description and some references.

Additional Southern California Beaches. *Herron* [29.22] has summarized the early placement of large quantities of sand as beach nourishment on several beaches including Santa Monica Bay. This sand was usually derived as a by-product of coastal projects rather than for beach nourishment. He states:

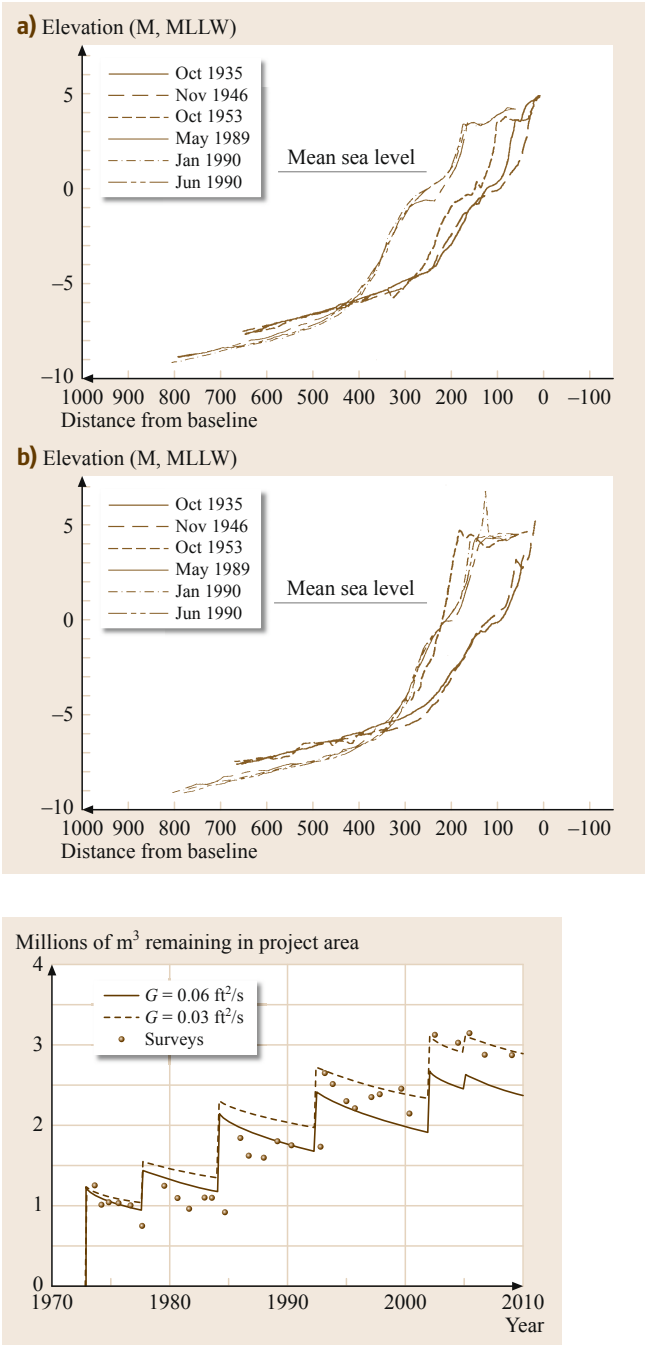


Fig. 29.17 Comparison of measured and calculated volume remaining within the project area, Delray Beach Nourishment Project. Two values of longshore diffusivity G. No background erosion considered

Between 1919 and 1978 about 60 *coastal projects have supplied* 83 million m^3 *of sand to improve the*

beaches of southern California. Three projects were solely beach replenishment projects. All of the others were coastal oriented projects in which the deposition of sand on the public beaches was simply a by-product or secondary benefit of the total consideration.

Of this 83 million m^3, Herron only lists 21 million m^3 placed in the Santa Monica Bay cell leaving 62 million m^3 placed on beaches other than those in Santa Monica Bay. Based on personal observation, many of these beaches continue to perform well with only minor maintenance required.

Miami Beach, FL. This project, also known as the Dade County Beach Nourishment Project, is approximately 16 km long and was constructed over the period 1976 to 1981 with approximately 7.7 million m^3 of sand from offshore borrow areas. Prior to the nourishment, in many places the beach was located along the seawalls making it risky to walk along the shoreline at high tide and energetic wave conditions. An excellent review of this project is provided by *Wiegel* [29.23]. The effects of this project in revitalizing the economy of the area are quite substantial and are presented in *Houston* [29.24, 25].

Atlantic Beach and Jacksonville Beach, FL. The St. Johns River Entrance north (updrift) of Atlantic Beach and Jacksonville Beach has been stabilized by long jetties and deepened for navigational purposes. This entrance has essentially severed the substantial net southerly sediment transport estimated at approximately 420 000 m^3/year resulting in severe erosion of the beaches of Atlantic Beach and Jacksonville Beaches in the early 1960s. The major northeaster of 1962 and Hurricane Dora in 1964 devastated these beaches destroying seawalls and other structures located along the shore. Following these events, nourishment projects were constructed with sand obtained from the entrance channel shoal material and ebb tidal shoal of St. Johns River. This combination of projects has successfully maintained the beaches in this area since the 1970s, best illustrated in the beach performance during the 2004 severe hurricane and Nor'easter season, where the buildings and properties adjacent to nourished beaches were relatively undamaged.

Since 1963, a total of 10.3 million m^3 of sand has been placed south of the St. Johns River Entrance. Ten projects were constructed with sand dredged for navigation (4.05 million m^3), seven were Federal shore protection projects (5.2 million m^3) and two were joint navigation and shore protection projects.

Although a thorough documentation of this project is warranted, none has been conducted. However, an excellent Powerpoint presentation by *Howard* et al. [29.26] is available.

29.7 Other Beach Nourishment Projects

The nourishment projects discussed in this section are examples of projects of different scales and settings and are generally considered to be performing quite effectively.

29.7.1 Seabright to Manasquan Inlet, NJ

This beach has one of the longest histories as a recreational beach destination in the United States, which led to the early development of the coastline. The shoreline was protected early and often with a wide variety of coastal protective structures, including substantial seawalls, groins, and revetments, all considered the latest methods at the time.

Initial project construction occurred between January 1994 and June 2000 and the first and second renourishments were completed in November 2002 and early 2012. Initial construction placed 17.0 million m^3 of sand along most of the 34 km ocean shoreline of Monmouth County. The beach has performed well and exceeded the expected nourishment performance despite some predictions to the contrary.

29.7.2 Captiva Island, FL

Captiva Island is a 8 km long island on the southwest coast of Florida bordered by two coastal inlets; Redfish Pass to the north and Blind Pass to the south. Since Redfish Pass was opened by a hurricane in 1923 the island has been starved of the net southerly sediment transport and has eroded significantly. In 1981 beach nourishment took place along the northern 3.2 km of the island placing 500 000 m^3 of sand from the ebb shoal of Redfish Pass. Island-wide nourishment followed in 1989 which placed 1.2 million m^3 of sand was also taken from Redfish Pass ebb shoal. There have been two subsequent renourishments of the island using offshore sands in 1996 and 2005 which placed an additional 1.4 million m^3. The southern limit of the project has been extended to

the adjacent Sanibel Island. Captiva Island has two terminal groins to help reduce losses from the island.

29.7.3 Treasure Island, FL

Treasure Island is a 5.3 –km barrier island bordered by Johns Pass to the north and Blind Pass (a different Blind Pass than adjacent to Captiva Island) to the south. Treasure Island grew to its present size after a tidal inlet at the north end of the island closed and the southern migration of Blind Pass caused the island to lengthen. Development of resorts on the island began in the 1920s followed by a surge in development in the 1950s. By the 1960s, the island was facing serious erosion problems. In response, the city of Treasure Island constructed 56 concrete (originally timber) groins and a rubble mound jetty on the north side of Blind Pass. At the time, the average erosion rate was determined to be 52 000 m^3/year [29.27].

The first nourishment project on the west coast of Florida was on Treasure Island in 1969. After the initial construction and repair following Hurricane Gladys in 1968, Treasure Island was renourished every 2–3 years. The material used for the nourishments was sourced from Blind Pass, Johns Pass, and an offshore sand pit parallel to the island. Sunshine and Sunset beaches required frequent nourishment.

Sunshine Beach, at the north end of Treasure Island, is historically erosional as a result of inlet effects. The only times that Sunshine Beach has been accretional was after a nourishment in the 1970s when dredge material deposited in shallow offshore waters migrated onshore. In 1989, an angled structure was constructed to protect development to the south [29.28]. No evidence was found on why or when it was removed. A terminal groin was constructed on the south side of Johns Pass in 2000 to limit the end losses from Sunshine Beach to the inlet. Prior to its construction, the shape and beach width at Sunshine Beach fluctuated significantly. After the groin construction, the beach was fixed and straight, limited by the effective length of the groin.

29.8 Summary and Conclusions

Large-scale designed beach nourishment projects are the result of a relatively young technology and thus issues expected to be dominant in the next century or so are still evolving. However, the major emerging issues are discussed below.

Energy Costs

Dredging is a process requiring fairly large amounts of energy. High velocities must be maintained in the pipeline delivering the sand–water mixture to prevent sand from settling and blocking the pipe. Some projects are located at substantial distances from the borrow areas and may require a booster pump which increases the energy costs. Currently (2012), in the United States, energy costs are approximately one-third of the total. Although future innovations in dredging technology may reduce energy costs per volume of sand delivered a specific distance, these improvements are not expected to reduce energy requirements by more than 25%.

Environmental Issues

Beach nourishment has been shown to both temporarily impact and to be a long-term benefit to the beach environment. Certainly a nourished sandy beach provides more benefits to the biota than an armored beach which could be either a seawall or a revetment. In the State of Florida where many sea turtles nest, it has been found that, compared to prenourishment nesting density, usually a reduction in nesting density will occur for several years after which the turtles will return to their prenourishment nesting activities. Beach nourishment and society would benefit by a more balanced environmental assessment of the beneficial and adverse impacts of beach nourishment including the adverse impacts of a *do-nothing* approach.

Limited Sand Resources

The availability of offshore sand resources varies widely with location depending on its geologic history. The relatively large deposits of sand in ebb tidal shoals have been utilized for nourishment in some projects. In some areas, the volumes of good quality offshore sand resources are quite abundant whereas at others, quantities are quite limited. In southeast Florida, limited sand resources tend to be located between offshore reefs that are generally approximately parallel to the shoreline. Along some counties in this area, the offshore sources of large quantities of readily available, good quality sand have been nearly depleted. In some areas, sand from previous nourishments has evolved to uneven distributions along the project area, resulting in programs to redistribute the sand along the project rather than provide additional sand. In other areas, additional sand has been provided by truck haul from inland sources as discussed earlier. The limitations of truck haul and damage to roads have been discussed. One possibility would be to deliver sand from inland quarries to the shore by dedicated pipelines and then transport this sand by dredge

along the shoreline to areas of need. Sand delivered in this manner could also act as long-term *feeder beaches* to the adjacent shorelines.

Public Perception

The public often perceives beach nourishment as benefitting only those adjacent to the beaches although the benefits may extend well inland and accrue to both state and federal coffers (see the excellent articles by *Houston* [29.24, 25]). However, most, if not all, projects that receive partial public funding require public access and parking, restroom facilities, and must meet other public benefit criteria. Often, the projects play a substantial role in protecting emergency evacuation routes and provide storm protection to valuable real estate that generates significant tax revenue.

29.9 Nomenclature

α_b Azimuth of the direction from which the breaking waves originate.
B Berm height
d Sediment grain diameter
G Longshore diffusivity
g Gravitational acceleration
h_* Depth of closure, seaward limit of nourishment sand distribution
H_b Breaking wave height (root-mean-square)
K Sediment transport coefficient
κ Breaker index, or ratio between breaking wave height and breaking depth
ℓ Project length
$M(t)$ Proportion of beach nourishment sediment remaining at time t
p Porosity
σ Sorting, or range, of sediment sizes
S Specific gravity
t Time after placement
θ Azimuth of the outward normal from the unaffected shoreline
v Sediment volume density placed per unit length of beach
x Longshore distance
x_o Alongshore location of beach nourishment project center
Δy Equilibrated additional dry beach width
y Cross-shore distance
Y Beach nourishment project width

References

29.1 P. Bruun: Sea level rise as a cause of shore erosion, J. Waterw. Harb. Coast. Eng. Div. **88**(1), 117–130 (1962)
29.2 R.G. Dean: *Beach Nourishment: Theory and Practice* (World Scientific, Singapore 2002) p. 396
29.3 T.J. Campbell, M.G. Jenkins: Design considerations for hot spot erosion areas on beach nourishment projects, Proc. 28th Int. Conf. Coast. Eng. (2002) pp. 3642–3648
29.4 R.G. Dean: Terminal structures at ends of littoral systems, J. Coast. Res. **SI18**, 195–210 (1993)
29.5 R.G. Dean, R. Chen, A.E. Browder: Full scale monitoring study of a submerged breakwater, Palm Beach, Florida, USA, Coast. Eng. **29**, 291–315 (1997)
29.6 R. Pelnard–Considère, Essai de théorie de l'evolution des formes de rivage en plages de sable et de galets, 4th Journees de l'hydraulique, Les Energies de la Mer **II**(1), 289–298 (1956)
29.7 H. Hanson, M. Larson: Comparison of analytical and numerical solutions of the one-line model of shoreline change, Proc. Coast. Sediments (ASCE, Reston 1987) pp. 500–514
29.8 R.G. Dean, R.A. Dalrymple: *Coastal Processes with Engineering Applications* (Cambridge Univ. Press, Cambridge 2002) p. 475
29.9 B. Le Mehaute, A. Brebner: *An Introduction to Coastal Morphology and Littoral Processes*, Civil Engineering Department Report 14 (Queen's University, Canada 1961)
29.10 H. Hanson, N.C. Kraus: *GENESIS: Generalized Model for Simulating Shoreline Change: Report 1*, Technical Report CERC 89-19 (US Army Waterways Experiment Station, Coastal Engineering Research Center, Vicksburg 1989)
29.11 H. Hanson: Genesis: A generalized shoreline change numerical model, J. Coast. Res. **5**(1), 1–27 (1989)
29.12 A.J. Combe, C.W. Soileau: Behavior of man-made beach and dune, Grand Isle, Louisiana, Proc. Coast. Sediments (ASCE, Reston 1987) pp. 1232–1242
29.13 L. Benedet, J.H. List: Evaluation of the physical process controlling beach changes adjacent to nearshore dredge pts, Coast. Eng. **55**, 1221–1236 (2008)
29.14 S.L. Douglass: Estimating landward migration of nearshore, constructed, sand mounds, J. Waterw. Port Coast, Ocean Eng. **121**(5), 247–250 (1995)
29.15 S.L. Douglass: Nearshore placement of sand, Proc., 25th Int. Conf. Coast. Eng. (1996) pp. 3708–3721, Chapter 286
29.16 E. Hands, M.C. Allison: Mound migration in deeper water and methods of categorizing active and sta-

ble depths, Proc. Coast. Sediments (1991) pp. 1985–1999
29.17 E.O. Otay: Long-Term Evolution of Nearshore Disposal Berms (Department of Coastal and Oceanographic Engineering, University of Florida, Gainesville 1994)
29.18 C.J. Andrassy: Monitoring of a disposal mound at silver strand park, Proc. Coast. Sediments (ASCE, Reston 1991) pp. 1970–1984
29.19 A.E. Browder, R.G. Dean: Monitoring and comparison to predictive models of the perdido key beach nourishment project, Florida, USA, Coast. Eng. **39**(2–4), 173–192 (2000)
29.20 R.G. Dean, C.H. Yoo: Beach-nourishment performance predictions, J. Waterw. Port Coast. Ocean Eng. **118**(6), 567–586 (1992)
29.21 C.J. Leidersdorf, R.C. Hollar, G. Woodell: Beach enhancement through nourishment and compartmentalization: The recent history of Santa Monica Bay. In: *Beach Nourishment Engineering and Management Considerations*, ed. by D.K. Stauble, N.C. Kraus (ASCE, Reston 1993) pp. 71–85
29.22 W.J. Herron: Artificial beaches in Southern California, Shore Beach **48**, 3–12 (1980)
29.23 R.L. Wiegel: Dade county, Florida beach nourishment and hurricane surge protection project, J. Shore Beach **60**(4), 2–28 (1992)
29.24 J.R. Houston: International tourism and U.S. beaches, Shore Beach **64**(2), 3–4 (1996)
29.25 J.R. Houston: The economic value of beaches – 2002 update, Shore Beach **70**(1), 9–12 (2002)
29.26 S. C. Howard, K. R. Bodge, T. R. Martin: Beach renourishment in Jacksonville, http://www.fsbpa.com/2011TechPresentations/Howard_Bodge_Martin_fsbpa%201-2011r.pdf, Powerpoint Present. 2011 FSBA Conf., Jacksonville
29.27 USACE: *Beach Erosion Project Review Study and Environmental Impact Statement for Pinellas County, Florida* (US Army Corps of Engineers, Jacksonville 1984)
29.28 J. Krock: Historical Morphodynamics of Johns Pass, West-Central Florida, M.S. Thesis (University of South Florida, Tampa 2005), http://scholarcommons.usf.edu/etd/731

30. Storm Hazard Mitigation Structures

David R. Basco

Hard structures have long been the method of shore protection from coastal storms. In this chapter, we first introduce the design criteria, design philosophy, and design constraints. Both armored shorelines (seawalls, dikes, revetments, etc.) and stabilized shorelines (headland breakwaters, near-shore breakwaters, groins, etc.), are possible alternatives. However, stabilized shorelines coupled with beach nourishment (Chap. 29) have become the preferred choice for many reasons at many locations. Rubble-mound (rock) structures and monolithic (concrete) structures are the most common types. Their functional design (wave run-up, overtopping, plan form layout, etc.) and structural stability (armor layer, sliding, overturning, etc.) are the focus of this chapter summarizing the latest and most trusted probabilistic design methods employed in the United States, Europe, and Japan. There exists a very large literature on these topics. Design practice details are quite extensive; we include website links to major reference sites, and a responsible engineer must consult these references for all the details.

In the future, the potential for the accelerated, relative sea level rise (including local subsidence) will increase the risk for more damage and reduce the resilience of existing shore protection systems. The design of hard structures for shore protection continues to evolve over time. As sea levels rise more quickly, the need for more basic research and more knowledgeable coastal engineers will also increase.

30.1 **Design Criteria, Philosophy, and Constraints** 655
30.1.1 Water Levels 655
30.1.2 Wave Conditions 656
30.1.3 Surf Conditions on the Structure 656
30.1.4 Probabilistic Design 657
30.1.5 Coastal Risk 657
30.1.6 A Coastal Storm Severity Index 657
30.1.7 Design Constraints 657

30.2 **Coastal Armoring Structures** 658
30.2.1 Types and Purpose 658
30.2.2 Functional Design 658
30.2.3 Structural Design 663

30.3 **Shoreline Stabilization Structures** 674
30.3.1 Types and Purposes 674
30.3.2 Functional Design 675
30.3.3 Structural Design 681

30.4 **Websites and Sea Level Rise Trends** 681
30.4.1 Websites 681
30.4.2 Sea Level Rise Trends 681

References 681

Armoring the shoreline with a hard structure has long been a traditional way of coastal shore protection from the excessive forces of the sea. Since about 1950, coastal population increases, tourism, and concerns for the environment have caused a shift away from the traditional armoring solution. Shoreline stabilization structures and the addition of natural sediments (Chap. 29) have often become the preferred choice for shore protection. In this chapter, *hard structures* refers to both (1) coastal *armoring structures* (seawalls, dikes, revetments, etc.) and (2) shoreline *stabilization structures* (head lands, breakwaters, groins, reefs, sills, etc.). These structural alternatives are changes to the natural, physical system for shore protection. Nonstructural alternatives (flood proofing, zoning, retreat, etc.) are changes to the built environment by adaptation and accommodation. There are many alternatives for shore protection as illustrated in Fig. 30.1. The correct choice depends on the design constraints for the particular site and is the responsibility of the design coastal engineer.

The excessive forces are the result of:

1. Coastal storms that produce elevated water levels, high waves, and strong currents over the storm duration, and

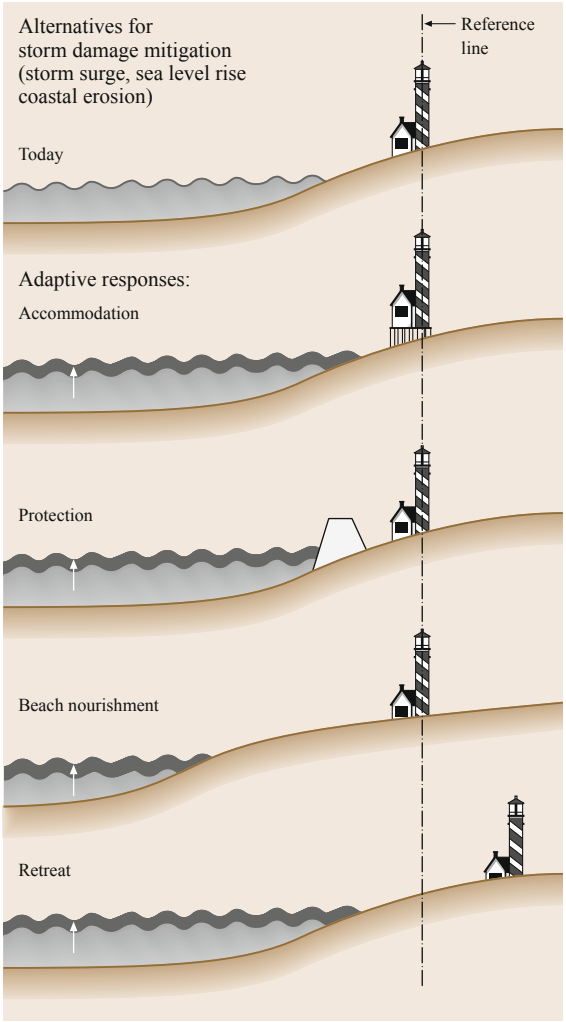


Fig. 30.1 Alternatives for shore protection (after [30.1, Fig. V-3-2])

2. Coastal tsunami events. These excessive hydrodynamic forces may also cause
3. Coastal erosion.

In this chapter, the term *shore protection* refers to the mitigation of the damage hazard consequences to people, property, infrastructure, and transportation systems from these types of excessive hydrodynamic forces and resulting coastal erosion. These forces are much different for open coasts than for limited fetch, bays, and estuaries so that the shore protection alternative selected for design is always site specific.

A discussion of the needed design criteria is presented in Sect. 30.1. Design practice has evolved from deterministic to probabilistic so that water wave spectra and statistics are now required in design. A coastal engineer is a member of an interdisciplinary team and always limited by the five major design constraints (science/technology, economics, the environment, institutional-political-social, and aesthetics). The key to a long term, sustainable solution is to first clearly define the problem and the purpose of the design solution for the general public (or client) responsible for paying the life-cycle costs for initial construction and long-term maintenance. Practice of the coastal engineering profession is much more than modeling and applying the design formulas in the next sections of this chapter.

In Sect. 30.2, the types, functional design, and structural design of coastal armoring structures are briefly summarized. Then, Sect. 30.3 addresses the types, functional and structural design of shoreline stabilization structures. There exists a vast amount of literature on these topics. The main references used herein are the US Army, Corps of Engineers, Coastal Engineering Manual [30.1], the Wave Overtopping Manual [30.2] funded by the UK, Germany, and the Netherlands, the Rock Manual [30.3] funded by CIRIA (Construction Industry Research and Information Association)/CUR, the textbook by *Goda* [30.4] that summarizes the Japanese literature, the results of the PROVERBS project for vertical breakwater design [30.5], and results of PIANC Working Groups [30.6, 7]. We can only present a very limited overview and summary of the literature. Section 30.4.1 lists the latest website links to some references at the time of publication of this book. Design practice details are quite extensive so that the responsible engineer must consult these references for the details. Space did not permit the discussion of dikes (levees), flood walls, and moving gate structures in tidal estuaries. The information presented on wave run-up, overtopping, and wave forces on vertical walls is applicable for these structures.

In the future, the potential for the accelerated, relative sea level rise (RSLR) rates (including local rates of subsidence) will increase the risk for more damage and reduce the resilience of existing shore protection systems. Section 30.4.2 concludes with some suggestions for the *best* alternatives for man's sustainable living at the coast. The design of hard structures for shore protection continues to evolve over time. As sea levels rise more quickly in the future, the need for more basic research and more knowledgeable coastal engineers will also increase.

30.1 Design Criteria, Philosophy, and Constraints

30.1.1 Water Levels

Water depth is arguably the most important independent variable in coastal structure design. By definition, the structure is at or near the land–water interface (the coast) where shallow water depths limit the wave height due to wave breaking. Low pressure storms and hurricanes with strong winds create storm surge that can significantly increase the local water depths during the storm event. Larger storm waves result before breaking depths are again attained and result in higher loads on the structure. Figure 30.2 displays the water elevations (above the National Geodetic Vertical Datum, 1929) versus the recurrence interval, T_r and the exceedance probability, $P_>$ for that elevation within the Chesapeake Bay (data from the National Oceanic and Atmospheric Administration (NOAA)/National Ocean Service (NOS) Tide Gauge at Sewells Point, Norfolk, VA). The elevation of Hurricane Isabel (Sep 18, 2003) was 2.38 m (7.82 ft) and the second highest water level recorded since records began in 1927. The mean tidal range is 0.82 m (2.7 ft) at this location. Clearly, the water depths at the structure used in design must consider the probabilistic variation of the water elevations.

Storm surge is the major difference between the measured and predicted, gravitational tide. If Hurricane Isabel (Sep 18, 2003) had occurred 10 days later during spring tide conditions, the water surface elevation measured at the Sewells Point (Norfolk, VA) tide gauge would have been 0.26 m (0.85 ft) higher and the highest ever recorded since 1927. There are several methods that have been historically applied for a flood frequency analysis. They are:

1. Estimates based only on historical storm data
2. The empirical simulation technique (EST)
3. The joint probability method (JPM), and
4. The Monte Carlo simulation method.

The length of record (years) for the historical storm data is generally short for many tide gages stations in the world but the total number of surge events can be extended. The storm events may be caused by both tropical and extra-tropical storms. See *Coles* [30.8] for a practical reference on statistical modeling of extreme flood events.

The EST has been developed *Scheffner* et al. [30.9] to generate a synthetic sequence of storms. The historical storm record is extended by assuming that each storm could have occurred at various combinations of the tidal phase and tidal range as in the example above for Hurricane Isabel. See *Scheffner* et al. [30.10] for complete details. The EST method can be applied for both tropical and extra-tropical storms.

The JPM is primarily for tropical storms (hurricane) since the joint, wind variable probabilities within the hurricane are reasonably well know that produce the storm surge along the coast. A modified JPM-OS method has been developed [30.12]. It has also been applied for extra-tropical storm events.

The Monte Carlo method [30.13] is similar to the JPM and relies on estimates of the probability distributions established for the parameters needed to characterize a storm. But a random number generator is employed to select values for each parameter. Each storm selected by this Monte Carlo procedure is equivalent to simulating a long period of history and many thousands of storms can be generated with high speed computers to improve the extrapolations to extreme storm events.

For hurricanes, [30.14, 15] and [30.12] recommend the JPM method.

Coastal sediment transport processes are discussed in Chap. 29 (erosion and accretion) and may also alter the bathymetry to modify the water depths and wave conditions near the coast and at the structure. Numerical modeling of bathymetric change near coastal structures is now possible for consideration in design.

Some level of the recurrence interval is often still selected in the deterministic design method. The risk-based approach employs the entire exceedance proba-

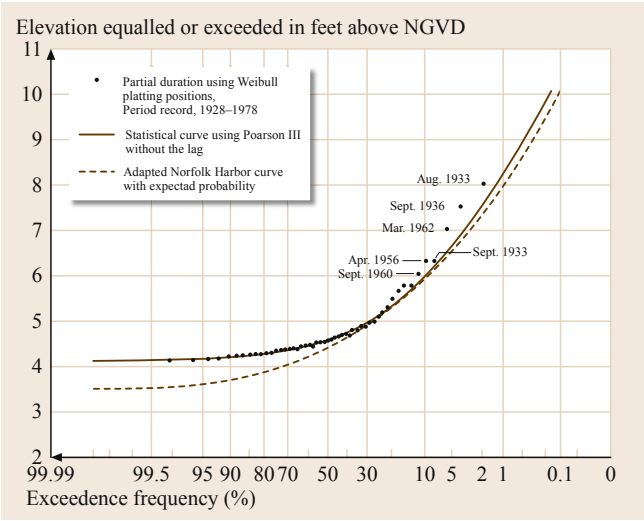


Fig. 30.2 Exceedance probability curve for water levels, Norfolk, Virginia, Tide Gauge at Sewells Point for 61 years data set (after [30.11])

bility curve and is the modern approach as discussed below.

Finally, the wave conditions at the structure can produce wave setup, $\bar{\eta}$ which is the time averaged increase in water level due to the presence of waves at the coast. Local beach slope is important. *Dean* [30.16] suggests using $\bar{\eta}/H_s = 0.191 \pm 0.100$ as a useful guide for the maximum setup at the coast. See *Dean* [30.16] for a detailed review of the wave setup. However, the state-of-practice is to use a cross-shore wave transformation model to calculate setup because the water level in the surf zone is a complex nonlinear process that depends on the details of the near-shore morphology and wave breaking. Empirical relations are not as accurate.

30.1.2 Wave Conditions

The mechanics of ocean waves and the modeling of coastal waves and hydrodynamics (wave-induced currents and tidal currents) are discussed in Chaps. 4 and 27, respectively. The real ocean surface variation is irregular in space and time so that some statistical, characteristic wave height and wave period must be employed for design. The most common choice is the significant wave height, H_s, which is the average of the highest one-third of the wave heights in a 17–20 min record of water surface variation as determined by the zero down-crossing analysis method in the time domain. The down-crossing method correctly captures the true wave height (vertical distance taken from trough in front and tail in rear) and should be used in the zero-crossing analysis method. The mean time period, T_m, is then used for the representative wave period.

An alternative, common choice is the result of a spectral analysis of the water surface record in the frequency domain to determine the spectral, significant wave height, H_{mo}, and the spectral peak wave period, T_p (see Chap. 4 for the definition of H_{mo}). A third wave period definition, the spectral wave period (or negative, first-moment wave period), $T_{m-1.0}$ ($= m_{-1}/m_o$) has been found useful because it gives more weight to the longer periods in the spectrum and is independent of the type of spectrum. It is commonly employed for wave run-up and overtopping calculations on coastal structures and is discussed in full detail below. The wave spectrum may contain two peaks indicating waves from different wave directions that produce both sea and swell conditions needed in structural design. The wave direction and storm duration are additional variables required for both functional and structural design.

The risk-based design approach also requires the analysis of many years of wave data or wave information. Figure 30.3 displays the Weibull distribution,

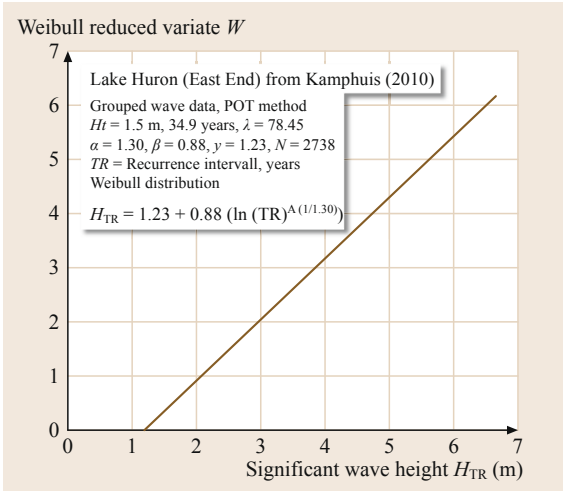


Fig. 30.3 Weibull distribution for extreme, significant wave heights for Lake Huron, Eastern Shore for 34.9 years (wind information after [30.17, Chap. 4])

reduced variate, W curve for extreme, significant wave heights, H_s as determined by the peak-over-threshold (POT) method for one site at the far eastern end of Lake Huron, Canada. The POT method separates related wave heights into independent *storm* events and the *reduced variate* is simply a coordinate transformation of the exceedance probability of storms to produce a straight line for extrapolation to extreme, wave heights ([30.17, Chap. 4], for a thorough explanation of how to apply the POT method with worked examples). The extreme wave heights for the very high recurrence interval conditions may never be reached if the water depths limit the wave heights due to wave breaking. The symbol $\hat{H}_S^T$ (hat for central estimate, subscript s, and super script T) will be employed for the central estimate of the significant wave height, H_s which on average is exceeded every T years [30.1, 2011, Part VI, Sep. 28, 2011].

Finally, the full wave climate (heights, periods, and directions) as experienced over the average, annual, seasonal cycle is needed for the functional design of shoreline stabilization structures (near-shore breakwaters, groin fields, etc.). In the United States, the US Army Corps of Engineers, Wave Information Studies (WIS) provide the full wave climate for the Atlantic, Pacific, Gulf of Mexico, and Great Lakes coasts.

30.1.3 Surf Conditions on the Structure

Different types of wave breaking (spilling, collapsing, plunging, and surging) occurs on structure slopes, $\tan\alpha$. The wave steepness is the ratio of the wave height, H to wavelength, L, and is usually defined for the deepwater

wavelength, $L_o = (g/2\pi)T^2$ so that clearly the definition of wave steepness depends on what irregular wave characteristics are employed for H and T.

The surf similarity parameter, ξ, is the ratio of the structure slope to the square root of the wave steepness and employed in many of the empirically developed design formulas for the functional and structural design of coastal structures because the different types of wave breaking produces different results as explained further below.

30.1.4 Probabilistic Design

Three levels of probabilistic design can be considered. We herein apply the level I (lowest) method which simply employs partial safety factors related to a characteristic load, γ_H, and characteristic resistance, γ_Z for the stability equations utilized in design. The designer selects the appropriate probability of the failure level, P_f and the degree of uncertainty as determined by the coefficient of variation, σ' is defined as the standard deviation, σ divided by the mean. Commonly employed partial safety factors for implementing reliability in design are discussed in detail below. Complete details are found in the CEM [30.1, VI, Sect. 6].

30.1.5 Coastal Risk

Fragility curves are functions that describe the probability of failure that is dependent on the *load* (force) over the full range of the *loads* to which the coastal system might be exposed [30.18]. These authors discuss four methods (judgemental, empirical, analytical (models) and hybrid combinations) to develop fragility curves. How to define the *load* for coastal systems is not discussed. Herein, we define the *load* as the coastal storm.

The probability of failure is the convolution of the probability of exceedance for the hazard (coastal storm) and the probability of coastal system damage from the hazard [30.17]. Coastal risk is then simply the summation of the probability of failure times all the consequences (economic, structural, functional, loss of life, environmental, etc.), over the full range of coastal storms.

Resilience is the ability of a system to maintain and recover its structural and functional performance following a disturbance [30.19]. The *disturbance* is the short-term excess of forces (the load or coastal storm) acting on the coastal system components and processes that may impair the system function. These authors discuss three types of resilience (ecological, engineering, community) for coastal systems, focus on engineering resilience, and cite the Saffir–Simpson wind speed scale for hurricanes as an example of a coastal storm disturbance scale. If the level of disturbance (coastal storm) exceeds a critical level, both the level of performance impairment and the duration of performance impairment may exceed management objectives for the resilience of the coastal system.

Fragility curves are needed to quantify risk and hence resilience in coastal systems design. The intensity of the *load* or *disturbance*, the severity of the coastal storm must be quantified to develop fragility curves. Excess water levels (storm surge), wave conditions (height, period, direction), and storm duration all contribute to the intensity of a coastal storm. How to combine these three factors has long been a concern for coastal scientists and engineers.

30.1.6 A Coastal Storm Severity Index

The definition of a coastal storm may depend on wave height, or water level (site specific), and sometimes depends on the consequences of the storm event. For example, the Corps of Engineers, Field Research Facility (FRF) at Duck, NC, on the mid-Atlantic, east coast of the US uses the mean, significant wave height plus two standard deviations of the mean (as measured at an offshore, deepwater buoy) to define the start of a coastal storm (*Birkemier* [30.20], personal communication). However, flood levels in the Piazza San Marco in Venice, Italy, result from meteorological conditions in the Adriatic Sea that trigger a coastal *storm* event for the Italian government. The consequences are property damage, interruptions of tourism, etc. Therefore, a storm is defined in Venice by the *flood level*. There presently exists no known, universally accepted, hydrodynamic definition of a coastal storm. One potential method that combines waves, water levels, currents, and storm duration into one number is the coastal storm impulse (COSI) parameter [30.21].

30.1.7 Design Constraints

One good definition of engineering is *design under constraint*. Coastal engineering design is constrained by our scientific and engineering understanding of nature; by economics (costs); by concerns of environmental impact; by social, political, and institutional issues; and possibly by aesthetics. See the CEM [30.1, Part V, Chap. 3, Sect. 1c] for a detailed discussion of design constraints. This chapter only considers our scientific and engineering understanding of armoring and stabilization structures for coastal protection design.

30.2 Coastal Armoring Structures

30.2.1 Types and Purpose

There are many different types of seawalls and dikes, bulkheads, and revetments as illustrated in Fig. 30.4. Their functional purposes are different. The seawall mitigates damage to what lies behind from flooding and wave forces; the bulkhead prevents the soil behind from sliding into the sea; and the revetment protects the bank from erosion. They all have, however, the following in common:

1. They are located at the land/water interface.
2. They harden (armor) the shoreline to fix its location, and
3. They have some configuration and crest elevation that determines the height of the wave run up and the amount of wave overtopping for the range of water levels and wave conditions at their location on the coast.

30.2.2 Functional Design

The functional design of coastal armoring structures involves calculations of wave run-up, wave overtopping, wave transmission, and reflection. These technical factors together with economic, environmental, and other constraints all combine to determine the crest elevation of the structure. We only consider run-up and overtopping due to space limitations.

Wave Run-Up

The wave run-up, R_u is defined as the vertical distance above the still water level that can be reached by the waves as illustrated in Fig. 30.5. Common practice is to consider the irregular wave heights following the Rayleigh distribution and define a run up level exceeded by i % of the waves as $R_{ui\%}$. One popular, dimensionless, formulation [30.23] is

$$\frac{R_{ui\%}}{(H_s)_{toe}} = (A\xi + C)\gamma_r\gamma_b\gamma_h\gamma_\beta \,, \tag{30.1}$$

where:

- H_s = the significant wave height at the toe of the structure
- ξ = the surf similarity parameter
- A, C = empirical coefficients from laboratory and field data depending on the value of i, and
- γ = reduction factors (subscripts) for roughness (r), fronting berm (b), shallow water depths (h), and wave approach angle (β).

The key independent variable is the surf similarity parameter, ξ, and three definitions have been employed in the literature depending on which irregular wave period, T is specified to define the deep water wavelength L_o and the wave steepness $S_o = H_s/L_o$. The mean wave period T_m, the spectral peak wave period T_p; and the spectral wave period $T_{m-1.0}$ all have been employed by various researchers. And, since the structure slope and wave steepness trigger different types of breaking waves (spilling, plunging, collapsing, and surging, see *Battjes* [30.23]), the coefficients A and C vary with ξ for a given structure. *Hunt* [30.24] was the first to introduce (30.1).

For the design of crest elevations with nonovertopping wave conditions, the 2% value for run-up has long been the choice. See the EurOtop Manual [30.2] for the history of this choice for design dating back to 1932 in the Netherlands.

Impermeable embankments (dikes and levees) with smooth, rough, and rock cover were studied by *de Waal* and *van der Meer* [30.25]. Using the peak wave period T_p to define ξ_{op}, the run-up at the $i = 2\%$ exceedance level, $R_{u2\%}$ depended on the breaker type. For $\xi_{op} \le 2$, $A = 1.5$ and $C = 0$. When $2 < \xi_{op} < 3$–4, $A = 0$ and $C = 3.0$. The coefficient of variation $\sigma' = 0.085$. When the impermeable surface is roughened, γ_r was in the 0.5–0.6 range. See Table VI-5-3 in the CEM [30.1] for other reduction factors in (30.1).

Rock-armored slopes with impermeable and permeable cores (rubble-mound breakwaters) were studied in great detail by *van der Meer* [30.26]. The permeable structure resulted in a slightly more complicated form for the dimensionless, wave run-up formulation

$$\frac{R_{ui\%}}{(H_s)_{toe}} = A\xi + B\xi^C + D \,. \tag{30.2}$$

And, the mean wave period T_m was employed to calculate the surf similarity parameter ξ_{om} for use in correlating the experimental data. For the impermeable core, the permeability parameter $P = 0.1$ and the permeable structure $P = 0.5$ (see Fig. 30.12 in the next section for a graphical depiction of P).

At the 2% exceedance level for the impermeable core case $A = 0.96$, and $B = D = 0$, when $\xi_{om} \le 1.5$; and $A = 0$, $B = 1.17$, with $C = 0.46$ and $D = 0$ when $1.5 < \xi_{om} \le 3.1$. For the permeable core, these coefficients for A, B, and C are the same except now $D = 1.97$ for all $\xi_{om} \ge 3.1$. The CEM [30.1] also presents coefficients for A, B, C, and D for other values of the exceedance level i for use in (30.2). See Table VI-5-5 in the CEM [30.1].

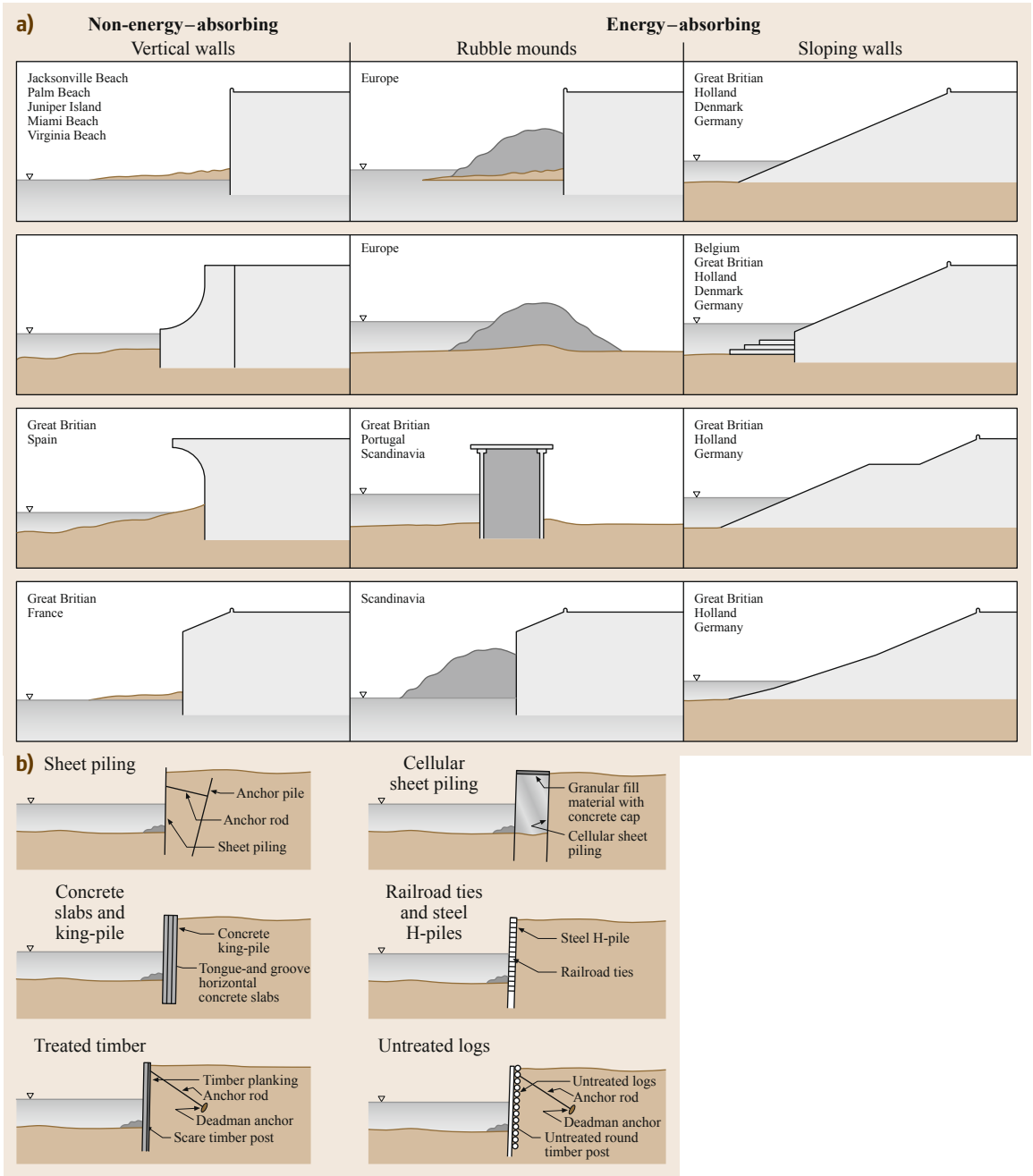


Fig. 30.4 (a) Typical seawall types (after [30.22]); (b) typical cantilever and tie-back bulkheads; (c) typical embankments with revetment armor

But all the above is for some level of the significant wave height H_s at the toe of the structure. A full probabilistic treatment for wave run-up on rock-armored slopes with impermeable and permeable cores using $\hat{H}_S^T$, the mean significant wave height at some recurrence interval can now be determined for a range of partial safety factors and levels of probability of failure P_f. See Table VI-6-17 as developed from the work of *de Waal* and *van der Meer* [30.25] in the CEM [30.1] for complete details.

Many other formulations are available. See the EurOtop Manual [30.2] for the calculation of wave run-

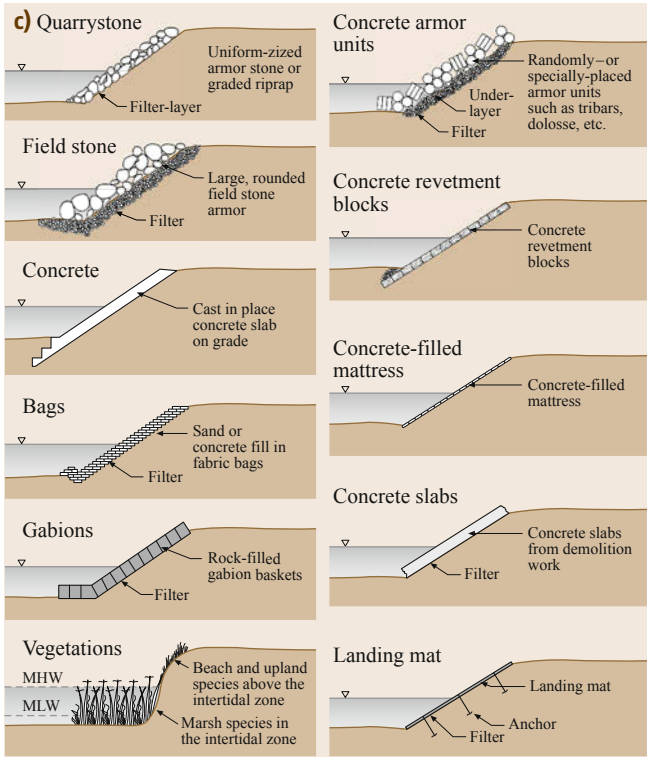


Fig. 30.4 (continued)

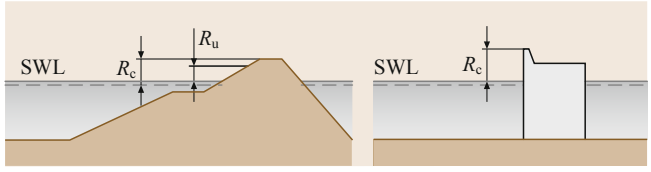


Fig. 30.5 Wave run-up R_u and wave freeboard R_c definitions

up heights under the range of experimental conditions employed to develop EurOtop. These authors have employed the spectral wave period $T_{m-1.0}$ to define $\xi_{m-1.0}$ in all their work to better incorporate the longer period waves in the empirical, wave run-up formulations, and coefficients.

The US Army, Corps of Engineers, has developed a hydrodynamic, numerical model to predict wave run up for coastal structures and beaches called CSHORE. The Federal Emergency Management Agency (FEMA) in the US sponsored a study [30.27], which concluded that EurOtop *is likely to have better skill than CSHORE* for structures that were within the range of experimental conditions (bathymetry and wave-water level). However, for complex near-shore bathymetry and complex structures with waves and water levels outside the range of EurOtop, the *CSHORE is likely to show better skill*. CSHORE also predicts cross-shore beach morphology change during storm events that can significantly impact wave run up.

Wave overtopping is related to wave run up and the crest elevation of the structure.

Wave Overtopping

Wave overtopping occurs when the highest wave run-up R_u exceeds the crest freeboard R_c as defined in Fig. 30.5. For almost all coastal structures illustrated in Fig. 30.4, the crest elevation will be overtopped by some waves during extreme storm events. The functional purpose of the structure and the design constraints discussed above will dictate the design crest elevation such that some wave overtopping will occur over the life of the structure. Individual waves overtopping a structure vary considerably in space and time, so that the average discharge rate may only be a small fraction of the peak value from a single wave. However, the *average*, allowable wave overtopping discharge rate q (m^3/s per meter length of structure) in most cases determines the design crest elevation of the structure.

Field studies of critical, tolerable overtopping limits have been performed by many investigators over the years. Figure 30.6 is a composite of US, Dutch, Italian, and Japanese researchers as complied in the CEM [30.1, Table VI-5-6]. The designer must first start with a target goal for the allowable value of q which may come from Fig. 30.6 or institutional criteria, local experience, or an educated guess of the client. For example, the US Government, Federal Emergency Management Agency (FEMA) publishes *Suggestions for Interpretation of Mean Wave Overtopping Rates* [30.28, Table D.2.8-6], for flood insurance risk zones behind coastal structures. For Zone AO (1 m depth) the order of acceptable q is 9.3–93 l/s per meter. Figure 30.6 can then be consulted for safety of structures, vehicles, and pedestrians. If unacceptable, the crest elevation of the structure can be increased to reduce the average overtopping rate by applying a suitable formula for calculation.

A review of the structural profiles in Fig. 30.4 reveals four basic types of geometrical patterns: straight slope, bermed slope, slope with crown wall, and vertical wall (with and without recurved top). As a result, many different wave overtopping formulations have appeared in the literature (see the CEM [30.1], Table VI-5-7 for eight popular formulae dating from *Owen* [30.29, 30] to *Pedersen* [30.31]). Space limitations permit only the case for straight and bermed impermeable slopes [30.32] to be presented here. Their formulation, however, includes the influence of surface roughness, shallow water, oblique, and short-crested waves so that it is quite useful. The surf similarity parameter ξ_{op} is employed.

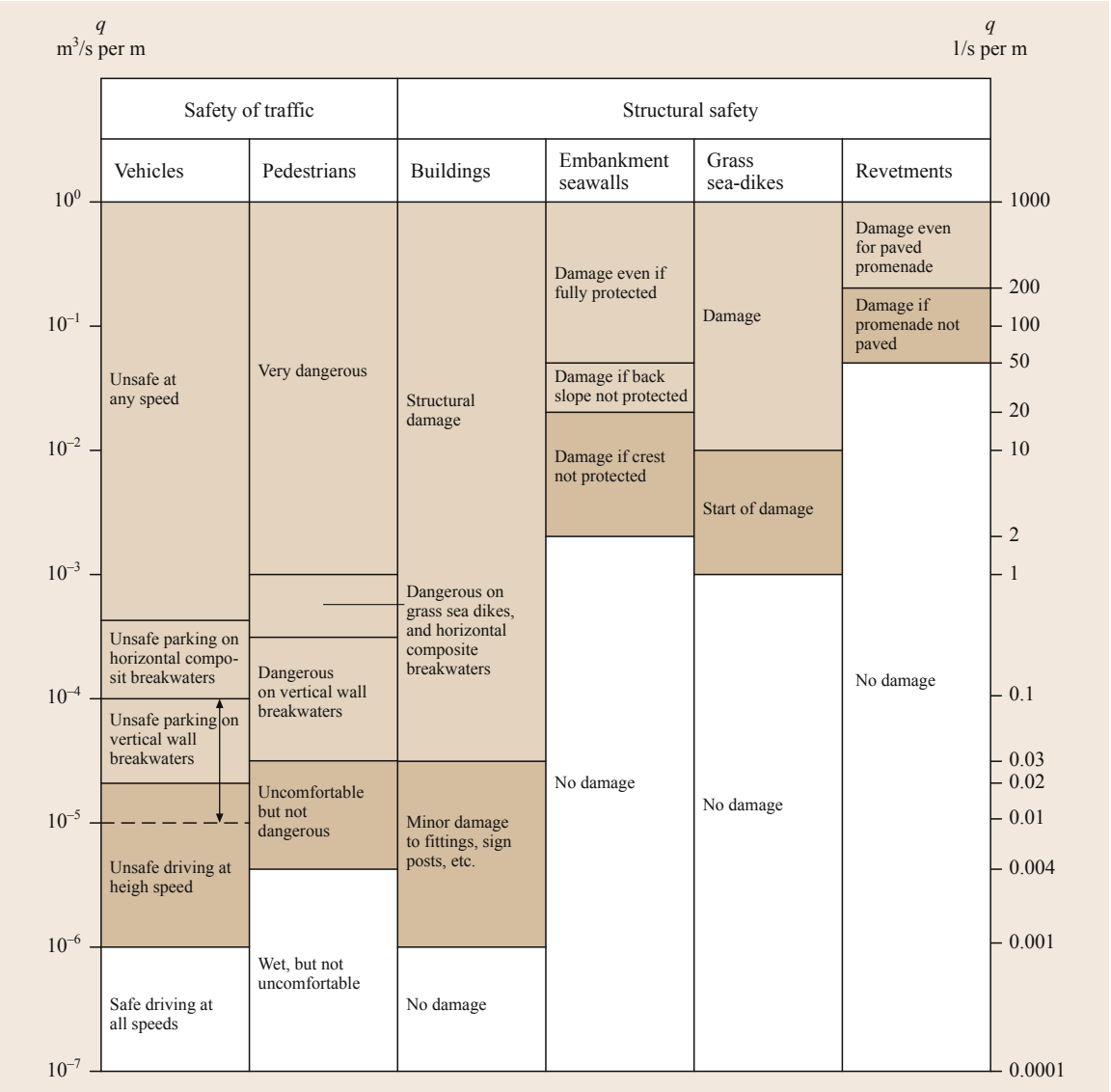


Fig. 30.6 Critical values of average wave overtopping discharge rates (after [30.1])

When $\xi_{op} < 2$

$$\frac{q}{\sqrt{gH_s^3}}\sqrt{\frac{s_{op}}{\tan\alpha}} = 0.06\exp\left(-5.2\frac{R_c}{H_s}\sqrt{\frac{s_{op}}{\tan\alpha}}\frac{1}{\gamma_r\gamma_b\gamma_h\gamma_\beta}\right) . \tag{30.3}$$

The application range is

$$0.3 < \frac{R_c}{H_s}\sqrt{\frac{s_{op}}{\tan\alpha}}\frac{1}{\gamma_r\gamma_b\gamma_h\gamma_\beta} < 2 . \tag{30.4}$$

The uncertainty $\sigma' = 0.55$ on the factor 5.2.

When $\xi_{op} \geq 2$

$$\frac{q}{\sqrt{gH_s^3}} = 0.2\exp\left(-2.6\frac{R_c}{H_s}\frac{1}{\gamma_r\gamma_b\gamma_h\gamma_\beta}\right) . \tag{30.5}$$

The uncertainty σ' equals 0.35 on the factor 2.6.

All the symbol definitions are the same as that for wave run-up as are the reduction factors for roughness, water depth, and wave obliqueness. See Table VI-5-11 in the CEM [30.1] for adjustments for short- and long-crested waves. The minimum value of any combination of γ-factors is 0.5. Wave overtopping formulas for ver-

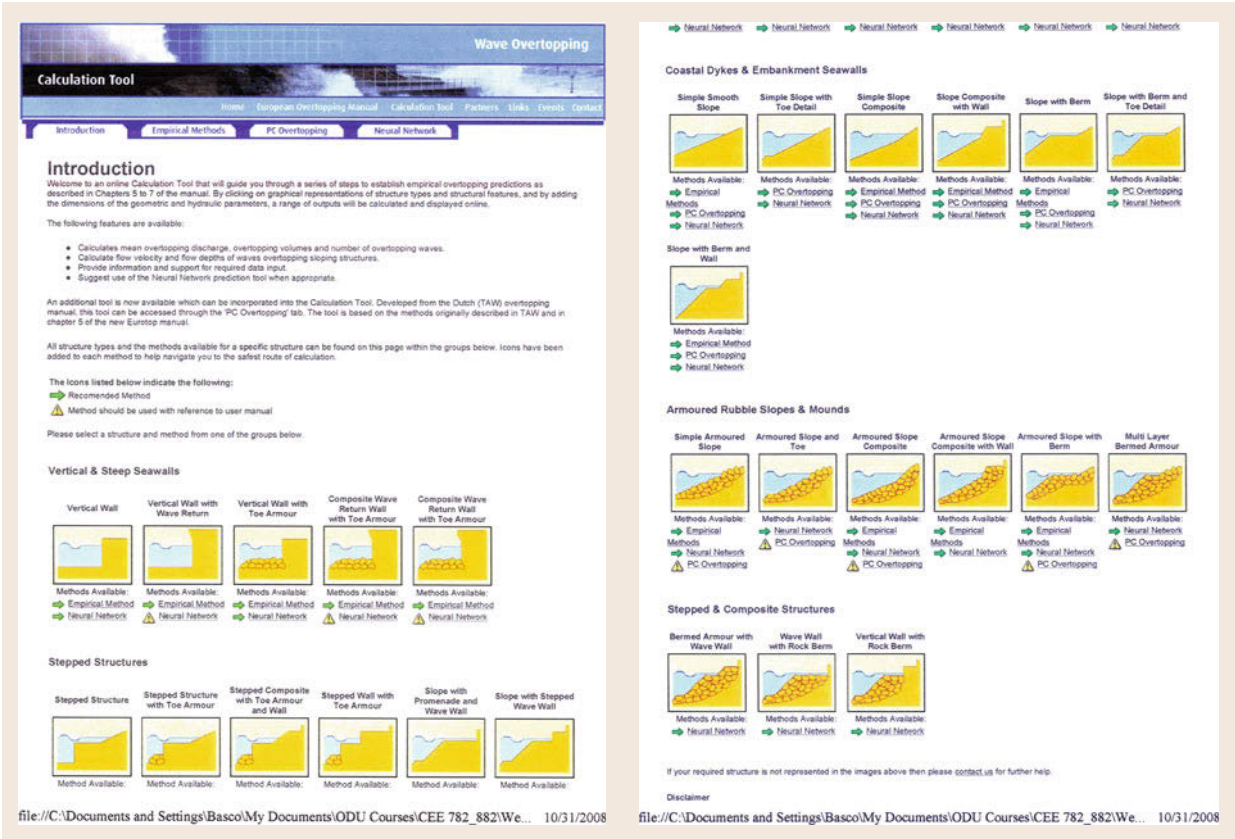


Fig. 30.7 Structure types as treated in the online, EurOtop calculation tool (after [30.2])

tical walls are not well covered in the CEM [30.1] and recent research efforts in northern Europe have provided a more extensive treatment of wave overtopping in the EurOtop Manual.

In August 2007, a research team from the United Kingdom, Germany, and the Netherlands, published the *Wave Overtopping of Sea Defences and Related Structures: Assessment Manual* commonly referred to as the EurOtop Manual. The publication is No. 73 in the Archive for Research and Technology on the North Sea and Baltic Coast [30.2]. It replaces previous manuals in these three countries and provides significant new information obtained from many European Union (EU) and national projects (OPTICREST, PROVERBS, SHADOW, VOWS, BIG-VOWS, ComCoast, etc.) and continues in the CLASH project. The acronym *CLASH* stands for

> *crest level assessment of coastal structures by full scale monitoring, neural network prediction and hazard analysis on permissible wave overtopping*

and is a fifth framework program in the European Community. The key final product of this major international effort is the *EurOtop calculation tool* which is an extremely useful, online, and freely available tool for use by the world's coastal engineering community [30.33].

The structure types analyzed are: vertical and steep seawalls; stepped structures; coastal dikes and embankment revetments; armored rubble slopes and mounds; and stepped and composite structures as displayed in Fig. 30.7. Three calculation methods are available (empirical, overtopping formulas, and neural networks) but not all for each structure type.

One advantage of the *EurOtop calculation tool* is the ease of application for a large range of structure types and knowledge that the latest international, field research efforts have been incorporated for the results. One disadvantage is that each water level and wave condition of interest must be entered manually. Documentation of the formulas employed in the calculation tool is discussed in the assessment manual available at www.overtopping-manual.com. Some formulas are the same as in the CEM [30.1] such as those by *van der Meer* and *Janssen* [30.32] for impermeable slopes (30.3).

30.2.3 Structural Design

The structural design of coastal armoring structures involves the water levels and wave forces (and possibly ice loads) that may produce movements (sliding, rotation, overturning, etc.) of complete sections of the structure or loss of stability for individual units of rock (rubble)- and concrete-armor layers of the structure. Structural stability is the goal under design storm events. Damage is probable under more extreme storm conditions for the given design constraints (economic, environmental, etc.) so that maintenance costs for repair must always be considered in the initial design phase of the project. We herein consider, monolithic, caisson-type, vertical-walled structures and rock (rubble) revetments. Water level and wave loads on vertical *seawalls* are similar as considered herein for caissons. The stability of rubble-mound (rock) breakwaters is also included here with further discussion in Sect. 30.3. The origin of the use of the word *rubble* for rock structures is unknown.

The choice between use of a vertical-walled, caisson-type structure or a rubble-mound breakwater depends on many factors with water depth being one of the most important. In deep water, the cross-sectional area (hence cost) of the caisson-type is lowest where as in shallow water the material cost often favors the rubble-mound type. No general design rules apply so that both types should be considered in the feasibility study.

Oumeraci et al. [30.5] cite the following reasons for research on monolithic, caisson-type structures: catastrophic failures in Europe, need for breakwaters at greater depths, need for environmentally friendly structures, need for multipurpose structures, and the potential for application in sea-wall designs. The result was the PROVERBS project as discussed below.

Vertical-Walled, Caisson-Type Structures

With no waves present, the water pressure distribution on a vertical wall is hydrostatic being zero at the surface and $-\rho z$ at any distance z below the surface ($+z$ is positive above the surface). The mass density of the fluid is ρ. The horizontal, water force per unit width f_h on the vertical wall is then $\frac{1}{2}\rho g h^2$, where h is the fluid water depth at the wall. Storms increase the water depth and bring waves that are both reflected by and break on the wall to greatly increase the horizontal forces on the wall. A long history of research efforts throughout the world has produced many formulas over the years for the calculation of wave pressures and forces on vertical walls.

Pioneering efforts of Japanese coastal engineers have led the way to develop practical formulas for the wave pressures and forces on caisson-type structures under both nonbreaking and breaking wave conditions. We focus on these herein.

Figure 30.8 (from the CEM [30.1]) presents the key pressure distribution and other variables for irregular waves as formulated by *Goda* [30.34] and *Tanimoto* [30.35] and herein called the *Goda* formula

$$\eta^* = 0.75(1+\cos\beta)\lambda_1 H_{\text{design}} \,, \tag{30.6}$$

$$p_1 = 0.5(1+\cos\beta)(\lambda_1\alpha_1 + \lambda_2\alpha_* \cos^2\beta) \times \rho_w g H_{\text{design}} \,, \tag{30.7}$$

$$p_2 = \begin{cases} \left(1-\dfrac{h_c}{\eta^*}\right) p_1 & \text{for } \eta^* > h_c \\ 0 & \text{for } \eta^* \le h_c \end{cases} \,, \tag{30.8}$$

$$p_3 = \alpha_3 p_1 \,, \tag{30.9}$$

$$p_u = 0.5(1+\cos\beta)\lambda_3\alpha_1\alpha_3\rho_w g H_{\text{design}} \,, \tag{30.10}$$

where

$$\beta = \text{angle of incidence of waves} \,,$$

$$\alpha_* = \alpha_2 \,,$$

$$\alpha_1 = 0.6 + 0.5\left[\frac{4\pi h_s/L}{\sinh(4\pi h_s/L)}\right]^2 \,, \tag{30.11}$$

$$\alpha_2 = \text{the smallest of } \frac{h_b - d}{3h_b} \times \left(\frac{H_{\text{design}}}{d}\right)^2 \text{ and } \frac{2d}{H_{\text{design}}} \,, \tag{30.12}$$

$$\alpha_3 = 1 - \frac{h_w - h_c}{h_s}\left[1 - \frac{1}{\cosh(2\pi h_s/L)}\right] \,, \tag{30.13}$$

- L = wavelength at water depth h_b corresponding to that of the significant wave $T_s \simeq 1.1T_m$, where T_m is the average period.

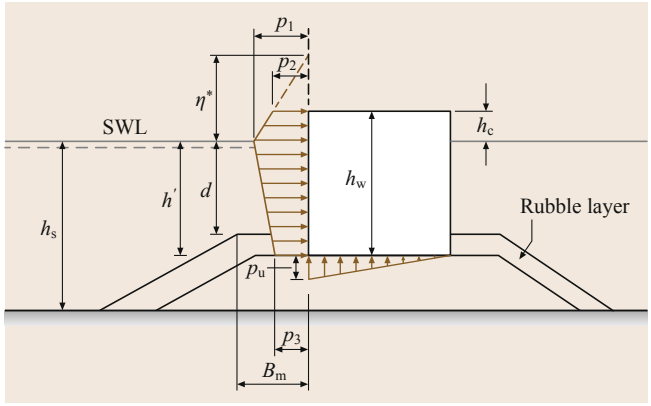


Fig. 30.8 Definitions of terms in Goda formula for irregular waves (after [30.1])

- h_b = water depth at a distance of 5 H_s seaward of the breakwater front wall.
- λ_1, λ_2, and λ_3 are the modification factors depending on the structure type. For conventional vertical wall structures, $\lambda_1 = \lambda_2 = \lambda_3 = 1$. Values for other structure types are given in related tables.

The key pressure is p_1 at the design still wave level (SWL) and its calculation depends on many factors in the formulation listed above including the design wave height H_{design}. What to use for the design wave height in the Goda formula to calculate p_1 is subject to debate as discussed further below. The water level on the leeward side of the structure is the same so that the hydrostatic forces on the structure are balanced out and not shown.

The wave pressure and uplift pressure distributions produce forces and moments per unit length of structure that are readily calculated by the following formulas (from the CEM [30.1, Table VI-5-55]). The gravity force F_G is simply the difference between the caisson weight in air W and the buoyant force B

$$F_H = U_{FH}\left[\frac{1}{2}(p_1+p_2)h_c + \frac{1}{2}(p_1+p_3)h'\right], \tag{30.14}$$

$$F_U = U_{FU}\frac{1}{2}p_u B, \tag{30.15}$$

$$F_G = \rho_c g B h_w - \rho_w g B h', \tag{30.16}$$

where:

- ρ_c = mass density of the structure
- ρ_w = mass density of the water
- U_{FH} = stochastic variable signifying the bias and the uncertainty related to the horizontal force in the Goda formula

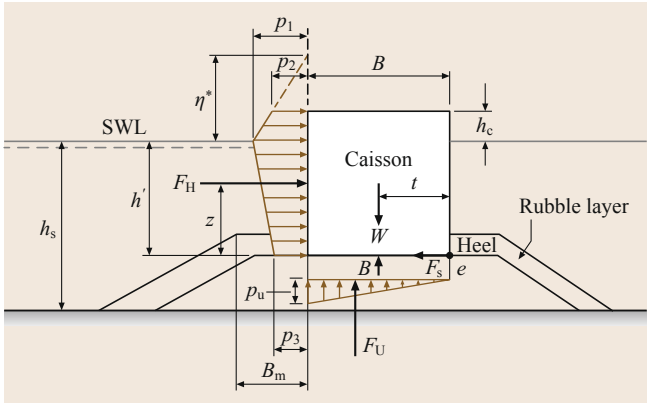


Fig. 30.9 Forces on the Caisson structure that cause sliding, overturning and bearing pressures

Table 30.1 Correction factors for moments in Goda formulae

Stochastic variable X_i	Mean value μ_{X_i}
U_{FH}	0.90
U_{FU}	0.77
U_{MH}	0.81
U_{MU}	0.72

- U_{FU} = stochastic variable signifying the bias and the uncertainty related to the uplift force in the Goda formula
- h' = submerged height of the wall from the toe to the still water level
- B = structure width.

The corresponding moments at the heel of the structure are

$$M_H = U_{MH}\left[\frac{1}{6}(2p_1+p_3)h'^2 + \frac{1}{2}(p_1+p_2)h'h_c + \frac{1}{6}(p_1+2p_2)h_c^2\right], \tag{30.17}$$

$$M_U = U_{MU}\frac{1}{3}p_u B^2, \tag{30.18}$$

$$M_G = \frac{1}{2}B^2 g(\rho_c h_w - \rho_w h'). \tag{30.19}$$

Uncertainty and bias in the Goda formulae were reanalyzed by *van der Meer* et al. [30.36] and resulted in correction factors U for the horizontal and uplift forces and moments (per unit length of structure) in these equations. These correction factors are all less than unity and are related to the H_{design} value employed to calculate p_1.

The correction factors for moments are defined as follows and their values are indicated in Table 30.1:

- U_{MH}: Stochastic variable signifying the bias and the uncertainty related to the horizontal moment in the Goda formula.
- U_{MU}: Stochastic variable signifying the bias and the uncertainty related to the uplift moment in the Goda formula.

See the CEM [30.1, Table VI-5-55] for standard deviations of the mean and the coefficients of variation for the uncertainty factors U listed above in the Goda formula.

Three stability tests are normally conducted (sliding, overturning, and heel bearing pressure) to determine the design caisson width B as illustrated in Fig. 30.9. The gravity force F_G is simply the difference between the caisson weight in air W and the buoyant force B and when multiplied by a sliding friction coeffi-

cient μ produces the sliding friction resistance force F_S that opposes the horizontal wave force F_H. The wider the caisson width B the greater the sliding friction resistance force to produce a stable structure against sliding failure.

The uplift force F_U is the result of an assumed triangular uplift pressure distribution with p_u maximum on the front edge. Note that p_u is not equivalent to p_3. The uplift moment arm is $2/3B$ and the horizontal wave force moment arm is the distance z for clockwise moments about the heel of the caisson at location e. The gravity force $(W-B)$ times its moment arm, t produces the counter-clockwise moment about the heel location e to prevent overturning of the caisson. Here again, the wider the caisson width B, the greater the resistance to the overturning moments to produce a stable structure against overturning failure.

The third *failure* test normally considered is the soil-bearing pressure at the heel of the caisson structure. Space limitations do not permit a discussion here. See the CEM [30.1].

Global safety factors can be applied for the sliding, overturning, and bearing pressure formulations but they are not recommended because they are not related to the probability of failure of the structure. Modern, probabilistic design methods now employ partial safety factors. For sliding failure, conditions for deep and shallow water, and design with and without model tests are presented for both the load partial safety factor γ_H for wave height and the resistance partial safety factor γ_Z for the friction coefficient. See Table VI-6-24 in the CEM [30.1]. Note that two levels of knowledge of the wave height coefficient of variability ρ' (0.05 for good knowledge and 0.2 for relatively poor knowledge) are also presented for use to estimate the appropriate five levels (0.01, 0.05, 0.10, 0.20, 0.40) of probability of failure P_f. Similar summary tables of partial safety factors for overturning (Table VI-6-25) and for bearing pressure foundation failure for both sand subsoil (Table VI-6-22) and clay subsoil (Table VI-6-23) are found in [30.1].

In all the tables of partial safety factors in [30.1], the design wave height, H_{design}, is taken as the mean (carrot symbol) of the long term, statistical probability distribution for the significant wave height (subscript s) at the recurrence interval of choice (superscript T) noted as $\hat{H}_S^T$. *Goda* [30.4] recommends using H_{design} as 1.8 H_s if seaward of the surf zone and *the highest of the breaking waves* when within the surf zone. The use of the 1.8 factor corresponds to 0.15% exceedance value in the Rayleigh distribution included in 1/250 total number of waves. Use of $H_{design} = 1.8H_s$ in the specification of the design wave height is equivalent to using a higher recurrence interval wave height. For example, in Fig. 30.3, if the design wave height was 4 m ($T = 50$ yrs), then 1.8 H_s is about 7 m at a recurrence interval, T of 5000 yr. Tables of partial safety factors [30.1, Tables VI-6-22, 23, 24, 25], all specify using $\hat{H}_S^T$.

The original Goda method is normally for design of relatively long structures where the structure length in effect spreads the wave loads along the length of the structure. This horizontal spreading of the wave loads calculated per unit length of structure is an added factor of safety in applying the original Goda formulations for the pressure p_1 at the water surface as shown in Fig. 30.8 for both nonbreaking and breaking waves [30.37]

However, under some wave-breaking conditions, very high wave impact (aka, impulsive, slamming, hammering) loads of short duration occur that produce greater wave loads than estimated by the original [30.34] formula. *Takahashi* [30.38] has modified the pressure p_1 at the still water level (SWL) to take this slamming force into account. [30.1] advises that these relatively small-scale model results do not correctly include air pockets trapped between the wall and the slamming wave. The use of the Minikin method as found in [30.39] is recommended [30.1, p. VI-5-162] even though it is well know that the Minikin method gives very high results.

The need for improved knowledge for the wave impact loads, geotechnical aspects, and structural analysis aspects of vertical breakwater design resulted in the PRObabilistic Design Tools for VErtical BreakwaterS (PROVERBS) project in the European Union (1994–1998) as summarized below.

The results of the PROVERBS project are summarized by *Oumeraci* et al. [30.5]. Only the wave impacts are presented here. The interested reader should consult this reference for geotechnical aspects, structural aspects, and probabilistic design tools that are applicable for vertical, monolithic structure design.

Three major advances have resulted from the PROVERBS project, namely:

1. A parametric decision map has been developed (Fig. 30.10) to provide a calculation procedure to identify the breaker-type in front of the vertical structure.
2. The breaker-type depends on the structure (four structure types are identified and studied) as well as the wave parameters approaching the structure.
3. A 12-step design procedure has been developed that includes new formulas to calculate the wave impact

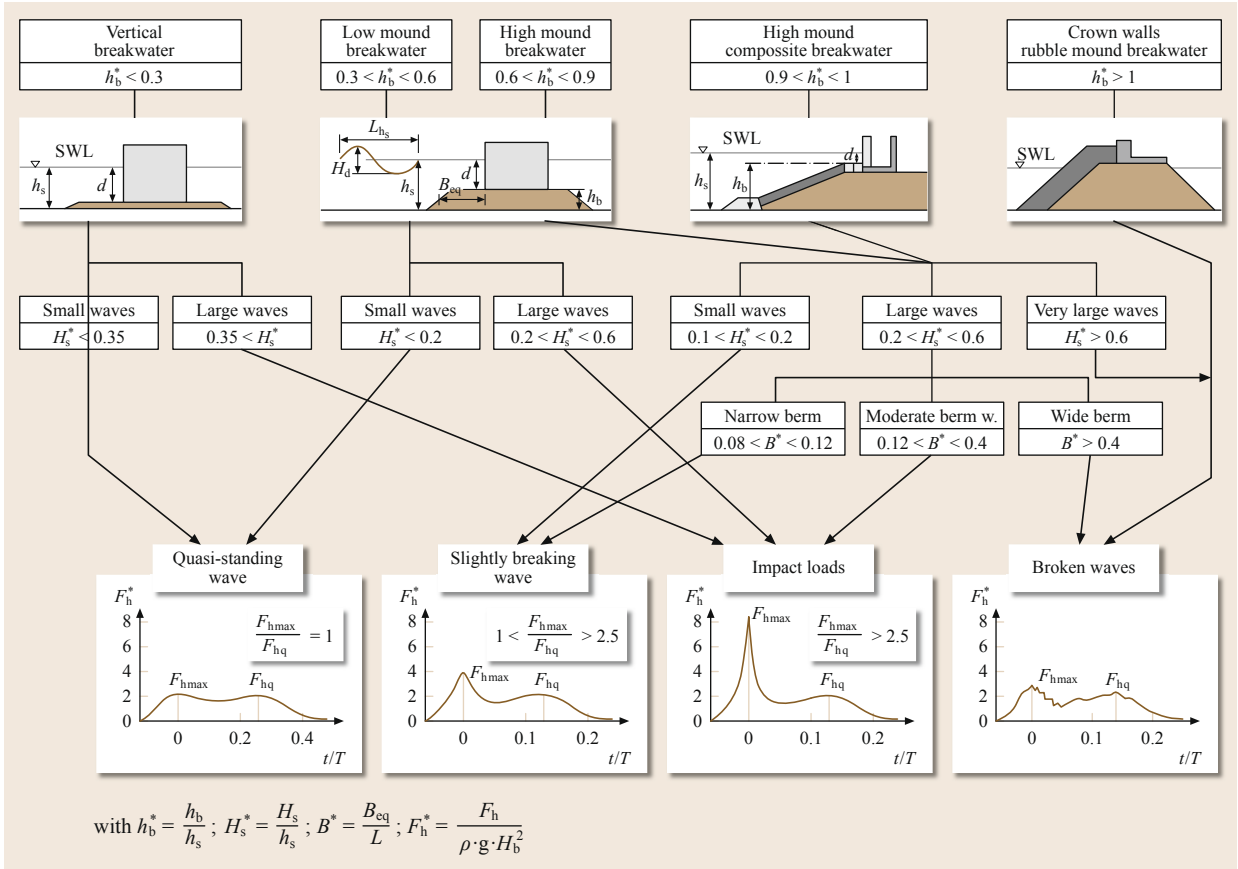


Fig. 30.10 The parametric decision map for four structure-types and four classes of irregular wave types that result in different wave loadings on the structures (after [30.40])

forces on the four types of vertical-walled structures considered.

As shown in Fig. 30.10, the four structure types are: a *vertical* breakwater, a low- or high-mound breakwater (Fig. 30.9), a composite breakwater, and a rubble-mound breakwater with crown wall. They are identified by the relative berm height, h_b^*, and the relative berm width, B^*, which can trigger various kinds of wave breakings on the structure. The key wave parameter is the relative (depth-limited), significant wave height H_s^* and results in four different breaker types: quasi-standing waves, slightly breaking waves, impacting waves, and broken waves. Small, large, and very large waves can produce all four breaking wave types.

The bottom of Fig. 30.10 shows the relative, horizontal force F_h^* (t/T) that is produced by the four kinds of wave breaking over the wave period (t/T). Clearly, the wave impact case produces the largest maximum force on the structure.

The 12 steps in the recommended design procedure are as follows:

- *Step 1*: Identify the main geometric and wave parameters. Various possible water levels should be considered since they produce different values of the relative berm height h_b^* and relative wave height H_s^*.
- *Steps 2–4*: Estimate wave force and pressure distributions. Use *Goda* [30.34], outlined above to provide the first estimate of structure size. Use $H_{max} = 1.8\, H_s$.
- *Step 5*: Identify the loading case using the parameter map (Fig. 30.10). The key decision parameters are h_b^*, H_s^*, and B^*. For example, if there is a low-mound breakwater $(0.3 < h_b^* < 0,6)$ with large waves $(0.2 < H_s^* < 0.6)$ then the impact load formulas must be applied to check the design dimensions.
- *Step 6*: Initial calculation of impact force. If the results from step 5 indicate a *slightly breaking* or *im-*

pact load, then estimate the horizontal force, $F_h = 15\rho w g d^2 (H_s/d)^{3.134}$ [30.41]. This can be used as a reference value.

- *Step 7*: Estimate the percentage of breaking waves ($P_{i\%}$) that produce impacts. Use results to decide and confirm the loading case type from the following levels of Pi%:
  - Pi% < 2%: Little breaking. Wave loads are primarily nonbreaking.
  - 2% < Pi% < 10%: Breaking waves give impacts.
  - Pi% > 10%: Heavy breaking may give impacts or broken wave forces.

  Unfortunately, space does not permit the procedure and formulas needed to compute Pi% and the reader must see *Oumeraci* et al. [30.5, pp. 78–82].
- *Step 8*: Estimate impact force. This is a key result within the PROVERBS project that gives a new procedure based on the statistical distribution of forces [30.42]. The relative wave force F_h^* is found from

$$F_h^* = \frac{\alpha}{\gamma}\{1 - [-\ln P(F_h^*)]^{\gamma}\} + \beta \,. \qquad (30.20)$$

  The wave force on the structure F_h divided by $\rho g H_b^2$ gives the relative wave force F_h^*. And H_b is the *individual or maximum wave height at breaking* [30.5, p. 99] calculated taking into account local refraction, shoaling, and reflections from the structure that modify a version of the *Miche* breaking formula [30.43] adapted for irregular, significant wave heights by *Vrijling* and *Bruinsma* [30.44]. The calculation of H_b is first used in Step 7 to find Pi%. Again space does not permit us to explain all the details.

  This expression for F_h^* is a generalized extreme value (GEV) distribution [30.8] and dependent on three statistical parameters α, β, and γ with the key variable $P(F_h^*)$, the probability of non exceedance for the impact force. The values for α, β and γ depend on the bed slope and number of tests conducted to measure the wave force F_h. The summary PROVERBS report [30.5] recommends using $\alpha = 3.97$, $\beta = 7.86$, and $\gamma = -0.32$ as determined from large-scale model tests closest to prototype conditions.
- *Step 9*: Impact rise time and duration. The spike in the impact load (force) over the wave period (Fig. 30.10, bottom) can be estimated by formulas representing a triangular force distribution in time with (F_h) maximum taken at the 90% non exceedance level.
- *Step 10*: Estimate uplift forces under impacts. Uplift force, uplift pressure distribution, and uplift moment arm location formulas are presented to calculate the structure stability.
- *Step 11*: Scale corrections. When conditions in steps 6 and 7 warrant impact loads, the PROVERBS includes some scaling corrections in the force calculation.
- *Step 12*: Pressure distributions. Formulas are presented to estimate the pressure distribution over the vertical face of the wall and used to find the location for the moment arm of the horizontal force. Again, when the Steps 6 and 7 reveal that the impact loads are important, these formulas should be employed.

Finally, space does not permit us to provide the details needed to follow steps 9–12 herein and the reader must refer to *Oumeraci* et al. [30.5] to implement the parameter map (Fig. 30.10) and apply the results for vertical, monolithic structure design. A section is included for the case of broken waves hitting the vertical wall.

When the vertical breakwater is made up of relatively short, individual sections, the sudden, direct impact of a plunging, breaking wave may be along the length of one, short, caisson section. Then clearly, the impact load formulas from the PROVERBS project should be used to check the design.

The development of spreadsheets greatly facilitates the design process. The key design variable is the caisson width B which dictates cost. The crest elevation can also be varied (for the same width) so that the minimum, caisson cross-sectional area can be determined.

Oumeraci et al. [30.5] also included a section on what are labeled *alternative, low-reflection structures*, namely perforated vertical walls.

In summary, for many designs, use of Fig. 30.10 to determine the loading case will reveal that the *Goda* [30.34] method is adequate. And when it is not, following Steps 6–12 for the impact loading case will provide a safe design.

Seawalls and Bulkheads

As illustrated in Fig. 30.4, (a) seawalls and (b) bulkheads are backed by soils which provide substantial resistance to water level forces and wave forces. The geotechnical loadings of the soils on the wall or bulkhead often governs the design. and are beyond the scope of discussion in this chapter. Proper seawall design requires the combined efforts of coastal engineers (wave and water level forces on the vertical wall, toe scour at the base of the wall to locate the lowest beach elevation, potential down drift impacts, etc.); geotechnical engineers (soil properties behind the wall, active and passive soil pressures on the wall, active and passive

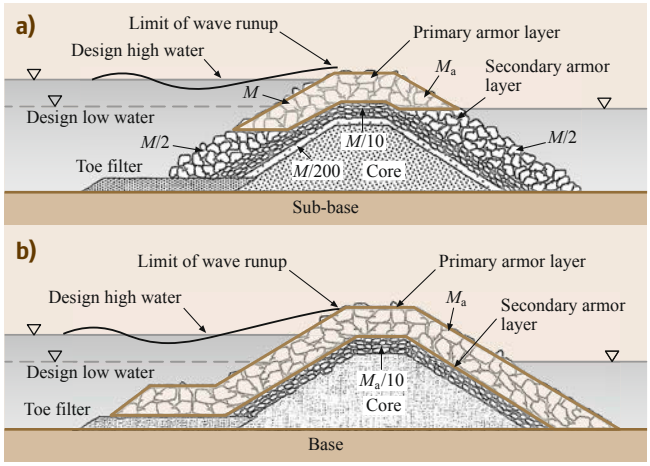


Fig. 30.11 (a) Cross-section of breakwater in deep water (b) Shallow-water depth

soil forces and moments on the wall, etc.) and structural engineers (loads, stresses, strength of materials, etc.) for concrete, steel, and wood (bulkhead) structures. The design of the Virginia Beach seawall by the Norfolk, District Office of the US Army, Corps of Engineers is an example [30.45]. The design of seawalls and bulkheads would require a complete, separate chapter in this Handbook.

Rock Structures

Natural rock of sufficient size (weight or mass) is the common material to provide protection for shorelines around the world. The rock is employed as the outer, *armor layer* (revetment) for sloping embankments and dikes composed of earth materials (Fig. 30.4c, and for breakwaters, groins and jetties where the entire cross-section may be constructed of rock (Fig. 30.11). Determining the stable size (weight or mass) of the *armor layer* under the design wave conditions is the responsibility of coastal engineers.

There are many variables that control the required size and hence stable weight (or mass) of the armor layer for a stable rock structure. Table 30.2 provides a list as found in [30.47]. This initial discussion is for structures with crest elevations of sufficient height to prevent little of no wave overtopping.

In the early 1980s, the Delft Hydraulics Laboratory in the Netherlands conducted an extensive series of laboratory experiments using irregular wave spectrums with the following *variables*:

1. Spectral shapes
2. Core permeability
3. Structure slopes
4. Rock grading, and
5. Relative mass density of the rock structures at the test section.

The results were published as *Design Formulae-(for) Stability of Breakwater Armour Layers* by *van der Meer* [30.47], which has become a standard reference work [30.3]. It was learned that the type of wave breaking as quantified by the surf similarity parameter ξ_m (using T_m, the average wave period) was critical to separate the data set into that for plunging-type wave breaking and that for surging-type waves on the structure. In other words, for irregular, head-on waves impacting the rock test section, two separate dimensionless formulations were employed [30.26, 47], namely:

1. Plunging waves: $\xi_m < \xi_{mc}$

$$\frac{H_s}{\Delta D_{n50}} = 6.2 S^{0.2} P^{0.18} N_z^{-0.1} \xi_m^{-0.5} , \qquad (30.21)$$

and

2. Surging waves: $\xi_m < \xi_{mc}$,

$$\frac{H_s}{\Delta D_{n50}} = 1.0 S^{0.2} P^{-0.13} N_z^{-0.1} (\cot\alpha)^{0.5} \xi_m^P , \qquad (30.22)$$

where the critical, dividing surf similarity parameter ξ_{mc} is

$$\xi_{mc} = [6.2 P^{0.31} (\tan\alpha)^{0.5}]^{1/(P+0.5)} , \qquad (30.23)$$

and:

- H_s: significant wave height in front of breakwater
- D_{n50}: equivalent cube length of median rock
- ρ_s: mass density of rock
- ρ_w: mass density of water

Table 30.2 Primary variables affecting armor stability (after [30.47])

Primary variables affecting armor stability		
Variable	**Symbol**	**Dimension**
Nominal diameter	D_{n50}	m
Relative mass density	Δ	–
Significant wave height	H_s	m
Average wave period	T_z	s
Slope angle	α	degr.
Damage level	S	–
Number of waves	N	–
Armour grading	D_{85}/D_{15}	–
Spectrum shape	$\epsilon_{5\%}$, Q_p^*	
Groupiness of waves	GF, j_1, j_2^*	–
Permeability of core	P	–
Gravity	g	m/s^2

* Described in detail by [30.46]

- Δ: relative mass density $(\rho_s/\rho_w) - 1$
- S: relative eroded area
- P: permeability parameter
- N_z: number of waves
- α: structure slope angle
- s_m: wave steepness, $s_m = H_s/L_{om}$
- L_{om}: deepwater wavelength corresponding to mean wave period, T_m.

These two formulae can be converted into design equations with partial safety factors γ_H and γ_Z; where the hat symbol indicates mean values of the variable, and each equation has its own table of partial safety factors for the full range of failure probability P_f and wave height uncertainty σ'. The partial safety factors are from [30.6] and [30.48]:

- Plunging waves

$$G = \frac{1}{\gamma_Z} 6.2 \hat{S}^{0.2} \hat{P}^{0.18} \hat{\Delta} \hat{D}_n \hat{f} (c\hat{o}t\,\propto)^{0.5} (\hat{s}_{om})^{0.25} \hat{N}_Z^{-0.1} - \gamma_H \hat{H}_S^T \,, \tag{30.24}$$

with partial safety factors as displayed in Table 30.3.

- Surging waves

$$G = \frac{1}{\gamma_Z} \hat{S}^{0.2} \hat{P}^{0.13} \hat{\Delta} \hat{D}_n \hat{f} (c\hat{o}t\,\propto)^{0.5-p} (\hat{s}_{om})^{-0.5P} \hat{N}_Z^{-0.1} - \gamma_H \hat{H}_S^T \,, \tag{30.25}$$

with partial safety factors as displayed in Table 30.4.

In these formulations for design use, the reduction factor $\hat{f}$ models the effect of low crested, overtopped breakwaters [30.49] and is given by the expression

$$\hat{f} = \left(1.25 - 4.8 \frac{R_c}{H_s} \sqrt{\frac{s_{op}}{2\pi}}\right)^{-1} , \tag{30.26}$$

$$0 < \frac{R_c}{H_s} \sqrt{\frac{s_{op}}{2\pi}} < 0.052 \,. \tag{30.27}$$

Note that when the freeboard R_c is zero, $\hat{f}$ reduces to a minimum of 0.8 and when the argument $\frac{R_c}{H_s}\sqrt{\frac{s_{op}}{2\pi}}$ reaches 0.052, there is no reduction in the stone size since $\hat{f} = 1.0$.

See also *van der Meer* [30.49] for the formulation when the rock breakwater is completely submerged. The overtopped and submerged cases do not always give a smaller, lighter rock size for the armor layer. *Burchardt* et al. [30.48] give an alternate method to estimate the static stability of a detached, low-crested breakwater.

For the G = Resistance – Load to be zero, the primary variables in these design equations are ($\hat{H}_S^T$, $\hat{P}$, $\hat{S}$, N, T_{om}, and $c\hat{o}t\,\propto$) for a given, stable rock size $\hat{D}_n$. The rock mass is then simply $\rho\,(\hat{D}_n)^3$. In what follows, the mean value (hat symbol) is assumed and the hat symbol is omitted.

The permeability parameter, P, as defined by *van der Meer* [30.47] is shown in Fig. 30.12. At one extreme are impermeable structures ($P = 0.1$) such as revetments on dikes and natural earth slopes. Rubble-mound, breakwater structures as depicted in Fig. 30.11 with a core use $P = 0.4$ or 0.5. The most permeable structure is made up of all one size rocks with no core and has a permeability parameter, $P = 0.6$, in the *van*

Table 30.3 Partial safety factors (after [30.1, Table VI-6-5])

	$\sigma'_{FH_s} = 0.05$		$\sigma'_{FH_s} = 0.2$	
P_f	γ_H	γ_Z	γ_H	γ_Z
0.01	1.6	1.04	1.9	1.00
0.05	1.4	1.02	1.5	1.06
0.10	1.3	1.00	1.3	1.10
0.20	1.2	1.00	1.2	1.06
0.40	1.0	1.08	1.0	1.10

Table 30.4 Partial safety factors (after [30.1, Table VI-6-6])

	$\sigma'_{FH_s} = 0.05$		$\sigma'_{FH_s} = 0.2$	
P_f	γ_H	γ_Z	γ_H	γ_Z
0.01	1.7	1.00	1.9	1.02
0.05	1.3	1.10	1.6	1.00
0.10	1.3	1.02	1.4	1.04
0.20	1.1	1.10	1.2	1.08
0.40	1.0	1.08	1.1	1.00

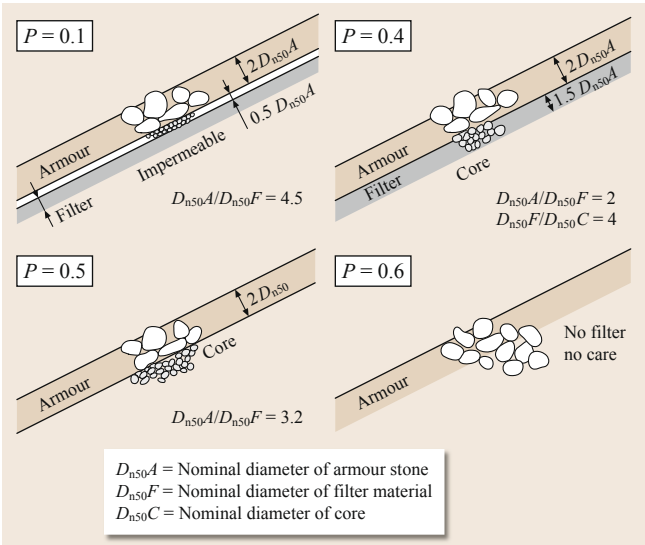


Fig. 30.12 Permeability parameter P as defined in [30.47]

der Meer [30.47] formulation. The structure configuration for $P = 0.4$ was assumed and never tested by *van der Meer* [30.26].

The acceptable level of damage is parameterized by the dimensionless damage factor S which is the ratio of the area eroded to the square of the nominal rock diameter, that is, $S = A_e/(D_n^2)$. The zero damage level is $S = 2$ (for cot $\alpha = 1.5$, 2, and 3) and $S = 3$ (for cot $\alpha = 4$, 5, and 6). Failure levels (for a $2D_{n50}$ armor-layer thickness) are when the secondary layer is exposed beneath the armor layer. These failure levels are $S = 8$ (cot $\alpha = 1.5$, 2), $S = 12$ (cot $\alpha = 3$) and $S = 17$ (cot $\alpha = 4$, 5 and 6). The *van der Meer* [30.47] equations can be used to develop damage (fragility) curves for a given design, stable rock size D_n by calculating S as the wave heights increase above the design wave height. See [30.3] for *intermediate* damage levels S.

The surf similarity parameter ξ_m brings the mean wave period T_m into the formulation. For normal, single peaked spectrums, $T_m \approx 0.82\ T_p$, where T_p is the spectral peak wave period.

Finally, the number of waves N_z is a variable in the *van der Meer* [30.47] formulation. For short storms lasting 4–7 h, $N_z = 2500$ is recommended for mean wave periods from 5 to 10 s. The maximum N_z is 7500 giving 15–30 h storms for waves in the 5–15 mean period range.

The development of spreadsheets with the structure slope angle α as the key design variable greatly facilitates the application of these equations and tables for design of the stable size, or weight (mass) of the armor layer.

There are many variables, coefficients, and exponents in the original [30.26, 47] formulations as seen in (30.21)–(30.23). Graduate students at the Delft Technical University and engineers at Delft Hydraulics (now Deltares) have been conducting research investigation over the past 15–20 years to further understand and improve the original vdM formulas. The following summarizes some past and recent developments as found in the literature with more (probably) to follow.

The publication by *van der Meer* [30.26] was his Ph.D. dissertation at the Technical University of Delft. In it, *van der Meer* [30.26] acknowledges the earlier work of *Thompson* and *Shuttler* [30.50] as the starting point for this basic research using irregular, random waves in the laboratory. He recognized that the coefficient 6.2 (defined as coefficient C_{pl} for the plunging breaker (30.21)) and coefficient 1.0 (defined as coefficient C_s for the surging breaker (30.22)) were stochastic, mean values with standard deviations of mean of 0.4 and 0.08, respectively. Hence, the coefficient in the transition (30.23) separating both formulas was $C_{pl}/C_s = 6.2$. *van der Meer* [30.26] also recognized that waves moving into shallow water would be depth-limited in height and suggested using the Rayleigh distribution ratio of $H_{2\%}/H_s = 1.4$ to adjust the C_{pl} and C_s coefficients and to use $H_{2\%}$ in the formulas. Finally, *van der Meer* [30.26] called for more research to understand the permeability parameter P and stone shape (roundness) on static stability.

It has been recognized that most of the over 300 tests conducted by *van der Meer* [30.26] were with relatively deep water at the toe of the structure. Additional tests (over 200) of rock stability with shallow foreshores have been conducted at Delft Hydraulics [30.51], *Smith* et al. [30.52] as reported by *van Gent* et al. [30.53]. These additional tests demonstrated that:

1. Less scatter was evident if the mean wave period T_m was replaced by $T_{m-1.0}$, the spectral wave period to find the surf similarity parameter $\xi_{m-1.0}$.
2. Less scatter was evident if the ratio of $H_{2\%}/H_s$ is employed because of depth-limited wave breaking.
3. The additional data reduced the standard deviations of the mean values of the modified C_{pl} coefficient in the original formulations of *van der Meer* [30.26].

For *shallow water* wave conditions, *van Gent* et al. [30.53] recommended using $C_{pl} = 8.4\ (H_s/H_{2\%})$ and $C_s = 1.3\ (H_s/H_{2\%})$. These values are also given in the Rock Manual [30.3] and require/allow the user to specify the appropriate $(H_s/H_{2\%})$ ratio. Note that these modified values for C_{pl} and C_s for shallow water will result in smaller (lighter) rock for stability than the original, deep water [30.26] values since the $(H_{2\%}/H_s)$ ratio will be less than 1.4. This is in keeping with shallow water wave breaking and truncation of the wave spectrum from that of the deep water, Rayleigh distribution. The Rock Manual [30.3] defines *shallow water* when the water depth h is less than 3 H_s at the toe of the structure. *Van Gent* and *Pozueta* [30.54] suggested using $H_{s_{toe}} = 0.7H_{so}$ (deep water wave) as the *shallow water* condition. No partial safety factors have yet been published for the shallow water, van der Meer formulations.

More recent research efforts have focused on the permeability parameter P. The values originally assigned by *van der Meer* [30.26] have no physical definition except the general trend that as P increases, the structure has more pore space volume to absorb wave energy. *Verhagen* et al. [30.55] used a numerical model (volume of fluid, VOF) method to calculate wave run-up on permeable and impermeable cores and correlated P as a function of run-up ratios and the surf similarity parameter. *Kik* et al. [30.56] showed with new laboratory experiments that the $P = 0.4$ configuration in Fig. 30.12 is correct in the *van der Meer* [30.26] formulas. Note that P is an important variable in (30.23)

separating plunging, and surging wave breaking and identifying which formulation to employ in the static stability analysis.

The Corps of Engineers has a long history of research on the static stability of rubble-mound breakwaters. The *Hudson* [30.57] formulation was the standard for many years and is still applied for artificial, concrete armor-layer units as discussed below.

Melby and *Kobayashi* [30.58] proposed new, static stability formulations for plunging and surging waves on sloping, armor stone structures that depart from the traditional approach where the primary load variable is the significant wave height H_s at the toe of the structure. In their approach, the maximum, nonlinear wave momentum flux $(MF)_{max}$ as defined by *Hughes* [30.59] is employed as the key load variable in the static stability formulations. The results in *Melby* and *Kobayashi* [30.58] expanded and improved the earlier work in *Melby* and *Hughes* [30.60]; utilized the small-scale physical data from *van der Meer* [30.26], and employed the same definitions for S, N_z, and P as originally proposed by *van der Meer* [30.26]. Two new coefficients were needed and defined. One coefficient, a_m appears in the final equation, but has two separate formulations, one for plunging and one for surging waves. The second new coefficient, K_s, is added for an analysis of damage progression and will be further discussed below. The new, dividing equation separating the plunging and surging formulations was formulated to only depend on the structure slope and does not include the permeability parameter P as found in *van der Meer* [30.26] and (30.23). The analysis and formulas presented by *Melby* and *Kobayashi* [30.58] are relatively new and are not included herein.

We now have over 25 years of experience applying the *deep water*, *van der Meer* [30.26] formulations for the static stability of rock structures. And, slight adjustments have been added for the *shallow water*, van der Meer versions as summarized in Table 30.5.

Equation (30.23) for finding the critical, dividing line surf similarity parameter ξ_{mc} is still valid but uses the C_{pl}/C_s values in Table 30.5 for shallow water applications.

The Rock Manual [30.3] presents further design details surrounding the van der Meer static stability formulas including: the influence of rock gradations, the influence of rock shapes, the influences of packing and rock placement, and the influence of steep approach slopes. All these variables along with a sensitivity analysis using the standard deviations σ in Table 30.5 should be considered for design.

Artificial, Concrete Armor-Layer Units

When it is not possible to use natural rock for the armor layer (quarry size limitations, transport distance, etc.) then some manufactured, artificially shaped, concrete unit must be employed. Figure 30.13 displays a sample of the some shapes that have been employed (from *Kamphuis* [30.17]). The most common are tetrapods, dolos, Accropode, Core-Loc and cube.

The US Army, Corps of Engineers, has developed the CORE-LOC unit (not shown) [30.61], which is similar in shape to the Accropod. Many of these shapes have superior interlocking ability so that smaller, lighter weight units are more stable than natural rock. But

Table 30.5 Key coefficients in the van der Meer formulas

Breaker type	Deep water mean value μ	Deep water standard deviation σ	Shallow water mean value μ	Shallow water standard deviation σ	Equation
Plunging C_{pl}	6.2	0.4	$8.4(H_s/H_{2\%})$	0.109	(30.21)
Surging C_s	1.0	0.08	$1.3(H_s/H_{2\%})$	0.109	(30.22)
Surf parameter ξ	ξ_m		$\xi_{m-1.0}$		
Wave period T	T_m		$T_{m-1.0}$		

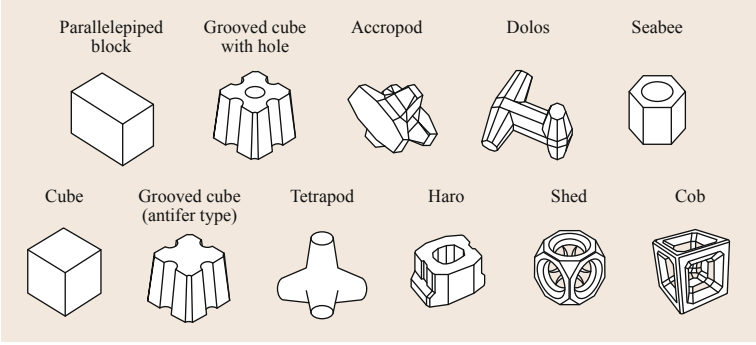


Fig. 30.13 Sample of some concrete armor units (after [30.17])

their cost is generally more due to special forms, concrete mixtures, curing times, etc., for their manufacture. Some contain steel reinforcement but most are completely made of concrete.

The Hudson formula [30.57] is often applied to determine the stable size (mass, or weight) for design

$$\frac{H}{\Delta D_{n50}} = (K_D \cot\alpha)^{1/3} \quad \text{or}$$

$$M_{50} = \frac{\rho_s H^3}{K_D \left(\frac{\rho_s}{\rho_w} - 1\right)^3 \cot\alpha}\,, \tag{30.28}$$

where

- H: characteristics wave height (H_s or $H_{1/10}$)
- D_{n50}: equivalent cube length armor unit
- M_{50}: medium mass of armor unit, $M_{50} = \rho_s D_{n50}^3$
- ρ_s: mass of density of armor units
- ρ_w: mass of density of water
- Δ: $(\rho_s/\rho_w) - 1$
- α: slope angle
- K_D: Hudson stability coefficient.

The only new variable is the Hudson stability coefficient K_D. The Hudson formula is popular for the design of concrete armor units because the K_D values are available for many of the common shapes as illustrated in Table 30.6. Unfortunately, however, the Hudson formula does not include the influence of the wave period T, the structure permeability P, the level of damage S, and the number of waves N_z. And, unfortunately, it is not clear which wave height to apply for design (H_s or $H_{1/10}$) and different K_D values appear in the literature for breaking and nonbreaking waves. The K_D values in Table 30.6 come from many sources including CEM [30.1], and *Reeve* et al. [30.62]. Note that all the concrete units have K_D values greater than natural rock. Some are licensed and have trademarks for application.

Table 30.6 Hudson stability coefficients for nonbreaking waves on concrete units, zero damage, nonovertopping, structure trunk

Armor unit	Hudson stability coefficient K_D
Natural rock	4 (nonbreaking waves)
Natural rock	2 (breaking waves)
Tetrapod	8
Dolos	32 (8)
Accropod	9.5–15
CORE-LOC	13–16
Antifer (grooved)	6.0
Cube, modified	7.5
Cube	6.5
Tribar	10
XBloc	13–16
Stabit	10

A *top-down* design is followed for the standard, cross-sectional breakwater shapes illustrated in Fig. 30.11. Ten *rules* for design have been developed over the years by experience and model testing in the laboratory. These rules cover the following topics:

1. Mass (weight) of the lower layer relative to upper layer ($W_{lower} \approx 1/10 W_{upper}$ as shown in Fig. 30.11)
2. Extent of armor layer below the water surface in deep water
3. Armor-layer thickness
4. Number of concrete units per unit structure length (if applicable)
5. Crest width
6. Crest elevation
7. Need for filter cloth and bedding stone at the bottom
8. Toe protection
9. Design of the head relative to the structure trunk, and
10. Materials and construction methods.

Space does not permit a full discussion and the designer must consult the CEM [30.1] or the Rock Manual [30.3] for complete details.

All the above is for the seaside, static stability of the breakwater shape under the design conditions. Consideration must also be given to the stability of the armor stones on the crest and rear-side of the structure as shown in Fig. 30.11. The maximum velocity on the rear side of the crest during a wave overtopping event (exceeded by 1% of the waves) was given by *van Gent* [30.63] and is used in the stability equation for the required, stable stone size, D_{n50} as found in [30.54]. Complete details can be found in the Rock Manual [30.3, Sect. 5.2.2.11]. No partial safety factors are presented.

When the quarry cannot produce the size of armor stones needed and the cost of manufacturing concrete armor units is prohibitive, then a dynamically stable (*berm* or *reshaping* or *unconventional*) breakwater cross-section may be employed as illustrated in Fig. 30.14 (from *Baird* and *Hall* [30.64]). The upper cross-section is the conventional design and the other cross-sections show a typical berm breakwater design using smaller rocks and a larger volume of armor for subsequent dynamic profile development (reshaping) during storm events. A model for the dynamic shape can be found in *Pilarczyk* [30.22] as proposed by *van der Meer* [30.13] (in *Pilarczyk* [30.22, p. 157]). A state-of-the art report on designing and constructing berm breakwaters has been compiled in [30.7]. Many other recent references can be found in *Torum* et al. [30.65, 66].

Life-Cycle Cost Analysis and Balanced Design

All breakwaters will receive some *damage* (i. e., rock movements that are greater than that in the original design and resulting shape changes that may expose the vulnerable under-layers) when the wave heights are greater than the design wave height. Repairs can be made immediately or after the storm season ends when favorably wave conditions return to the site. These repairs are the *maintenance* costs to keep the breakwater shape functional. These maintenance costs are probabilistic in that they are the product of the exceedance probability curve for the wave heights at the site (Fig. 30.3) and the damage (fragility) curve as found from the van der Meer formulation as discussed above or actual model tests. These annual repair costs must then be multiplied by the present worth factor (pwf) to determine the total, present worth maintenance costs for the design life of the project. (See any standard engineering economic text for the formula to calculate the pwf for the design interest rate i and estimated design-life T) The total life-cycle costs for the given design are then simply the sum of the initial costs and the expected maintenance costs over the design life of the project.

The above procedure should be repeated for a range of design wave heights to determine the optimum or *balanced-design*. Lower design wave heights will produce lower initial costs but higher, long-term maintenance costs. Very high design waves will increase the initial construction costs and lower the maintenance costs. Ideally, a classical, U-shaped, total cost curve will result with the *balanced-design* being the lowest total cost for the structure. This will be different than the deterministic design as dictated by a selected recurrence interval (e.g., the 1% chance wave height) for the design.

In many cases, annual repairs of damaged breakwaters are not carried out for a variety of reasons. *Melby* and *Kobayashi* [30.67–70] have developed formulas to estimate the progression of damage in time from varying wave and water-level conditions for use in the life-cycle analysis process. Their recent work [30.58] extends these earlier efforts to include the nonlinear wave momentum flux [30.43] as the forcing function in the time-varying, $S(t)$, equation for damage progression. The second new variable, K_s, discussed above is needed to handle the initial adjustment of stones places in unstable patterns. Other formulations for damage progression are possible, for example, *van der Meer* ([30.71], who made the direct use of his static stability formulations (30.21) and (30.22)). Space does not permit more details on this relatively new and interesting topic.

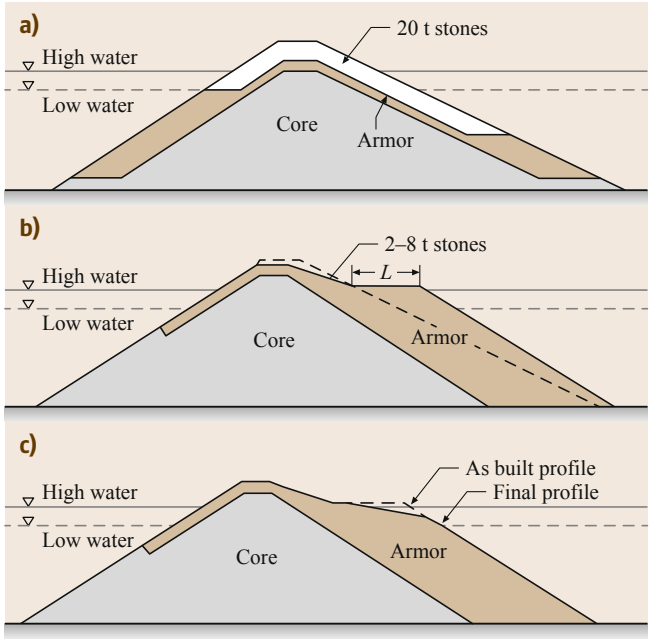


Fig. 30.14 **(a)** Conventional breakwater shape and **(b,c)** berm breakwater with reshaping armor-layer profile (after [30.64])

Functional Costs and Risk

Added to the maintenance costs for structure repair are the functional costs for the economic loss (if any) due to the inability to perform the functional needs of the breakwater. For example, if fuel, oil, or other materials cannot be safely unloaded from a docking facility behind the breakwater, then these potential economic losses (costs) must also be considered in the evaluation of the *risk* for the selected breakwater design.

The basic risk equation for coastal engineering design is

$$R = \Sigma P_E P_F C_i \,. \tag{30.29}$$

where:

- R: The risk
- P_E: The exceedance probability curve for the wave heights
- P_F: The damage (fragility) curve for the structure, and
- C_i: The consequences (i. e., the costs) with $i = 1, 2, 3$, etc. for all the consequences.

In the simple example above, C_1 were the maintenance repair costs and C_2 were the functional (loss) costs for not being able to unload materials. Many other consequences can be considered including environmen-

tal, loss of life, etc. when using the risk equation in coastal engineering design. The final choice for the selected design should consider all the possible consequences of damage and the potential failure of the selected design. The one with the lowest risk should be selected for construction.

30.3 Shoreline Stabilization Structures

Coasts with marine deposits (barrier islands and sandy beaches) are one of the most widely found geomorphic forms in the world. Figure 30.15a displays an example of a dynamically stable shoreline shape between two rocky headlands (from [30.72]) and Fig. 30.15b the concept of a minimum dry beach width, Y_{min}, in the profile cross-section. The dry beach width is much narrower behind the headland than that receiving direct wave attack during storms. This volume of sediment protects the uplands (cliffs, dunes, vegetation, infrastructure, or structures) from damage under normal storm conditions. In 1960, *Silvester* wrote [30.73, p. 469]:

> *... to allow for storm-cycles and the short term reversals of drift, a sufficient width of beach should be allowed as working capital on which the sea can operate. Once the coast has been stabilized by preventing the net movement of sediment, no long-term erosion need be anticipated and the* active *beach width can be minimized.*

However, many shorelines are eroding due to both natural causes (e.g., sea level rise) and for anthropogenic reasons (e.g., down drift of jetties that block the natural sand movements past tidal inlets). Shoreline stabilization structures can provide upland protection beyond the reference baseline for storm events and hence can mitigate the impacts of coastal erosion. At a minimum, these structures should be designed to provide the minimum, dry beach width, Y_{min}, for shore protection.

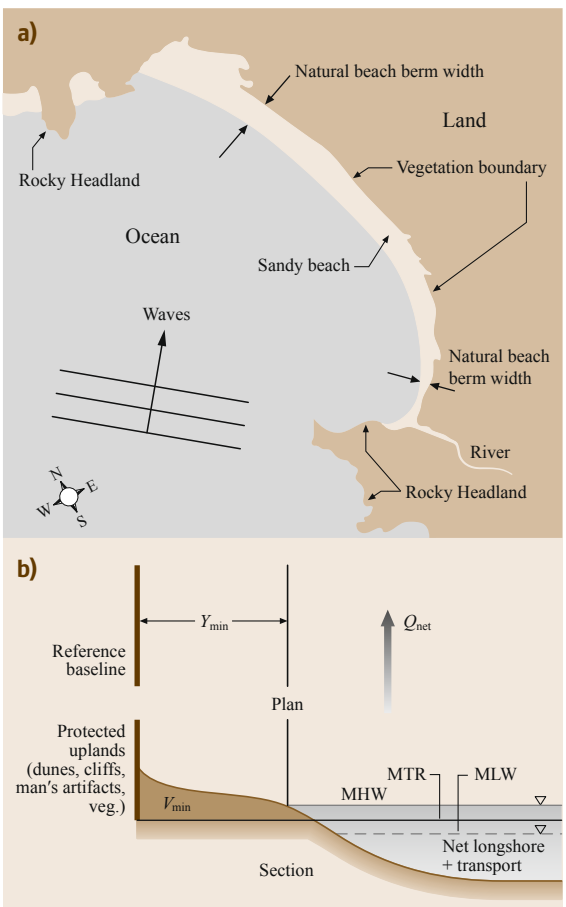


Fig. 30.15 **(a)** Plan form of naturally stable shoreline (after [30.72]) with the *natural beach width* distance from mean sea level (MSL) to upland vegetation; and **(b)** concept of minimum dry beach width, Y_{min} (after [30.1, V-3-3])

30.3.1 Types and Purposes

Figure 30.16c–e displays the three most common types of beach erosion mitigation structures, namely headland breakwaters, near-shore breakwaters, and a groin field. Their primary purpose is to retain sand on the beach during storm events. Present design practice is to always combine their construction with beach renourishment (Chap. 29) to minimize their down drift impact. Clearly, headland and near-shore breakwaters work best when cross-shore sediment transport processes are dominant and a groin field where long-shore sediment transport processes are most important.

As illustrated in Fig. 30.16, the response of the shoreline behind the structure due to waves, currents and resulting sediment transport can be either a tombolo (Fig. 30.16c, shoreline attached to structure) or a salient (Fig. 30.16d, no shoreline attachment). Field data published by *Pope* and *Dean* [30.74] are presented in Fig. 30.17 and reveal that the key variables are:

- L_s: The length of the structure
- L_g: The length of the gap between the structures
- Y: The structure distance offshore from the design, renourished beach shoreline, and

- d_s: The mean water depth offshore at the toe of the structure.

When the L_s/L_g ratio is large (long structure/small gap) so that less wave energy reaches the shoreline and/or when the Y/d_s ratio is small (structure close to shore in shallow water) tombolos form behind the structures which are called *headland* breakwaters. When the opposite is true with these key ratios, more wave energy reaches the shoreline to maintain long-shore sediment transport processes and only salients are present behind what are called *near-shore* breakwaters. Clearly, as illustrated in Fig. 30.17, when short structures with large gaps are located far offshore in relatively deep water, they may have little or no influence on the shoreline (no sinuosity) and offer little shore protection.

30.3.2 Functional Design

Headland Breakwaters

The shoreline configuration between headland-type breakwaters is modeled after the equilibrium plan form shape of natural beaches between natural, rocky outcropping called *headlands* along the coast. As illustrated in Fig. 30.18, wave rays (solid lines 1,2,3, etc.) shoal, refract and diffract in the shallow water contours (dashed lines) and strike the shoreline. If the wave rays break at right angles to the shoreline, no long-shore sediment transport can take place. This is the equilibrium shoreline position for the given offshore wave direction. When the shoreline is in nonequilibrium as shown in Fig. 30.19 (dotted lines) it will continue to evolve (erode or fill) until again reaching equilibrium. The furthest nonequilibrium position is after a coastal storm when the design Y_{min} is reached. Major storm events may produce breaching.

For design, the ratio of $L_S/Y = 1.8$ is recommended for a headland breakwater system [30.75]. Over 40 headland breakwater systems have been built around the Chesapeake Bay [30.76] in Virginia, USA. They recommend using $L_s/L_g = 1$, and setting $Y_g/L_g = 0.6$, where Y_g is the gap distance from a line connecting the structures and the re nourished beach shoreline. These shore parallel structures have been combined with a beach fill at the time of construction and the berm elevation set so that the subaerial beach is visible (tombolo) at high tide.

An empirical approach developed from shoreline data taken from bays known to be in static plan form equilibrium and physical hydraulic models has been developed by *Hsu* et al. [30.77] and is called the *parabolic* model. Figure 30.20 illustrates the two, key independent variables (1) R_o, the distance of a control line from the upland coast headland (tip) to the down coast headland (tip) and (2) the approach wave angle β, between the wave crest and the control line, R_o ($\beta = 24°$ in Fig. 30.20). These variables are site specific and part of the headland breakwater design. The two dependent variables are (R, θ), where R is the radius and θ the angle as defined in Fig. 30.20. Using 10–12 (R, θ) pairs defines a parabola given by the formula [30.77]

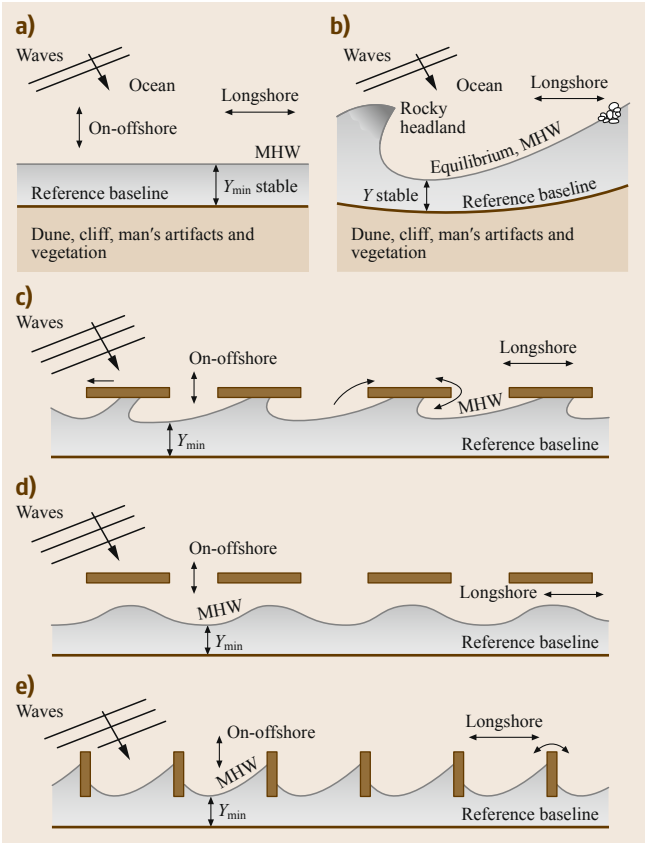


Fig. 30.16a–e Natural **(a–b)** and artificial stable shorelines with minimum dry beach width, Y_{min} for **(c)** headland breakwaters, **(d)** nearshore breakwaters and **(e)** groin field

$$\frac{R}{R_o} = C_o + C_1\left(\frac{\beta}{\theta}\right) + C_2\left(\frac{\beta}{\theta}\right)^2 . \qquad (30.30)$$

The coefficients C_o, C_1, and C_2 in the parabolic model are shown in Fig. 30.21 as a function of the approach wave angle, β, developed empirically by *Hsu* et al. [30.77]. Simple, second-order polynomials have been applied to curve-fit these coefficients as a function of the approach wave angle, β and give

$$C_o = -0.0002\beta^2 + 0.0079\beta - 0.0209 , \qquad (30.31a)$$

$$C_1 = 0.0002\beta^2 + 0.0004\beta + 0.9437 , \qquad (30.31b)$$

$$C_2 = -0.00005\beta^2 - 0.0093\beta + 0.0879 . \qquad (30.31c)$$

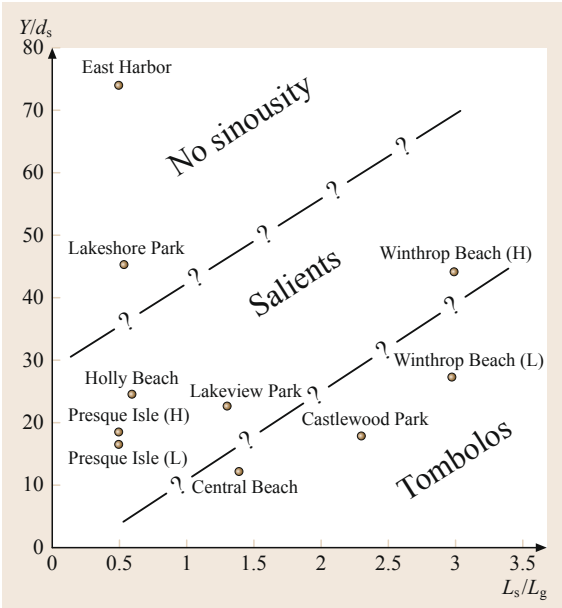


Fig. 30.17 Field data from the Corps of Engineers studies of breakwaters constructed around the United States (after [30.74])

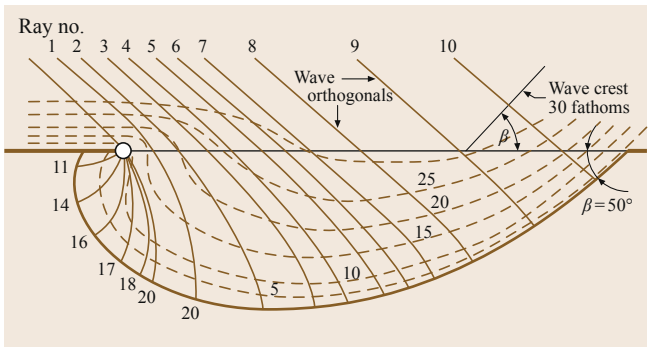


Fig. 30.18 Wave rays shoaling, refracting and diffracting to break at right angles so that no long-shore sediment transport results on the equilibrium plan form shape

The accuracy of the C_2 expression can be improved using a higher order polynomial. To improve the efficiency of application and facilitate the design process, a software package MEPBAY has been developed from computation to graphic display of the results [30.78]. The accuracy of (30.30) to model equilibrium-shaped bays of natural and man-made beaches in Spain has been confirmed by *Gonzalez* and *Medina* [30.79].

Most sites have dominant waves approaching from more than one direction as schematically illustrated in Fig. 30.22 showing wave roses from the Northeast (blue) and the Southwest (green) directions. The parabolic model can then be applied twice and the

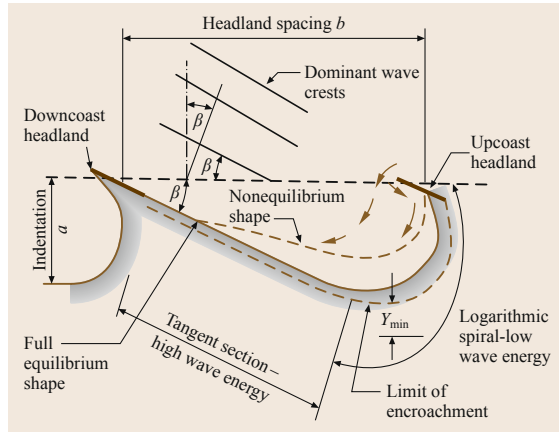


Fig. 30.19 Schematic of headland breakwaters and beach plan forms that are in equilibrium (*solid line*) and nonequilibrium (*dotted lines*) shape

resultant, stable plan form shape (red) be estimated (weighted average position) as the result. In Italy, *D'Alessandro* and *Frega* [30.80] proved that (30.30) modeled the shoreline behind a series of nearshore breakwaters along the Tyrrhenian sea coast. And, they also showed that the formula to calculate the surface area in the tombolo created by the breakwater could be modeled by the method (space herein not available) given by *Ming* and *Chiew* [30.81].

When placed further offshore, the system of nearshore breakwaters will create salients as discussed in the next section.

Near-Shore Breakwaters

Near-shore breakwaters are found mainly in the US, Japan and in the Mediterranean Sea. These locations are primarily along coasts with small (micro-tidal, 0–2 m) daily variations in the tidal range. The empirically derived relationships for the primary variables shown in Fig. 30.17 and listed below are for micro-tidal coasts and from the US experience as reported in the CEM [30.1].

Near-shore breakwater systems are generally shore-parallel and detached structures that are located and spaced such that only salients form behind the structure. Figure 30.23 presents the key variables in the plan form and profile views. Their purpose is the same as that for headland breakwater systems, namely to:

1. Increase the fill life (longevity) of the beach nourishment project behind the breakwaters
2. Provide upland protection from storm damage
3. Provide a wide recreational beach, and in some cases
4. Create or stabilize an upland wetland area.

The key geometric variables are the structure length L_s and the distance offshore Y from the renourished beach at the design beach width. *Dally* and *Pope* [30.75] recommend $L_s/Y = 0.5$ to 0.67 for the shore system and even further offshore (i. e., $L_s/Y = 0.125$) for very long systems of near-shore breakwaters as for example constructed at Presque Isle, Pennsylvania, USA.

A detached, near-shore breakwater system with salient formation is the preferred, breakwater shoreline stabilization system recommended by the US Army Corps of Engineers as stated in [30.1]. This design permits long-shore sediment transport to continue to move through and past the project area to down drift shorelines and minimize down drift impacts. Wave action and long-shore currents are strong enough to keep the salient from connecting to the structure. An extensive literature review of both near-shore and headland- breakwater systems can be found in *Chasten* et al. [30.82] as applied on micro-tidal beaches.

Note in Fig. 30.23 that the salient distances from the design beach width shoreline position, W, are not symmetrical ($Y_s \neq Y_g$) and that the storm erosion distance, e in the gap must be included to predict Y_{min} for the design storm. The plan form configuration of the salient-shaped beach, and the storm erosion distance, e have been studied extensively over the years, but no general analytical procedure exists. See *Chasten* et al. [30.82] for some examples. The breakwater crest elevation and crest width, the structure permeability, the structure slope, and other factors also influence the functional performance of near-shore breakwater systems. For these reasons, numerical models of waves, wave-induced currents, sediment transport, and bathymetric change have been employed to study the shoreline modifications created by near-shore breakwater systems.

On coasts with tides in the range of 2–4 m (meso-tidal) or greater than 4 m (macro-tidal), design guidance is limited. Because over 75% of the UK coastline is classified as meso- or macro-tidal, *Rodgers* et al. [30.83] concluded that the UK needed to study and develop guidelines for breakwater performance on sandy, macro-tidal coasts. Two morphological, numerical models were employed (PISCES and MIKE 21-CAMS) in the effort with conflicting results as reported in Environmental Agency [30.84]. This is not surprising since the numerical models require four coupled, submodules (waves, currents, sediment transport, and bathymetric change) and use different empirical formulations for sediment transport. In fact, the design guidelines cited above for micro-tidal coasts have been used to confirm numerical model results [30.85].

However, results of the numerical modeling study did reveal some important differences that have been summarized in Environmental Agency [30.86]. Diurnal and semidiurnal types, and tide range; the effect of oblique wave incidence; and the effect of breakwater crest elevation were all found to be important.

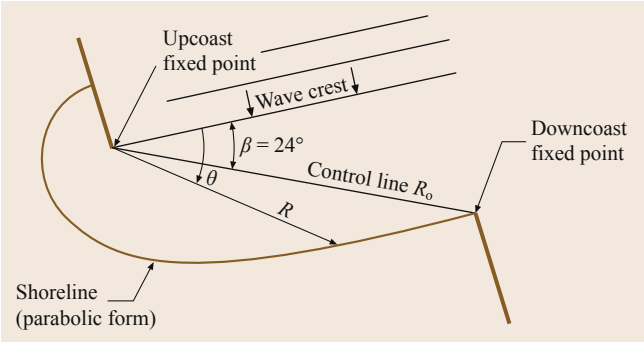


Fig. 30.20 Schematic of parabolic shoreline form defined by four variables R_o, β, R, and θ

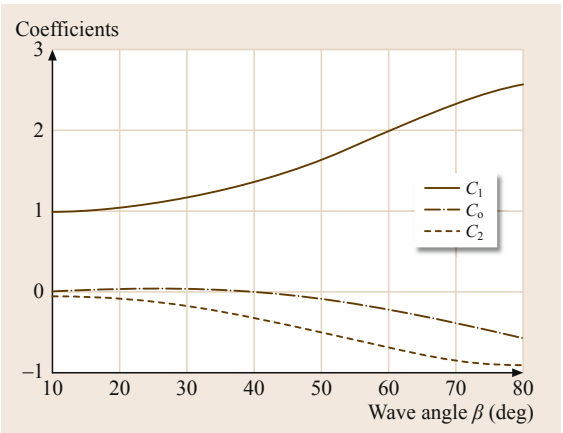


Fig. 30.21 Coefficients C_o, C_1, C_2 in the parabolic model (after [30.77])

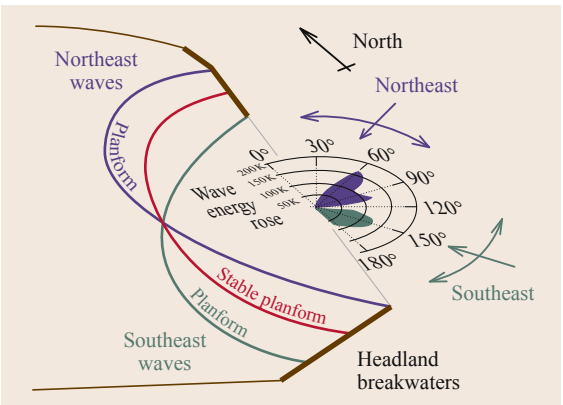


Fig. 30.22 Stable plan form (*red*) estimated from two wave directions using the parabolic model

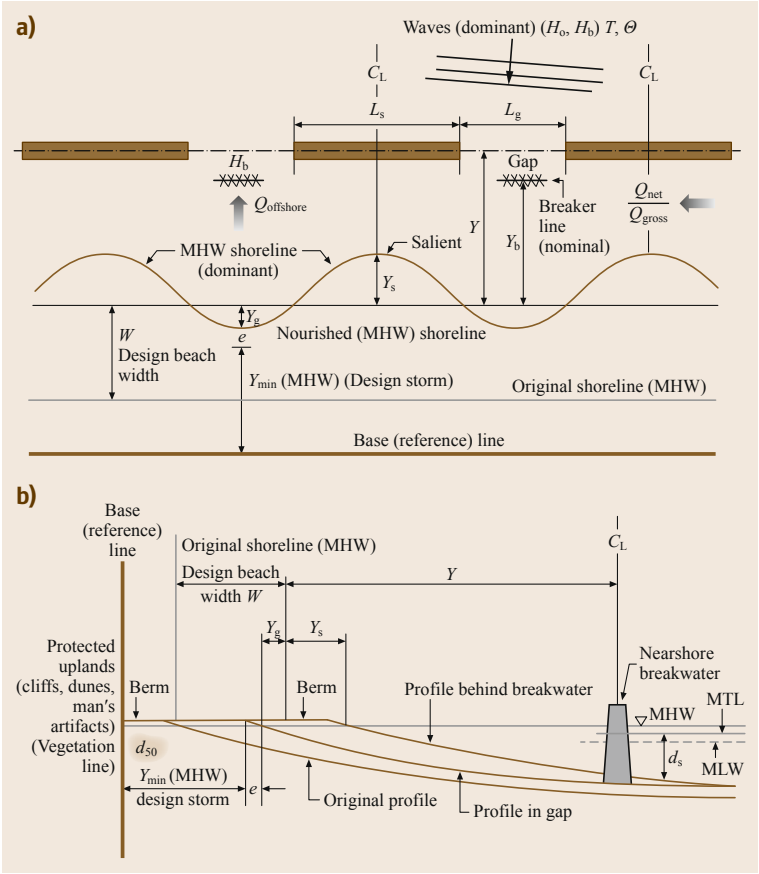


Fig. 30.23a,b Near-shore breakwater. **(a)** Plan view, **(b)** profile view (after [30.1])

As a result the existing design guidance (*Fleming* and *Hamer* [30.87] and similar to that above) was replaced with a new design procedure for macro-tidal coasts. Unfortunately, space does not permit all the details to be presented herein.

At some shoreline locations with large long-shore sediment transport rates in one dominant direction, an initially constructed, detached, near-shore breakwater system will become filled with sediment and be transformed to a headland system with tombolos. This is what took place at Fort Story, Virginia, on the Atlantic Ocean near Virginia Beach, Virginia, as shown in Fig. 30.24. Sediments moving to the North (bottom left to upper right) have filled in behind the breakwaters and trapped a large volume of sediment that was not transported to downdrift beaches and is producing down drift impacts. The designer must always be aware of the potential for sand retention behind a system of detached, near-shore breakwaters that could cause down drift erosion.

Fig. 30.24 The system of nineteen near-shore breakwaters at Fort Story, Virginia that have been transformed into headland breakwaters

Groin Fields

Groins are the oldest, the most commonly applied, and arguably the most misused and improperly designed of all coastal structures. They are normally shore-perpendicular structures of relatively short lengths (when compared with navigational jetties at tidal inlets) as illustrated in Fig. 30.16e. A groin interrupts the

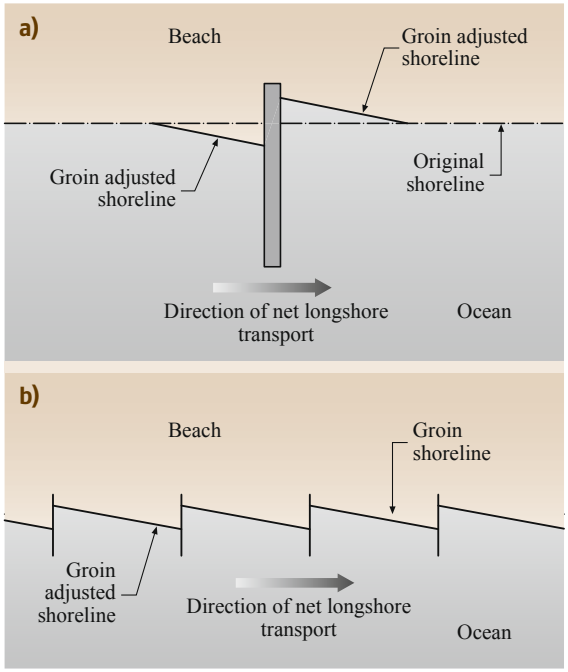


Fig. 30.25a,b Schematic of (a) single groin and (b) groin field displaying down drift impacts ◄

long-shore transport such that sand is deposited on the up-drift side and consequently absent on the down-drift side as illustrated in Fig. 30.25.

The net long-shore transport direction is revealed (sand accumulates on the side of the groin from the up-drift direction). How to maintain a minimum width of beach at the down-drift side has always been of concern in coastal engineering design. As illustrated in Fig. 30.26, the present-day philosophy of groin design is to combine the structure with a beach nourishment project to:

1. Minimize down-drift impacts
2. Maintain some basic level of coastal hazard mitigation as defined by Y_{min} after storm erosion e, and
3. Allow some sand bypassing around the end.

The basic rule of thumb for the ratio of groin spacing X_g to the groin length Y_g is $X_g/Y_g = 2-3$ as found from many years of experience to satisfy the three design goals listed above [30.88]. Note that the groin

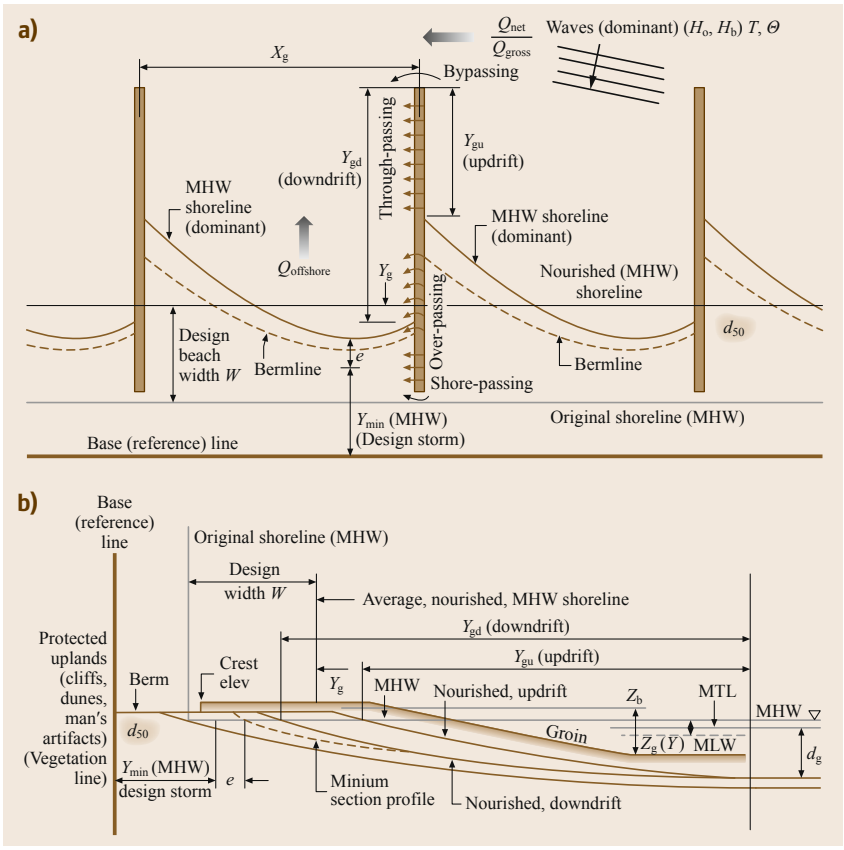


Fig. 30.26a,b Definitions of variables in groin field design

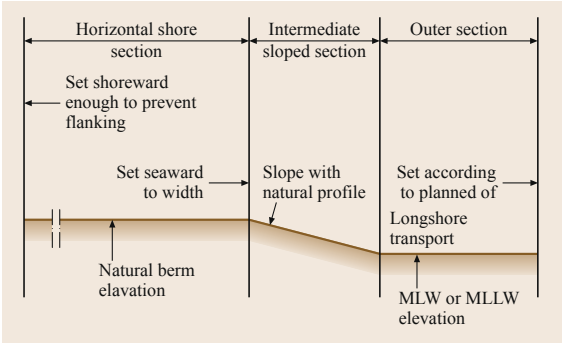


Fig. 30.27 Recommended groin profile with sloping section to permit sand bypassing both over the groin and around the outer end to mitigate down-drift impacts

length, Y_g is measured from the nourished (mean high water, MHW) shoreline position to the end of the groin. The up-drift, Y_{gu}, and down-drift, Y_{gd}, distances shown in Fig. 30.26 are after shoreline adjustment within the groin compartment.

The cross-shore profile of the top of the groin should mimic the local, normal beach profile as illustrated in Fig. 30.27.

There are ten basic *rules* commonly recommended for the functional design of groins [30.88]:

- *Rule 1*: Conservation of mass for the transport of sediment alongshore and cross-shore means groins neither create nor destroy sediment.
- *Rule 2*: To mitigate the erosion of adjacent beaches, always include a beach fill in the design.
- *Rule 3*: Agree on the minimum, dry beach width Y_{min} for upland protection during storm events as a measure to judge success.
- *Rule 4*: Begin with $X_g/Y_g = 2–3$ at time of construction.
- *Rule 5*: Use a numerical, one-line model to estimate shoreline change around single groins and groin fields (e.g., the GENESIS model of *Hanson* and *Kraus* [30.89]).
- *Rule 6*: Use a cross-shore, profile-change model to estimate the minimum, dry beach width Y_{min} during storm events (e.g., the SBEACH model of *Hanson* and *Kraus* [30.89]).

  In the near future, Rules 5 and 6 may be combined by using a numerical, beach evolution model that combines both long- and cross-shore transport processes in one model. For example, GenCade [30.90].
- *Rule 7*: Bypassing, structure permeability, and the balance between net and gross long-shore transport rates are the three key factors in the functional design. Use model simulation to iterate a final design to meet the Y_{min} criterion.
- *Rule 8*: Consider tapered ends, alternate plan forms, and cross-sections to minimize impacts on adjacent beaches. For example, consider T-groins, permeable groins, notched groins, submerged groins, and other shapes [30.88]
- *Rule 9*: Establish a field-monitoring effort to determine if the project is successful and to identify adjacent beach impacts.
- *Rule 10*: Establish a *trigger* mechanism for decisions to provide modification (or removal) if adjacent beach impacts found not acceptable.

A Rule 0 could be added to emphasize the fact that groins are only useful where long-shore sediment transport processes are dominant. If cross-shore transport processes control, then near shore or headland breakwaters should be considered first.

Reference [30.91] focused on the functioning and design of coastal groins. It included papers on the background and literature review, management, design, case studies, and numerical and physical modeling of groin fields. It contains a wealth of information and should be consulted for both functional and structural design.

Summary

From the above, it is clear that our understanding of the functional, plan form layout of coastal structures is relatively primitive. We only have some design guidance and rules of thumb that have been obtained from field experience and physical model experiments in the laboratory. The use of numerical models will signifi-

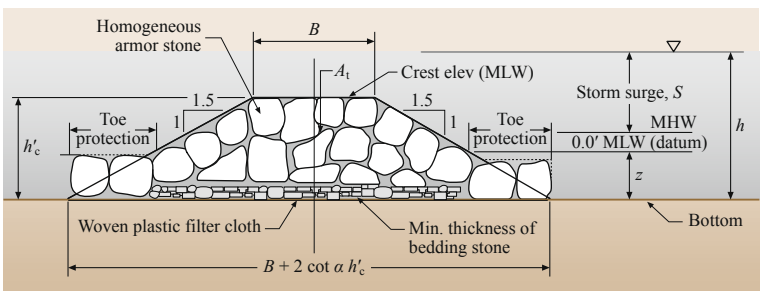


Fig. 30.28 Typical cross-section of a low-crested rock structure constructed with one stable-sized armor stone as determined by *van der Meer* (after [30.47])

cantly increase our understanding of the physics in the future.

30.3.3 Structural Design

The predominant material of construction for shoreline stabilization structures is rock. Timber, steel, and concrete materials have been employed for groins but are not considered here.

The structural design of rock structures is discussed above (Sect. 30.2.3) for coastal armoring structures. However, in many cases where headland breakwaters, near-shore breakwaters, and groins are built in shallow water, only one rock size is employed as illustrated in Fig. 30.28. The crest elevation is above the mean high water (MHW) tidal elevation but the structure is overtopped and often completely submerged during storm surge events. The van der Meer formulations for overtopped and completely submerged conditions should be employed. These structures are highly permeable so that the permeability parameter, $P = 0.6$.

30.4 Websites and Sea Level Rise Trends

30.4.1 Websites

- The US Army Corps of Engineers, The Coastal Engineering Manual (CEM): http://chl.erdc.usace.army.mil/cem.
- The Construction Industry Research and Information Association (CIRIA) The CIRIA/CUR/CETMEF Rock Manual: http://www.ciria.org.
- The EurOtop Team The Assessment Manual: http://www.overtopping-manual.com.
- The Calculation Tool: http://www.overtopping-manual.com/calculation_tool.html.

30.4.2 Sea Level Rise Trends

Global climate change is real. The potential for accelerated, RSLR rates at most coastal locations poses a threat to existing shore protection hard structures around the world. Severe storms (low-pressure events and hurricanes) may occur more often to increase the frequency of storm surges and shift the exceedance frequency curves for high water elevations (Fig. 30.2). More severe storms may produce higher offshore wave heights. Consequently, breaking waves near shore will be larger due to deeper water depths closer to shore. These consequences of climate change and sea level rise on water elevations and wave heights will produce larger stresses on coastal structures. The exceedance probability for the loads (water levels and wave heights) will increase the probability of failure and resulting risk. In addition, wave run-up and overtopping will increase to reduce the functionality of the structures and add to the flooding damage.

A review of the typical shore protection structures shown in Fig. 30.4 and shoreline stabilization structures presented in Fig. 30.16 reveals that rock structures are most often the *best* alternative (lowest life-cycle costs) for shore protection with hard structures. Rock structures can be more readily raised in crest elevation and repaired when damaged than all the other types. Rock structure crests can be elevated to keep up with potentially accelerating sea level rise rates to maintain functionality for wave run-up and overtopping. And, rock structure damage can be repaired to keep up with larger waves producing more damage more often. The resilience of the coastal shore protection effort can be maintained with rock structures and hybrid structure combinations (rubble mounds with frontal berms) will be important in the future.

However, complete failure of the hard, shore protection system must always be considered in modern, contemporary, coastal engineering design. See *Kamphuis* [30.17, Chapter 17] for a discussion of resilient systems design that must also consider the social, political, and institutional constraints in the complete process of resilient systems design.

References

30.1 CEM: *The Coastal Engineering Manuel* (US Army Corps of Engineers, Engineering Research and Development Center, Vicksburg 2006, 2011)

30.2 Kuratorium für forschung im küsteningenieurwesen: EurOtop, wave overtopping of sea defences and related sturctures: Assessment Manual, Küste **73**, 1–178 (2007)

30.3 CIRIA/CUR/CETMEF: *Rock Manual, 2007. The Use of Rock in Hydraulic Engineering*, 2nd edn. (CIRIA, London 2007)

30.4 Y. Goda: *Random Seas and Design of Maritime Structures*, 3rd edn. (World Scientific, Singapore 2010)

30.5 H. Oumeraci, A. Kortenhaus, W. Allsop, M. de Groot, R. Crouch, H. Vrijling, H. Voortman: *Probabilistic*

Design Tools for Vertical Breakwaters (CRC, Boca Raton 2001)
30.6 PIANC: *Analysis of Rubble Mound Breakwaters*, Report of Working Group No. 12 (PIANC, Brussels 1992)
30.7 PIANC: *State-of-the-Art of Designing and Constructing Berm Breakwaters*, Report of Working Group 40, Supplement to Bulletin No. 78179 (Maritime Navigation Commission, Brussels 2003)
30.8 S. Coles: *An Introduction to Statistical Modeling of Extreme Values*, Springer Series in Statistics (Springer, London 2001)
30.9 N.W. Scheffner, L.E. Borgman, J.E. Clasusner, B.L. Edge, P.J. Grace, A. Militello, R.A. Wise: *Users Guide to the Use and Application of the Empirical Simulation Technique*, Techn. Rep. CHL-97-00 (US Army Engineer Waterways Experiment Station, Vicksburg 1997)
30.10 N.W. Scheffner, J.E. Clausner, A. Militello, L.E. Borgman, B.L. Edge, P.E. Grace: *Use and Application of the Empirical Simulation Technique: Users Guide*, Techn. Rep. CHL-99-21 (US Army Engineer Waterways Experiment Station, Vicksburg 1999)
30.11 Norfolk District Office, U.S. Army Corps of Engineers: General Design Memorandum, Virginia Beach Hurricane Protection Project (1988)
30.12 G.R. Toro, D.T. Resio, D. Divoky, A.W. Niedoroda, C. Reed: Efficient joint probability methods for hurricane surge frequency analysis, Ocean Eng. **37**, 125–134 (2010)
30.13 P.J. Vickery, P.F. Skerjl, L.A. Twisdale: Simulation of Hurricane risk in the U.S. using empirical track model, J. Struct. Eng. **126**, 1222–1237 (2000)
30.14 S. Agbley, D.R. Basco: An evaluation of storm surge frequency of occurrence estimators, Proc. Solut. Coast. Disasters 2008 (ASCE, Reston 2008) pp. 185–197
30.15 D.T. Resio: *White Paper on Hurricane Surge Frequency Analysis* (Corps of Engineers, Vicksburg 2007), unpublished manuscript
30.16 R.G. Dean, T.L. Walton: Wave setup. In: *Handbook of Coastal and Ocean Engineering*, ed. by Y.C. Kim (World Scientific, Singapore 2010)
30.17 J.W. Kamphuis: *Introduction to Coastal Engineering and Management*, 2nd edn. (World Scientific, Singapore 2010)
30.18 M.T. Schultz, B.P. Gouldby, J.D. Simm, J.L. Wibowo: *Beyond the Factor of Safety: Developing Fragility Curves to Characterize System Reliability*, ERDC-SR-10-1 (US Army Corps of Engineers, Vicksburg 2010)
30.19 M.T. Schultz, S.K. McKay, L.Z. Hales: *The Quantification and Evolution of Resilience in Integrated Coastal Systems*, ERDC-TR-11-XX (US Army Corps of Engineers, Vicksburg 2012), draft
30.20 W. Birkemier: *Definition of a Coastal Storm at the Corps of Engineers* (Field Research Facility, Duck 2010), personal communication
30.21 D.R. Basco, N. Mahmoudpour: The modified coastal storm impulse (COSI) parameter and quantification of fragility curves for coastal design, Proc. Coast. Eng., Vol. 1 (ASCE, Reston 2012)
30.22 K.W. Pilarczyk (Ed.): *Coastal Protection* (Balkema, Rotterdam 1990)
30.23 J.A. Battjes: Surf similarity, Proc. 14th Int. Coast. Eng. Conf., Vol. 1 (ASCE, Reston 1974) pp. 466–479
30.24 A. Hunt: Design of seawalls and breakwaters, J. Waterw. Harb. **85**(3), 123–152 (1959)
30.25 J.P. de Waal, J.W. van der Meer: Wave run-up and overtopping on coastal structures, Proc. 23rd Int. Coast. Eng. Conf., Vol. 2 (ASCE, Reston 1992) pp. 1758–1771
30.26 J.W. van der Meer: *Rock Slopes and Gravel Beaches Under Wave Attack*, Ph.D. Dissertation Ser. (Delft University of Technology, The Netherlands 1988), also Delft Hydraulics Publication No. 396
30.27 J.A. Melby: *Wave Runup Prediction for Flood Hazard Assessment*, ERDC/CHL TR-12-24 (Corps of Engineers, Vicksburg 2012)
30.28 FEMA: *Atlantic Ocean and Gulf of Mexico Coastal Guidelines: Guidelines and Specifications for Flood Hazard Mapping Partners* (Dept of Homeland Security, Washington 2007)
30.29 M.W. Owen: *Design of Seawalls Allowing for Wave Overtopping, Rep. No. 924* (Hydraulics Research Station, Wallingford 1980)
30.30 M.W. Owen: The hydraulic design of seawall profiles, Proc. Coast. Prot. Conf. (Thomas Telford Publishing, London 1982) pp. 185–192
30.31 J. Pedersen: *Experimental Study of Wave Forces and Wave Overtopping on Breakwater Crown Walls*, Paper 12 (Department of Civil Engineering, Aalborg University, Aalborg 1996)
30.32 J.W. van der Meer, W. Janssen: Wave run-up and wave overtopping at dikes. In: *Wave Forces on Inclined and Vertical Wall Structures*, ed. by ASCE (ASCE, Reston 1995) pp. 1–27
30.33 HR Wallingford Ltd.: Waves Overtopping, Online Calculation Tool, http://www.overtopping-manual.com/calculation_tool.html
30.34 Y. Goda: New wave pressure formulae for composite breakwaters, Proc. 14th Int. Coast. Eng. Conf., Vol. 3 (ASCE, Reston 1974) pp. 1702–1720
30.35 K. Tanimoto, K. Moto, S. Ishizuka, Y. Goda: An investigation on design wave force formulae of composite-type breakwaters, Proc. 23rd Jpn. Conf. Coast. Eng. (1976) pp. 11–16, in Japanese
30.36 J.W. van der Meer, K. de Angremond, J. Juhl: Probabilistic calculation of wave forces on vertical structures, Proc. 24th Int. Coast. Eng. Conf., Vol. 2 (ASCE, Reston 1994) pp. 1754–1769
30.37 J.A. Battjes: Effects of short-crestedness on wave loads on long structures, J. Appl. Ocean Res. **4**(3), 165–172 (1982)
30.38 S. Takahashi, K. Tanimoto, K. Shimosako: A proposal of impulsive pressure coefficient for design of composite breakwaters, Proc. Int. Conf. Hydro-Techn. Eng. Port Harb. Constr. (1994) pp. 489–504
30.39 US Army, Corps of Engineers, Coastal Engineering Research Center (US): *Shore Protection Manual* (Dept. of the Army, Waterways Experiment Station, Washington 1984)
30.40 A. Kortenhaus, H. Oumeraci: Classification of wave loading on monolithic coastal structures, Proc. 26th

Int. Conf. Coast. Eng. (ICCE) (ASCE, Kopenhagen 1998) pp. 867–880
30.41 N.W.H. Allsop, M. Calabrese: Impact loadings on vertical walls in directional seas, Proc. 26th Int. Conf. Coast. Eng., Vol. 2 (ASCE, Reston 1998) pp. 2056–2068
30.42 H. Oumeraci, A. Kortenhaus: Wave impact loading-tentative formulae and suggestions for the development of final formulae. Proc. 2nd Workshop, MAST III, PROVERBS-Project: Probabilistic Des. Tools Vertical Breakwaters, Edinburgh (1997)
30.43 R. Miche: Mouvements ondulatoires de la mer en profondeur constante ou decroissante, Ann. Ponts Chaussées **2**, 285–319 (1944)
30.44 J.K. Vrijling, J. Bruinsma: Hydraulic boundary conditions. Hydraulic aspects of coastal structures: Developments in hydraulic engineering related to the design of the Oosterschelde storm surge barrier in the Netherlands, Part II: Des. Philos. Strategy Proj. Relat. Res. (Delft University Press, Delft 1980) pp. 109–133
30.45 J. Gaythwaite, D. Pezza, L. Topp: Beach erosion control and hurricane protection for Virginia Beach, Proc. 6th Symp. Coast. Ocean Manag. (ASCE, Charleston 1989) pp. 791–805
30.46 J.W. van der Meer, K.W. Pilarczyk: Stability of rubble mound slopes under random wave attack, Proc. 19th Int. Conf. Coast. Eng. (ICCE), Houst. (ASCE, New York 1984) pp. 2620–2634
30.47 J.W. van der Meer: Stability of breakwater armour layers – Design formulae, Coast. Eng. **11**, 219–239 (1987)
30.48 H.F. Bucharth, J.D. Sorensen: Design of rubble mound breakwaters using partial safety factors, Proc. Coast. Eng., Vol. 1 (ASCE, Reston 2000)
30.49 J.W. van der Meer: *Stability and Transmission at Low-Crested Structures*, Delft Hydraulics Publication No. 453 (Delft Hydraulics Laboratory, Delft 1991)
30.50 D.M. Thompson, R.M. Shuttler: *Riprap Design for Wind Wave Attack: A Laboratory Study in Random Waves HRS*, Rep. EX 707 (HR Wallingford, Wallingford 1975)
30.51 M.R.A. van Gent: *Physical Model Investigations on Coastal Structures with Shallow Foreshores: 2D Model Tests with Single and Double Peaked Wave Energy Spectrums*, Rep. H3608 (Delft Hydraulics Laboratory, Delft 1999)
30.52 G.M. Smith, I. Wallast, M.R.A. van Gent: Rock slope stability with shallow foreshores, Proc. Int. Conf. Coast. Eng. (ASCE, Reston 2002) pp. 1524–1536
30.53 M.R.A. van Gent, A.J. Smale, C. Kuiper: Stability of rock slopes with shallow foreshores, Proc. Coast. Struct. (ASCE, Reston 2004) pp. 100–112
30.54 M.R.A. van Gent, R. Pozueta: Rear-side stability of rubble mound structures, Proc. Int. Conf. Coast. Eng. (ASCE, Reston 2005)
30.55 H.J. Verhagen, D. Jumeler, A.V. Domingo, P. van Broekhoven: Method to quantify the notational permeability, Proc. Coast. Struct. (ASCE, Reston 2011)
30.56 R. Kik, J.P. van den Bos, J. Maertens, H.J. Verhagen, J.W. van der Meer: Notational permeability, Proc. Int. Conf. Coast. Eng. (ASCE, Reston 2012)
30.57 R.Y. Hudson (Ed.): *Concrete Armor Units for Protection Against Wave Attack*, Miscellaneous Paper H-74-2 (US Army Engineer Waterways Experiment Station, Vicksburg 1974)
30.58 J.A. Melby, N. Kobayashi: Stone armor damage initiation and progression based on the maximum wave momentum flux, J. Coast. Res. **27**(1), 110–119 (2011)
30.59 S.A. Hughes: Wave momentum flux parameter: A descriptor for near shore waves, Coast. Eng. **51**, 1067–1084 (2004)
30.60 J.A. Melby, S.A. Hughes: Armor stability based on wave momentum flux, Proc. Coast. Struct. Reston, Va. (ASCE, New York 2004) pp. 53–65
30.61 J.A. Melby, G.F. Turk: The CORE-LOC: Optimized concrete armor, Proc. 24th Int. Coast. Eng. Conf., Vol. 2 (ASCE, Reston 1994) pp. 1426–1438
30.62 D. Reeve, A. Chadwick, C. Fleming: *Coastal Engineering, Processes, Theory, and Design Practice*, 2nd edn. (Spoon, London 2012)
30.63 M.R.A. van Gent: Wave overtopping events at dikes, Proc. Int. Conf., Vol. II (ASCE, Reston 2003) pp. 2203–2215
30.64 W.F. Baird, K. Hall: Breakwater breakthrough, ASCE Civ. Eng. **57**(1), 45–47 (1987)
30.65 A. Torum, F. Kuhnen, A. Menze: On berm breakwaters. Stability, scour, and overtopping, Coast. Eng. **49**, 209–238 (2003)
30.66 A. Torum, M.N. Moghim, K. Westeng, N. Hidayati, O. Arntsen: On berm breakwaters: Recession, crown wall wave forces, and reliability, Coast. Eng. **60**, 299–318 (2012)
30.67 J.A. Melby, N. Kobayashi: Progression and variability of damage on rubble mound breakwaters, J. Waterw. Port Coast. Ocean Eng. **124**(6), 286–294 (1998)
30.68 J.A. Melby, N. Kobayashi: Damage progression on breakwaters, Proc. 26th Int Conf. Coast. Eng. (ASCE, Reston 1998) pp. 1884–1897
30.69 J.A. Melby, N. Kobayashi: Damage progression and variability on breakwater trunks, Proc. Coast. Struct. (Balkema, Rotterdam 1999) pp. 309–316
30.70 J.A. Melby, N. Kobayashi: Damage development on stone-armored rubble mounds, Proc. 27th Int. Conf. Coast. Eng. (World Scientific, Singapore 2000) pp. 1571–1584
30.71 J.W. van der Meer: Design of concrete armor layers, Proc. Coast. Struct. (Balkema, Rotterdam 2000) pp. 213–221
30.72 R. Silvester, J.R.C. Hsu: *Coastal Stabilization: Innovative Concepts* (Prentice Hall, Englewood Cliffs 1993)
30.73 R. Silvester: Stabilization of sedimentary coastlines, Nature **188**, 467–469 (1960)
30.74 J. Pope, J.L. Dean: Development of design criteria for segmented breakwaters, Proc. 20th Int. Conf. Coast. Eng. (ASCE, Reston 1986) pp. 2144–2158
30.75 W.R. Dally, J. Pope: *Detached Breakwaters for Shore Protection*, Techn. Rep. CERC-86-1 (US Army Engineer Waterways Experiment Station, Vicksburg 1986)

30.76 C.S. Hardaway, G.R. Thomas, J.-H. Li: *Chesapeake Bay Shoreline Study: Headland Breakwaters and Pocket Beaches for Shoreline Erosion Control*, *Final Rep. No. 313* (Virginia Institute of Marine Science, Gloucester Point 1991)
30.77 J.R.C. Hsu, R. Silvester, Y.M. Xia: Static equilibrium bays: New relationships, J. Waterw. Port Coast. Ocean Eng. **115**(3), 285–298 (1989)
30.78 A.H.F. Klein, J.T. Menezes: Visual assessment of bayed beach stability with computer software, Comput. Geosci. **29**, 1249–1257 (2003)
30.79 M. Gonzalez, R. Medina: On the application of static equilibrium bay formulations to natural and man-made beaches, Coast. Eng. **43**, 209–225 (2001)
30.80 F. D'Alessandro, F. Frega: A verification of static-equilibrium parabolic formulation at the protected shoreline of pizzo calabro (Italy), Proc. 30th Int. Conf. Coast. Eng. (ASCE, Reston 2007)
30.81 D. Ming, Y.-M. Chiew: Shoreline changes behind detached breakwater, J. Waterw. Port Coast. Ocean Eng. **126**(2), 63–69 (2000)
30.82 M.A. Chasten, J.D. Rosati, J.W. McCormick, R.E. Randall: *Engineering Design Guidance for Detached Breakwaters as Shoreline Stabilization Structure*, Techn. Rep. CERC-93-19 (US Army Engineer Waterways Experiment Station, Vicksburg 1993)
30.83 J. Rodgers, T. Chester, B. Hamer: *LEACOAST2: Practical Guidance Scoping Study*, Tech. Rep. (Department for Environment, Food, and Rural Affairs, London 2006)
30.84 Environmental Agency: *Modeling the Effect of Nearshore Detached Breakwaters on Sandy Macro-Tidal Coasts*, Project SC 060026/R2 (Environmental Agency, London 2009)
30.85 T.V. Karambas: Design of detached breakwaters for coastal protection: Development and application of an advanced numerical model, Proc. 33rd Int. Conf. Coast. Eng. (World Scientific, Singapore 2012)
30.86 Environmental Agency: *Guidance for Outline Design of Nearshore Detached Breakwaters on Sandy Macro-Tidal Coasts*, Project SC 060026/R1 (Environmental Agency, London 2009)
30.87 C. Fleming, B. Hamer: Successful implementation of an offshore reef scheme, Proc. 27th Int. Conf. Coast. Eng. (ASCE, Reston 2000)
30.88 D.R. Basco, J. Pope: Groin functional design guidance from the coastal engineering manual, J. Coast. Eng. **33**, 121–130 (2004)
30.89 H. Hanson, N.C. Kraus: *GENESIS: Generalized Model for Stimulating Shoreline Change*, Techn. Rep. CERC-89-19 (US Army Engineer Waterways Experiment Station, Vicksburg 1989)
30.90 A.E. Frey, K.J. Connell, H. Hanson, M. Larson, R.C. Thomas, S. Munger, A. Zundel: *GenCade Version 1 Model Theory and User's Guide*, Techn. Rep. ERDC-CHL TR-12-X (US Army Engineer Research and Development Center, Vicksburg 2012)
30.91 N.C. Kraus, K.L. Rankin: *Functioning and Design of Coastal Groins*, *Journal of Coastal Research*, *Special Issue 33* (CERF, Weit Palm Beach 2003)

31. Port and Harbor Design

Andrew Cairns, John M. Carel, Xiao Li

31.1 Port and Harbor Layout and Design 685
31.1.1 Types of Marine Terminals 687
31.1.2 Vessel Overview 689
31.1.3 Harbor Operational Limits 690
31.1.4 Harbor Navigation 691
31.1.5 Sediment Transport Considerations 692
31.2 Structure Types 692
31.2.1 Breakwaters and Wave Attenuation 692
31.2.2 Open Pile Platforms 693
31.2.3 Solid Structures 694
31.2.4 Drydock Facilities 698
31.2.5 Floating Structures 701
31.2.6 Swinging Moorings 701
31.2.7 Ice Breakers 703
31.3 Loads on Structures due to Vessel Mooring and Berthing 703
31.3.1 Vessel Berthing Loads 703
31.3.2 Mooring Loads 704
31.3.3 Loads on Structures due to Ocean Environmental Conditions 706
31.3.4 Tide and Storm Surge 707
31.3.5 Ice 707
31.3.6 Tsunamis 707
31.4 Suggested Reading 707
31.5 Notation 709
References 709

A harbor is defined as a haven, and in the context of this handbook, it is a body of water that provides shelter to vessels. A harbor can be natural or artificial, situated along the shore with water deep enough for anchoring a ship in an area that provides protection from wind, waves and currents. When Henry Hudson arrived in the new world in 1609, he discovered what would become New York, and the success of New York can be traced to its excellent natural harbor. Many of today great seaports are located at natural harbors that have been developed into world-class port facilities. This handbook will deal exclusively with artificial harbors and the port facilities they protect.

Ports can be defined as a city, town, or place where ships are moored, loaded, or unloaded. They typically reside within a harbor and are made up of one or more individual terminals that handle a particular cargo including passengers, bulk cargo, or containerized cargo. This chapter in the handbook will define the various types of marine terminals and the structures that can be found in a port.

This chapter provides the coastal and ocean engineer an understanding of the types of port and harbor facilities that they may be involved in, including the individual structures that make up these facilities and the loads imposed on them. It is intended to provide an overview of these structures, so that the coastal and ocean engineers can better provide the design support to the marine structural engineers in the planning and designing of such facilities.

31.1 Port and Harbor Layout and Design

In the design of ports and harbors, the design team must be aware of the site-specific environmental conditions (such as tide and storm surge, current, waves, sediment transport), the type of marine terminal that is being constructed, and the various vessel types that may call on the terminal. Ports can be offshore terminals or within naturally or artificially protected harbors, bays, estuarine bodies, or rivers. Offshore terminals may be on either reclaimed or natural islands joined to the land by trestles or pipelines or using buoys and pipelines to shore. The function and capacity of the terminal have an effect on the types of vessels that will call on the terminal and in turn dictate factors in the design of the navigation channel, berth depth and length, environmental conditions that impact berthing and mooring, and the overall layout of the facility. This section will

discuss the geometric layout of various harbor facilities as well as the vessels that call on them. These factors will dictate the design effort that coastal and ocean engineers will have to put in during the overall design of such facilities and harbors. Figure 31.1 illustrates typical features found within a harbor and includes the design of the following elements:

- Navigation channel – A channel deep enough and wide enough to allow for vessels to access the port and pass if two-way traffic is required or demanded.
- Turning basin – An area at the end of the navigation channel that allows for the vessels to turn around and maneuver into and out of the harbor. Typically, the basin is at a preferred minimum, twice the length of the largest vessels calling on the terminal.
- Berthing basin – A dredged area adjacent to the wharf structure that allows for the berthing and mooring of vessels.
- Berth – A dock structure with sufficient length and lateral capacity to support berthing and mooring loads. Principal types include:
  - Wharf, quay, pier, or jetty structure: Wharves or quays are typically oriented parallel to shore while piers or jetties project out from shores and include decks for loading and unloading commodities. Terminology varies with locale. These types of berthing structures typically handle container (Fig. 31.2), break bulk, or general cargo, that is, noncontainerized cargo shipped in bags, barrels, drums, palletized, loose, etc., and handled by stevedores, or roll-on/roll-off cargo, pushed, pulled, or driven directly off the ship requiring laydown space for loading and unloading until moved off the dock. Mooring dolphins at ends of the berth, described below, are often provided to tie up bow and stern lines obviating the need to extend the deck beyond the limits required for loading and unloading.
  - Dolphin-type docks include multiple freestanding structures (breasting and mooring dolphins), including pile clusters, multiple piles supporting a cap equipped with mooring hardware, and fenders or solid-filled structures, such as circular cofferdams. An unloading platform and a shore access trestle are typically provided to support pipes and equipment for transferring product handled to shore. Narrow walkways or *catwalks* are usually provided between access platforms and the dolphins to facilitate mooring line handling. These types of facilities typically handle liquid (Fig. 31.3) and dry-bulk materials (Fig. 31.4) that are transported using pipelines or conveyors.

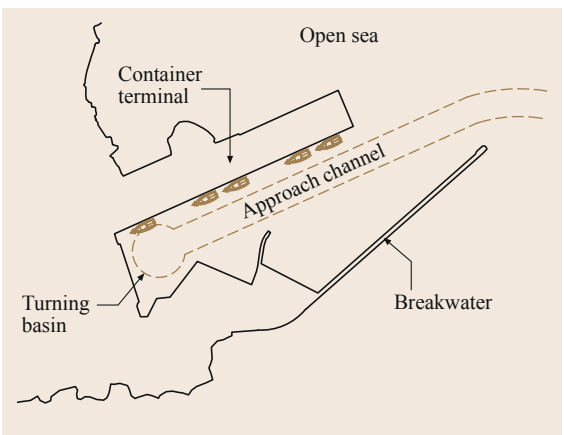


Fig. 31.1 Typical port features

Fig. 31.2 Container terminal – Antwerp (courtesy of AECOM)

Fig. 31.3 LNG terminal cove point Chesapeake Bay (courtesy of AECOM)

Fig. 31.4 Dry bulk berth (courtesy of AECOM)

- Fender systems – Elements that allow for the absorption of berthing loads to prevent damage to the wharf structure and the vessel hull.
- Mooring hardware – Elements, such as bollards, bitts and cleats to safely tie up vessels during various environmental conditions. It also includes motorized capstans to assist longshoremen or stevedores with line handling.
- Upland facilities – Depending on the type of facility, the area required to perform loading and offloading operations as well as to provide temporary storage of goods or product. It also provides connection to intermodal transportation of goods.

31.1.1 Types of Marine Terminals

Throughout history, waterborne transportation has been an integral part of the movement of goods and people. Today, waterborne transportation accounts for the largest volume of goods transportation in the world. With the advent of containerization in the mid-to-late 1950s the cost of moving large quantities of goods around the world has allowed for the continued economical shipment of goods. Additionally, the shipment of bulk materials, such as oil, natural gas, coal, grain, and minerals has continued to grow. Although air travel has taken over from vessel travel for passengers, there continues to be significant passenger trips by vessel in pleasure cruises and in ferry movements. Ports can be dedicated to a single commodity or handle multiple cargo types. Ports range in size from small single-berth facilities to large ports containing dozens of berths. Typical port facilities can be classified into the following categories.

Container Terminals

These ports handle cargo that is shipped in specialized containers. To unload these vessels, the port must have specialized cranes designed to lift the containers from the vessels to the wharf.

Most container terminals use ship-to-shore cranes (Fig. 31.2) that are fixed on rails and that can gantry on the rails along the length of the wharf to unload the vessels. This, however, can also be accomplished with mobile harbor cranes traveling on wheels or tracks and lifting with outriggers that can be fitted with a spreader to handle containers or other equipment to unload the noncontainerized cargo. Mobile harbor cranes are typically more common on smaller or multipurpose terminals. One of the critical factors in the design of these facilities is the wave condition at the berth, as the vessel must remain stable during the loading and unloading of containers. Once unloaded, imported containers are transferred from dockside using straddle carriers or onto trailers/bombcarts and then into the yard. Straddle carriers, rubber tire gantries (RTGs), rail-mounted gantries (RMGs), or reach stackers, stack the containers for short-term storage in the container yard until shipped to final destination either by rail, truck, or smaller vessels (sometimes call short sea shipping). Exports work in reverse.

Liquid Bulk Terminals

These port facilities handle the loading and unloading of liquid materials, such as oil, gas, and liquefied natural gas (LNG) (Fig. 31.3). To facilitate the transfer of liquid material from the vessel, these facilities typically have loading arms installed on the dock structure that maneuver flexible hoses to the ship deck. These hoses are connected to pipelines that allow for the transfer of material to and from storage tanks located upland of the dock. Wave conditions that these facilities can operate under are typically greater than that of container facilities due to the flexibility of the transfer hoses. With certain products, such as LNG, additional space is required in the navigation channel and turning basins for ship maneuvering due to safety concerns. The product may be shipped into or from the terminal either by pipeline, truck, or rail.

Dry Bulk Terminal

These port facilities handle the loading and unloading of dry bulk material, such as coal, iron ore, minerals, grain, and sugar (Fig. 31.4). Its basic and original form, a derrick or crane with a clamshell or other bucket is used for loading and unloading assisted by bobcats or front-end loaders to pile material. To facilitate and automate the transfer of material to and from the vessel, these facilities are typically equipped with specialized bulk-handling equipment, such as conveyor belts, screw conveyors, stackers, reclaimers, shiploaders, hoppers,

and dumpers. The upland facilities typically consist of storage facilities, such as stockyards, storage silos, or stockpiles. The operational conditions tolerable at these facilities are dependent upon the type of material handling equipment used, as some operations can allow vessel movement, while others require a very stable environment. Bulk materials are distributed via trucks or rail to their ultimate destination.

Military Terminals

These port facilities support military operations and are typically used to homeport naval vessels. The sizes of vessels calling on these ports can vary greatly from small craft to destroyers and submarines to the largest aircraft carries. As such, the navigation channel must be adequate to allow for the largest vessels to access the dock facilities. The design of the harbor must also take into account security measures. Typically, the wharf facilities are outfitted with lifting equipment to load materials and supplies to and from the vessels and to support minor repairs. Additionally, the wharf structures are equipped to supply power, communications, water, and sewage utilities to the vessels while they are in port. Special care is required for berthing of submarines and aircraft carriers typically involving the use of camels to hold the vessels off the face of the dock, in the case of aircraft carriers due to the overhanging decks and in the case of submarines which lay up against the dock below waterline. Some specialized facilities include double-deck piers that allow servicing the ship on a lower deck and loading/unloading operations on an upper deck.

Ship Repair Facilities (Drydocks)

These port facilities are designed to support the building and repair of vessels of various sizes. Typically, these facilities have wharf structures that facilitate working on the vessels in the dry. This can be accomplished in a variety of ways, including graving docks (dry docks), floating dry docks, ship lifts, and marine rail facilities. As these facilities may require ships to remain at berth during heavy weather conditions, consideration must be made for heavy weather mooring in their design.

Cruise Ship Terminals

These port facilities provided for the loading and unloading of passengers embarking on pleasure cruises. As this industry has grown in the recent past, the size of the vessels in the cruise fleet has continued to grow and the facility designs have had to accommodate these changes. The wharf facilities typically are outfitted with movable gangways to allow for the safe boarding of passengers as well as security zones for customs and immigration processing. In some harbors where adequate berthing facilities are not available, offshore

Fig. 31.5 Whitehall ferry terminal (courtesy of AECOM)

mooring locations are used and passengers are shuttled to and from the vessel via ship tenders.

Ferry Terminals and Transfer Bridges

These port facilities (Fig. 31.5) can either accommodate just passengers, or passengers and vehicles. To accommodate vehicles, the wharf facility must include a platform or ramp that can support a ramp from the vessel, or provide a movable ramp for the vehicles to travel over. The end of the ramp is suspended and supported from a gallows tower or a pontoon. Similar transfer bridges (Fig. 31.6) serve similar function allowing rail cars to be transported across a harbor or river. Fendering and mooring elements are designed to accommodate the unique characteristics of the ferries that call on the facility.

Fishing Ports and General Cargo Terminals

These port facilities are typically deigned to support a wide variety of vessels and the design of berthing, and mooring structures must accommodate these requirements. Fishing wharves typically do not have heavy lifting equipment but the upland facilities must support

Fig. 31.6 Rail transfer bridge (courtesy of AECOM)

a wide variety of processing and logistics operations. General cargo wharfs may or may not provide lifting equipment, and if they do, it is usually mobile harbor cranes that require heavy live load capacity to handle the variety of material that the port services. Cargo arrives and is typically transported from these terminals either by rail or truck.

Small Craft Harbors and Marinas

These port facilities are typically smaller in nature than the previously described facilities but have their unique design considerations. These port facilities typically include floating docks that are fixed in place by anchors or guide piles. For their design, the tidal range, currents, and wave conditions are critical.

31.1.2 Vessel Overview

When designing for and dealing with ports, it is important to first have an understanding of the vessels that utilize them. The world shipping fleet contains hundreds of types of ships of many different sizes and measurements, with many different distinct uses and functions. Despite the variance in ship type, there are measurements and size classifications common to all shipping vessels.

Some of the most common cargo ships today include general cargo, container, tanker, bulk, car carrier, and cruise ships. Each of these types of ship varies in the cargo it transports. *General cargo vessels* have been plying the seas for thousands of years and can handle a multitude of goods and products. In order to speed and facilitate vessel loading and unloading, dedicated purpose vessel were developed in the 20th century. *Container ships* transport goods in stacked containers. Tankers carry petroleum products in tanks. *Bulk carriers* transport bulk items such as ores and raw materials that are held in bays on board. *Tankers*, *ATBs*/*ITBs* (articulated and integrated tug barges, a barge with a notch at the stern that a tug noses into an articulated connection or is rigidity connected (ITB)) carry oil and chemicals in tanks. Liquefied natural gas (*LNG*) and liquefied petroleum gas (*LPG*) vessels transport fuel in bulk, liquified, and under pressure. *Car carriers* or *RO-RO vessels* transport vehicles that are unloaded by driving across ramps incorporated into the vessel, hence the name, Roll-on, Roll-off. *Cruise ships* transport people. In river and estuarine bodies, smaller versions of these vessels may operate or cargo may be transported in or on *barges* either self-propelled or towed or pushed by *tugs* singly or lashed together in groups. Barges include *deck barges* that cargo is placed on the deck, *scows* where principally dry bulk commodities are placed compartments or bays (a smaller version of a bulk carrier) or *carfloats* which is a barge with tracks on deck for transporting railroad cars. Figure 31.7 illustrates some of these different types of vessels.

Additionally, the world's fleet of vessels includes military and coast guard vessels, pleasure craft, fishing and exploration vessels, ice breakers, construction equipment, and service craft. Military vessels include aircraft carriers, battleships, destroyers, submarines, cruisers, and support/supply vessels often similar to commercial fleet vessels. Service craft include pilot boats to guide a large vessel from the open sea to port, tug boats to assist vessels maneuvering in confined waterways and during docking, barges to fuel and offload waste and bilge water from ships, as well as floating

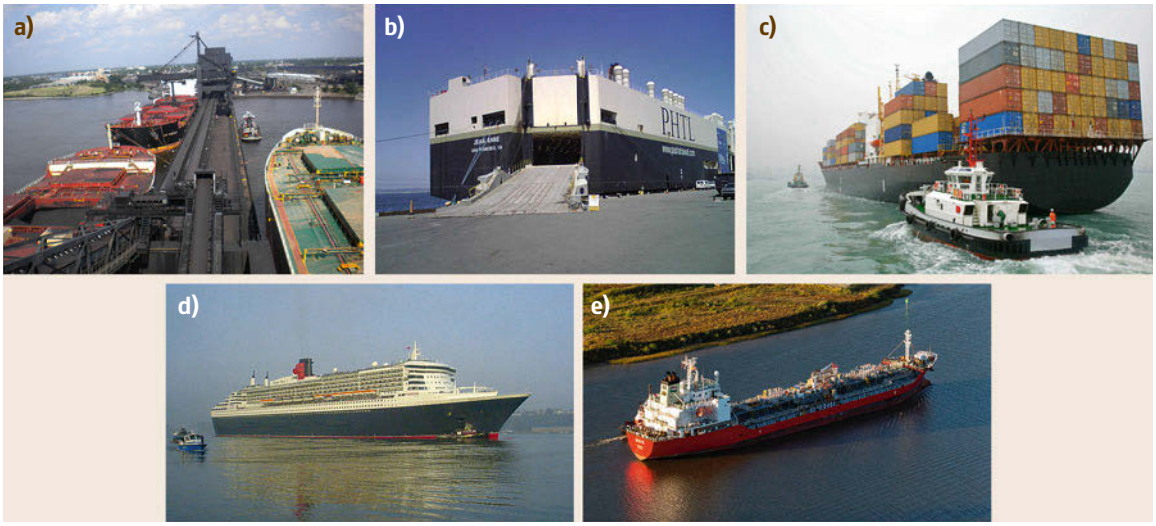


Fig. 31.7 Vessel types (courtesy of AECOM)

drydocks and other repair-type vessels. Construction equipment includes derricks, barges, tugs, dredges, and various specialty craft.

While designed for different purposes and each having unique characteristics, there are common dimensions and measurements between all vessel types of which engineers and designers should be aware.

The most commonly used measurements for ships assess both the physical dimensions of ships and the amount of cargo that they can carry. For harbor or port design, the principal dimensions of a ship include length overall (LOA), length between perpendiculars (LBP) beam (width), averaged draft, freeboard, molded depth, longitudinal (broadside) wind area, and transverse (head-on) wind area. The beam is the measure of the breadth of a ship at its widest point. LOA is the measure of a distance from the bow of a ship to the stern. The draft of a vessel is the vertical distance between the waterline and the deepest part of the ship at any point along the length of the vessel. Other measurements of vessel which are important to the harbor/port design include *displacement* (the displacement is the weight of the water of the displaced volume of the ship), *DWT* (deadweight tonnage, the total carrying capacity of the vessel including cargo, crew, water, fuel, etc. DWT is typically the measurement used to place ships into different classes but is not needed for design rather used only for classification), center of gravity, center of buoyancy, metacenter, and center of flotation. Important coefficients of form of vessel include block coefficient (the ratio of the immersed hull volume at a particular draft to that of a rectangular prism of the same length, breadth, and draft as the ship), midship section coefficient (the ratio of the area of the immersed midship section at a particular draft to that of a rectangle of the same draft and breadth as the ship), the waterplane coefficient (the ratio of the area of the waterplane to that of a rectangle of the same length and breadth as the ship), and longitudinal prismatic coefficient (the ratio of the immersed volume to the volume of a prism with length equal to the ship's and cross-section area identical to the midship section). For container ships, vessels may also be classified by TEU (twenty-foot equivalent unit) capacity. A TEU refers to the container used to ship goods at sea which typical measure 20 ft long by 8 ft high by 8 ft wide (1 TEU) or 40 ft long by 8 ft high by 8 ft wide (2 TEUs). For mooring analysis, the layout of onboard mooring equipment will be important. For berthing analysis, in addition to ship's mass and principal dimensions, hull radius is needed which is sometimes difficult to obtain. For ship navigation studies, the vessel parameters (or key components) that affect the vessel maneuverability will be required; these parameters or components include vessel wind coefficients, current coefficients, wave hydrodynamic coefficients, engine type and power, rudder type specifications, bow/stern propellers, etc.

By using many of the measurements above, ships can be placed into different classes. Classes vary by the type of a ship, but most are broken up into DWT ranges. Handi or Handymax vessels are traditionally the backbone of the dry bulk vessels. Some classes are also derived from trade route restrictions that limit where certain sized ships can travel in the world. Canals, such as the Panama and Suez, offer ships a more direct route to destinations, but the lock systems used in these canals limit the maximum size of vessels that can pass through. The *Panamax* class is an example of a class that was created due to size restrictions in the Panama Canal. The name Panamax refers to the maximum size ship that can fit through the locks in the Panama Canal. *Capesize* vessels are too large for the Suez or Panama canals and must travel around the capes of South Africa or South America when making long voyages. Malacamax refers to the largest that can pass the Strait of Malacca between Malaysia and Indonesia. While different types of cargo ships may share class names, the ranges and limits of the defining measurements may vary. Table 31.1 gives example ranges of some of the different classes.

The world's fleet of ships has changed greatly over the past decades and will continue to do so in the future. In 2011, the world cargo carrying fleet was comprised of over 60 500 ships with a total gross tonnage of a trillion tons. It is predicted that as ships continue to get bigger, a larger percentage of total fleet DWT will come from a fewer number of ships. Ports worldwide are looking at ways to accommodate larger ships, as the size of new vessels keeps increasing. The Panama Canal hopes to have its third set of locks operational by 2016, which would allow post-Panamax ships to make use of the canal. Figure 31.8 illustrates the increase in vessel size that the Panama Canal expansion will allow it to handle.

With a basic understanding of ship types and sizes, it is easier to understand what must be considered when working on a port-type project. Different types of cargo ships have different characteristics and measurements that must be taken into account to design the most effective marine environment.

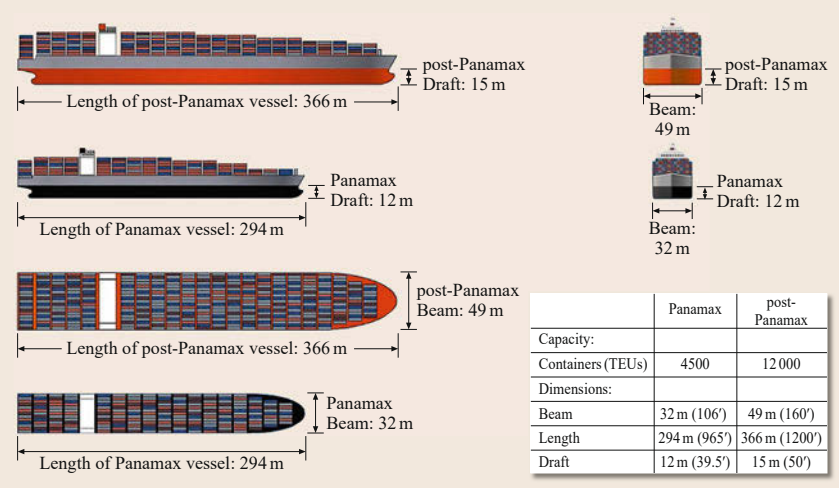
31.1.3 Harbor Operational Limits

Operational limits can be divided into different levels, from cargo handling limit based on maximum allowable vessel movements at berth to the upper limit that a vessel can stay at berth, but with cargo handing ceased, then to the ultimate limit that a vessel can stay at berth

Table 31.1 Example of vessel class

Vessel class	DWT [t]	Other capacity measurement	Beam [m]	Length [m]	Draft [m]
Handymax	30 000–50 000	–	27–	150–200	10–
LNG[a]	60 000–160 000+	75 000–175 000 m^3	35–48	260–300	10–12
Panamax	65 000–80 000	5000 TEU	32.3	294.1	12
Post-Panamax	–	12 000 TEU	49	366	15
Suezmax	150 000	10 000 TEU	48	274	16
Capesize	150 000+	–	> 32	–	> 20
VLCC (very large crude carriers)	200 000–300 000	–	58	330	22
ULCC (ultra large crude carriers)	300 000–55 000	–	68	380	–
Malaccamax	–	–	60	470	24
ULCS (ultra large container ship)	200 000+	18 000 TEU	57	–	16.4

[a] Representing approximately 80% of world's fleet in 2009



	Panamax	post-Panamax
Capacity:		
Containers (TEUs)	4500	12 000
Dimensions:		
Beam	32 m (106′)	49 m (160′)
Length	294 m (965′)	366 m (1200′)
Draft	12 m (39.5′)	15 m (50′)

Fig. 31.8 Panamax vs. post-Panamax

during an extreme storm event before departing to the open sea. There are also environmental limitations for vessel navigation safety (including limiting wave conditions for tugs) and vessel berthing. For some old terminal facilities, the structural capacity of the pier or wharf may limit conditions when the moored vessel may stay at berth. In general, the harbor operational limits will depend on the type and size of vessel, the mooring facilities on board and at the terminal, and the cargo handling equipment. The most commonly used criteria is [31.1] and when published the updated criteria from PIANC MarCom 52: Criteria for the (un) Loading of Container Ships. Some countries also develop their own criteria for harbor operational limits, such as BS 6349 Part 1 [31.2].

31.1.4 Harbor Navigation

For harbor navigation design, the fundamental elements to be considered include design vessels and their maneuverability, layout of navigation components (navigation channel, harbor entrance, stopping area, turning basin, berthing basin, and anchorage area), aids to navigation, requirement/availability of tugs (Fig. 31.9) and pilots, and limiting environmental limit conditions in which maneuvers can be performed. The approach channel and harbor basins are sized for the design vessel. In most cases, due to the variation of types and sizes of vessels calling on the terminal, several design vessels will be needed to ensure that all other vessel using the channel and basin can safely navigate. The navigational design usually includes channel alignment, cross section, and depth. During the detailed design phase, fast-time ship simulations and real-time ship simulations are typically employed to reduce the uncertainty and improve the design. The assessment of marine risk and safety of operation are very important issue in the modern harbor navigation design and usually include traffic simulation for the estimation of marine risk, to employ aid to navigation and ves-

Fig. 31.9 Tanker berthing under tug assist with concrete dolosse breakwater in foreground (courtesy of AECOM)

sel traffic service, and establishment of harbor safety criteria. For navigation channel design, [31.1], [31.3] and [31.4] may be used. For aids to navigation design, the International Association of Lighthouse Authorities (IALA) Aids to Navigation Manual [31.5] can be referenced.

31.1.5 Sediment Transport Considerations

In general, the evaluation of maintenance dredging is an important consideration in the planning and design of a harbor. The methodology for the evaluation of maintenance dredging is significantly different for mud transport and sand transport. In most situations, the reasonable evaluation of maintenance dredging is much more complicated than most people realize. Typically, a numerical model is employed and, in some special cases, a physical model may be needed. Regardless of the modeling technique employed, the wave–current climate and sediment source must be studied first. It should be realized that though extreme storm events could in some cases dominate sediment transport, moderate waves and current could also have a large contribution to the long-term sediment transport because these events occur very frequently. Maintenance dredging of navigation channels and harbor facilities restrict or complicate ship movement operations, and dredging costs are a burden for most operators of harbors. These burdens have become worse due to stricter environmental restrictions on dredged material disposal. To cope with this increasing challenge, the most important step is to perform a study of key coastal processes (tide, current, wave, and sediment transport) and to have a sound understating of the siltation mechanism before any attempt is made to develop mitigation strategies. Typical mitigation strategies include basic structural approaches to promoting the passage of suspended sediments through facilities to limiting the entrance of water with suspended sediment from entering the harbor area including measures such as training walls, jet array systems, and moveable curtains, etc.

31.2 Structure Types

In the design of ports and harbors, there are various types of structures that need to be considered. The configuration of these structures responds differently to the wave, wind, current, and tides that coastal and ocean engineers evaluate. This section describes the various types of structures as well as the loads on structures due to vessel mooring and berthing as and due to ocean environmental conditions.

31.2.1 Breakwaters and Wave Attenuation

A breakwater is a coastal defense structure employed to reflect and dissipate the wave energy and thus prevent or reduce wave action within the harbor. Breakwaters can also sometimes serve to guide the current and improve navigation. In other cases, a breakwater can reduce the amount of dredging required in harbor entrance by blocking the longshore sediment transport. In some harbors, breakwaters serve dual functions for wave protection and quay facilities. There are three main types of breakwaters, namely, rubble mound breakwater, vertical breakwater, and composite breakwater (Figs. 31.9 and 31.10). For the majority of cases, the primary func-

Fig. 31.10 Concrete block breakwater at port of Haina – Dominican Republic (courtesy of AECOM)

tion of a breakwater is wave protection. The degree of protection required (or allowable wave conditions within the harbor) will depend on the vessel characteristics, harbor operational and downtime requirements. In practice, a numerical wave model with diffraction capability needs to be employed to investigate the optimum layout of breakwater and wave climate within the harbor. For breakwater design, bathymetry and geotechnical considerations are very important because it will determine the applicability of various types of breakwaters, the volume of material needed to build the breakwater, and scour requirements. Wave height and their frequency of occurrence (both during normal conditions and extreme storm events) are the most important inputs for the optimum design of a breakwater. Wave run-up and overtopping analysis also need to be conducted to determine the elevation of the breakwater. Guidance for the design of breakwaters may be found in [31.6, Chap. 5]. For navigation safety, aids to navigation are typically needed at the head and along the trunk of breakwater. During design, the availability and cost of construction materials and equipment need to be considered. Revetments often found adjacent to or beneath in the case of pile supported marginal wharfs. Design principles for revetments are similar to breakwaters and in addition to [31.6] guidance may be found in [31.7].

31.2.2 Open Pile Platforms

Open pile platforms allow water to flow underneath and are supported on steel, concrete, timber, or composite piles or drilled shafts. Typically, a solid deck is provided, except for liquid and dry bulk facilities, which use pipelines, conveyors, or other methods to transfer bulk products, where only a frame may be provided. Deck types can include either concrete two-way flat plate slabs or concrete pile caps with one-way deck slabs with the pile caps typically oriented normal to the pier's or wharf's length to take advantage of the pile cap's greater strength to resist lateral loads due to berthing or mooring. One-way deck systems may be cast in place but more commonly use precast planks (solid or voided) or less frequently beams (box, tee, etc.) since such deck components can be more easily precast and erected rapidly. Two-way deck systems with or without pile caps in one or both directions are also common; these usually employ longitudinal beams/girders to accommodate heavy wheel loads, such as container and other travelling cranes, often on rails. Two-way slabs without beams are less common since a cast-in-place deck is usually necessitated requiring a high-cost formwork over the water; drop panels over piles if used further increase costs. A deck cast on ground and then soil below removed is also possible, but there is a challenge in removing the soil below. Two main types of open pile platforms include high- and low-level platforms. In addition to berthing and mooring loads discussed hereafter, the design of open pile platforms must consider current and wave loads on substructure elements. Guidance may be found in [31.6, Chap. 5].

Low-Level or Relieving Platforms

Relieving platforms (Fig. 31.11) are structures where fill is placed atop a structural deck set several feet below the paved surface, often just above mean low water. Historically, these types of platforms were usually constructed on timber piles whose butts were set just above mean-low water to minimize rot and replacement of the pile in the tidal zone area and areas subject to wetting and drying, highly rot susceptible. As waterways became highly polluted at the end of the 19th century, resulting in the demise of marine borers, their popularity increased in coastal waterways since timber piers no longer were subject to this type of attack and a long-life achievable with rot concern eliminated by virtue of their continual submergence. When clean water regulations were developed and the waterways became cleaner, borers returned making timber less desirable in coastal waterways; however, with the advent of modern higher capacity materials including steel and concrete piles, low-level platforms remained viable and economical allowing the use of fewer higher capacity piles that are out of the highly corrosive splash zone by virtue of a low deck. Relieving platforms are highly effective in distributing heavy concentrated loads through the soil to the deck resulting in a more uniform load that can be more economically constructed. They are also useful as platforms over highly compressible soils, *relieving* superimposed material storage loads which

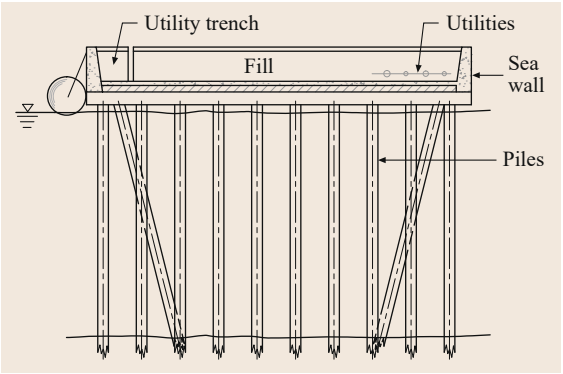


Fig. 31.11 Low-level relieving platform

would otherwise result in an unacceptable settlement of the facility. Finally, low-level platforms offer the advantage of a no cost space for utilities within the fill atop structural deck, especially useful for military, service, and outfitting piers. Their main disadvantage is their intrinsically higher cost since they support higher dead loads and their construction is limited to periods of lower water.

High-Level Platforms

High-level platforms (Fig. 31.12), where the deck is the finish pavement, are the most common type of open-pile platforms. They are usually less costly than low-level platforms since construction can proceed uninterrupted by tide cycles and, with the possible exception of heavy moving concentrated loads, the deck structure need not be as robust as a low-level platform. Hence, they can usually be constructed at lower cost than a comparable low-level structure. The height of the pier deck of high-level platforms allows easier access and more space below deck; hence, they are easier to maintain. However, the piles in the tidal and splash zone are more susceptible to deterioration and either requires special and more costly design consideration and protection or increased maintenance.

Hybrid Platforms

These include both high- and low-level platforms, most frequently as existing structures supported on timber piles. Usually, this type of platform will include piers with high-level aprons whose piles have long exposed lengths and an inner core on a low-level platform with piles of short exposed length. Typically, the mudline is mounded toward the center for this type of structure to take advantage of the higher capacity of short piles. Another variation of a hybrid platform is where piles are near or at the water level and concrete walls or truncated pyramid pedestals are constructed up to a high-level deck.

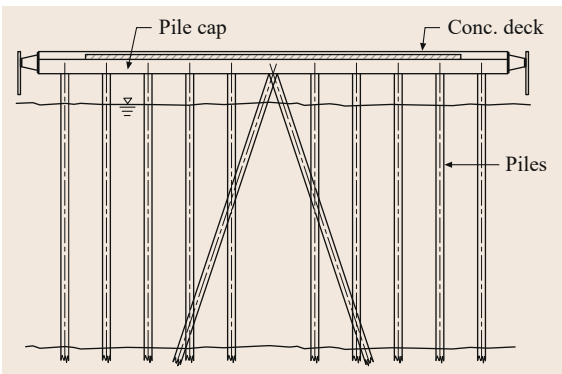


Fig. 31.12 High-level platform

31.2.3 Solid Structures

Solid structures use a wall to retain fill that supports the working structure. Solid structures do not allow water to flow beneath and interrupt or disrupt water circulation patterns. Solid structures additionally provide a berthing cushion which temporarily traps water between the vessel and the berth. If bulbous bow ships will use the facility, the bulkhead is typically set back from the berthing line. Lateral forces due to wave loading on solid structures can be found in [31.6, Chap. 5].

Sheet Pile Bulkheads

Sheet pile bulkheads include cantilever bulkheads, anchored bulkheads, and cellular cofferdams. Sheet piling consists of rolled steel, Z- or U-shape sheets or flat sheets both with interlocks.

Cofferdams

Cofferdams (Fig. 31.13) have been around for thousands of years and the Roman Architect Marcus Vitruvius Pollio wrote of them in *De Architectura* in 15 BCE. In modern day cofferdams, flat sheets are used in cellular cofferdams, either circular with connecting arcs or diaphragm type with connecting straight cross walls. Cloverleaf types are also used. Cofferdams attain their strength as gravity structures relying on hoop tension in the cell or diaphragm to contain the earth fill and lateral mooring loads.

Cantilever Bulkheads

These rely solely on embedment in the earth for their support. Since embedment is typically in the range of one to two times exposed wall height and stresses increase rapidly, cantilever walls are generally economical for shorter heights, most typically used on rivers and small craft facilities. Cantilever bulkheads are also used in locations where anchoring the bulkhead is not feasible.

Anchored Bulkheads

Anchored walls use either deadmen and tiebacks or soil anchors to hold back the top of the wall, thereby significantly reducing the embedded depth of the bulkhead with the most common embedment of 40–60% of the exposed height. This makes anchored bulkheads practical and economical generally up to about 45–50 ft exposed height using Z sheeting. However, modern terminals can typically exceed these heights and *combi-* or *king pile* bulkheads (Fig. 31.14) have been developed for these facilities that use large diameter pipe- or H-pile sections, by themselves or in combination with Z sheets, allowing bulkhead heights to accommodate

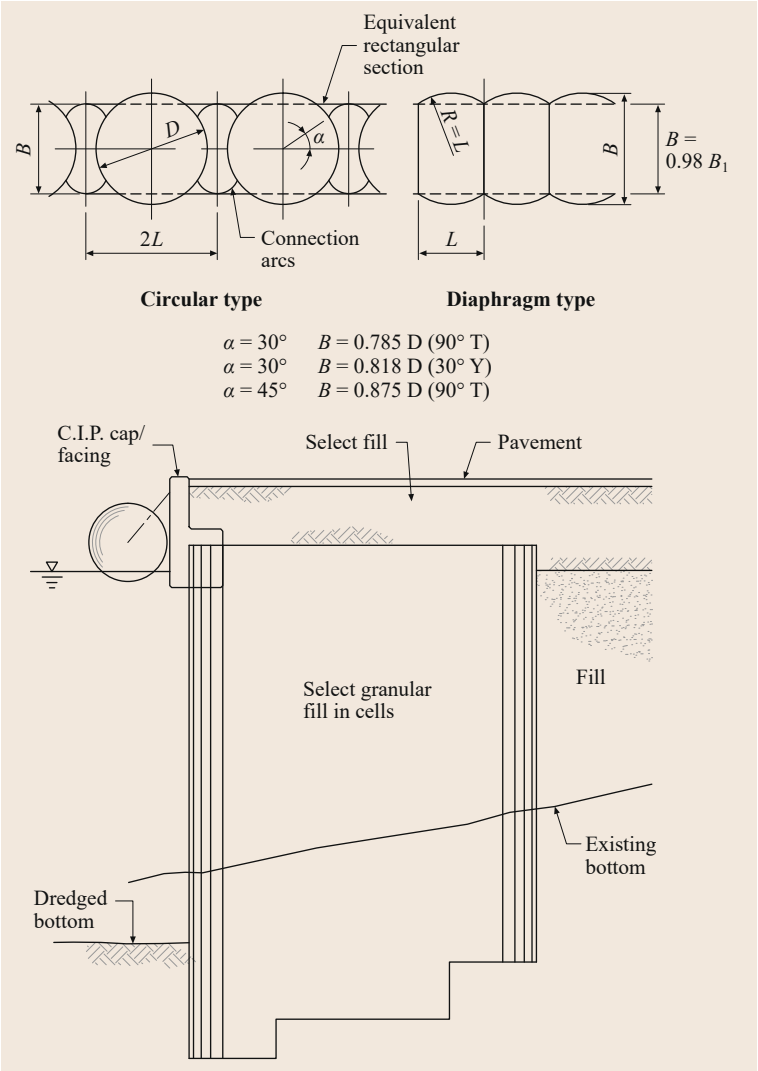


Fig. 31.13 Cellular cofferdam

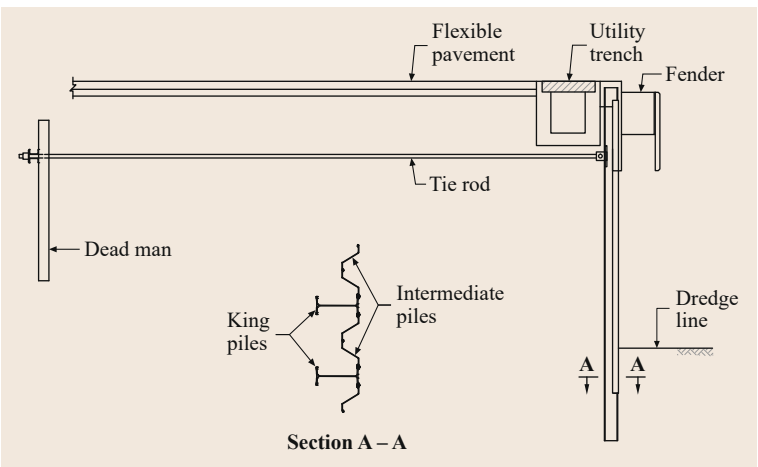


Fig. 31.14 Anchored king pile bulkhead

deepest draft vessels. Sheet pile structures are usually built in the waterway (in the wet); however, they can and have been constructed upland and the berth excavated afterward. For the lowest structural demand on the wall, the anchor should be as low as possible but above water for practicality of installation. Typically, tie rods are set around the mean tide level allowing installation for most of a usual 8 h work day, roughly half a diurnal tide cycle.

All sheet pile bulkheads in salt water ports have the ubiquitous problem of corrosion, especially in the tidal zone. Even the latest coating systems cannot be expected to last more than 20–30 years and therefore should have provision for cathodic protection (CP) as coatings degrade or a corrosion allowance. Since CP systems are only partially effective in the tidal zone, and ineffective above in the splash zone, alternative future provisions, such as concrete facings should be considered. Aluminum, fiberglass, vinyl and other composite plastics are also available; however, they are used for smaller marine bulkhead structures.

In addition to earth pressure, the principal forces acting on bulkheads include wave forces, tidal lag, ice thrust, mooring and berthing loads, scour and propeller swash. Guidelines for the design of sheet pile bulkheads are found in [31.8] for cofferdams [31.9] and the commonly referred to [31.10].

Gravity Wall Structures

Gravity-retaining structures typically constructed of concrete rely on their weight to resist lateral loads of the soil and mooring forces. The most common types of gravity wall structures include block walls, retaining walls, caissons, and crib structures.

Block and Retaining Walls

Block walls (Fig. 31.15) consist of massive concrete blocks stacked into place either in the wet or in an excavated dry basin. Similarly, massive retaining walls are possible but uncommon. The concrete blocks are often un-reinforced and are not subject to typical corrosion found in reinforced concrete structures.

Caisson Walls

Caissons (Fig. 31.16) are box- or egg-like concrete structures with a bottom; typically, they are floated into position, filled with rock or earth, and then sunk into place. Block wall and caisson structures are extremely durable but uncommon in the United States due to their high initial capital cost, especially massive material quantities. They are common in Europe, Middle, and Far East, where long-term viability and sustainability are desirable, and especially in countries with emerging economies where the cost of labor and materials can be significantly less.

Cribs are similar to caissons but constructed of timber or concrete *Lincoln log style*. Cribs are used on smaller marine structures, more commonly in the past than present. Concrete cribs are typically erected in place without a bottom, whereas timber cribs often had a bottom and were floated into place and filled with rock.

Secant Pile, Diaphragm, or Concrete Slurry Wall

These structures are basically modern adaptations anchored sheet pile bulkheads using reinforced concrete instead of steel sheets, developed in Europe and used there and elsewhere since the 1940s. These types of walls typically avoid the corrosion issues of steel bulk-

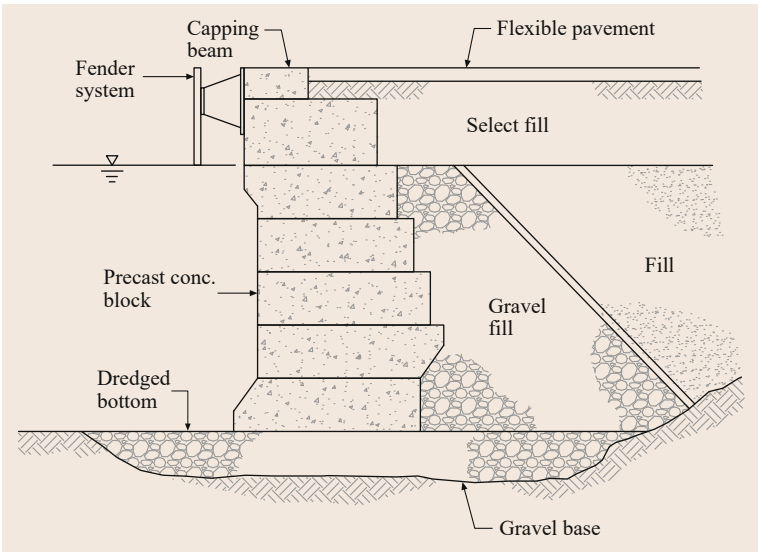


Fig. 31.15 Concrete block wall

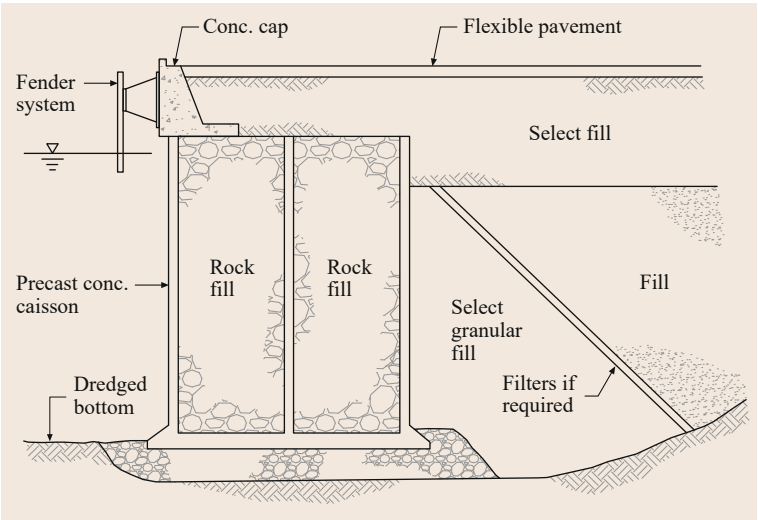


Fig. 31.16 Concrete caisson wall

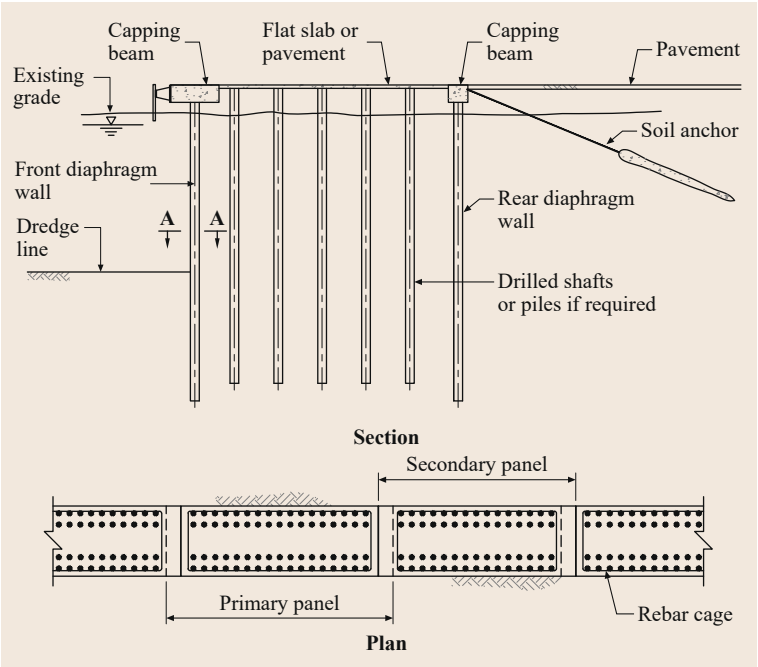


Fig. 31.17 Concrete diaphragm wall

heads, although concrete design mixes, cover, and rebar protection must be carefully considered. They are typically used in *dig-out* ports where the bulkheads are constructed in the upland and the berth excavated later. Secant walls consist of overlapping cast-in-place circular piles or drilled shafts, whereas diaphragm walls (Fig. 31.17) consist of overlapping rectangular piles or panels. Drilled shafts are sometimes referred to as caissons in foundation engineering. This terminology should be avoided in marine and coastal engineering to avoid confusion with caissons described above. These walls are typically drilled in, with a casing if conditions warrant, which is then filled with a bentonite slurry; a reinforcing cage is then lowered into the slurry and the hole concreted by displacing the slurry and the casing if provided is withdrawn. They are typically installed in a skip-step pattern filing in and overlapping the skipped piles or secondary panels on a second pass. A continuous slurry wall can also be used.

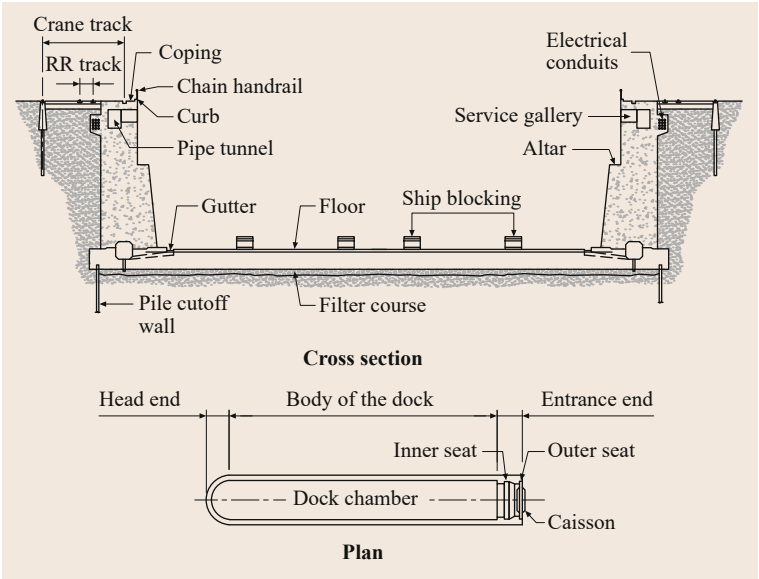


Fig. 31.18 Graving dock

Summary

While some type of structures may be more common in a particular locale, region, or area, there is no one-size-fits-all choice for the selection of structure type. Table 31.2 presents some of the considerations in the selection of structure types. Local practices, regulations, contractor familiarity, design life, and costs can heavily influence structure selection.

31.2.4 Drydock Facilities

Dry docks [31.11, 12] are used for the repair, overhaul, and painting of ships above and below the waterline. Principal types include graving docks, floating drydocks, and marine railways.

Graving Docks

Graving docks (Fig. 31.18) are large, fixed basins including a floor, sidewalls, head wall, and a gate. Altars or steps are often incorporated into the side walls for structural stability. Since they are at water's edge and below groundwater, graving drydocks must either be massive heavy structures or employ dewatering systems to reduce hydrostatic pressure. The choice is usually economic one, balancing the high operational cost of a relieved drydock vs. the higher initial capital cost of nonrelieved docks. The principal critical loads acting on drydocks include:

- Dock under construction
- Dock empty; maximum hydrostatic uplift
- Maximum ship load; minimum hydrostatic uplift
- Dock full of water

The drydock is separated from the waterway by a gate which resists the water pressure when the dock is dewatered. There are four principal types of gates, each with their own advantages, disadvantages, and design considerations. They include *miter gates* that consist of a pair of gate leaves hinged at the dock walls that swing open horizontally, sides and bottom fit against seats in the drydock walls and floor when closed. *Set-in-place gates* are available in various forms and may be built in one piece or multiple sections of beam and plate construction. These gates are extensively used for small docks. Their placement and removal must be done by weight-handling equipment and large gates render them impractical. The *floating caisson gate* is of the most common type, consisting of a watertight box with flooding and dewatering chambers and systems. After the drydock is flooded, the caisson is dewatered to raise it up off its seat. It can then be towed clear of the dock entrance. *Sliding or rolling caissons* consist of builtup box sections with a sliding or rolling surface at the base. The gate slides or rolls into a notch built into the side of the dock. Flap gates consist of a rigid one-piece gate hinged at the bottom that swings downward and upward. Both miter and sliding gates are similar in design and operation to canal lack gates. In addition to potential wave loads, an important consideration in the design of a caisson gate and graving dock entrance is the siltation potential at the entrance. Table 31.3 provides some guidance on merits of each type of gate.

Floating Drydocks

Floating drydocks (Fig. 31.19) consist of two main parts – pontoon and wingwalls. Pontoons are the main

Table 31.2 Merits of open platforms and solid structures

Factor	Open pile platform	Solid structures	Advantage
Environmental impacts	Minimizes filling of waterway; maintains fish habitat	Reduces habitat significantly	Open pile
Deep depth to competent bearing strata, moderately compressible material above	Long piles costly; splicing is time consuming and cuts productivity	Soil improvements techniques can be used	Solid
Moderate depth of very soft to soft soil strata	Easily penetrated by piles to competent materials below	Requires time consuming soil improvements or costly dredging out and replacement	Open pile
Wide structure or marginal wharf type	High per square foot cost	Pavement and fill relatively inexpensive if fill readily available but costly bulkhead or wall construction	Solid – wide to very wide open pile – narrow to moderate
Maintenance	High long term labor intensive maintenance required/replacement	Steel bulkhead types can be faced with concrete or cathodically protected; block or caisson type virtually forever. Diaphragm and secant type walls fall between	Solid

Table 31.3 Advantages and disadvantages of drydock gate types

Gate type	Advantages	Disadvantages
Miter gates	• Fast operations	• Creates heavy loads on wall supports. Although loads can be relieved by roller and track arrangements, these are not satisfactory for operational and maintenance reasons. The same objections apply to relief of loads by controlled buoyancy in the gates. • Operating mechanism is costly, susceptible to breakdown and requires costly maintenance • Major repairs require removal of gate • Recesses must be built into dock walls to maintain ship clearances with open gates adding to drydock length and complexity
Set in place gates	• Low capital costs	• Heavy and impractical for larger docks • Time consuming to install and require weight handling equipment
Floating caisson gate	• Low cost • Can be towed to other sites or drydocks • May be used in more than one dock • Reversible, exposed side can be repaired in place	• Slow operating • Needs shore power to operate (electric and sometimes air); it is not operative too far from the caisson seat
Sliding or rolling caissons	• Fast operations	• Cleaning and maintenance of rollers or slide paths is difficult • Operating mechanism is expensive • Major repairs require removal of gate • Recesses must be built into Walls
Flap gates	• Fast operations	• Creates heavy loads on wall • Operating mechanism is costly & requires costly maintenance • Major repairs require removal of gate • Recesses must be built into walls to maintain clearances

supporting bodies that displace the weight of the vessel and drydock in order to lift the vessel using buoyancy. The pontoon must distribute the concentrated load of the ship along the dock's centerline to the uniform buoyant support of the water pressure by its transverse strength. The wing walls provide stability when the pontoon is submerged and the longitudinal strength to distribute the irregular ship weight to the uniform

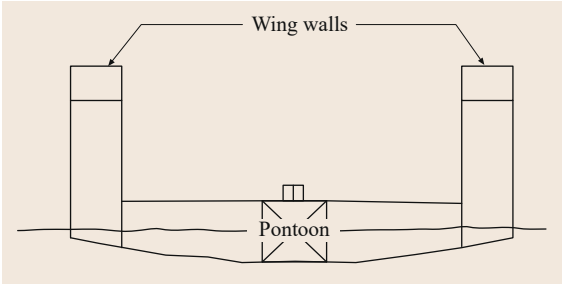


Fig. 31.19 Floating drydock

buoyant support. Some floating drydocks have portal cranes on tracks on the wingwalls and some floating drydocks use cranes from the pier side. The dredge depth at these facilities must accommodate the floating drydock when submerged. Floating dry docks have capacities from a few hundred tons to over 100 000 t. In general, the most economical range for floating docks is about 1000–100 000 t. Advantages of a floating drydock are:

- Does not require valuable waterfront real estate.
- Can be built at the yard of low bidder and towed to the site keeping construction costs low by increasing competition.
- Can be sold on the world market keeping resale values high and making obtaining financing easier.
- Vessels can be transferred to and from shore relatively easily.
- Can be operated with a list or trim when docking vessels with a list or trim. This can reduce block loading and reduce or eliminate vessel stability problems when landing.
- Permits docking of vessels longer than the dry dock by overhanging the bow and/or stern.
- Can be moved for dredging.
- Minimal landslide construction works with potentially easier permitting.
- Can be moved to deeper water for docking and undocking operations, reducing or eliminating need for dredging.
- The dock can be lengthened relatively easily.

Disadvantages of a floating drydock are:

- High maintenance is required on pumps, valves, and steel structure.
- Routing of men and material is restricted to gangplank and/or crane service.
- Large tidal variations can complicate gangways, mooring, etc.

Floating drydocks are normally moored using two or more vertical spuds or mooring arms, sometimes called Crandall arms after the Crandall Drydock Company that pioneered their development, which hold the dock in position throughout its full range of vertical movement from fully submerged to fully dewatered and during normal tidal changes. The pier/wharf structure must be designed to accommodate spud or mooring arm placement and loading.

Marine Railway

A marine railway [31.13] (Fig. 31.20) is a mechanical means of hoisting a ship out of the water along an inclined plane. It consists of an inclined track on a foundation, a cradle, or bogies that move on rollers or wheels to support the vessel and ride up and down the track, and a hauling mechanism to pull and control release the cradle. Two types include endhaul and sidehaul. Endhaul type is easier, safer, and less complicated to pull up and occupy less valuable waterfront frontage than sidehaul or broadside type. A sidehaul marine railway may be the only type that can be used safely on a nontidal river bank if there is no slack water period to permit docking a vessel normal to the current, width, and location, and the traffic up and down the river may preclude the use of an endhaul docking. The sidehaul type is particularly adapted to hauling out vessels with flat bottoms and a shallow draft, such as barges and other river craft. Operationally, the cradle is lowered into the water along the inclined track. The ship is floated over the cradle and tied to the uprights, the cradle is hauled up the track, and the vessel grounds onto the blocks. After complete grounding on the keel blocks, the side blocks are brought to bear and the hauling continues until the cradle is full up. Some facilities employ a lateral transfer system once upland to accommodate multiple vessels using a single railway. Ideally, the slope of the track would be selected to fit the natural slope of the particular site to minimize dredging and yet provides the required drafts over the blocks for docking the vessel; however, tradeoffs usually have to be made between the length and slope. Lift capacities range from 100 to 6000 t, or theoretically, even larger but the floating drydock usually becomes a more economical alternative. Advantages of a marine railway over drydocks are as follows:

- Low initial construction cost
- Fast operating
- The track slope can fit the natural slope of the shore in many cases. This eliminates or reduces dredging or bulk-heading requirements.

Disadvantages of a marine railway are as follows:

- It is a mechanical system that requires the periodic replacement of some moving parts (hauling chains, rollers, etc.).

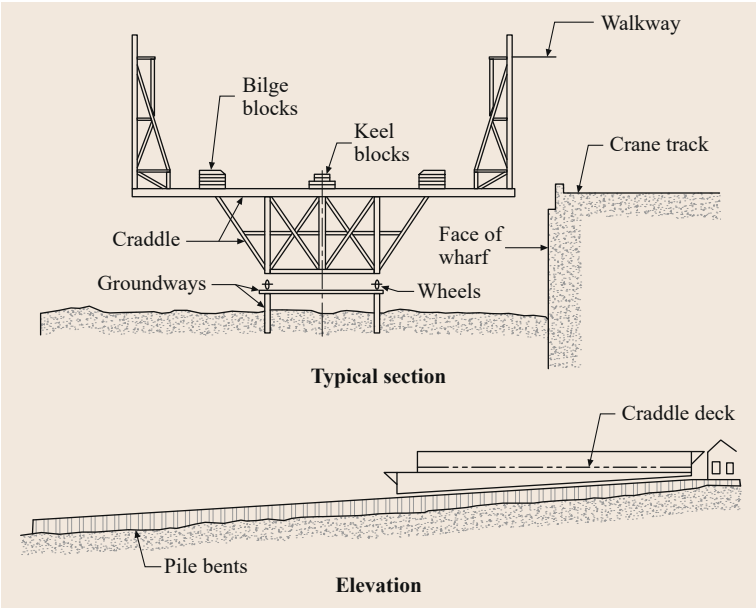


Fig. 31.20 Marine railway

- Handling and transition from tugs to cradle can make handling difficult.
- Underwater maintenance is required.
- The vessels can damage the track.
- Ice can form on the tracks, rollers, and chains, potentially delaying or endangering the docking. Ice can build up at the graving dock entrance, but removal is easy.

31.2.5 Floating Structures

Floating structures use a pontoon structure anchored to the seabed through spud piles or mooring lines and connected to the shore by bridges or ramps. Floating piers are most commonly found in small-craft marinas and include foam-filled concrete floats, steel pontoons, or plastic encased foam-filled pontoons joined and connected with a steel, aluminum, or timber frame and deck. Modular steel pontoon systems are also used for marine construction including shore landings. The Third Lake Washington Floating Bridge in Seattle is a 105 ft wide 1 mile bridge that includes a series of pontoons. The US Navy has developed concepts [31.14] for large concrete floating pier; however, a prototype has not yet been built. Some of the largest floating structures are concrete breakwaters including the world's largest at La Condaime Port in Monaco 352 m long by 28 m wide.

A variation, semibuoyant structure, utilizes a pontoon or an underwater chamber for the partial support of the pier. Similar to floating structures, caissons are floated into place then sunk onto a pile or other foundation. They are useful where firm-bearing material is at great depth and buoyancy provided by caisson offsets the pile loads, reducing the number of piles required. An example includes Pier 57 in New York City.

31.2.6 Swinging Moorings

Swinging moorings, referred to as fleet moorings in Navy parlance, include moorings where the ship is not berthed against a structure; rather, it is allowed to swing freely on a mooring in a basin, river, harbor, or sea. In its simplest form, a ship-dropping anchor is a temporary mooring under benign conditions. A large area or a ship-swing watch area is required (Fig. 31.21). Two or

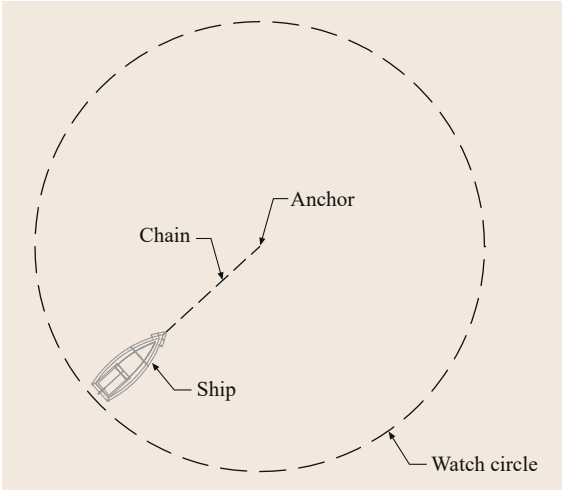


Fig. 31.21 Single-point mooring watch circle

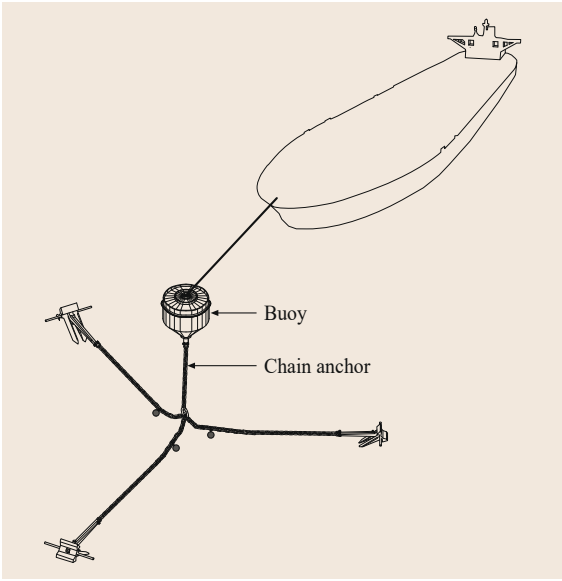


Fig. 31.22 Single-point mooring

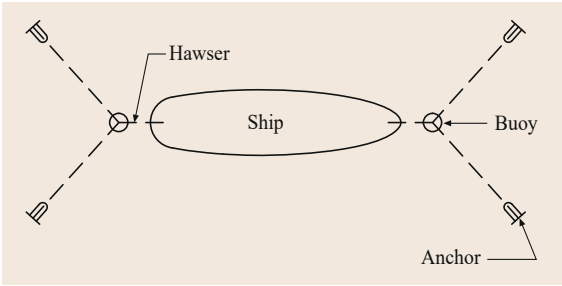


Fig. 31.23 Bow-stern mooring

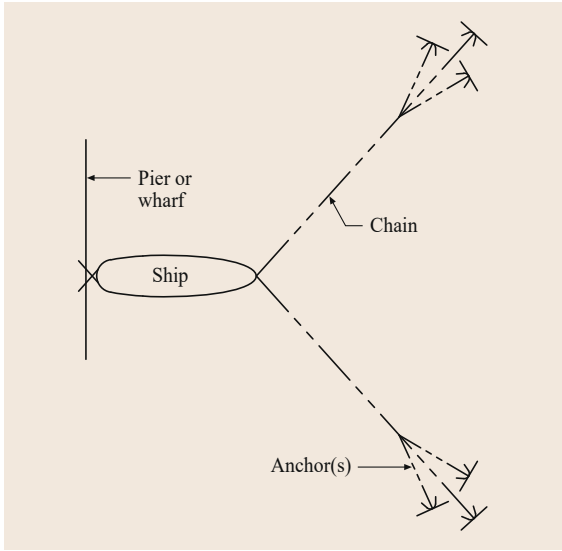


Fig. 31.24 Mediterranean mooring

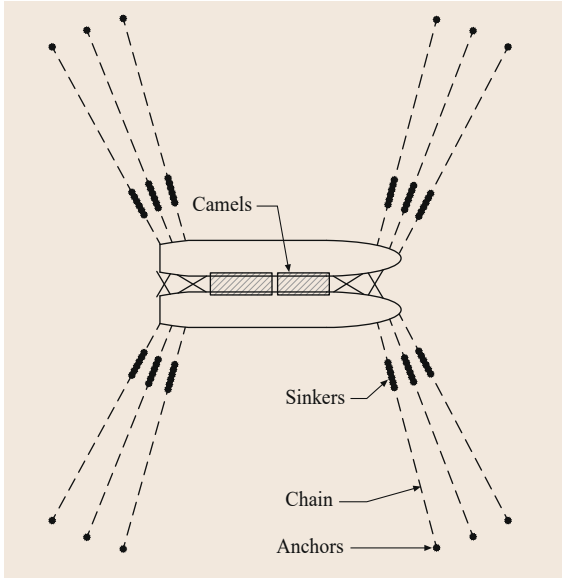


Fig. 31.25 Spread mooring

more anchors may be used to decrease the watch area or deal with shifting wind, crowded anchorages, twisted ropes, etc., including Bahamian mooring, dual anchor mooring, fore and aft mooring, hammerlock mooring, dual tandem mooring, star, and modified star but are usually not the subject of marine engineering, and the readers should, if interested, do their own research on subject. The other type of swing moorings is structural elements to which a vessel is moored including anchors, ground legs, a riser chain, a buoy, and other mooring hardware. These include single-point mooring (SPM), bow-stern mooring, Mediterranean mooring, and spread mooring.

Single-Point Mooring (SPM)

An SPM consists of a buoy anchored to a base on the seabed (Fig. 31.22). There are literally dozens of types of SPMs depending on the method of anchoring the buoy to the seabed, which can include chains acting as catenaries, pretensioned chains, or underwater towers structures. SPMs include riser type where a single chain anchors the buoy and whose bottom is secured with one or more chains. Telephone-type moorings include buoys secured with multiple chains to the seabed. A ship moors to the buoy using an anchor chain or hawser and weathervanes about the buoy depending on the wind and current direction and speed. In its simplest form, SPMs consist of an anchor, catenary chain, and buoy that are used to moor small craft boats on rivers and bays. The largest SPMs are used to moor VLCCs and ULCCs and transfer liquid petroleum prod-

Fig. 31.26 North river ice breakers (courtesy of J. Carel)

uct from the mooring to shore. The design of SPMs is a specialized area of marine engineering. Two references include [31.15] and [31.16].

Bow–Stern Moorings

Bow-Stern Moorings aka double- or twin-buoy moorings (Fig. 31.23) are similar to SPMs except buoys are provided and the ship is secured to buoys at both ends of the vessel. This system has a much smaller watch circle than a vessel at a single mooring buoy. Also, two moorings share the load. However, the mooring tension can be much higher if the winds, currents, or waves have a large broadside component to the ship.

Mediterranean Mooring

Med mooring or Tahitian mooring (Fig. 31.24) the vessel bow is secured to two mooring buoys and the stern is moored to the end of a quay, pier, or wharf. This type is the most common for tenders or where the harbor space is limited.

Spread Moorings

Spread moorings (Fig. 31.25) use multiple mooring legs to secure a vessel. Spread moorings are especially useful for securing permanently or semi-permanently moored vessels, such as floating drydocks and inactive vessels. With spread moorings, the ships are usually oriented parallel to the current. Marine and Coastal Engineers involvement typically involves developing oceanographic and environmental loads (wind, waves, current, and ice) on the vessel.

31.2.7 Ice Breakers

Ice breakers (Fig. 31.26) are commonly provided to break up large ice floes during spring thaw and protect structures along riverbanks. Their design is governed by current speed and ice sheet size flowing downriver. They may be constructed of any material but typically utilize steel rails or beams to actually break up ice by both lifting and separating the ice sheets.

31.3 Loads on Structures due to Vessel Mooring and Berthing

31.3.1 Vessel Berthing Loads

Berthing loads are imparted onto a structure as a vessel docks at a berth due to its kinetic energy which reduces to zero as it comes to rest at a berth. This kinetic energy $KE = \frac{1}{2}MV_b^2$ of the moving vessel is a function of the mass of the vessel M and its berthing velocity V_b normal to the berth where:

- M = the vessel's displacement tonnage divided by gravity g plus an added mass of water moving along with the vessel. Displacement tonnage may not be the vessel's maximum tonnage and will depend on the type of facility (import, export, both) and its order of call (is the vessel lighter due to unloading at other port of call). Do not confuse ship's displacement tonnage with dead weight tonnage (DWT), the ship's carrying capacity or other terms, such as gross or net tonnage.
- V = the Vessel's approaching speed which is perpendicular to the berth.

The ship's velocity is a function of vessel size, method of berthing (tug assist or self), exposure of the dock, ease of navigation, and physical exposure conditions, including wind, waves, and currents. This kinetic energy on berthing must be resisted by the work done by the pier or wharf during impact. The ship's kinetic energy is transferred and absorbed by the fender in the form of work done as the fender is compressed. For detailed descriptions of the phenomena involved, see *Costa* [31.17].

A fender is usually provided to resist vessel berthing energy and distribute resulting loads to the pier or wharf structure. Such systems have evolved from simple timber piles with wales and chocks to various rubber and synthetic rubber compound shapes and forms with and without fender piles. Some of the most common types currently used include cell fenders (Fig. 31.27) and foam-filled fenders and low-budget ubiquitous old truck tires (Fig. 31.28). Table 5 of [31.2] and Section 5.3 of [31.12] provide illustrations of various fender types.

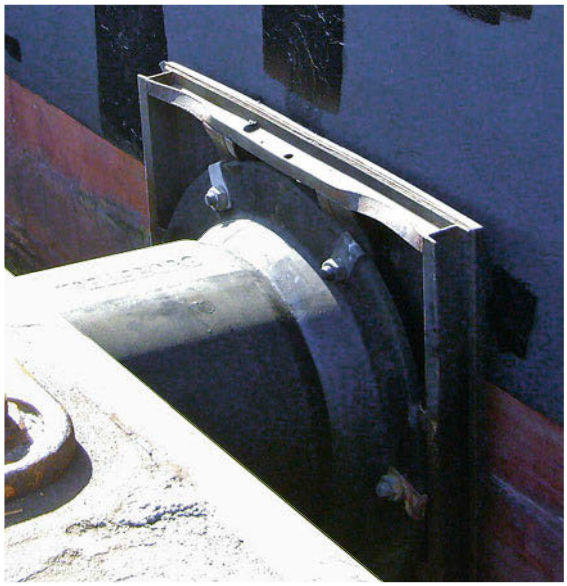


Fig. 31.27 Cell type fender (courtesy of AECOM)

Fig. 31.28 Used tire fendering (courtesy of AECOM)

The required design energy for a fender could be calculated by the following equation,

$$KE_{\text{design}} = \frac{1}{2} M V_{\text{b}}^2 C_{\text{m}} C_{\text{e}} C_{\text{s}} C_{\text{c}} \,,$$

in which:

- C_{m} = added mass factor of vessel
- C_{e} = eccentricity factor added mass factor of vessel (depending on the vessel characteristics, berthing, contact point, and berthing angle)
- C_{s} = softness factor (depending on the ratio between the elasticity for the fender system and that of the vessel's hull)
- C_{c} = berth configuration factor (depending on the pier structure type, keel clearance, berthing pattern, and vessel's hull, and fender's thickness).

It should be noted that abnormal impact should be considered in most fender designs. The abnormal factor applied to the design will depend on the size (the smaller the vessel, the larger the abnormal factor should apply); type of the design vessel, and the site-specific conditions; it is usually in the range from 1.25 to 2.0.

Fender device design is based on the above vessel energies, and the method of berthing. The area of the fender is dependent on the number of fenders, the vessel hull radius, the vessel's angle of approach, etc. Detailed guidance on calculating vessel berthing loads and required fender energy absorption can be found in [31.18], and [31.2] provides detailed guidance on this subject as does [31.12, Chap. 5]. Most major fender manufacturers also provide guidance based on these standards. These standards also provide guidance for vessel sizes based on the class of vessel. Guidance on confidence levels for ship dimensions can be found in [31.19]. Abnormal higher than usual impacts resulting from mishandling or accidents should also be considered. Another source for obtaining vessels characteristics is [31.20].

In most scenarios, the vessel berthing load on the structure is a complicated process. However, for practical design purpose, this load can be approximately estimated from the reaction–deflection curve of the design fender (rated–reaction).

31.3.2 Mooring Loads

Mooring loads are important for the lateral capacity design of pier and wharf structures. In port and harbor projects, vessel mooring loads are usually caused by external environmental conditions acting on the ship, such as wind, current, waves/swell, tidal variations, and passing vessel.

For most protected docks, if the effect of waves is minor and the tidal range is not significant, it may be sufficient to design the mooring system to withstand wind and current forces only. In this situation, the mooring loads can be estimated either by hand/spreadsheet calculation (see related OCIMF references) or a static mooring analysis program. The basic governing equations for static mooring analyses are as follows

$$\left(\sum \frac{\partial F_x}{\partial x}\right)\Delta x + \left(\sum \frac{\partial F_x}{\partial y}\right)\Delta y + \left(\sum \frac{\partial F_x}{\partial \theta}\right)\Delta \theta = -F_{x\text{a}} - \sum F_x \,,$$

$$\left(\sum \frac{\partial F_y}{\partial x}\right)\Delta x + \left(\sum \frac{\partial F_y}{\partial y}\right)\Delta y + \left(\sum \frac{\partial F_y}{\partial \theta}\right)\Delta\theta = -F_{ya} - \sum F_y \,,$$

$$\left(\sum \frac{\partial M_{xy}}{\partial x}\right)\Delta x + \left(\sum \frac{\partial M_{xy}}{\partial y}\right)\Delta y + \left(\sum \frac{\partial M_{xy}}{\partial \theta}\right)\Delta\theta = -M_{xa} - \sum M_{xy} \,,$$

where:

- F_x and F_y = x-direction and y-direction components of forces exerted on vessel by mooring lines
- M_{xy} = yaw moment on vessel due to load in mooring lines
- F_{xa} and F_{ya} = x-direction and y-direction components of total applied load on vessel due to wind and current
- M_{xya} = yaw moment on vessel due to applied wind and current loads
- Δx, Δy, and $\Delta\theta$ = finite vessel motion increment.

The above set of vessel equations can be solved by the iteration method to obtain a method for Δx, Δy, and $\Delta\theta$. The vessel is moved to $x + \Delta x$, $y + \Delta y$, and $\theta + \Delta\theta$, and the process is repeated until the computed total force components are all within the specified tolerance. Once the final displacement of vessel (Δx, Δy, and $\Delta\theta$) is determined, the corresponding mooring loads on mooring lines could be estimated based on the load-elongation/excursion curve of mooring lines/chains.

At offshore, nearshore, or unprotected berths, sites where waves are considerable or significant tidal variations or berths along the narrow channel where vessels pass closely, the effects of waves, tides, and passing vessel must be included in the analysis. As a commonly accepted practice in the marine engineering community, under this situation, a dynamic mooring analysis has to be conducted, which is much more complicated than the static mooring analysis.

In general, there are two categories of numerical models for dynamic mooring analysis: frequency-domain analysis model and time-domain analysis model.

The motions of a moored vessel in three-dimensional space can be described by six degrees of freedom (surge, sway, heave, roll, pitch, and yaw) in a right-handed coordinate system (x, y, z). Initially, the vessel is at rest with its center of gravity (COG) in the origin of the space-fixed coordinate system. The motion of a moored vessel can be represented by displacements in the x-, y-, and z-directions (surge, sway, and heave, respectively), and rotations about the x-, y-, and z-axes (roll, pitch, and yaw, respectively).

The motion of the moored vessel in the frequency domain can be represented by the following set of equations,

$$\sum_{j=1}^{6}\{(M_{kj} + A_{kj})\ddot{x}_j + B_{kj}\dot{x}_j + C_{kj}x_j\} = F_k^{\text{wave}} + F_k^{\text{other}}, \quad k = 1, 2, \ldots, 6 \,,$$

where

Vessel mass and inertia M_{kj} :

$$= \begin{bmatrix} m & 0 & 0 & 0 & 0 & 0 \\ 0 & m & 0 & 0 & 0 & 0 \\ 0 & 0 & m & 0 & 0 & 0 \\ 0 & 0 & 0 & I_{44} & 0 & -I_{46} \\ 0 & 0 & 0 & 0 & I_{55} & 0 \\ 0 & 0 & 0 & -I_{46} & 0 & I_{66} \end{bmatrix}.$$

- A_{kj} = vessel frequency-dependent added mass coefficient matrix
- B_{kj} = vessel frequency-dependent damping coefficient matrix
- C_{kj} = hydrodynamic restoring force matrix
- x_j = vessel displacement or rotation in surge, sway, heave, roll, pitch, and yaw
- F_k^{wave} = wave-induced excitations (force or moment)
- F_k^{other} = all other external forces (such as wind and current) and mooring constraints (such as mooring lines and fenders).

It must be realized that a major assumption made for a frequency-domain approach is that all the right-side items in the above-mentioned equation (wave-induced excitations, all other external forces and mooring constraints) vary sinusoidally at a single frequency and the coefficients on the left-hand side of the equation are constant (or nearly constant) for that frequency. However, in most cases, this assumption is invalid because of the nonlinear behavior of the mooring system.

The shortcomings of a frequency-domain model can be overcome by using a time-domain model at the expense of added complexity and computational effort.

The vessel motion in the time domain could be described by a set of equations as follows

$$\sum_{j=1}^{6}\left\{(M_{kj} + m_{kj})\ddot{x}_j + \int_{-\infty}^{t} R_{kj}(t-\tau)\dot{x}_j + C_{kj}x_j\right\} = F_k^{\text{wave}} + F_k^{\text{other}} \,, \quad k = 1, 2, \ldots 6 \,,$$

in which, the retardation function

$$R_{kj}(t) = \frac{2}{\pi} \int_0^\infty B_{kj}(\omega) \cos(\omega t)\, \mathrm{d}\omega \,,$$

where:

- ω = the wave frequency (rad/sec)
- m_{kj} = vessel-frequency-independent added mass coefficient matrix and can be obtained from the following equation
- $m_{kj} = A_{kj}(\omega') + \frac{1}{\omega'} \int_0^\infty R_{kj}(t) \sin(\omega' t)\, \mathrm{d}t$
- ω' = an arbitrarily chosen value of ω. The result for m_{kj} is independent of the value of $\omega' A_{kj}(\omega')$ and is the frequency-dependent added mass of vessel.

The set of vessel motion equations in time domain can be solved for x_j for each time step, and the corresponding mooring loads can be obtained from the known load-elongation curve for the mooring line or the load-excursion curve for the chain and anchor, or the reaction-deflection curve for the fender (if applicable). The wave forces (first order of wave load, the mean drift wave force, and oscillating drift wave force) can be computed though utilizing the corresponding vessel hydrodynamic database.

Commercially available programs have been developed for mooring analysis both by static and dynamic methods. These include [31.21] for static analysis, [31.22] for a time-domain dynamic method, and [31.23] for both frequency and time-domain dynamic models.

Depending on site-specific conditions (metocean, type, and size of vessel, and mooring layout) and the project phase (feasibility study, conceptual design, preliminary design, or final design), and available budget, the complexity of this analysis varies from very simple hand calculations to complicated dynamic computer-aided mooring models. For a vessel moored at a pier or wharf, typical input data for a dynamic mooring analysis model include the vessel characteristics (length overall, length between perpendiculars, beam, loading condition, draft, vessel displacement, windage areas, fairlead coordinates, mooring line properties, winches/bollard capacity), basin water depth, metocean conditions at berth, fender or dolphin locations and properties, and safe working capacity of on-shore mooring fittings. The output of analysis includes mooring loads and, if applicable to vessel unloading requirements, vessel responses in surge, sway, heave, roll, pitch, and yaw. References [31.2], [31.16], [31.24] and [31.25] can be used as references.

Vessel collisions with marine structures are rare and infrequent but have occurred in the past. There are no definitive standards or requirements for design for such events, typically an errant barge, but occasionally for a ship that has lost power. Coastal engineers may occasionally be required to recommend design requirements that can typically be based on actual vessel traffic which in the US is available through the US Coast Guard. The authors of [31.26, Sect. 3.14] provides some guidance for bridge structures, which can serve as guidelines for marine structures.

31.3.3 Loads on Structures due to Ocean Environmental Conditions

The load sources in the ocean environment include wind, wave, current, tide and storm surge, ice, tsunamis caused by earthquakes, and vessel propeller wash. Among these load sources, for the majority of situation, the wave load is the most complicated and the most important load that needs to be analyzed for the port and harbor structure design.

Wind

Wind load could be critical in some cases. Strong winds can be generated by hurricanes, typhoons, tropical storms, and locally generated storms. For port and harbor design, long-term wind records are usually required to derive the extreme wind for different return periods. When such records are not available, numerical wind models are often employed. It needs to be realized that it is difficult to predict the wind with sufficient accuracy with numerical models. Due to its dynamic nature, the accurate calculation of wind load on marine structures is difficult. In most situations, appropriate approaches are employed. Depending on the input of wind to the formula or model, there are two methods commonly used for wind load calculation. The first method for wind load calculation is to use constant speed. The second method for it is to use constant wind speed plus gusting wind. The commonly used wind spectrums for wind gusting include API, Ochi-Shin, and Harris-DNV. When performing the wind load calculations, it is important to apply the concept of wind duration. For rigid marine structures, it is reasonable to use the 3 s gust, as is used by most building codes. However, for the vessel, a much longer wind duration needs to be used for the wind load calculation. Depending on the size of the vessel and corresponding anchor or mooring system, it could be of the range from 15 to 60 s. Many references are available for wind load calculation, these include [31.27–29], etc.

Waves

For port or harbor design, the wave prediction and wave loads are a primary concerns for the designer. For many

sheltered ports or harbors, wave diffraction analysis is required to derive the appropriate design wave parameters for wave-loading calculation. In general, the accurate prediction of the wave load is extremely difficult due to the dynamic and nonlinear nature of waves and the fluid–structure interaction. The magnitude of wave loads depends not only on the wave height and wave period, but also on the ratio of wavelength to the dimension of the structure. If the structure is relatively slender compared to the wavelength, the drag and inertia forces dominate and the Morison equation can be used to calculate the wave load. If the structure's dimension is larger than the wavelength, the wave reflection will become important and the effect of wave reflection should be considered in the wave load calculation. If the structure's dimension is between these two cases, the effects of wave diffraction will become important, and the wave-loading calculation will become more complicated. For wave loadings on floating structures, in additional wave height, wave period, and structure's dimension, the wave load will also depend on the anchor system and the natural frequency of the floating structure. It needs to be realized that the wave load and corresponding calculation are different for breaking and nonbreaking waves. The impact force on structures caused by wave breaking is significant larger than the quasi-static wave force used in the most wave load formulas. For more detailed wave-loading analysis, the major references include USACE (United States Army Corps of Engineers) Coastal Engineering Manual (CEM) [31.6] and British Standard 6325 [31.28]. National codes and guidelines, and related ocean and coastal engineering design books can provide further guidance.

Current

In the marine environment, the magnitude and direction of a current at a site vary not only with time, but also with the vertical water-depth profile. The current could be tidal current, river discharge currents (within the estuary), wind-generated surface currents, wave-induced currents, and currents caused by vessel propeller wash. The current load acting on a structure usually consists of a static component and an oscillatory component. However, for port or harbor design, in majority cases, the maximum depth-averaged current under the design storm condition is employed in the current force formula. For more detail on the current-loading calculations, the reader may refer to [31.6, 28].

31.3.4 Tide and Storm Surge

The water-level change by tide and storm surge will have effects on structure loads. These effects include hydrostatic pressure change, soil surcharge change, change on the mooring line loads and fender load, change in the wave overtopping rate and corresponding impact force, and change in the wave uplift force on the pier deck due to the change of the air gap. For port or harbor structure design, the effects of tide and storm surge on the load calculation should be considered in the design.

31.3.5 Ice

In some ports or harbors, the load caused by the sea ice on the marine structures includes the impact (or thrust) load from floating drift ice, increased weight and area (for current load), uplift due to the decrease of the air gap for pier deck, etc. The above effects need to be considered when performing the port or harbor structure design.

31.3.6 Tsunamis

Tsunamis are actually one type of long waves. They are mainly caused by earthquakes. Though heights of these waves are small in deep water, they can be amplified significantly when these long waves propagate close to the shore if the sea bottom slope is steep. Tsunamis could be disasters for port or harbor facilities and moored vessels. For more details on the mitigation of tsunami disasters in port planning and design, readers can refer to PIANC [31.30].

31.4 Suggested Reading

This chapter is not intended to be a comprehensive manual on the planning and design of ports and harbors. For more detailed information on the civil and structural design of ports and harbors, it is suggested to refer to more detailed international design standards and publications, such as the following:

- Department of Defense, United States of America, UFC 4-152 *Design Piers and Wharves*, 2012: This Unified Facilities Criteria (UFC) document contains descriptions and design criteria for pier and wharf construction, including subsidiary, contiguous, and auxiliary structures. Loading details, regulations,

furnishings, appurtenances, and other information are discussed when applicable. This UFC provides minimum facility planning and design criteria for efficient homeporting facilities of Naval vessels. Existing ports, facilities, and berths may not meet all criteria and may, therefore, perform less efficiently, but they do not necessarily require to be upgraded. This UFC focuses on the entire homeport operation. UFC documents are available through Whole Building Design Guide, National Institute of Building Sciences (NIBS), 1090 Vermont Avenue NW, Suite 700, Washington, DC 20005 can be only be downloaded for free at their web site: http://www.wbdg.org

- Overseas Coastal Area Development Institute of Japan (OCDI), *Technical Standards and Commentaries for Port and Harbor Facilities in Japan*, 3-2-4 Kasumigaseki, Chiyoda-ka, Tokyo. This manual covers a wide variety of port and harbor facilities and structure types. It provides the detailed design guidance on foundations, navigation channels and basins, protective structures, and mooring facilities available at http://www.ocdi.or.jp/en/public.html.
- Oil Companies International Marine Forum (OCIMF): The OCIMF, 9 Queen Anne's Gate, London SW1H 9BU, England provides a series of publications related to the design of marine oil terminals and moorings. These include: Guidelines for the Design, Operation and Maintenance of Multi-Buoy Moorings (MBM); Jetty Maintenance and Inspection Guide; Mooring Equipment Guidelines (MEG3); and the Single Point Mooring Maintenance and Operations Guide. OCIMF documents, published by Witherby and Co, London can be ordered from their web site: www.ocimf.com/Library/Books.
- Permanent International Association of Navigation Congresses (PIANC) is a global organization providing guidance for sustainable waterborne transport infrastructure for ports and waterways. It is a forum where professionals around the world join forces to provide expert advice on cost-effective, reliable, and sustainable infrastructure to facilitate the growth of waterborne transport. Established in 1885, PIANC continues to be a leading partner for the government and private sector in the design, development, and maintenance of ports, waterways, and coastal areas. As a nonpolitical and nonprofit organization, PIANC brings together international experts on technical, economic, and environmental issues pertaining to waterborne transport infrastructure. Members include national governments and public authorities, corporations, and interested individuals. With expert guidance and technical advice, PIANC provides guidance to public and private partners through high-quality technical reports. Their international working groups develop regular technical updates on pressing global issues to benefit members on shared best practices. PIANC is headquartered at Bâtiment Graaf de Ferraris – 11ième étage, Blvd. du Roi Albert II, 20 – Boîte 3 B-1000 Bruxelles (la Belgique) and Technical Reports can be ordered on their website: www.pianc.org.
- British Standards Institution (BSI), 389 Chiswick High Road, London, W4 4AL, England, provides guidance on the criteria relevant to the planning, design, construction, and maintenance of structures in a maritime environment and located at or close to the shore and covers environmental factors, operational requirements, sea state, loadings, geotechnics, materials, and protective measures. The various standards under BS 6349 Maritime Structures include BS 6349-1 Code of Practice for General Criteria; BS 6349-2, Design of Quay Walls, Jetties, and Dolphins; BS 6349-3 Design of Dry Docks, Locks, Slipways and Shipbuilding Berths, Shiplifts and Dock and Lock Gates; BS 6349-4 Code of Practice for Design of Fendering and Mooring Systems; BS 6349-5: Code of Practice for Dredging and Land Reclamation; BS 6349-6: Design of inshore moorings and floating structures; BS 6349-7: Guide to the Design and Construction of Breakwaters; BS 6349-8: Code of Practice for the Design of Ro-Ro Ramps, Linkspans and Walkways. These standards can be ordered from the BSI website: http://shop.bsigroup.com/en/.

Reference books covering the full range of planning and design of marine structures include [31.31–38], as well as the following titles:

- Task Committee on Marinas 2020 of the Ports and Harbors Committee of the Coasts, Oceans, Ports, and Rivers Institute (COPRI) of the American Society of Civil Engineers (ASCE) *Planning and Design Guidelines for Small Craft Harbors*, ASCE Manuals and Reports on Engineering Practice No. 50, American Society of Civil Engineers, Reston, Virginia, 2012.
- Drydock Asset Management Task Committee of the Ports and Harbor Committee of the COPRI of the ASCE, *Safe Operation and Maintenance of Dry Dock Facilities*, ASCE Manuals and Reports on Engineering Practices No. 121, American Society of Civil Engineers, Reston Virginia, 2010.

31.5 Notation

A_{kj} vessel frequency-dependent added mass coefficient matrix
B_{kj} vessel frequency-dependent damping coefficient matrix
C_c berth configuration factor
C_e eccentricity factor during berthing
C_{kj} hydrodynamic restoring force matrix
C_m added mass factor of vessel
C_s softness factor during berthing
DWT deadweight tonnage of a ship, total carrying capacity
F_{ax} x-direction component of total applied load on vessel due to wind and current
F_{ay} y-direction component of total applied load on vessel due to wind and current
F_x x-direction component of forces exerted on vessel by mooring lines
F_y y-direction component of forces exerted on vessel by mooring lines
LBP length between the perpendicular sides of a vessel
LOA length overall of a ship
M vessel mass
m_{kj} vessel-frequency-independent added mass coefficient matrix and can be obtained from equation below
M_{xy} yaw moment on vessel due to load in mooring lines
M_{xya} yaw moment on vessel due to applied wind and current loads
SPM single point mooring
TEU twenty foot equivalent unit container
V vessel approach velocity
x_j vessel displacement or rotation in surge, sway, heave, roll, pitch, and yaw
ω wave frequency [rad/s]
F_k^{other} all other external forces and mooring constraints.
F_k^{wave} wave-induced excitations (force or moment)

References

31.1 Working Group II-30 PIANC-IAPH: *Approach Channels: A Guide for Design*, Final Report (PIANC, Brussels 1997)
31.2 British Standards Institute: *Maritime Structures. Code of Practice for Design of Fendering and Mooring Systems*, BS 6349-4 (BSI, London 1994)
31.3 B.L. McCartney, L.L. Ebner, L.Z. Hales, E.E. Nelson (Eds.): *Ship Channel Design and Operation* (ASCE, Reston 2005) p. 272
31.4 PIANC-IAPH: *Joint PIANC-IAPH Report on Approach Channels – Preliminary Guidelines (Volume 1)* (PIANC, Brussels 1995)
31.5 IALA ANM Committee: *Aids to Navigation Manual (IALA NAVGUIDE)* (IALA-AISM, Saint-Germain-en-Laye 2010) p. 190
31.6 U.S. Army Corp of Engineers: *Coastal Engineering Manual – Part VI*, EM-1110-2-1100 (U.S. Army Corp of Engineers, Washington 2002)
31.7 U.S. Army Corp of Engineers: *Design of Coastal Revetments, Seawalls and Bulkheads*, EM 1110-2-1614 (U.S. Army Corp of Engineers, Washington 1995)
31.8 U.S. Army Corp of Engineers: *Design of Sheet Pile Cellular Structures, Cofferdams and Retaining Structures*, EM 1110-2-2503 (U.S. Army Corp of Engineers, Washington 1989)
31.9 U.S. Army Corp of Engineers: *Design of Sheet Pile Walls*, EM 1110-2-2504 (U.S. Army Corp of Engineers, Washington 1994)
31.10 United States Steel: *USS Steel Sheet Piling Design Manual* (U.S. Department of Transportation, Washington 1984)
31.11 R. Heger: *Dockmaster's Training Manual* (Heger Dry Dock, Holliston 2005)
31.12 Design: Piers and Wharves UFC 4-752-01 (Department of Defense, Washington DC 2005)
31.13 Naval Facilities Engineering Command: *Marine Railways*, MIL-HDBK-1029/2 (Department of the Navy, Washington DC 1989)
31.14 M.W. LaNier, M. Wernli, R. Easley, P.S. Springston: New technologies proven in precast concrete modular floating pier for U.S. Navy, PCI J. **50**(4), 76–99 (2005)
31.15 American Bureau of Shipping: *Rules for Building and Classing Single Point Moorings* (ABS, Houston 2014)
31.16 Naval Facilities Engineering Command: *Design: Moorings*, UFC 4-159-03 (Department of Defense, Washington DC 2005)
31.17 F.V. Costa: Dynamics of berthing impacts. In: *NATO Advanced Study Institute on Analytical Treatment of Problems in the Berthing and Mooring of Ships*, (NATO Advanced Study Institute, Wallingford 1973)
31.18 P. Lacey, P.D. Stebbings, P. Vallander, H.W. Vollstedt, H.W. Thoresen, M.L. Broeken, S. Meijer, A.G. Traffers, S. Uda, M. Tartaglini, M. Faeth, J.E.P. Serras, C.S.S. Hill, C.N. van Schaik, H. Smitz, J. Villaneuve, J. Uzcanga, H.F. Burcharth, P. Acton, P. Levreton: *Guidelines for the Design of Fender Systems: 2002* (PIANC General Secretariat, Brussels 2002)
31.19 Y. Akahura, H. Takahashi: *Ship Dimensions of Design Ship Under Given Confidence Limits* (The Port and Harbour Research Institute, Ministry of Transport, Japan, Kanagawa 1998)

31.20 Clarkson Research Services Limited: World Fleet Register, http://www.crsl.com (Clarkson Research, London)
31.21 U. S. Navy: *FIXMOOR [Computer Program]* (National Technical Information Service, Alexandria 1989)
31.22 Maritime Researach Institute: TERMSIM II [Computer Program], http://www.marin.nl
31.23 ANSYS: AQUA [Computer Program], http://www.ansys.com
31.24 W.E. Cummins: *The Impulse Response Function and Ship Motions*, David Taylor Model Basin Report No. 1661 (US Department of the Navy, Bethesda 1962)
31.25 G. van Oortmerssen: *The Motions of a Moored Ship in Waves*, MARIN Publication No. 510 (Wageningen, the Netherlands 1976)
31.26 American Association of State Highway and Transportation Officials: *AASHTO LRFD Bridge Design Specifications* (AASHTO, Washington 2010)
31.27 American Society of Civil Engineers: *ASCE 7: Minimum Design Loads for Buildings and Other Structures* (ASCE, Reston 2010)
31.28 British Standards Institute: *Maritime Works. Code of Practice for Planning and Design for Operations*, BS 6349-1-1 (BSI, London 2013)
31.29 I.C. Council: *2012 International Building Code* (ICC, Country Club Hills 2012)
31.30 Permanent International Association of Navigation Congresses: Mitigation of Disasters in Ports, PIANC Report No. 112-2010
31.31 J.W. Gaythwaite: *Design of Marine Facilities for the Berthing, Mooring, and Repair of Vessels*, 2nd edn. (American Society of Civil Engineers, Reston 2004)
31.32 G.P. Tsinker: *Marine Structures Engineering – Specialized Applications* (Chapman Hall, New York 1995)
31.33 H. Agerschou (Ed.): *Planning and Design of Ports and Marine Terminals*, 2nd edn. (Thomas Telford, London 2004)
31.34 C.A. Thoresen: *Port Designer's Handbook; Recommendations and Guidelines* (Thomas Telford, London 2010)
31.35 G.P. Tsinker: *Port Engineering: Planning, Construction, Maintenance, and Security* (Wiley, New York 2004)
31.36 G.P. Tsinker: *Handbook of Port and Harbor Engineering: Geotechnical and Structural Aspects* (Chapman Hall, New York 1997)
31.37 G.P. Tsinker: *Marine Structures Engineering: Specialized Applications* (Chapman Hall, New York 1995)
31.38 B.C. Gerwick Jr.: *Construction of Marine and Offshore Structures*, 3rd edn. (CBC, Boca Rotan 2007)

32. Marine Outfalls

Peter M. Tate, Salvatore Scaturro, Bruce Cathers

Marine outfalls are used to discharge treated liquid waste to the environment. Not all contaminants in liquid waste can be removed by treatment. A properly designed, constructed, and operated marine outfall effectively dilutes the discharged waste which then substantially reduces the concentration of contaminants in the wastewater. In turn, this reduces the risk to biota and human users of the marine environment. An introduction to some of the main aspects of marine outfalls is provided.

Five areas are covered, commencing with the main influences associated with the decision to build a marine outfall. Included is an overview of the wastewater treatment process. Near-field numerical modeling is described and it is demonstrated how this tool can be used to assist with the design of a marine outfall. Outfall hydraulics is discussed, detailing a range of features including head losses, manifolds (or diffusers), seawater intrusion, and air entrainment. A very brief summary of the construction of a marine outfall is provided. The final area covered describes environmental monitoring that should be undertaken to confirm the putative impacts associated with a marine outfall.

32.1 **Terminology** 712

32.2 **Governance** 713
32.2.1 Drivers for a Marine Outfall 713
32.2.2 Wastewater Treatment 714
32.2.3 Data Collection for Outfall Design 715

32.3 **Predicting Near-Field Dilutions** 716
32.3.1 Physical Models 717
32.3.2 Positively Buoyant Jets and Plumes 717
32.3.3 Negatively Buoyant Jets 718
32.3.4 Model Validation 718
32.3.5 Far-Field Numerical Modeling 721
32.3.6 Data for Running the Models 721
32.3.7 Conceptual Design for Positively Buoyant Discharges 722

32.4 **Hydraulic Analysis and Design** 723
32.4.1 Governing Hydraulics 723
32.4.2 Diffusers – Hydraulic Design 725
32.4.3 Flow Variability 729
32.4.4 Hydraulic Integration 730
32.4.5 Air Entrainment 732
32.4.6 Sedimentation 733

32.5 **Outfall Construction** 734
32.5.1 Construction Materials 734
32.5.2 Construction Methods 734
32.5.3 Some Considerations 735

32.6 **Environmental Monitoring** 736
32.6.1 Change Versus Impact 736
32.6.2 Pre- and Post-construction Monitoring 736
32.6.3 Long-Term Monitoring 738
32.6.4 Summary 739

References 739

Work presented in this chapter concentrates on the discharge of wastewater to the environment through marine outfalls. Marine structures are required for intakes for drinking water (e.g., extraction of some types of minerals, cooling of thermoelectric power equipment, desalination plants) and water for industrial or commercial use (e.g., flushing of toilets). The focus here is on marine outfalls; marine intakes are not considered further.

The primary objective of this chapter is to provide practitioners with an overview of the fundamentals of marine outfalls and to outline some initial considerations to help those new to the subject area. Understandably, the present chapter does not cover all areas in detail; the focus is on the design and monitoring aspects of marine outfalls. Information on some of the problems drawn from experience with marine outfalls is provided and reference material with additional detail is identified.

32.1 Terminology

The following terminology is used throughout this chapter. Sewage is the raw input to a municipal wastewater treatment plant, the product of which is effluent. Seawater is the raw input to a desalination plant, the product of which is brine. Wastewater refers to either effluent or brine. An outfall refers to the disposal system from the treatment plant to the discharge outlets. The manifold is that part of the outfall from which there is a series of offtakes, termed risers. A diffuser comprises that section of the outfall which includes the manifold and the risers.

Two generalized marine outfalls are shown in Fig. 32.1. Figure 32.1a shows the discharge of positively buoyant effluent from a wastewater treatment plant and Fig. 32.1b shows the discharge of negatively buoyant brine from a desalination plant. Both schematics show an inclined tunnel from the wastewater treatment plant, the outfall tunnel, a diffuser comprising several risers, and the wastewater plume emanating from outlet nozzles on the top of each riser. The outfall tunnel is inclined upward to ensure any air trapped in the declined tunnel exits though the outlet nozzles (or ports) and does not remain in the tunnel. The number of risers, separation distance between risers, length of each riser and number of outlet nozzles on each riser will depend on the specific needs for each outfall.

Some points to note are provided later:

- It is advantageous to locate the diffuser in fast flowing ambient waters. This will enhance dilution of the wastewater and rapidly transport the wastewater away from the diffuser.
- For effluent discharge, the outlet nozzles are usually horizontal. Effluent is less dense than the ambient marine waters and will rise to the sea surface or, if the stratification is sufficiently strong, become trapped below the surface. In contrast, outlet nozzles associated with a brine discharge are angled toward the surface (often, an angle of 60° to the horizontal is used). The density of brine is greater than that of seawater and it will fall toward the sea bed. Angling the outlet nozzles toward the surface and discharging the brine with high velocity will maximize its dilution.
- A pipeline may replace the tunnel as shown in Fig. 32.1. The pipeline is anchored to the sea bed and discharge is through outlet nozzles fixed to the pipeline. Risers are not used in these con-

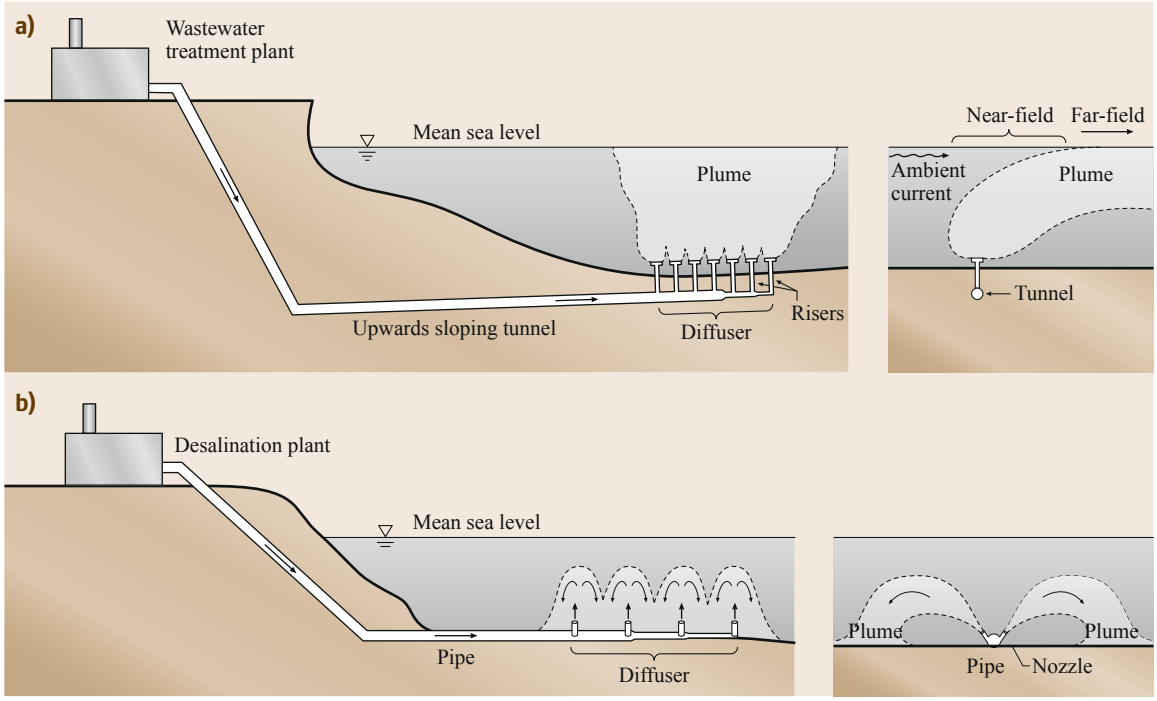


Fig. 32.1a,b Schematics of **(a)** a wastewater treatment plant outfall, and **(b)** a desalination plant outfall, showing side views and end views

figurations. (Risers are vertical structures used to transfer the wastewater from an outfall tunnel to the outlet nozzles. They may be tens of metres in length).
- The number of outlet nozzles attached to each riser is usually restricted to eight or less. If more than eight outlet nozzles are used, the plumes from adjacent nozzles interfere with each other and reduce the effective dilution of the wastewater.
- The outfall pipeline and diffuser may be tapered to ensure the velocity of the wastewater remains sufficiently high to prevent sediments from settling in the pipeline.
- Outlet nozzles may be fitted with nonreturn check valves (also called duckbill valves). These valves are closed when the wastewater flow is zero and prevent the ingress of seawater into the pipeline. One advantage of check valves is that they enhance dilution, compared with a round nozzle of the same cross-sectional area [32.1]. However, they may be fouled by biota or fishing nets, rendering them permanently open or closed.

32.2 Governance

There are many factors affecting the decision to build a marine outfall.

Municipal wastewater collects at the bottom of the catchment. For a coastal city, this is at the edge of the marine environment. There are large costs associated with the movement of wastewater to the top of a catchment for potable reuse, including construction of a pipe network, pumps, and energy required to operate the pumps. Furthermore, there may be high costs associated with the conversion of wastewater to potable water. The disposal of wastewater through a marine outfall may be the best overall use of resources. Despite this, the decision to proceed with a marine outfall should first examine other options and maximize the beneficial uses of recycled wastewater.

32.2.1 Drivers for a Marine Outfall

SPHERE is an acronym we use to describe the main factors overlying the need for government sponsored development (social, public health, environmental, regulation, economic). The first three elements of SPHERE represent the main aspects in which a marine outfall has an impact (i. e., the community values). The last two elements of SPHERE represent the constraints on the marine outfall – *regulation* tending toward high treatment and consequential high cost and *economic* tending toward low cost and consequential low treatment. Below, some of the considerations of SPHERE are described in the context of a marine outfall.

Most countries have environmental guidelines that need to be met during the design of a marine outfall. These guidelines are unique to each country and all cannot be detailed here. Suffice to say that they include meeting concentrations of contaminants which may include pathogens, nutrients, metals, and organics. These guidelines usually apply at the boundary of a mixing zone, which needs to be clearly defined prior to construction of the marine outfall.

Social

What does the community expect from a marine outfall? What are the values that are important to the community? This will vary among and within different geographical regions and cultural groups. Some communities will comprise a large number of beach users. To them, the concept of a marine outfall may not be palatable unless it can be clearly demonstrated that the marine outfall poses minimal risk to their use of the marine environment.

Public Health

Is it safe to swim in the marine waters? What are the types and concentrations of substances that will be discharged to the marine environment? Will they be of harm to us? Much information is available to inform us about of the potential harm of substances that may be discharged through a marine outfall. Most countries synthesize this information into a set of guidelines applicable to their marine environment. There is a tacit assumption that, provided the concentrations of the substances are kept below the guideline levels, the health of the users of the marine environment will be maintained. This does require knowledge of the types, concentrations, and variability of the substances in the wastewater. It should be noted here that all the substances are potentially toxic given sufficiently high concentrations and the environment into which they are discharged.

Environmental

Will the discharge of substances through the marine outfall cause harm to marine organisms? Will the marine environment be degraded into the future? Will the beaches and marine waters be free from visible pol-

lution – oil, grease, rags, etc.? As noted under *Public Health*, most countries have environmental guidelines. Provided these guidelines are met, it is assumed that the marine environment will be protected. The guidelines are usually in the form of concentrations of substances (e.g., metals, nutrients, and bacteria) that should be met at a specific distance from the outfall (this distance defines a mixing zone). This implies that there will be a region inside the mixing zone in which the guidelines may not be met. The consequences are that the biological diversity inside the mixing zone may not be the same as that in reference areas.

Regulation

What are the regulations that govern the discharge of any substance to the marine environment? Regulations are often in the form of licence conditions restricting the types, concentrations, and/or loads of substances that can be discharged to the marine environment. As noted above, there is a tacit assumption that keeping within these restrictions will ensure the safety of humans, and the protection of flora and fauna in the marine environment.

Economic

Governments will invest a large amount of money for the construction of a marine outfall. Ultimately, this money is raised through taxes and governments are accountable for the wise use of the taxes they collect. Outfall dollars will be competing with funding areas as diverse as education, security, and care for the aged. What does the community value? What is the community willing to pay to protect both humans and the environment? The marine outfall is just one of many options that should be considered. Ultimately there is a balance between the level of protection offered and the cost incurred by each option. It is the responsibility of the engineer and scientist to evaluate each option and provide the government with the most effective solution.

32.2.2 Wastewater Treatment

Our main focus in this section is on municipal outfalls. Critical to a marine outfall is knowledge of what is being discharged, particularly the types, concentrations, and variability of contaminants in the wastewater. Discharge of contaminants from other sources including private outfalls, rivers and estuaries, atmospheric inputs, discharges from vessels, and illegal dumping are not considered. The reader is referred to *Tchobanoglous* et al. [32.2], which provides considerable detail on wastewater treatment.

Wastewater discharges from domestic, commercial, and industrial sources. Often, the wastewater systems are not isolated from the environment and infiltration of water during storms may also occur. The composition of wastewater depends on the relative contribution of these three main sources and on the type and size of industry and/or commercial activity. Each wastewater system is unique and treatment plants are designed to deal with the quantity and quality of wastewater produced by a specific system.

Wastewater comprises particulate matter, pathogens, nutrients, organic, and inorganic material. Severe environmental damage can result if wastewater is discharged undiluted or without treatment. Therefore the main objective of sewage treatment is the elimination or reduction in concentration of these materials.

Different concentrations of substances will invoke different responses in different species. Metals may be adsorbed onto particulates that may be ingested by fish and shellfish. Organics are often adsorbed by the fatty tissues in aquatic animals. Reducing the concentrations of suspended solids, oil, and grease during the wastewater treatment process, reduces the quantity of metals and organics that may affect marine organisms.

Wastewater treatment can be broadly divided into three levels: primary, secondary and tertiary (or advanced). The levels are modular, subsequent treatments being *bolted* onto lower levels of treatment. Within each level of treatment there are multiple options that produce wastewater of similar quality. The distinction among the treatment levels themselves is blurred and will depend on how individual levels are operated and maintained. Usually, concentrations of suspended solids, biochemical oxygen demand (BOD), and indicator bacteria in the effluent are used to distinguish the levels of treatment. The type of wastewater treatment plant adopted is often based on the collective experience of the engineers and process workers within an organization.

Primary Treatment

Primary treatment removes debris that could damage the wastewater treatment system. This is done by passing the sewage through trash racks and screens. Sewage then flows through sedimentation tanks at low velocities ensuring residence times of 2–3 h or more [32.2]. This allows sufficient time for negatively buoyant solids to settle at the bottom of the tank and positively buoyant oils and greases to rise to the surface of the tank. Chemicals can be added to the sewage to accelerate the settling process. Both the solids and oil and grease can then be easily removed. Primary treatment also helps regulate the flow of sewage to subsequent levels of treatment. Primary treatment may be used in isolation, but this usually depends on the environment into which the wastewater is discharged.

Table 32.1 Median concentrations of substances in sewage and after various levels of treatment. The numbers are indicative only and may vary in time and between sewage treatment plants

Substance	Units	Raw sewage	Primary	Secondary	Tertiary
Faecal coliforms	cfu/100 ml	10^7	10^6	10^4	10
Suspended solids	mg/l	250	100	10	< 5
BOD	mg/l	200	100	10	< 5
Oil and grease	mg/l	50	20	< 5	< 5
Total nitrogen	mg/l	50	40	20	10
Total phosphorus	mg/l	10	7	5	3

Secondary Treatment

Secondary treatment covers a wide range of biological processes including: activated sludge, trickling filters, rotating biological contactors, aerated lagoons, oxidizing beds, and membrane bioreactors. The basic objective of all of these processes is the removal of organic material and suspended solids. Secondary treatment may also include disinfection to reduce the concentrations of bacteria in the wastewater. A common form of secondary treatment is activated sludge in which microorganisms are mixed with the wastewater under aerobic conditions for about 4–8 h. The microorganisms metabolize the organic matter in the wastewater, ultimately producing inorganic materials.

Tertiary Treatment

Tertiary treatment often involves the further removal of suspended materials using sand filters. High levels of nitrogen and phosphorus may remain in the wastewater after secondary treatment, which can contribute to excessive primary production and eutrophication. The basic premise for nitrogen removal is to convert nitrate to nitrogen gas, which is then discharged to the atmosphere. Biological processes and chemical precipitation are two methods used to remove phosphorus from the wastewater. Once removed, phosphate can be used as a fertilizer.

Microfiltration and reverse osmosis are specific forms of advanced wastewater treatment. The wastewater is forced through a fine membrane. The size of the membrane mesh is sufficient to allow the passage of water, but larger materials are captured and removed from the wastewater. Increasingly, micro- or milli-filtration are added to primary or secondary treatment processes. When combined with an effective outfall diffuser, the diluted wastewater may achieve licence requirements.

An indication of the median effluent concentrations of selected substances after treatment is given in Table 32.1. It is stressed that these are general values that will differ for specific wastewater treatment plants and are highly variable.

A desalination plant discharges brine in which the primary contaminant is salt. Median concentrations of brine are about 60 psu, although it may vary between 40 and 80 psu. The median salt content in seawater is about 35 psu. Marine organisms can tolerate salt concentrations to about 39 psu [32.3], although this value varies with different organisms. Therefore, the configuration of an outfall discharging brine to the marine environment should ensure a rapid reduction in salinity to less than 39 psu.

32.2.3 Data Collection for Outfall Design

While the preliminary design of a marine outfall can be undertaken using minimum data, the detailed design usually requires considerable data. The main aim for these data collection programs is identification of the site (or sites) for the marine outfall. That is, for the lowest cost, identifying the level of treatment and site that best meets the environmental (and other) guidelines.

The type and volume of data required depends on the marine outfall being considered. Broadly, data include: volume and flow rates of effluent to be discharged, water quality (both in the treated effluent and in the marine waters), ocean currents, and stratification. A critical aspect of monitoring, often overlooked, is the variability of these data. Our favored approach is to use the data variability in a Monte Carlo approach, running the models for many different combinations of input values. This results in a statistical distribution of the concentrations of contaminants in the marine waters, which can be synthesized in, for example, a probability of exceedance plot.

Historical data or data collected from different projects can be used. The difference between the data needs and the historical data defines a gap that a data collection program needs to fill. Much of the data collected as part of these studies can also be used in Sect. 32.6. Some of the main data collection programs are outlined below.

The volume of effluent flow can be estimated from human population projections. This information allows assessment of when environmental guidelines are likely to be exceeded; hence when upgrades of the treatment plant are likely to be needed.

Water quality in the treated effluent will be a function of the level of treatment. Estimates can be obtained from other, similar treatment plants or from the indicative values provided in Table 32.1. Ongoing monitoring of the effluent quality after construction will help ensure maintenance of environmental standards.

Measuring water quality in the marine waters (into which the wastewater is discharged) provides background concentrations of contaminants. The background concentrations must be added to the modeled concentrations to estimate the total concentration of contaminants in the marine waters. Background concentrations may already exceed environmental guidelines, in which case, they may need to be relaxed or another outfall location sought. To obtain a representative picture of marine water quality, sampling should take place over large spatial and temporal scales and should include replication. Instruments that can be moored in the field for long periods of time are increasingly being used to obtain water quality measurements, although the accuracy of such results is less than can be achieved in the laboratory.

Current speed and direction are important for plume dilution. Moored current meters can provide detailed temporal information at a point in space (or a profile throughout the water column). However, they are expensive to deploy, maintain and retrieve, and careful consideration needs to be made in regard to the number and location of such moorings. Spatial information can be obtained by profiling currents from a vessel underway, drifters drogued at specific depths and remotely sensed data (e.g., via satellites or airborne scanners). The number and duration of moored current meters will depend on the size of the outfall under consideration. For a moderately sized outfall, a single profiling current meter moored for 12 months and serviced monthly, provides the minimum data requirements. A roving current meter (deployed at different locations for one month at a time) may provide a compromise between the number of instruments and spatial coverage.

Density stratification of the water column largely governs the height of rise of the wastewater. (The effect is much reduced for brine discharges). In coastal marine waters, density is a function of both temperature and salinity, both of which should be measured. For shallow outfalls (outfalls in water depths less than about 10 m), stratification has little effect. However, for outfalls in deep waters, relatively small stratification may produce a submerged plume resulting in lower dilutions and a nonvisible plume (at least to a surface observer). Moored temperature/salinity strings provide a profile of density throughout the water column, although marine fouling will reduce the quality of data from salinity sensors. Such data can also be collected during the servicing of moored instruments, which may be monthly over a period of 12 months.

Other data such as surface waves and tides may also be important, particularly for shallow outfalls where the changes in water depth resulting from such processes may represent a significant proportion of the water column.

32.3 Predicting Near-Field Dilutions

The design of a marine outfall centers on the dilution required to meet the relevant guidelines. Occasionally, guidelines may be met after an appropriate level of sewage treatment. However, many substances will rely on the dilution with marine waters to meet these guidelines. Dilution depends on:

- Wastewater flowrate
- Depth of water into which the wastewater is discharged
- Length of the diffuser
- Outlet diameter (and whether a single or multiple outlets will be used)
- Configuration of the diffuser (e.g., whether T-section outlets or gas-burner type rosettes are used, whether nonreturn check valves are used)
- Ocean conditions (e.g., currents, stratification of the water column, tides, and ocean turbulence).

Together with cost, the above factors are used to optimize the location and configuration of the marine outfall. This is further discussed in Sect. 32.3.7

After discharge from the marine outfall, effluent rises (whereas brine descends) due to buoyancy (Fig. 32.1). The wastewater (effluent or brine) then mixes with the ambient currents and is diluted. Two types of models are used to quantify this process: near-field and far-field. This separation is made because the time and space scales of the processes in each model are substantially different.

In the near-field, the motion of the wastewater is dominated by its initial momentum and buoyancy; the velocities and rates of dilution are high. Up to 90% of wastewater dilution takes place within the near-field at the end of which most regulations apply. The engineer can configure the outfall design to maximize dilution in the near-field.

In the far-field, the wastewater is passively transported by the ambient currents and the rates of dilution are much lower than in the near-field. Far-field mixing is dominated by natural processes, over which the design engineer has little control.

While the use of both near-field and far-field models may be necessary for the detailed design of a marine outfall, we argue that near-field modeling alone may be adequate for the initial design and emphasis in the following sections is placed on near-field modeling.

This section provides a broad introduction to near-field modeling. *Wood* et al. [32.4] provide considerable detail on near-field modeling and many of the problems that may be encountered in the design of a marine outfall.

32.3.1 Physical Models

While the focus of this section is on near-field numerical modeling, it is recognized that physical modeling can also play an important role in the design of a marine outfall. Scaled physical models of prototype marine outfalls are sometimes constructed in the laboratory and used to examine the behavior of jets and plumes in the near-field. They provide good visualization of the plumes, particularly interactions between multiple plumes and include effects that are not found in most numerical models.

The fluids used in the model are typically fresh and saline water. The scales for such models are expressed as a ratio of a prototype quantity to a model quantity. Model design requires the selection of (i) the fluids which yield the reduced gravity ratio (g' defined below), (ii) the length scale to ensure that the model Reynolds numbers are sufficiently high to guarantee turbulent model flows and (iii) a scaling criterion which, in this case, is the densimetric Froude number (FR), i. e., there is a point-to-point correspondence of FR in prototype and model. This scaling criterion, together with the length scale, yields the velocity scale. Other scales for time, pressure and buoyancy force can then be determined. Although the inclusion of ambient currents in physical models is possible, it places greater demands on laboratory facilities and data acquisition systems.

Results from physical models may include information on dilutions, trajectories, the velocity field, and interactions between neighboring plumes.

32.3.2 Positively Buoyant Jets and Plumes

Two basic approaches to near-field numerical modeling are available: Eulerian and Lagrangian. A Lagrangian approach is followed in both *Lee* and *Cheung* [32.5] and *Tate* and *Middleton* [32.1, 6]. Central to either approach are the conservation equations for mass, momentum, and buoyancy. In a Lagrangian framework, they are:

- Mass conservation
$$\frac{\partial(\rho V)}{\partial t} = \rho_a f_{ent} U_{ent} A$$
- Momentum conservation
$$\frac{\partial(\rho V u_i)}{\partial t} = U_i \frac{\partial(\rho V)}{\partial t} + \rho g' V$$
- Buoyancy conservation
$$\frac{\partial(g' V)}{\partial t} = -N^2 u_i V ,$$

where ρ is the density of the jet/plume, ρ_a is the density of the ambient fluid, f_{ent} is an entrainment function, U_{ent} is the entrainment velocity, V is the volume of a buoyant fluid element, A is the cross-sectional area through which ambient water is entrained, u_i is the velocity of the buoyant fluid, U_i is the velocity of the ambient fluid, g' is the buoyancy modified gravity $(= (\Delta\rho g)/\rho_{ref}$, where ρ_{ref} is a reference density), and N is the Brunt–Väisälä frequency

$$N = \sqrt{-\left(\frac{g}{\rho_{ref}}\right)\frac{\partial \rho_a}{\partial z}} .$$

These governing equations are also applicable to negatively buoyant jets and plumes.

If the buoyant jet/plume (a) lies well away from its source (i. e., beyond the influence of the initial momentum), (b) is moving with the ambient fluid, and (c) the Boussinesq approximation is applied, then the above equations can be solved analytically to give what is known as the asymptotic results. These equations (Table 32.2) are equivalent to the *advected thermal* equations in *Wood* et al. [32.4] and the corresponding flow classifications are detailed in *Jirka* and *Akar* [32.7] and *Jirka* and *Doneker* [32.8].

Solutions to the asymptotic governing equations for positively buoyant plumes emerging from round (i. e., axisymmetric) outlet ports and from a slot (i. e., line source), in a flowing ambient fluid, with both linearly stratified or nonstratified marine waters, are presented in Table 32.2. It should be noted that these asymptotic solutions below should only be used at the conceptual stage of outfall design. For preliminary and detailed design, the full set of conservations equations above should be used and solved numerically.

The entrainment function has evolved from the constants used in *Morton* et al. [32.9] to a complex function of the densimetric Froude number, plume geometry, the velocity of the fluid inside the plume, and the velocity of the ambient current [32.5]. *Wood* et al. [32.4] use

Table 32.2 Solutions to the governing asymptotic equations for positively buoyant plumes when the ambient current speed is nonzero and the marine water density is linearly stratified (after [32.6]). Solutions to the asymptotic equations are applicable only at the end of the near field and do not include the outlet port diameter, angle of discharge, or the exit velocity

Axisymmetric source	Line source
$z(x) = \frac{0.98}{f_{ent}^{2/3}} \left(\frac{B_S[1-\cos(\frac{Nx}{U})]}{UN^2} \right)^{1/3}$	$z(x) = \frac{1.00}{f_{ent}^{1/2}} \left(\frac{B_S[1-\cos(\frac{Nx}{U})]}{UN^2} \right)^{1/2}$
$2b(z) = 2.00 f_{ent} z$	$2b(z) = 2.00 f_{ent} z$
$(z) = \frac{C_S}{C(z)} = 3.14 f_{ent}^2 U z^2 \frac{n_{ports}}{Q}$	$S(z) = \frac{C_S}{C(z)} = 2.00 f_{ent} U z \frac{L_D}{Q}$
$z_{max} = \frac{1.24}{f_{ent}^{2/3}} \left(\frac{B_S}{UN^2} \right)^{1/3}$	$z_{max} = \frac{1.41}{f_{ent}^{1/2}} \left(\frac{B_S}{UN^2} \right)^{1/2}$
$2b(z_{max}) = 2.48 f_{ent}^{1/3} \left(\frac{B_S}{UN^2} \right)^{1/3}$	$2b(z_{max}) = 2.82 f_{ent}^{1/2} \left(\frac{B_S}{UN^2} \right)^{1/2}$
$S(z_{max}) = \frac{C_S}{C(z_{max})} = 4.84 f_{ent}^{2/3} \left(\frac{UB_S^2}{N^4} \right)^{1/3} \frac{n_{ports}}{Q}$	$S(z_{max}) = \frac{C_S}{C(z_{max})} = 2.83 f_{ent}^{1/2} \left(\frac{UB_S}{N^2} \right)^{1/2} \frac{L_D}{Q}$

z is the elevation of the plume above the outlet [m], z_{max} is the maximum elevation (i. e., height of rise) of the plume [m], $S(z) = \frac{C_S}{C(z)}$ is the average dilution at elevation z, C_S is the concentration at the source [kg/m^3], $C(z)$ is average concentration at elevation z [kg/m^3], x is distance downstream from the outfall [m], f_{ent} is the dimensionless entrainment function, U is ambient current velocity [m/s], N is the Brunt–Väisälä frequency [1/s], where $N^2 = \frac{-g}{\overline{\rho_a}} \frac{d\rho_a}{dz}$, and $\overline{\rho_a}$ denotes a representative seawater density, $2b(z)$ is the diameter (or thickness) of the plume [m], Q is the flow through the outfall [m^3/s], L_D is the length of the diffuser [m], n_{ports} is the total number of outlet ports on the diffuser and, B_S is the buoyancy flux at the source, $B_S = g \frac{\rho_a - \rho_S}{\rho_S} \frac{Q}{n_{ports}}$ for an axisymmetric source and $B_S = g \frac{\rho_a - \rho_S}{\rho_S} \frac{Q}{L_D}$ for a line source, where ρ_S is the density of the wastewater at the source.

a spreading function to model the entrainment of ambient fluid into the plume.

32.3.3 Negatively Buoyant Jets

Research conducted on negatively buoyant jets over the past several decades has sought to quantify jet behavior using a variety of analytical and experimental techniques. Results from these studies have led to the development of proportionality coefficients which relate the jet densimetric Froude number and nozzle diameter to trajectory and dilution. The particular points of interest along the jet trajectory are the centerline peak (z_m) and return point (x_r) which are both defined in Fig. 32.2. A range of experimentally derived values for each of the proportionality coefficients compiled from various experimental studies are presented in Table 32.3, as reported in *Lai* and *Lee* [32.10]. Note that these coefficients are only valid for single jets discharging into quiescent ambient conditions from a nozzle orientated at 45° to the seabed. Coefficients for other discharge angles can be found throughout the research literature [32.11–17]. Current research on negatively buoyant jets focuses on multiport diffusers and discharge into receiving waters with ambient currents.

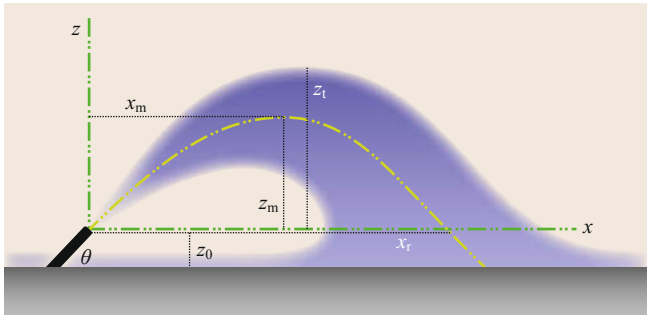


Fig. 32.2 Trajectory of a negatively buoyant jet

32.3.4 Model Validation

The information presented in Table 32.2 and Fig. 32.3 are based on asymptotic models i. e., results only at the end of the near-field and should only be used at the conceptual stage of outfall design. Full numerical models detail the movement of the wastewater from the outlet nozzle to the end of the near-field and include the nozzle size, the initial momentum of the wastewater and its trajectory. A limited set of results from the laboratory experiments of *Fan* [32.18] for a single outlet, discharging positively buoyant water into a flowing, unstratified ambient fluid are compared with several near-field models that have been used by the authors. The models are: IMPULSE [32.19], JETLAG [32.5], CORMIX [32.7, 8], OSPLM [32.4], and PLOOM [32.1, 6]. Fan's data set is used here because it is independent of the laboratory

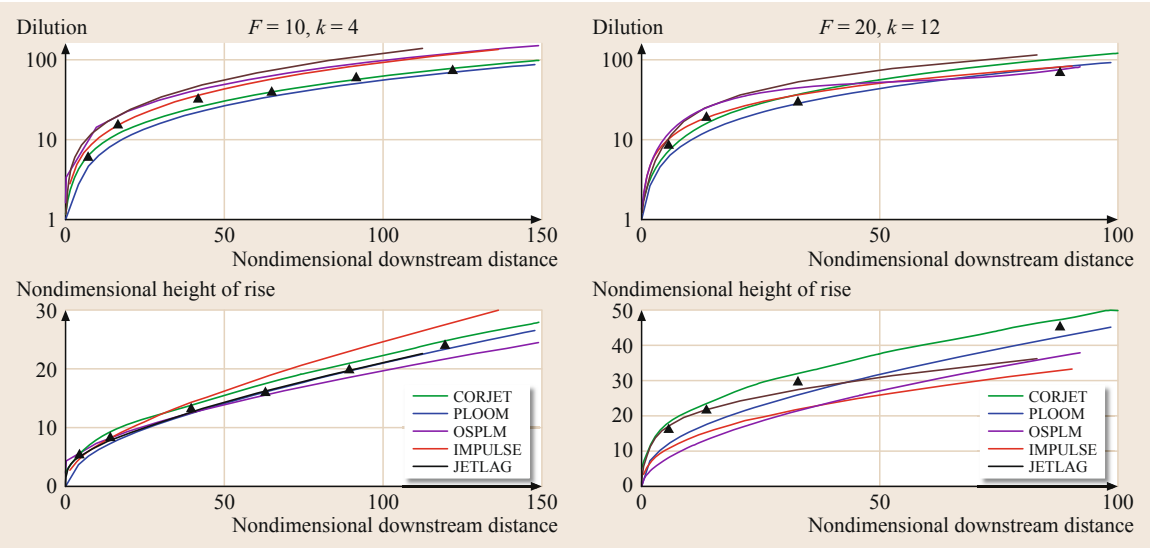


Fig. 32.3 Near-field model results for plume dilution and trajectory compared with laboratory data from *Fan* [32.18]

Table 32.3 Experimentally derived coefficients for a single negatively buoyant jet discharging into quiescent ambient conditions at an angle of 45° to the seabed (after *Lai* and *Lee* [32.10])

Description	Equation	Experimentally derived coefficients
Jet terminal rise height	$z_t = \frac{C_1}{D\mathrm{Fr}}$	$1.43 \le C_1 \le 1.61$
Horizontal location of return point	$x_r = \frac{C_2}{D\mathrm{Fr}}$	$2.82 \le C_2 \le 3.34$
Dilution at return point	$S_r = \frac{C_3}{\mathrm{Fr}}$	$1.09 \le C_3 \le 1.55$
Vertical location at jet trajectory centerline peak	$z_m = \frac{C_4}{D\mathrm{Fr}}$	$1.07 \le C_4 \le 1.19$
Horizontal location at jet trajectory centerline peak	$x_m = \frac{C_5}{D\mathrm{Fr}}$	$1.69 \le C_5 \le 2.09$

data collected by any of the above authors. The unique identifiers for Fan's experiments are FR, the densimetric Froude number

$$\mathrm{Fr} = \frac{u_{\mathrm{port}}}{\sqrt{g' d_{\mathrm{port}}}} ,$$

where u_{port} is the velocity through the outlet port, d_{port} is the diameter of the outlet port and

$$k = \frac{u_{\mathrm{port}}}{U}$$

and

$$g' = g\left(\frac{\rho_a - \rho_S}{\rho_S}\right) .$$

Compared with the laboratory data of *Fan* [32.18], these models all produce similar results (Fig. 32.3) providing confidence in the models themselves. However, it is recognized that different models may behave differently for different regimes and model selection needs to be appropriate to the problem under investigation. For example, a particular model may provide good estimates of dilution for outfalls comprising a single point discharge but poor estimates of dilution for outfalls comprising a long diffuser with multiple risers and outlet nozzles. It is stressed that there are other near-field models available [32.20–22] that would likely provide similar results.

While laboratory studies are often used to calibrate parameters within a model, field experiments are used to validate the model predictions for a specific marine outfall. This is undertaken using outfall dilution studies. Obviously, such validation studies can only be carried out after the marine outfall has been constructed. Outfall dilution studies involve the continuous injection of a tracer into the wastewater and the measurement of its concentration downstream from the discharge point. The tracer is injected at a known rate and concentration, and the flow of wastewater is also known. Therefore, by measuring the concentration of the tracer in the marine waters, the concentration of the wastewater can be determined.

Many tracers are available, for example: rhodamine water tracing (WT), fluorescein and the isotopes gold-198, technetium-99m, and tritium. Natural tracers such as salinity have also been used, but the variability in the data is usually too large to produce meaningful results. Preference is given to use a tracer that has little or no background signal; hence contact with the tracer will result in unambiguous readings. The tracer sensing device (such as a fluorometer or scintillation counter) may be towed behind a vessel and/or profiled through the water column to build a three-dimensional picture of the location and size of the plume. A critical element of the work is accurate position fixing, now usually done with differential GPS (global positioning system).

Simultaneously, the wastewater flow, ambient current speed and direction, and the density of the water column are measured. These data are used as input to the model. A direct comparison between the observations and model results can then be made. However, models are only approximations to the real world and there will be uncertainty associated with the results. Based on the results from many such experiments, predicted dilutions within a factor of two of actual dilutions are generally acceptable.

An example of the results obtained from a tracer experiment is shown in Fig. 32.4. The transect lines run parallel to the diffuser, 100 (lower panel) and 1000 m (upper panel) downstream from the outfall. Multiple transect lines are shown in each panel. In the lower panel, plumes from each of the nine risers comprising this outfall can be clearly identified. At a distance of 1000 m downstream from the outfall, plumes from the individual risers have merged, the overall width of the plume has increased and the concentration of the tracer (or plume) has markedly reduced.

Problems with Tracer Studies

Some problems encountered by the authors in conducting tracer studies are outlined below.

Some marine outfalls may have intermittent flow, particularly early in the life of the outfall when the design flow capacity is not yet reached. With intermittent flow, the time history of the patch of wastewater is unclear. However, it may be possible to temporarily store the wastewater, to enable a continuous and steady flow over the duration of the field experiment.

Locating the plume in the field may be difficult. The tracer may not be visible when the ambient waters are stratified and the plume is trapped below the water surface. A conductivity-temperature-depth probe can be used to identify stratification in the water column and hence the likely depth at which the effluent will reside.

Isotope tracers decay with time. The half-life of technetium-99m is 6 h which is comparable with the duration of many tracer experiments. If technetium-99m is used as a tracer, the initial signal will change significantly over time and needs to be accounted for in the data analysis. Tritium, with a half-life of about 12 yr, can be used for long duration tracer experiments or when there is considerable transport time between the nuclear facility that produces the isotope and the experiment site.

When a positive contact is made with the labeled plume, it is not possible to know where this contact oc-

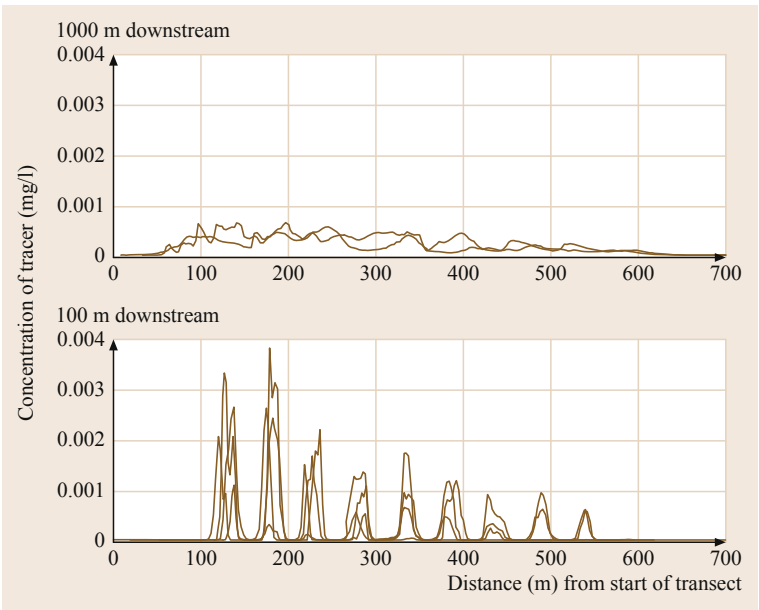


Fig. 32.4 Example of tracer concentrations obtained from field studies. Concentration data were collected from 1 m below the surface, at distances of 100 and 1000 m downstream from the outfall. Note the uneven distribution of concentration along the diffuser indicating an uneven distribution of flow

curs in the wastewater plume. One solution is to take many tracer readings closely separated in space and time to identify the plume boundaries and the region of highest tracer concentration.

The fluorescence of rhodamine WT is highly temperature dependent. It loses about 3% of its fluorescence for every one degree Celsius drop in water temperature. In very cold environments, it may not be possible to detect a signal at all – as happened to the authors when first using rhodamine WT in Antarctica.

32.3.5 Far-Field Numerical Modeling

The emphasis in this chapter is on near-field modeling rather than far-field modeling. The reason for this is because most of the dilution of the discharged wastewater occurs in the near-field, and environmental guidelines and licence conditions are usually applied at the end of the near-field. However, far-field modeling is important when assessing discharges into relatively shallow waters when mixing in the near-field is incomplete or when examining potential impacts at sites remote from the outfall, e.g., beach bathing waters or sensitive marine habitats or communities.

Far-field modeling usually includes hydrodynamic and water quality components. The hydrodynamic models are based on the principles of mass and momentum conservation of the marine waters; water quality models are based on mass considerations of the contaminant(s) or tracer(s) being discharged. Hydrodynamic models are usually based on a fixed mesh in space (i. e., an Eulerian formulation) and produce the depth and velocity fields as output. The water quality models require the velocity field as input; their formulation may be based on the same mesh as the hydrodynamic model. In another formulation (i. e., the Lagrangian formulation), many parcels of contaminant or tracer may be tracked as the velocity field transports and disperses them. The results from an Eulerian formulation yields the contaminant concentrations on the fixed mesh, while the Lagrangian formulation yields the number of contaminant parcels contained in each volume of fluid bounded by mesh points or nodes; these numbers can then be converted into contaminant concentrations.

The most common types of Eulerian models used are finite difference (i. e., point-wise approximations of the variables), finite element (piecewise approximations of the variables), or finite volume (based on fluxes of mass or momentum within each mesh cell). The meshes can be regular (i. e., structured) or irregular (i. e., nonstructured); they can be 2-D (2-dimensional; i. e., depth averaged) or 3-D (three-dimensional).

In a 2-D, hydrodynamic model the mesh is in the horizontal plane. At each mesh point or node, the unknowns consist of two velocity components and a depth. In a 3-D model, the mesh includes the 2-D horizontal plane and the mesh points or nodes in the vertical dimension. The unknowns at each 3-D mesh point or node, typically consist of two horizontal velocity components and pressure.

Usually, 2-D models require substantially less computational time and less data for calibration and running than 3-D models. In water depths exceeding about 20 m, the velocity vector at a single location in plan, may vary in magnitude and direction throughout the water column. If resolution of this variability is considered significant from the point of view of pollutant movement, a 3-D model may be preferred to a 2-D model. The horizontal spacing between mesh points or nodes will depend on the bathymetry and the presence of islands, headlands, and submarine canyons. The near-field model results need to be incorporated into the far-field model. To achieve this effectively, it may be desirable to refine the far-field mesh in the vicinity of the near-field.

Far-field models run under various flow scenarios can be used in the early stages of an investigation to guide data collection programs before, during, or after commissioning of an outfall. Such model studies may be conducted using a coarse mesh for quick turnaround of results.

32.3.6 Data for Running the Models

A range of information is required to run the numerical models. This includes: the outfall configuration, wastewater flow, and oceanographic data (currents and stratification of the water column). Over the long term, the model results can be used to examine changes in outfall performance. Below is a summary of the information required to run the models and how that information may be obtained.

Outfall Configuration

The concept outlined in the following section can provide a starting point for the outfall design. In the design phase, the outfall configuration can be changed and refined until the relevant environmental guidelines are met and engineering feasibility assessed. Once the outfall has been constructed, its configuration is essentially fixed. However, some flexibility may be enabled. For example, twin pipelines may be built and only one pipeline used for present wastewater flows (the second pipeline being saved for use when wastewater flow increases with future growth in population). Similarly, a multiport diffuser may have one or more outlet ports blanked, again in anticipation of future growth.

Information needed for the outfall configuration includes:

- Water depth in which the diffuser section is located.
- Length of the diffuser section.
- Configuration of the diffuser (e.g., a single or multiport outlet).
- Diameter of each outlet port.
- Whether the outlet ports are fitted with nonreturn check valves.

Wastewater Flow

Wastewater flow is usually measured in the outlet pipe at the end of the treatment processes. A range of flow measuring devices are available including flows based on electromagnetic, pressure, ultrasonic, or capacitance sensors. Also important is the density of the wastewater in relation to the density of the marine waters into which the wastewater is discharged. Usually it is safe to assume that the density of the wastewater is close to that of fresh water [32.2], although large amounts of particulate material in the wastewater may alter the density of the wastewater.

For numerical modeling purposes, wastewater flow is usually assumed to be uniform throughout each outlet port. This may not be necessarily the case. Energy loss may be significant over long diffuser sections, resulting in reduced flows through outlets lying further offshore. Low flows may result in the intrusion of seawater into the diffuser and a reduction in its performance. To help establish uniform flow, diffuser sections may be tapered (Fig. 32.1) and to help prevent the intrusion of seawater, outlet ports may be fitted with nonreturn check valves.

Currents

Currents determine the movement and dilution of the wastewater. Often a moored Doppler profiler is used to measure the current speed and direction throughout the water column. Doppler profilers can also be ship mounted, which allows a spatial picture of the currents to be obtained. Remote sensing and shore-based radar systems can provide detailed spatial coverage of the surface currents. However, it is subsurface current data which is critical for running the near-field models.

The choice of mooring location should be as close as possible to the diffuser. However, a compromise is often made, balancing the proximity of the diffuser to the mooring, with the health of workers who service the mooring (in waters that may be contaminated with diluted wastewater) and the security of the mooring itself.

Stratification

Stratification is a rapid vertical change in the density of the marine waters. In coastal waters, changes in the density are dominated by changes in temperature and salinity. The height to which a wastewater plume rises in the water column is largely governed by the strength of the stratification.

Measurements of temperature and salinity are often made using a conductivity, temperature, depth (CTD) probe. (Salinity is calculated from conductivity and temperature). The CTD probe can be lowered from a boat providing a continuous profile through the water column. CTD probes can also be moored, thereby providing a time series of density data at a fixed point. Historically, conductivity data from moored CTDs drift with time due to the gradual build-up of biofilm on the sensors. While there have been substantial improvements in the reliability of moored conductivity sensors in recent years, the quality of the data may still be highly variable. Temperature sensors (unless heavily fouled with marine growth) do not suffer the same problem. Hence changes in stratification using data from long-term moored systems are usually estimated from temperature sensors alone.

32.3.7 Conceptual Design for Positively Buoyant Discharges

Wilkinson [32.23] described a method by which the minimum length of a *simple outfall* could be determined. In his concluding remarks, *Wilkinson* [32.23] was careful to point out that this provides a preliminary estimate only, and he provided some suggestions on ways in which the outfall configuration could be further refined. This analysis is only intended as a starting point for outfall design. Detailed analyses are site specific and must be undertaken for final design. Some site specific factors include: the bathymetry, environmental guidelines, level of wastewater treatment, and the likelihood of plumes reaching the surface or sensitive ecological areas. *Wilkinson*'s [32.23] approach is modified here, by using the single set of equations (Table 32.2) and introducing construction cost as criteria for outfall design.

The following analysis is applicable to nonzero ambient currents, which is usually applicable to marine waters (e.g., currents near the Sydney, Australia deep water ocean outfalls exceed 0.05 m/s more than 90% of the time).

The total cost (T_c) of a marine outfall can be expressed as

$$T_c = lL_p + mL_D + nn_{ports} \quad (32.1)$$

where l is cost per meter of the outfall pipeline or tunnel [$/m], L_p = length of the outfall pipeline or tunnel [m], m is the cost per meter of the diffuser [$/m], L_D

is the length of the diffuser [m], n is the cost per outlet port [$], and n_{ports} is the number of outlet ports.

The basic premise used in *Wilkinson* [32.23] is that the profile of the water depths as a function of distance offshore (i. e., the length of the marine outfall, L_p) can be expressed as the power curve, $L_p = rz^s$, where r and s are constants that express the least-squares, best-fit shape of the across shelf bathymetry that may be obtained from navigational charts and z is the water depth.

Expressions for the length of the diffuser and the number of outlet ports can be obtained from Table 32.2 and the total cost can then be rewritten as

$$T_c = l(rz^s) + m\left(\frac{SQ}{2f_{ent}U}z^{-1}\right) + n\left(\frac{SQ}{3.14f_{ent}^2U}z^{-2}\right), \tag{32.2}$$

where S is the dilution required to comply with licence conditions or environmental guidelines.

To minimize the total cost, the above expression is differentiated with respect to the water depth (z), equated to zero and solved. The result gives the depth at which the minimum cost for the marine outfall is achieved. Substituting this value for depth into the equations in Table 32.2, gives the length of the diffuser and the number of outlet ports that comprise the marine outfall.

Actual costs do not need to be known. If the relative costs among l, m, and n are known, then the total cost of the marine outfall can be expressed in terms of a normalized cost, T_c/l. Again, it is stressed that this analysis is preliminary and is only intended as a starting point for outfall design.

32.4 Hydraulic Analysis and Design

It is often necessary to define the physical extent of a brine or wastewater outfall system for project planning and design purposes. While the ambient sea in the vicinity of an outfall structure typically defines the downstream boundary of an outfall system, defining the upstream boundary may not necessarily be as straightforward. For the purposes of this chapter, the upstream boundary is assumed to be a free surface which exists somewhere upstream of the outfall conduit entrance. Typical locations for this boundary could be the effluent level in outfall shafts, deaeration chambers, sedimentation basins, or pumping wet wells. For configurations without any free surfaces between the treatment process and outfall conduit, the upstream boundary may be taken at some hydraulically arbitrary point. Regardless of its physical location, this boundary represents a key design interface that must be properly integrated with the treatment plant as a whole. Determining the piezometric head at the upstream boundary of an outfall system is therefore a critical hydraulic design objective.

32.4.1 Governing Hydraulics

Consider the gravity-driven outfall system shown in Fig. 32.5 which includes an upstream shaft to capture plant effluents, as well as rosette style outfall structures installed on the seabed. (A rosette structure typically has multiple nozzles that are arranged around its perimeter.) The piezometric head at the outfall shaft, defined by point 0 on the fluid surface, can be obtained by applying the energy equation between this point and point 1 which is located precisely at the tip of the nozzle. It does not matter which outfall structure or nozzle is used to define point 1 because this multiple riser configuration is an example of parallel flow. For parallel flow, the total head loss must be equal through each parallel flow path.

The energy equation applied between (0) and (1) is written in terms of total head (i. e., energy per unit weight of effluent),

$$\frac{V_0^2}{2g} + \frac{p_0}{\rho_e g} + z_0 = \frac{V_1^2}{2g} + \frac{p_1}{\rho_e g} + z_1 + \sum H_{L(0\to1)}, \tag{32.3}$$

where V is the velocity [m/s], p is the pressure [Pa], z is the elevation above an arbitrary datum [m], z_{SL} is the elevation of sea level above datum [m], $\sum \Delta H_{L(0\to1)}$ is the total head loss between locations (0) and (1) [m], ρ_e is the density of effluent [kg/m^3], ρ_a is the density of ambient seawater [kg/m^3].

Working with gauge pressures and assuming negligible effluent velocity in the outfall shaft, the first two terms on the left-hand side of (32.3) are reduced to zero. Using the assumption that the pressure at (32.1) is hydrostatic based on the density of seawater, or $p_1 = (z_{SL} - z_1)\,\rho_a g$, this relationship is substituted into (32.3) to yield an expression for effluent level in the outfall shaft, z_0.

$$z_0 = \left(\frac{V_1^2}{2g}\right) + \left[\frac{\rho_a}{\rho_e}(z_{SL} - z_1) + z_1\right] + \sum \Delta H_{L(0\to1)} \tag{32.4}$$

Equation (32.4) demonstrates that the effluent level in the outfall shaft is a combination of (i) the head required to drive effluent out of the nozzle at the specified velocity, V_1, (ii) elevation to the center of the exit port (z_1)

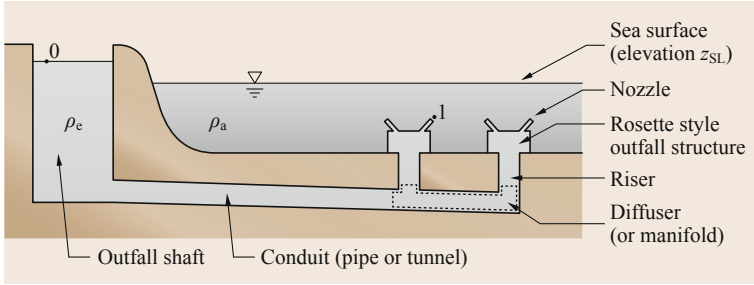


Fig. 32.5 Typical outfall system with rosette style outfall structures

plus discharge depth below sea level ($z_{SL} - z_1$) scaled by the ratio of fluid densities, and (iii) a summation of head losses through the system. Each of these components is briefly described in the following sections.

Nozzle Exit Velocity

The overall diffuser configuration including total number of ports or nozzles, and nozzle diameter (or exit velocity) is typically provided as an input to the hydraulic design based on the results of near-field modeling. The nozzle configuration affects the efficiency with which effluent is diluted in the near-field region and is generally selected based on the maximum outfall flow rate. In some outfall systems, such as those at desalination plants, maximum nozzle exit velocities of the order of 10 m/s may be required to ensure the brine is adequately diluted. The corresponding velocity head would likely be the largest component of the outfall shaft water level for exit velocities of this magnitude, especially for a relatively short outfall conduit with low conduit friction loss.

Sea Level

If seawater is discharged through the outfall system ($\rho_e = \rho_a$) the second term of (32.4) simplifies to z_{SL} and the outfall shaft fluid level is equal to sea level for the no-flow case. For sewage outfalls in which $\rho_e < \rho_a$, the effect of the density difference is to increase the outlet shaft effluent level. Conversely, the level in the outfall shaft is decreased when brine ($\rho_e > \rho_a$) is discharged. Density differences tend to be of the order of $(|\rho_e - \rho_a| / \rho_a) \times 100 \approx 3\%$ for most sewage and desalination applications. Although this difference corresponds to a relatively small change to outfall shaft level when discharging into shallow waters, (32.4) shows that the density ratio effect is amplified for deeper outfall discharges.

In addition to changes in plant operating conditions which could increase or decrease outlet flows, changes in sea level will also cause the outfall shaft effluent level to vary. The outfall system design should therefore consider the entire range of sea levels that could occur over the project design life, accounting for tidal fluctuations as well as storm surge. Statistical methods can be applied to sea level time series at the project location in order to determine exceedance probabilities and recurrence intervals. Using sound engineering judgment in conjunction with project requirements and/or local design standards for infrastructure design life, the results of the statistical analysis can be used to select design values for minimum and maximum sea level. To capture any seasonal trends which could include wind and barometric effects, sea level data used in the statistical analysis should include field measurements taken regularly throughout the year. It is imperative that data specific to the project location is used because sea level characteristics can vary greatly from one locale to another, regardless of the distance between them.

An allowance for sea level rise due to the effects of climate change should also be included because it could have a significant impact on maximum outfall shaft fluid level. Some statistical models estimate that the sea level will rise more than 1 m by the year 2100 (*Seneviratne* et al. [32.24]).

Head Losses

The total system head loss represented by the third term of (32.4) consists of the sum of conduit friction losses and the sum of local head losses through all fittings, system components (e.g., bends and contractions), and dividing flows in the manifold/diffuser. These losses can be expressed as

$$\sum \Delta H_L = \sum \frac{V_c^2}{2g}\left(f_c \frac{L_c}{D_c}\right) + \sum \frac{V_L^2}{2g}(K_L), \quad (32.5)$$

in which the first term on the right-hand side of the equation is the Darcy–Weisbach equation for conduit friction loss and f_c = conduit Darcy friction factor [–], D_c = conduit diameter [m], L_c = conduit length [m], V_c = velocity through the conduit [m/s], K_L = local head loss coefficient at fitting or component [–], and V_L = velocity through fitting or component [m/s].

In (32.5), the term *conduit* refers to the tunnel or pipe which delivers flow to the manifold and risers. The Darcy friction factor can either be determined from manufacturers' charts for particular wall roughnesses,

computed iteratively using the implicit Colebrook–White formula given in (32.6), or approximated using the explicit Swamee–Jain equation given in (32.7). The Swamee–Jain approximation is accurate to within a few percent of the value computed using the Colebrook–White equation over the typical ranges of roughness values and fully turbulent Reynolds numbers

$$\frac{1}{\sqrt{f_c}} = -2\log\left(\frac{k_S}{3.7D_c} + \frac{2.51}{\mathrm{Re}_c f_c^{1/2}}\right) \tag{32.6}$$

$$f_c \approx \frac{0.25}{\left[\log\left(\frac{k_S}{3.7D_c} + \frac{5.74}{\mathrm{Re}_c^{0.9}}\right)\right]^2} \tag{32.7}$$

where: k_S is the Nikuradse equivalent sand grain roughness of the conduit wall [m], Re_c is the conduit Reynolds number [−] $= \frac{V_c D_c}{\nu}$, ν is the effluent kinematic viscosity [m^2/s].

The kinematic viscosity of water for various temperatures and salinities can be obtained using the relationships provided in *Sharqawy* et al. [32.25]. Wall roughness values for common pipe materials can be found in any hydraulics data handbook, while roughness values for segmentally lined tunnels are presented in *Pitt* and *Ackers* [32.26]. It is customary to make an allowance for increased wall roughness over the conduit design life to account for aging and degradation.

Local head losses arise from flow through bends, tee or wye junctions, flow or pressure control devices (e.g., valves), and expansions or contractions in cross-sectional flow area. Local head losses will also occur at conduit entrances, at submerged discharges, and any location in the system where flow separation occurs. As shown in the second term of (32.5), local head loss is expressed as a multiple of the velocity head at the particular component of interest. The local loss coefficient, K_L, depends on the component geometry and is determined experimentally. Loss coefficients for common system components can be found in any hydraulics data handbook. Several references such as *Miller* [32.27] and *Idelchik* [32.28] are devoted entirely to local head loss coefficients and include many components present in marine outfall systems.

As local head loss coefficients are always based on some reference velocity, it is important to ensure consistency between a given K_L and the velocity, V_L, in the associated velocity head term ($V_L^2/2g$). For components with constant cross-sectional area such as certain bends, K_L is normally based on the average velocity through the bend. However, there is no standard reference velocity for components like nozzles or sudden expansions or contractions which have multiple cross-sectional areas; some sources may use the upstream velocity for reference, while others may use the downstream velocity. Using an incorrect reference velocity could result in significantly higher or lower head losses.

For complex hydraulic systems, it is important to note that the total local head loss may not simply be the sum of individual local loss components as (32.5) suggests. Rather, head loss coefficients are typically subject to certain limitations. For example, the coefficient for a single tee-junction can only be applied to a series of tee junctions (such as in a dividing manifold) if the separation distance between successive junctions is, say, 5 to 10 times the manifold diameter. Correctly applying local head loss coefficients will help to minimize under- or over-prediction of total local head loss. For cases in which the loss coefficient limitations are not clearly defined or the system configuration cannot easily be broken down into *standard* components (such as through a rosette-style outfall structure), physical and numerical modeling can be used to confirm head losses.

32.4.2 Diffusers – Hydraulic Design

Outfall systems usually consist of a manifold (also referred to as a diffuser) whereby a common pipe or tunnel supplies flow to multiple risers, ports, or branches. Although a manifold is an example of parallel flow in which the head loss between the outfall shaft and each exit port is the same, it is critical to note that the flow rate out of each port will not necessarily be the same. The variation in flow rate can be attributed to (i) decreasing flow and total head along the length of the manifold, (ii) changing depth along the manifold, and (iii) head loss coefficients for tee junctions which are a function of (a) the ratio of conduit diameter to branch diameter, and (b) the ratio of local flow through the conduit to flow through the branch. Head loss curves for a range of tee-junction configurations can be found in *Miller* [32.27].

Given that effluent dilution is directly related to port exit velocity, the overall design objective should be to achieve equal exit velocities (or as close to equal as possible) at each port to ensure consistent dilution levels over the length of the manifold. Traditional analytical methods of solving for the hydraulic performance of manifolds involve making an initial guess about the flow conditions at the most downstream port, then using an iterative approach to progressively work upstream, port by port, until a final solution is reached. An alternative approach using simultaneous equations is presented in this section. The resulting set of equations can be quickly solved using a spreadsheet application with built-in equation solver. The spreadsheet can be set up to optimize port velocities by varying known quantities such as the port and manifold diameters and/or port

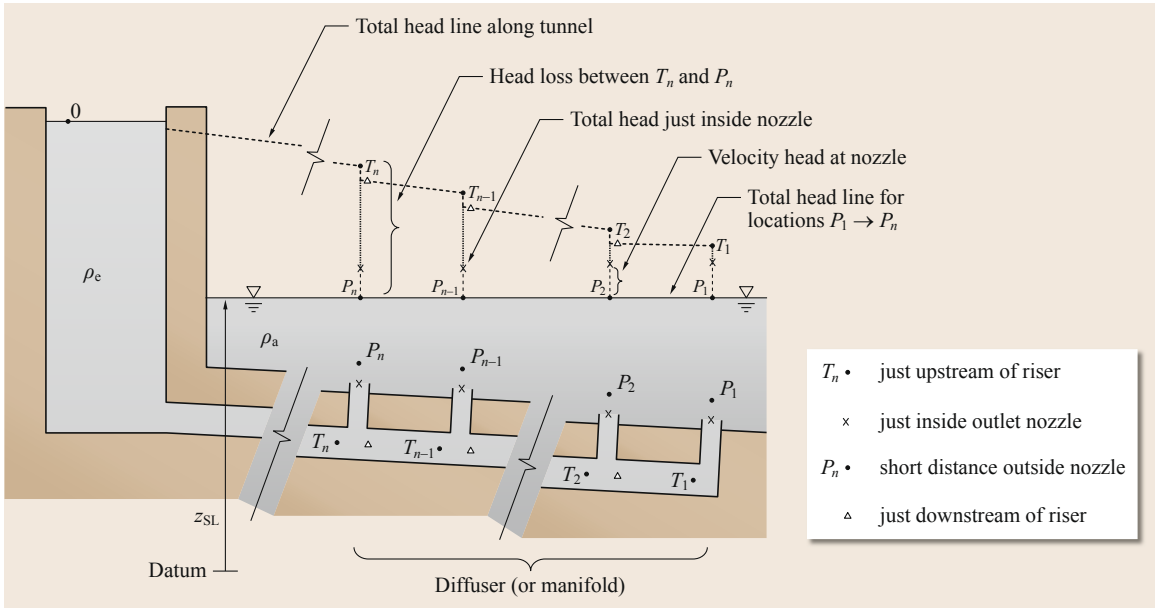


Fig. 32.6 Schematic for a manifold or diffuser flow calculation

spacing. Once the final manifold configuration is selected, the same spreadsheet model can also be used to determine the resulting port velocities and outfall shaft fluid levels over a range of flow rates and/or effluent densities.

Consider the diffuser with n ports arranged along a tunnel manifold as shown in Fig. 32.6. Points P_1 through P_n correspond to locations downstream of the individual port openings, while points T_1 through T_n are located along the tunnel centerline, just upstream of the port with the same subscript i. Flow conditions through this system – or any similar system – can be solved using the set of simultaneous equations provided in Table 32.4. Note that the values in brackets in the 3rd and 5th columns indicate that there is an equal number of equations and unknown variables ($8n + 1$). The parameters in the 6th column labeled *inputs* are assumed to be known values. The governing principles reflected in this set of equations are:

- *Continuity* – the sum of individual port flow rates (ΣQ_{P_i}) must equal the total outfall flow rate (Q_T). (Table 32.4, (8a))
- *Energy equation* – the total head at point P_n must be equal to the total head at point 0, less the total head loss between these two points $\left(\sum \Delta H_{L(0\rightarrow P_n)}\right)$. (Table 32.4, (8j))
- *Parallel flow* – the total head loss between points T_{i+1} and

$$P_i \left(\sum \Delta H_{L(T_{i+1}\rightarrow P_i)}\right)$$

must be equal to the total head loss between points T_{i+1} and

$$P_{i+1} \left(\sum \Delta H_{L(T_{i+1}\rightarrow P_{i+1})}\right)$$

(Table 32.4, (8k)). Note that exit loss should be included in the expression for the head loss between T_i and P_i (Table 32.4, (8f)) because the velocity head has been fully dissipated by the time the discharge has reached any point P_i. The total head at all points P_i are equal.

Subscript P_i denotes individual ports, subscript T_i denotes individual tunnel manifold sections, subscript P_n denotes the most upstream port, and subscript T_n denotes the tunnel section between the outfall shaft and port P_n (Fig. 32.6).

In cases where loss coefficients $K_{L_{T_i\rightarrow P_i}}$ in (8f) are functions of a flow ratio between the manifold and individual ports, the set of simultaneous equations in Table 32.4 may need several iterations in which adjustments are made to the loss coefficients after each iteration until the change in solution between successive iterations is negligible.

Note that the head loss coefficients in (8f) of Table 32.4 are presented in terms of the port exit velocity (i. e., the velocity at which the effluent is discharged into the sea). For cases in which the port geometry is more complex with varying diameters and multiple head losses, all loss coefficients used in (8f) of Table 32.4 must be converted such that they are based on the port exit velocity. To convert a head loss coefficient

Table 32.4 Set of simultaneous equations for manifold flow calculations

Description	Equations	No. of equations	Unknowns	No. of unknowns	Inputs	Equation
Continuity: total flow rate is equal to sum of individual port flow rates	$Q_T = \sum_{i=1}^{n} Q_{P_i}$	1	Q_{P_i}	n	Q_T	(8a)
Port exit velocity	$V_{P_i} = \frac{Q_{P_i}}{\frac{\pi}{4} D_{P_i}^2}$	n	V_{P_i}	n	D_{P_i}	(8b)
Velocity in manifold section between adjacent ports	$V_{T_i} = \frac{\sum_{1}^{i} Q_{P_i}}{\frac{\pi}{4} D_{T_i}^2}$	n	V_{T_i}	n	D_{T_i}	(8c)
Reynolds number in manifold section between adjacent ports	$\mathrm{Re}_{T_i} = \frac{V_{T_i} D_{T_i}}{\nu}$	n	Re_{T_i}	n	ν	(8d)
Darcy friction factor in manifold section between adjacent ports	$f_{T_i} = \frac{0.25}{\left[\log\left(\frac{k_{s_i}}{3.7 D_{T_i}} + \frac{5.74}{\mathrm{Re}_{T_i}^{0.9}}\right)\right]^2}$	n	f_{T_i}	n	k_{s_i}	(8e)
Head loss between manifold station i and just downstream of port i	$\sum \Delta H_{\mathrm{L}(T_i \to P_i)} = \frac{V_{P_i}^2}{2g}\left(\sum K_{\mathrm{L}_{T_i \to P_i}}\right)$	n	$\sum \Delta H_{\mathrm{L}(T_i \to P_i)}$	n	$\sum K_{\mathrm{L}_{T_i \to P_i}}$	(8f)
Head loss between outfall shaft and just downstream of the n-th port	$\sum \Delta H_{\mathrm{L}(0 \to P_n)} = \frac{V_{T_n}^2}{2g}\left(f_{T_n}\frac{l_{T_n}}{D_{T_n}} + \sum K_{\mathrm{L}(0 \to T_n)}\right) + \sum \Delta H_{\mathrm{L}(T_n \to P_n)}$	1	$\sum \Delta H_{\mathrm{L}(0 \to P_n)}$	1	l_{T_n} D_{T_n} $\sum K_{\mathrm{L}(0 \to T_n)}$	(8g)
Head loss between manifold station $(i+1)$ and just downstream of port i	$\sum \Delta H_{\mathrm{L}(T_{i+1} \to P_i)} = \frac{V_{T_i}^2}{2g}\left(f_{T_i}\frac{l_{T_i}}{D_{T_i}} + \sum K_{\mathrm{L}(T_{i+1} \to T_i)}\right) + \sum \Delta H_{\mathrm{L}(T_i \to P_i)}$	$n-1$	$\sum \Delta H_{\mathrm{L}(T_{i+1} \to P_i)}$	$n-1$	l_{T_i} $\sum K_{\mathrm{L}(T_{i+1} \to T_i)}$	(8h)
Pressure just downstream of port i, expressed in terms of the ambient seawater density	$p_{P_i} = \left(z_{\mathrm{SL}} - z_{P_i}\right)\rho_{\mathrm{a}} g$	n	p_{P_i}	n	z_{SL} z_{P_i} ρ_{a}	(8i)
Energy equation applied between point 0 and the most upstream port (n)	$\frac{V_0^2}{2g} + \frac{p_0}{\rho_{\mathrm{e}} g} + z_0 = \frac{V_{P_n}^2}{2g} + \frac{p_{P_n}}{\rho_{\mathrm{e}} g} + z_{P_n} + \sum \Delta H_{\mathrm{L}(0 \to P_n)}$	1	z_0	1	V_0 p_0, ρ_{e} z_{P_n}	(8j)
Parallel flow: head loss is the same between manifold station $i+1$ and just downstream of either port i or port $i+1$	$\sum \Delta H_{\mathrm{L}(T_{i+1} \to P_i)} = \sum \Delta H_{\mathrm{L}(T_{i+1} \to P_{i+1})}$	$n-1$	-	-	-	(8k)
Total:		$8n+1$		$8n+1$		

based on some velocity into an effective loss coefficient based on another velocity, the following equation can be used,

$$\frac{V_1^2}{2g} K_{\mathrm{L}_1} = \frac{V_2^2}{2g} K_{\mathrm{L}_1'} \Rightarrow K_{\mathrm{L}_1'} = \frac{V_1^2}{V_2^2} K_{\mathrm{L}_1} \,. \tag{32.8}$$

Manifold Section Diameters

The equations in Table 32.4 are kept general so that geometric parameters can be varied along the length of the manifold. In some cases, it may be necessary to progressively reduce the manifold diameter in order to maintain velocities that prevent settlement of solids

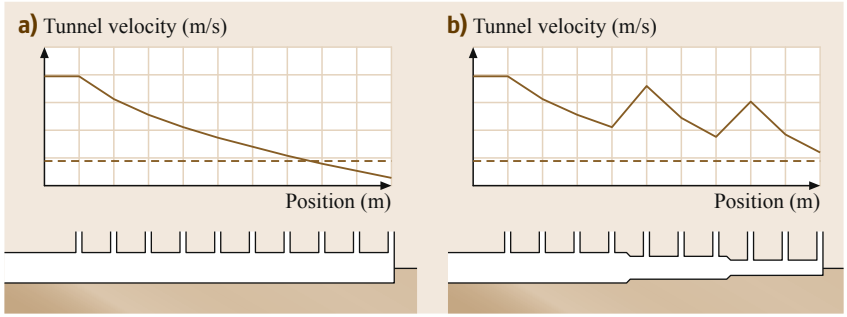


Fig. 32.7a,b Tunnel velocity variation along manifold – **(a)** constant diameter and **(b)** stepped diameter . The *dotted line* indicates a minimum self-cleansing velocity; refer to Sect. 32.3.7 for details

along the invert. (Due to the potential for excessively high head losses, it may not be possible to maintain sufficient self-cleansing velocities over the entire length of a conduit with a constant diameter.) Velocities through a typical diffuser are shown for two cases in Fig. 32.7 – constant diameter and stepped diameter. The dotted line represents a nominal minimum velocity required to prevent solids from accumulating. It should be noted that reducing the manifold section diameters is usually only feasible for applications involving piped outfalls. For tunneled solutions in which sedimentation can occur, an allowance for sedimentation build-up or a provision for periodic removal of accumulated sediments should be included in the design of the system. Refer to Sect. 32.3.7 for a discussion on sedimentation.

Port Diameters

The number of ports as well as the port exit velocity required to ensure adequate dilution are typically determined from the results of near-field modeling. These constraints effectively set the port diameter which, in turn, establishes hydraulic performance of the system. After the port diameter has been selected, the system should be checked for excessively high head losses, unbalanced flow distribution, and seawater intrusion (for sewage outfalls only) to minimize the potential for adverse operating conditions. In practice, an iterative design process is usually required in order to strike an appropriate balance between dilution performance and favorable hydraulic conditions.

Achieving a consistent discharge velocity at all ports along a line diffuser is not necessarily a simple task due to the varying head loss coefficients along the length of the manifold. An analysis of head loss coefficients for dividing flows shows that they are a function of both the flow ratio and the area ratio, between individual ports and the corresponding conduit section [32.27]. In general, loss coefficients are greater for larger flow ratios and smaller area ratios. Assuming constant conduit and port diameters, upstream ports will therefore have relatively low loss coefficients while downstream ports will have relatively high loss coefficients. In order to satisfy the principle of parallel flow such that head loss between the outfall shaft and each port is the same, the flow rate through each port will be different.

The flow variation among ports is not necessarily a problem from a hydraulic perspective. However, the near-field effluent dilution may become unbalanced if the flow differential among ports becomes too great. This condition may prevent near-field dilution targets from being met which could lead to adverse environmental consequences. In the case of sewage outfalls in which the effluent density is less than that of seawater, the variation in port discharges can create another undesirable condition – seawater intrusion. One way to prevent seawater intrusion from occurring is to use smaller port diameters to ensure adequate port discharge velocities. Modeling a less dense fluid flowing through an orifice into a more dense fluid shows that seawater intrusion can be prevented when the port densimetric Froude number is greater than approximately 1.6 [32.29],

$$\mathrm{Fr} = \frac{V_{P_i}}{\sqrt{g\left(\frac{\rho_a - \rho_e}{\rho_e}\right) D_{P_i}}} > 1.6\,. \tag{32.9}$$

In practice however, port densimetric Froude numbers are typically kept well above this threshold value. For example, the port densimetric Froude number at the Sydney deep water ocean outfalls is of the order of 20–30.

A sensitivity analysis of port diameters using the set of simultaneous equations in Table 32.4 shows that port flow rates along the manifold do not vary significantly after port diameters are reduced below some critical value. Moreover, reducing the port diameter will increase the overall system head loss and lead to higher outfall shaft effluent levels. In some cases, the maximum possible head that can be accommodated at the outfall shaft may dictate the smallest allowable port diameters. If these diameters still do not lead to sufficiently high velocities, an alternative configuration may be required.

One option is to use variable-orifice nozzles such as the duckbill valves described in Sect. 32.4.3.

Another option is to use rosette style outfall structures. A similar analysis of flow conditions through a manifold configured with rosettes shows that port discharges can be more evenly distributed. This better flow balance occurs because head loss coefficients along the manifold are nearly uniform along its length, provided the friction losses and local losses between the conduit and risers are kept low. In this case, flow rates into each rosette will be nearly equal. Furthermore, the flow rate through each port on a given rosette structure will be the same for axisymmetric rosette designs because the total head loss coefficient will be the same regardless of which port the effluent flows through. Rosettes can have any number of nozzles, however, there tends to be an upper limit beyond which the additional nozzles will impede individual jet mixing processes. This behavior occurs because neighboring jets tend to coalesce due to reduced pressures caused by entrainment of the ambient water between them. The result is to reduce overall dilution performance. Near-field modeling can be used to determine the maximum number of nozzles for a specific rosette structure configuration.

32.4.3 Flow Variability

The wide range of flows that may be experienced at treatment plants is one of the greatest challenges in designing an outfall system. The preceding discussion assumes that the nozzle diameters have been sized to provide the required dilution for the maximum design flow rate, while at the same time maintaining sufficient velocity to prevent salt water intrusion at the ports and avoiding problems arising from sedimentation. In reality, treatment plants operate over a wide range of flows, and it is quite possible that the design flow might not be achieved until many years after the plant is initially constructed. In addition, plant start-up and ramp-down operations as well as any major changes to process flow requirements could create situations in which the plant will operate well below the design flow rate. The frequency and duration of any such events that result in low dilutions should be considered. If low flow scenarios are likely to result in sustained periods of insufficient dilution, the following mitigation schemes for managing low-flow scenarios can be used. Each of these options has advantages and disadvantages, and a hybrid solution may turn out to be the most suitable approach.

Supplemental Flows

For desalination plants, a connection between the intake and outfall systems can be incorporated in the plant design to provide a way for seawater to be bypassed directly into the outfall system during periods of low outlet flow. The bypassed seawater effectively increases the outlet flow rate and nozzle exit velocity, thereby providing additional momentum for dilution by mixing. The bypassed seawater also acts as an initial diluting agent before effluent is discharged into the sea; as such, adequate levels of dilution may be obtained with lower nozzle exit velocities for a given brine flow. The amount of bypassed seawater required for any particular combination of outlet flow rate and salinity can be determined using results of the near-field modeling and basic conservation principles.

Although this option provides a near instantaneous level of control for managing the outfall system (when the appropriate flow control devices are included in the design), the addition of bypassed seawater will lead to increased energy use and operating costs due to additional pumping of seawater. These costs should be considered when deciding whether this option is an appropriate strategy for managing low flows. Depending on the frequency and amount of bypassing required, this solution may, in fact, be found to be the most practical or cost effective when compared with other alternatives. The incorporation of a seawater bypass system also provides a way to commission both the marine intake and outfall systems independently from the rest of the plant.

Staged Construction of Marine Works

Many treatment plants are designed to be initially operated at a low flow rate, then periodically upgraded until the maximum design capacity is attained. To manage the low flows associated with early stage plant operation, the marine construction works can be staged such that only some of the nozzles or rosette structures required for the ultimate flow case are initially installed. Additional nozzles, diffuser structures, or sections of a line diffuser can be installed each time the plant is upgraded; near-field modeling will inform the number of nozzles to be installed at each stage.

The main disadvantage with this solution is that additional marine construction works are required whenever the plant is upgraded. Moreover, it assumes that once the plant is upgraded to ultimate capacity, it will not operate at the lower early stage flows. This assumption may not be valid if demand can fall below the plant's maximum capacity after it is upgraded (for example, lower demand from a desalination plant due to a temporary period of increased rainfall). This solution also does not address periods of low flow that will occur even for ultimate capacity when the plant ramps up or shuts down due to operations or maintenance requirements. These periods of low flows may not be an issue if they occur infrequently and for short durations, and if environmental licences allow for temporary periods of

lower dilution. As such, plant processes and the feasibility of periodic marine construction works need to be considered should this option be pursued.

Nozzle Sizes and Blanks

If it is not feasible to stage the installation of marine outlet structures or sections of a line diffuser, another option for managing low flows is to install blanking plates at the nozzle openings. This solution provides greater flexibility than staged construction because each nozzle can be brought online or taken offline individually to suit the flow requirements. In addition, it allows the total flow rate to be evenly distributed since nozzles on multiple structures or at any location along a line diffuser can be blanked off. Near-field modeling is needed to confirm the number and location of nozzles to be blanked off for each long-term operating flow rate. Input from the plant operators should be sought when considering this option, as making adjustments to nozzle blanks more than two or three times per year may render this option infeasible.

A similar option is to install smaller diameter nozzles for the initial stages, then remove and replace with larger diameter nozzles each time the plant is upgraded. To minimize installation activities, each set of nozzles should be designed to have the same entrance diameter such that modification to the manifold or rosette structures is avoided. This option provides the same degree of flexibility as nozzle blanking if demand or process requirements lead to an extended period of low flow; the smaller nozzles can be re-installed at any time.

Although these options will also require marine works each time the plant is upgraded, the scope and duration of such works is likely to be considerably less than for staged installation of marine structures or the addition of more diffuser length. These options also carry some operational risk due to the fact that the plant could still be operated at any given flow rate, even with an incorrect number or size of nozzles. A nozzle configuration that is not aligned with the target operating flow could lead either to insufficient dilution (for lower flow), or excessively high head losses through the system (for higher flow). The operating procedures for these two options must ensure that plant capacity is linked to the installed nozzle configuration at any given time.

Variable-Orifice Nozzles

Variable-orifice nozzles, also known as *duckbill valves*, can be used instead of fixed diameter nozzles. These valves are made from a variety of flexible materials such as rubber or neoprene and are designed to close during periods of no flow. Due to stiffness of the valve material, the valves gradually open in an elliptical shape as the flow rate increases resulting in higher velocities at lower flow rates as compared to fixed-diameter nozzles. This eliminates or reduces the need for blanking nozzles or introducing bypassed seawater into the outlet flows is reduced or eliminated. The check valve feature of duckbill valves also prevents backflow from the sea into the outfall system during periods of zero or low flow. Seawater intrusion into wastewater outfalls can therefore be prevented with this option.

As material stiffness dictates the degree to which the valves will open for any given flow rate, these valves are usually custom designed for each particular application. Hydraulic testing of the valves is recommended prior to selection and incorporation into the final design, given the specificity of duckbill valves to each outfall design. Hysteresis of the material should also be investigated to ensure that the valve will perform consistently over time. This consideration is particularly important if flows are expected to be highly variable and/or prevail for extended periods (i. e., months or longer). The local conditions in which the valves will be installed is another important aspect that should be taken into account when deciding whether duckbill valves are an appropriate solution. If high levels of marine growth are expected, the valves may not close properly if organisms attach themselves to the area near the valve opening.

32.4.4 Hydraulic Integration

Following a hydraulic assessment, the range of piezometric heads at the upstream boundary of an outfall system must be considered with respect to the hydraulics through the entire treatment plant. Failure to properly integrate the outfall system with other plant systems could result in adverse operating conditions such as decreased plant capacity, spillage and flooding, and inefficiencies for pumped systems. The choice between gravity-driven or pumped outfalls depends on several factors including site topography and plant elevation, management of low flow scenarios and long-term operating or maintenance costs.

Gravity Systems

Gravity-driven outfall systems are often preferred because they are relatively simple to operate (few electrical and mechanical components) and have virtually no long-term operating costs. Apart from periodic inspection and maintenance, the passive nature of a gravity-driven system ensures continuous reliability and minimal operator intervention. The free surface that exists at an outfall shaft is convenient when outflows from different parts of the treatment plant must be collected and combined into a single outfall stream. In addition,

a hydraulic break is usually required for certain process schemes such as removal of entrained air or settlement of suspended particles; refer to Sects. 32.4.5 and 32.4.6 for a discussion of these issues.

The range of piezometric heads in the outfall shaft should first be considered in relation to the vertical alignment of the outfall conduit and ground surface level at the shaft location. Given the complexity associated with installing an outfall conduit, the key factors in selecting its vertical alignment typically include suitability of ground conditions, constructability, cost, and construction scheduling. For example, a trenched solution which follows site topography and bathymetry is often desired as it can be the cheapest and easiest to build. However, this alignment may cause adverse hydraulic conditions such as hydraulic jumps and accumulation of air at localized high points. The potential for such conditions should be assessed before finalizing the outfall conduit alignment.

In general, the entrance to the outfall conduit should be fully submerged for all flow rates and sea levels in order to prevent operational instabilities. Otherwise, hydraulic jumps would be likely to occur at the downstream end of unpressurized steep sections or where free surface flow conditions develop along an alignment that follows a ground profile containing intermediate peaks and valleys, as illustrated in Fig. 32.8. Hydraulic jumps would likely be dynamic, unstable, and unpredictable, and could lead to air entrainment (Sect. 32.4.5 discusses the effects of air entrainment in greater detail).

Regardless of whether the outfall shaft can be designed and constructed to accommodate the range of operating levels, these levels must also be compatible with the upstream treatment plant systems. Ideally, the plant is configured in such a way that shaft fluid levels do not influence the upstream hydraulics. At the very least, any backwater effect caused by the outfall shaft levels should not adversely impact the treatment plant processes or reduce the plant operating capacity.

Vertical Drops

For cases in which the treatment plant is located at a significantly higher elevation than the outfall shaft fluid levels, a means of transferring the outflows down a large vertical distance may be required. Allowing outlet flows to simply drop over a weir into the outfall shaft causes a great deal of air entrainment and results in rapid dissipation of energy which can cause operational instabilities and structural damage if not properly managed. A vortex, drop shaft system can be incorporated into the design using methods such as those outlined in *Hager* [32.30] in order to gradually dissipate energy and minimize impact on the structure, and reduce the potential for air entrainment.

As an alternative, a mini-hydroelectric scheme can be installed to recover the energy that would otherwise be *lost* by the vertical drop between the plant and the outfall shaft. Whether the recovered energy is used to power other parts of the plant or sold back to the power grid, a substantial portion of the costs associated with pumping of intake flows (e.g., at a desalination plant) may be recoverable. The potential cost savings associated with any such mini-hydro scheme should be compared with the initial cost of all necessary mechanical and electrical components, as well as their ongoing maintenance and replacement costs. The frequency and duration of flow/head combinations should also be considered when determining viability of a mini-hydro scheme. While a large amount of energy might be available for recovery during peak flow conditions, lower operating flow rates over the long term might render this option uneconomical. If a mini-hydro scheme is included, a passive system such as the vortex drop shaft may also be required as an emergency bypass or back-up in case the turbines need to be taken offline.

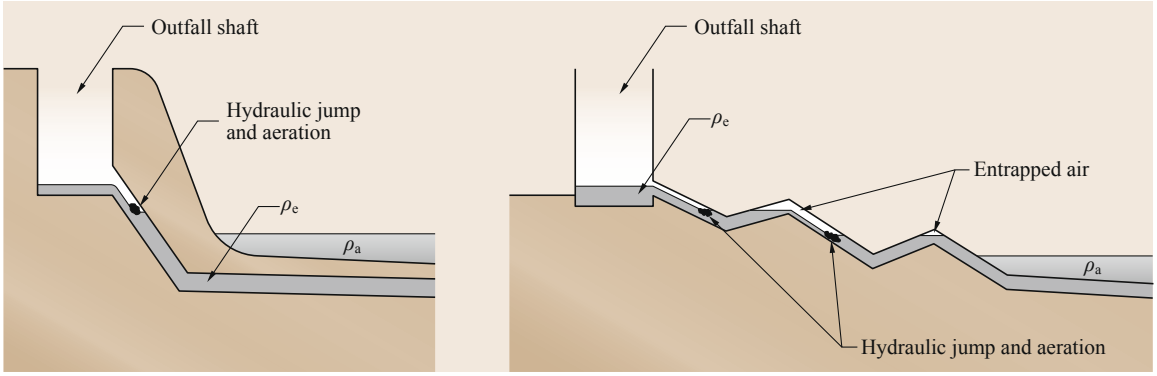


Fig. 32.8 Examples of configurations that could entrain or entrap air and cause hydraulic jumps

Pumped Systems

If specific site conditions render a gravity-driven system unfeasible, a pumped solution will be needed to provide the driving head required for maintaining the required nozzle exit velocities. The hydraulic analysis for a pumped system is the same as for a gravity-driven system because the pumps will need to deliver flow at a head equal to the fluid level in a hypothetical outfall shaft placed at the pumping station location. Pump selection should be based not only on the required piezometric head on the discharge side of the pump (which is a function of both flow rate and sea level), but also on any flow- or time-based fluctuations of the upstream water level or pressure. Selecting a single type and size of pump to efficiently cover the entire flow and head range may not be possible. However, using a single type and size of pump is generally desired from an operations and maintenance perspective because it adds flexibility and redundancy to the system.

Pumping can also be used as a way to manage low flow scenarios or ensure constant nozzle exit velocity over the entire flow range. This option provides near instantaneous control for managing nozzle exit velocities, but the associated operating costs will also be relatively high. Nonetheless, a pumped outfall system may be required to manage low flows at a desalination plant outfall when bypassing seawater is not possible.

32.4.5 Air Entrainment

Managing air that becomes entrained in the treatment plant effluent stream is an important consideration for outfall system design. If air bubbles are not captured and released before effluent enters the outfall conduit, the bubbles could eventually agglomerate and form air pockets. Depending on the particular conduit configuration, air pockets can become trapped at localized high points along the outfall conduit and alter the hydraulic performance of the system. Localized head losses can be induced at these locations due to the associated flow contractions and expansions. This condition could lead to reduced system capacity and the potential for excessively high outfall shaft water levels (in a gravity system) or inefficient pump operation (in a pumped solution). As it is typically not feasible to include air release mechanisms along an outfall conduit, trapped air pockets can remain in the system for extended periods of time.

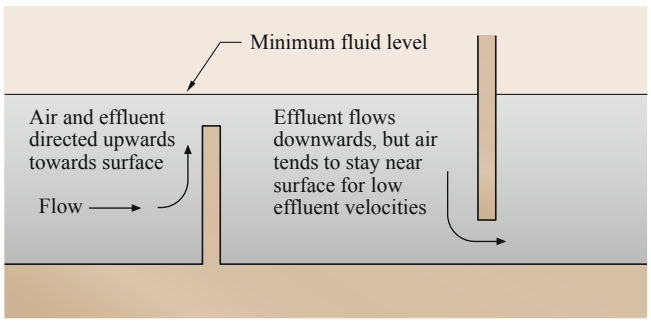


Fig. 32.9 Indicative deaeration system using baffles

Air pocket accumulation can also be problematic even if there are no high points along the alignment. In the case of a conduit with constant slope (including zero slope), entrained air will tend to come out of solution and accumulate along the conduit soffit. Depending on the conduit slope and diameter, and the effluent flow rate, the air pocket may move upstream or downstream. Regardless of direction, the air pocket may not start moving until it reaches some critical volume or pressure, at which point it may move suddenly and violently. This sudden release of air at either end of the conduit results in a condition called *blow-back* or *blow-out* which can cause strong vibrations and even structural damage in extreme cases. Operational instabilities in the outfall system and farther upstream may be experienced, especially in cases where the gradual accumulation of air followed by a sudden release (i. e., gulping), becomes repetitive.

It is important to recognize the processes by which air becomes entrained so that the system can be designed to minimize the adverse conditions described above:

- Flow conditions at the upstream end of the outfall conduit can be a major contributor to entrained air if the effluent is conveyed vertically through a drop shaft, over a weir, or by way of any other hydraulic structure which causes a jet to plunge through air.
- Hydraulic jumps can occur in open channels at locations where the cross-sectional flow area or depth changes abruptly. It is even possible to have hydraulic jumps occur within the outfall conduit if the upstream end is not fully submerged.
- For wastewater outfalls, gases can be generated over the entire conduit length due to biological processes in the effluent.

The two main approaches to removing entrained air are: (i) remove it upstream of the entrance to the outfall conduit, and (ii) configure the outfall conduit such that air bubbles either move upstream or downstream without forming large air pockets. A deaeration chamber or channel can be included immediately upstream of the outfall conduit entrance to capture and release air from the effluent in a controlled manner. The general idea behind these schemes is to provide a flow path

of sufficient length such that air bubbles have enough time to rise to the surface before entering the conduit. Indicative air bubble terminal rise velocities are provided in *Falvey* [32.31] and *Lauchlan* et al. [32.32]. To minimize the deaeration chamber or channel flow path length, effluent velocities can be reduced by increasing the cross-sectional flow area. Properly configured baffle walls can also be used to direct flow upward and facilitate air removal through the fluid–air interface, or to permit only the effluent at the bottom of a chamber or channel to continue downstream toward the outfall conduit; effluent near the invert is likely to have less entrained air than that near the surface for low effluent velocities. Refer to Fig. 32.9 that shows an indicative deaeration system using a baffled configuration.

In other cases where site constraints limit the amount of space available for upstream deaeration schemes, or where a majority of entrained air is the result of biological activity occurring in the outfall conduit, an alternative approach can be taken. The conduit can be designed to allow air bubbles to travel to either the upstream or downstream ends of the system. *Falvey* [32.31] presents a series of curves which summarize a range of conditions for air bubble and air pocket movement in a closed and fully flowing conduit; refer to Fig. 32.10 in which Q is the conduit flow rate, D is the conduit diameter, and θ is the angle of the conduit to the horizontal.

32.4.6 Sedimentation

Sediment management tends to be a more relevant issue at wastewater treatment plants than at desalination plants; desalination inflow (i. e., seawater) generally has fewer solids than sewage, and its pretreatment processes generally remove a wider spectrum of material. Indeed, the filtration membranes in reverse osmosis plants are very sensitive to fouling and require the removal of solids as well as organic compounds. Nonetheless, the solid matter removed in the desalination pretreatment processes may end up being added to the brine effluent downstream of the reverse osmosis filters if not disposed of by other means. The design of an outfall conduit for a desalination plant must therefore consider all plant processes that discharge an effluent stream into the outfall system.

The build-up of granular sediments, organic matter, and other solids along the invert of an outfall conduit will eventually lead to increased head losses and decreased hydraulic capacity. In extreme cases, the conduit section can become clogged and the function of the discharge nozzles impeded. There are two general approaches to reduce the risk of sediment build-up, each of which is described below. For some applications, a combination of these options may prove to be the most effective solution.

Minimize the Amount of Solids that Enter the Conduit

By providing a way for solids to settle out of the effluent stream upstream of the conduit entrance, deposition and removal becomes a matter that can be managed locally at the plant site, rather than along the entire length of the conduit. This approach using sedimentation basins or settling tanks may be the preferred solution for cases in which the outfall conduit is long and/or inaccessible for periodic inspections and cleaning. It may also be used if it is not possible to design the conduit to achieve self-cleaning velocities. The general design principle behind this approach is that the sedimentation basins/tanks reduce the effluent stream velocity to a point where the time it takes for suspended particles to settle is less than the residence time of the fluid in the basin or tank. Because particle settling time increases as the target particle size decreases, a potential drawback of this option is that a large basin or tank footprint may be re-

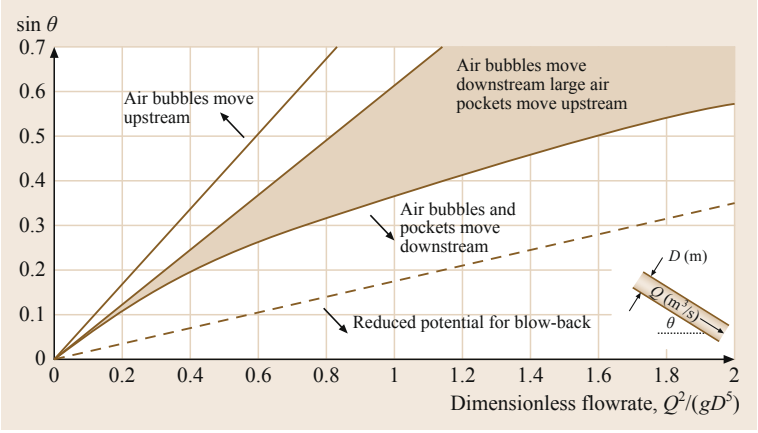


Fig. 32.10 Air movement in pressurized conduits (after *Falvey* [32.31, Fig. 29])

quired. As such, this option may only be feasible for removing larger solids. To decrease the basin/tank footprint and increase its efficiency, internal baffle walls can be installed to lengthen the flow path. However, the resulting effluent velocity should remain sufficiently low to allow sediments to settle out.

Design for Self-Cleansing Velocities

If upstream sedimentation control is not feasible, the outfall conduit diameter can be selected to yield velocities that are high enough to prevent deposition of solid materials. Given the wide flow variability associated with many types of treatment plants, the outfall conduit diameter should be based on the lowest-flow scenario. If, however, the chosen diameter leads to excessively high velocities and head losses during peak flow, it may not be possible to achieve self-cleaning velocities for all flow conditions. In this case, the design objective should aim for self-cleaning velocities on a frequent basis (i. e., at least once per day) and for sustained periods of time (i. e., hours) in order to prevent build-up of sediments and organic matter over time. The research literature contains a great deal of information on the velocities and shear stresses required to move solids through the conduit and to re-suspend solids that may have already settled. Velocities to keep solids in suspension range from approximately 0.5–2.0 m/s depending on conduit diameter, particle size and specific gravity, and particle concentration [32.33].

32.5 Outfall Construction

An effective design is not possible without input from those engaged to construct the outfall. The design must be detailed in conjunction with how the marine outfall will be built. *Grace* [32.34, 35] and *Wood* et al. [32.4] provide good information on outfall construction approaches and techniques.

There are two fundamental types of wastewater delivery: via tunnels or pipes (either on trestles or in trenches). They can be used in combination, a tunnel for the bulk of the outfall length and a pipeline on the sea bed comprising the outlet ports. The decision is often governed by the geography of the region – tunnels may be preferred in areas with rocky coastlines or where beach access is difficult, while pipelines may be preferred in areas with easy access such as a sandy shoreline.

32.5.1 Construction Materials

The materials used for outfall construction include steel, reinforced concrete, and high density polyethylene. The marine environment is corrosive and steel structures must be protected by coating the steel to prevent direct contact with sea water (often concrete is used) and/or using sacrificial anodes. Ongoing inspections are required to monitor the integrity of the protective coating or the anodes and replace them as necessary.

Steel reinforcement in concrete is susceptible to corrosion by chloride salts in marine waters. To help in preventing such corrosion, low permeability concrete is used, often in conjunction with additives that inhibit corrosion. Hydrogen sulphide in wastewater is corrosive to concrete and concrete pipes are lined (usually with plastic) for protection. Regular inspections are undertaken to ensure the lining has not peeled away from the concrete. The density of concrete is about one-third that of steel (although still about double that of seawater) and the concrete pipeline may need to be anchored to the sea floor.

High-density polyethylene is relatively light, with a density slightly less than that of seawater. Its big advantage is that it is flexible and can be relatively easy to deploy. Long sections are welded together on shore, the pipeline then towed into position and anchored to the sea floor. Purging of air from the pipe is critical to prevent the pipe from floating back to the surface.

32.5.2 Construction Methods

Pipelines are often manufactured in sections and assembled at an access point close to the proposed line of the outfall. When in the water, each section is buoyed to facilitate easy movement. Offshore, a vessel is anchored and pulls a section of the pipe into a trench along the line of the outfall. The next section is welded on and the pull continues. This is a popular technique for steel pipelines, which are very strong in tension. Concrete blocks or ballast rock may be used to anchor the pipe to the sea floor.

The trench is then backfilled either mechanically or by natural means. Techniques have been devised whereby trenching, pipe-laying and backfilling are all done in a single operation. The end of the pipeline terminates in a diffuser and the wastewater is discharged through outlet nozzles.

The active wave climate of the surf zone makes it the most critical region. Often, a temporary trestle and sheet piling is used to protect the pipeline through this region. Sometimes, wave action can damage the trestle and sheet piling (Fig. 32.11).

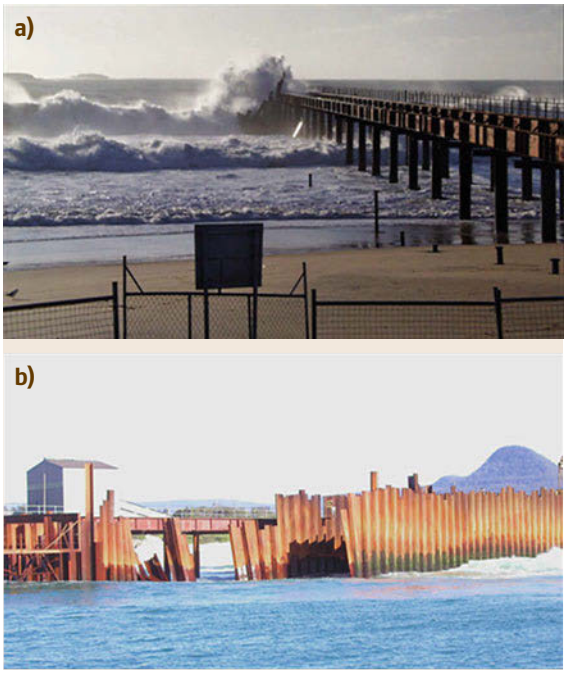


Fig. 32.11a,b Photographs from the same site, during **(a)** and after **(b)** a storm that caused extensive damage to sheet piling. The sheet piling was placed to protect the outfall pipeline (located in a trench) as it passes through the surf zone (photographs by P. Tate)

Tunnels are expensive to construct but their big advantage over pipelines is that they avoid the surf zone where damage to a pipeline is more likely to occur. This is particularly important along high wave energy coastlines.

One of the two tunnel construction techniques is usually adopted: full face tunnel boring or drill and blast. The latter technique is more commonly used on shorter outfalls where the cost of a tunnel boring machine cannot be justified. The drill and blast technique can be dangerous, releasing gases into the tunnel or weakening the tunnel structure.

Long risers (tens of metres or more in length) are drilled from the sea floor down to the tunnel. The risers are capped with high velocity nozzles through which the wastewater is discharged. On the three Sydney deep water outfalls, between four and eight outlet nozzles are fitted in a rosette configuration to each riser.

32.5.3 Some Considerations

Environmental Impacts of Construction

During the construction of a marine outfall there will inevitably be environmental damage caused by trenching and drilling activities. Habitats may be removed, infauna displaced and particulate material placed into suspension. Apart from the obvious physical destruction of habitats, suspended matter may reduce light in the water column (affecting photosynthesis), clog the gills of fish and, on resettling, may smother marine plants and infauna.

Route Selection

The shortest distance may not be the optimal route for the outfall. Other influencing factors include: the presence of rocky outcrops, wave and current climate, maritime activity, fishing zones, and ecological considerations. The route of the Wollongong outfall (Australia), constructed in 2005/2006, was changed in its early design phase to avoid the habitats of the weedy seadragon (*Phyllopteryx taeniolatus*) – a local protected species.

Wave and Current Climate

Waves can induce drag and lift forces on pipelines. Breaking waves are most prevalent in the surf zone. They can produce very large forces on a pipeline, although their duration is short lived. Abnormally large waves can be generated from storm surges, rogue waves and tsunamis. Currents over a pipeline exert a drag force. The weight of the pipeline needs to be sufficient to prevent such forces from moving the pipeline. *Grace* [32.34] provides some simple calculations that estimate the force exerted on a pipeline by waves and currents. In part, this problem can be largely overcome by burying the pipeline in a trench, which is then backfilled.

Sediment Movement

The current and wave climate may resuspend sediments. Once in suspension, their movement and distribution can be widespread. Of particular importance is the potential for sediments to erode from beneath the pipeline, potentially placing considerable stress on the pipeline itself. Conversely, an accretion of sediments may smother the pipeline inhibiting or preventing the discharge of wastewater from the outlet ports. This may also cause sediments to enter the outfall pipeline, reducing the efficiency of the outfall.

Head Loss Monitoring

The hydraulic grade line for an outfall can be determined as a function of the discharge. At each discharge, upper and lower limits for the hydraulic grade line can be established and by maintaining the discharge between these limits, we can optimize the hydraulic performance of the outfall. Below the lower limit, seawater intrusion into the outfall is likely to occur and operators can increase the wastewater flow accordingly (although

increasing the flow may not always be possible). The addition of salinity sensors on individual risers can confirm whether seawater intrusion has occurred.

Seawater Intrusion and Purging

A rule-of-thumb for preventing seawater from entering the outfall is to keep the port densimetric Froude number (32.9) well above unity (e.g., on Sydney's deepwater ocean outfalls, the port densimetric Froude number is of the order of 20–30). Occasionally, the port densimetric Froude number may become small and seawater may enter the outfall. One method to help prevent such an intrusion of seawater is to use nonreturn check valves. However, not all outfalls are so equipped and it may be necessary to periodically purge the seawater from the outfall. This is done by backing up the wastewater in the treatment plant and releasing the wastewater at high velocity. This also has the advantage of clearing the outfall of sediments.

Outfall Maintenance

Regular inspections of the marine outfall should be carried out as part of a regular maintenance program to ensure there is (a) no physical damage to any components and (b) no blockage of any outlet nozzles. This can be carried out by divers although this is accompanied by substantial health and safety concerns (e.g., diving in contaminated waters, wave action, and decompression requirements for deep diving). Our preference is to carry out inspections using remotely operated vehicles. The frequency of maintenance inspections will depend on a range of factors such as environmental conditions and construction materials. Often, annual inspections are adopted.

32.6 Environmental Monitoring

The fundamental objective of environmental monitoring is to quantify impacts that may arise as a result of the discharge of wastewater to the marine environment. Some of the questions that an environmental monitoring program should address include:

- Is it safe to swim?
- Is it safe to eat the seafood?
- Will the marine communities be protected into the future?
- Will the beaches be free from contamination?

Two distinct environmental monitoring programs are often implemented. The first is a pre- and post-construction monitoring program. This is usually an intense program of short duration (perhaps five years) aimed to quantify the initial impacts of the discharges from the marine outfall. The second is a long-term (ongoing) monitoring program, which usually takes a subset of the information obtained from the pre- and post-construction monitoring program, and uses it to define acceptable limits of change in the environmental indicators. Long-term monitoring is designed to determine whether a subsequent change lies outside these limits.

There is an assumption, often tacit, that the engineering aspects of a sewage treatment system and marine outfall are operating as designed. If there is a treatment bypass, breakdown or blockage of some of the outlet ports, the quality of the effluent (or its delivery to the marine environment) will be less than expected. Monitoring the operational performance is an important element of environmental monitoring.

32.6.1 Change Versus Impact

The marine environment is in a state of continual flux. The challenge of a monitoring program is to separate change that occurs via natural processes from change that is a direct result of the discharge from the marine outfall (i. e., impact). It is critical to define what constitutes an impact prior to the execution of the post-commissioning monitoring program. This places clear bounds on the interpretation of change that may be observed in the marine communities.

32.6.2 Pre- and Post-construction Monitoring

Broadly, the pre- and post-construction monitoring program compares conditions before and after the marine outfall is commissioned. It is critical that sufficient time is allowed to carry out the *before* monitoring. An *after* monitoring program can be conducted at any time after commissioning of the marine outfall. However, we only get one chance at the *before* phase of the monitoring program.

A monitoring program may be of several years' duration. Consequently, the cost of an environmental monitoring program may be large – perhaps 5% of the construction costs may be needed to properly address the main environmental issues.

Monitoring Philosophy

Some of the characteristics that constitute a good monitoring program are outlined below. These characteris-

tics combine to define our philosophy for environmental monitoring.

Environmental monitoring programs should be designed to establish cause-and-effect between the discharge and the environmental response. The approach we favor uses *weight-of-evidence*. This simply means applying different techniques and approaches to the same problem. If the answers that they provide are consistent with what might be expected from such discharges, then we have greater confidence in the overall result. Our weight-of-evidence is a three-pronged attack, which is detailed in the following section.

Most countries have guidelines (or licence conditions) that reflect the values that need to be protected. These values include social, public health and environmental aspects. While some values may be subjective, many can be quantified in terms of *safe* levels of contaminants that can be discharged. There is a tacit assumption, that by adhering to these concentration limits, there will be no irreversible damage done to the marine environment.

The most critical task in undertaking a monitoring program to quantify environmental impacts of a marine outfall is to ask the right questions. This requires an understanding of what is being discharged, what effects these discharges are likely to have on the marine environment and the level of change that we are willing to accept. An answer to the wrong question, no matter how accurate is that answer, will not allow impacts to be quantified.

Environmental monitoring programs must be scientifically robust and defensible. We must test a hypothesis and gather empirical evidence to support (or disprove) the hypothesis. The experiment must be repeatable, cover a range of conditions and, in theory, repeated experiments should arrive at the same conclusion. Marine outfalls may be contentious and the monitoring program may need to be defended. One way in which this can be done is via peer review of the work and publication of the results in a reputable journal.

Conditions of approval for a marine outfall are often reliant on predictions. The objectives of an environmental monitoring program should include the verification (or otherwise) of those predictions and whether the environmental values have been maintained. Further, the program should include a mechanism by which problems identified in the monitoring program can be rectified or mitigated.

The Three-Pronged Attack

An approach to collecting weight-of-evidence for impact assessment includes three components. These are detailed below.

Sources of Contaminants. There are many potential sources of contaminants discharged to marine waters, including rivers, sediments (as both a source and sink), private and industrial outfalls, illegal dumping, and coastal wastewater treatment plants. An important consideration to environmental monitoring is separating the relative contribution of an outfall from contributions from the other sources. It may be possible to characterize (and isolate) the different sources via the types, concentrations, and variability of contaminants being discharged from each source.

Marine organisms may be impacted if exposed to wastewater. Toxicity testing is a technique used to determine the concentration of substance which is likely to harm to marine organisms. Whole effluent toxicity (WET) testing examines whether a toxic response is identified when an organism is subjected to the complex wastewater matrix.

If toxicity testing is carried out using a single test organism, it is then inferred that all marine organisms will exhibit the same toxic response. This is not necessarily the case and multiple organisms are recommended for toxicity testing: different organisms display different responses to different substances at different stages in their life cycle. Ideally, a range of fish, invertebrates, and algae at different stages in their life cycle should be considered for toxicity testing. Toxicity tests can be either acute or chronic and both types of tests should be included. The former determine the concentration of wastewater that is lethal to the test organism. Chronic testing examines the reduced capability (e.g., impaired development or reproductive ability) of the test organism.

Movement of Wastewater. Once we know what is being discharged, we determine the path that the wastewater will take after discharge and how dilute it becomes. *Emery* and *Thompson* [32.36] describe many techniques to measure and/or monitor the physical properties of the marine environment.

Instrumented moorings provide excellent temporal coverage although the horizontal spatial coverage is limited by the number of such systems that the monitoring program can afford to deploy. Remote sensing provides good spatial coverage of the region, although in general, the temporal scales are much longer than those associated with the movement of wastewater plumes.

Numerical modeling enables us to predict how the wastewater will react to different marine conditions and to changes in the volume or method of discharge. It is critical to calibrate and validate any numerical model to ensure confidence in the results under existing conditions and for future scenarios. Coupled with this is

a need for confidence limits to be placed on model results. Too often model results are provided without a real appreciation for the confidence that can be placed on those results.

Quantifying Change. There are a number of approaches to quantifying change in response to a marine outfall. One favored by the authors is the *beyond BACI* (or multiple) approach [32.37]. BACI is an acronym for *b*efore, *a*fter, *c*ontrol, *i*mpact. Sites close to, and remote from, the outfall are monitored on several occasions both before and after commissioning of the marine outfall. Perhaps the main reason that this is a favored approach is its emphasis on statistical power and statistical error.

Statistical Errors. A Type I statistical error occurs when the results of our analysis incorrectly predicts that a change has occurred. We can protect against making such a mistake by specifying the level of significance. Usually this is set at 5% (equivalent to 95% confidence limits), i. e., there is a 5% chance that our analysis makes a Type I statistical error.

A Type II statistical error occurs when the results of the analysis incorrectly predicts that a change has not occurred. From an environmental point of view, this is more insidious than a Type I error because it incorrectly leads us to believe that there is no environmental problem. It is difficult to protect against a Type II error as this requires a priori knowledge of the variability in the system being measured. Such knowledge is acquired only after the monitoring is complete. Therefore, we need to use experience to estimate the system variability and design experiments accordingly. To protect against making a Type II statistical error, we usually design our experiments to target statistical power above 80%. However, if statistical power is too high, very small changes become statistically significant and we question whether such small changes are meaningful. It is critical to check the statistical power after the experiments have been completed.

Type III statistical errors occur when we arrive at the correct answer but have asked the wrong question.

Some of the marine components measured to detect change as a result of a new marine outfall include water quality, sediment quality and community studies (such as intertidal, plankton, pelagic, benthic, and sessile communities). Marine community studies should be tailored for specific outfalls and might include:

- Intertidal community studies for shoreline or short outfalls.
- Subtidal settlement panel studies for marine outfall in waters less than about 20 m depth (much deeper than this and the light attenuation starts to inhibit the growth of organisms on the settlement panels).
- Fish, shellfish, and planktonic communities move in, and with, marine waters. The variability among control sites may be as much as the variability between putatively impacted sites and control sites. Therefore, it may be difficult to isolate the marine outfall as the cause of change in such communities.
- Sediments have been used to assess the accumulation of contaminants and to examine infauna community variations. Our experience is that such studies have limited success. Unless the contaminant signal is very strong, it is unlikely to register in a sediment sample.
- Bioaccumulation studies are sometimes recommended as impact assessment indicators. However, caution is needed when interpreting the results of such studies because:
  - Fish move, and it is not always possible to know the region from where contaminants were accumulated. This necessitates fish home range studies, which can be expensive and, perhaps, inconclusive.
  - Species that are caught at one particular time and location may not be caught at other times or locations and pooling into higher biological levels may be required. Different species may accumulate different substances at different rates and the pooling process may mask potential impacts.
  - Moored systems comprising oysters or mussels are often used in bioaccumulation studies. However, this may involve removing the animals from their natural habitat thereby, adding stress to the organisms and confounding the results that are obtained.

32.6.3 Long-Term Monitoring

The pre- and post-construction monitoring program will identify whether there has been a step change in the baseline conditions as a result of the discharge from the marine outfall. Long-term monitoring is used to identify whether further change occurs well after the marine outfall has been commissioned. It can be used to extrapolate trends and, where necessary, design and implement an appropriate mitigation strategy to prevent, or reverse, the trend. The three-pronged attack described in Sect. 32.6.2 is also applicable to long-term monitoring.

However, there is a temptation to simply implement the bulk of the pre- and post-construction monitoring program as part of the long-term monitoring program. Apart from the expense associated in maintaining a de-

tailed monitoring program over a long period of time, environmental changes will be masked by the small time steps between consecutive sets of readings. It may be more effective to implement a low level, long-term monitoring program and revisit the detailed monitoring program, for example, over two consecutive years every 10 yr. If the long-term program indicates a potential problem, it provides the motivation for a more detailed investigation.

Where such an approach is adopted, source characterization combined with numerical modeling are used to estimate environmental impact. If the results indicate a possible impact, confirmation studies can be implemented.

32.6.4 Summary

This chapter provides a synopsis of the discharge of wastewater to the marine environment. Flows of effluent from municipal sewage treatment plants and brine from desalination plants are primarily considered within the context of near-field modeling and outfall hydraulics. The designers of outfalls are under increasing pressure from social, public health and environmental constraints, within a regulatory and economic framework, all of which need to be considered. This compact overview exposes the reader to fundamentals of outfall design and identifies some of the traps and problems that may arise.

References

32.1 P.M. Tate, J.H. Middleton: Buoyant jets of elliptic shape: Approximation for duckbill valves, J. Hydraul. Eng. **130**(5), 432–440 (2004)

32.2 Metcalf & Eddy Inc., G. Tchobanoglous, F.L. Burton, H.D. Stensel: *Wastewater Engineering: Treatment and Reuse*, 4th edn. (McGraw-Hill, New York 2003)

32.3 P. Palomar, I.J. Losada: Impacts of brine discharge on the marine environment. Modelling as a predictive tool. In: *Desalination, Trends and Technologies*, ed. by M. Schorr (InTech, Rijeka 2010) pp. 279–310

32.4 I.R. Wood, R.G. Bell, D.L. Wilkinson: *Ocean Disposal of Waste* (World Scientific, Singapore 1993)

32.5 J.H. Lee, V. Cheung: Generalized Lagrangian model for buoyant jets in current, J. Environ. Eng. **116**(6), 1085–1106 (1990)

32.6 P.M. Tate, J.H. Middleton: Unification of non-dimensional solutions to asymptotic equations for plumes of different shape, Boundary-Layer Meteorol. **94**(2), 225–251 (2000)

32.7 G.H. Jirka, P.J. Akar: Hydrodynamic classification of submerged multiport-diffuser discharges, J. Hydraul. Eng. **117**(9), 1113–1128 (1991)

32.8 G.H. Jirka, R.L. Doneker: Hydrodynamic classification of submerged single-port discharges, J. Hydraul. Eng. **117**(9), 1095–1112 (1991)

32.9 B. Morton, G. Taylor, J. Turner: Turbulent gravitational convection from maintained and instantaneous sources, Proc. R. Soc. A **234**(1196), 1–23 (1956)

32.10 C.C. Lai, J.H. Lee: Mixing of inclined dense jets in stationary ambient, J. Hydroenviron. Res. **6**(1), 9–28 (2012)

32.11 P.J. Roberts, A. Ferrier, G. Daviero: Mixing in inclined dense jets, J. Hydraul. Eng. **123**(7), 693–699 (1997)

32.12 A. Cipollina, A. Brucato, F. Grisafi, S. Nicosia: Bench-scale investigation of inclined dense jets, J. Hydraul. Eng. **131**(11), 1017–1022 (2005)

32.13 S. Nemlioglu, P. Roberts: Experiments on dense jets using three-dimensional laser-induced fluorescence (3DLIF), Proc. 4th Int. Conf. MWWD, Antalya (2006)

32.14 G. Kikkert, M. Davidson, R. Nokes: Inclined negatively buoyant discharges, J. Hydraul. Eng. **133**(4), 545–554 (2007)

32.15 D. Shao, A.W.-K. Law: Mixing and boundary interactions of 30 and 45 inclined dense jets, Environ. Fluid Mech. **10**(5), 521–553 (2010)

32.16 I.G. Papakonstantis, G.C. Christodoulou, P.N. Papanicolaou: Inclined negatively buoyant jets 1: Geometrical characteristics, J. Hydraul. Res. **49**(1), 3–12 (2011)

32.17 I.G. Papakonstantis, G.C. Christodoulou, P.N. Papanicolaou: Inclined negatively buoyant jets 2: Concentration measurements, J. Hydraul. Res. **49**(1), 13–22 (2011)

32.18 L.-N. Fan: *Turbulent Buoyant Jets into Stratified or Flowing Ambient Fluids*, Technical Report (CALTECH, Pasadena 1967)

32.19 V.H. Chu: L.N. Fan's data on buoyant jets in cross-flow, J. Hydraul. Div. **105**(4), 612–617 (1979)

32.20 P.J. Roberts, H.J. Salas, F.M. Reiff, M. Libhaber: *Marine Wastewater Outfalls and Treatment Systems* (IWA Publishing, London 2010)

32.21 B. Henderson-Sellers: *Modeling of Plume Rise and Dispersion-The University of Salford Model: USPR*, Lecture Notes in Engineering, Vol. 25 (Springer, Berlin Heidelberg 1987)

32.22 H.B. Fischer, E.J. List, R.C.Y. Koh, J. Imberger, N.H. Brooks: *Mixing in Inland and Coastal Waters* (Academic, New York 1979)

32.23 D.L. Wilkinson: Optimal design of ocean outfalls. In: *Environmental Hydraulics*, ed. by C.C. Lai, J.H. Lee (Balkema, Rotterdam 1991) pp. 275–279

32.24 S.I. Seneviratne, N. Nicholls, D. Easterling, C.M. Goodess, S. Kanae, J. Kossin, Y. Luo, J. Marengo, K. McInnes, M. Rahimi, M. Reichstein, A. Sorteberg, C. Vera, X. Zhang: Changes in climate extremes and their impacts on the natural physical environment. In: *Managing the Risks of Extreme*

Events and Disasters to Advance Climate Change Adaptation, ed. by C.B. Field, V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastrandrea, K.J. Mach, G.-K. Plattner, S.K. Allen, M. Tignor, P.M. Midgley (Cambridge Univ. Press, Cambridge 2012) pp. 109–230

32.25 M.H. Sharqawy, J.H. Lienhard, S.M. Zubair: Thermophysical properties of seawater: A review of existing correlations and data, Desalin. Water Treat. **16**(1–3), 354–380 (2010)

32.26 J. Pitt, P. Ackers: *Hydraulic Roughness of Segmentally Lined Tunnels* (CIRIA, London 1982)

32.27 D.S. Miller: *Internal Flow Systems* (Miller Innovations, Bedford 2009)

32.28 I. Idelchik: *Handbook of Hydraulic Resistance* (Jaico, Mumbai 2003)

32.29 D.L. Wilkinson: Avoidance of seawater intrusion into ports of ocean outfalls, J. Hydraul. Eng. **114**(2), 218–228 (1988)

32.30 W.H. Hager: *Wastewater Hydraulics: Theory and Practice* (Springer, Berlin Heidelberg 2010)

32.31 H.T. Falvey: *Air-Water Flow in Hydraulic Structures*, Technical Report (Water and Power Resources Service, Denver 1980)

32.32 C.S. Lauchlan, R.W.P. May, R. Burrows, C. Gahan: *Air in Pipelines: A Literature Review* (HR Ltd., Wallingford 2005)

32.33 D. Butler, R.W.P. May, J.C. Ackers: Sediment transport in sewers part 1: Background, Proc. ICE-Water Marit. Energy, Vol. 118 (1996) pp. 103–112

32.34 R.A. Grace: *Marine Outfall Systems. Planning, Design and Construction* (Prentice-Hall, Upper Saddle River 1978)

32.35 R.A. Grace: *Marine Outfall Construction: Background, Techniques, and Case Studies* (ASCE, Reston 2009)

32.36 W.J. Emery, R.E. Thomson: *Data Analysis Methods in Physical Oceanography* (Pergamon, Oxford 1998)

32.37 A. Underwood: Beyond BACI: Experimental designs for detecting human environmental impacts on temporal variations in natural populations, Mar. Freshw. Res. **42**(4), 569–587 (1991)

Offshore Part D

Part D Offshore Technologies

Ed. by R. Cengiz Ertekin, Dominique Roddier

33 Offshore Platforms
Arisi S.J. Swamidas, St. John's, Canada
Dronnadula V. Reddy, Boca Raton, USA

34 Stability of Offshore Systems
Alexia Aubault, Berkeley, USA
R. Cengiz Ertekin, Honolulu, USA

35 Wave, Current and Wind Loads
R. Cengiz Ertekin, Honolulu, USA
George Rodenbusch, Dripping Springs, USA

36 Vortex-Induced Vibrations
Michael S. Triantafyllou, Cambridge, USA
Rémi Bourguet, Toulouse, France
Jason Dahl, Narragansett, USA
Yahya Modarres-Sadeghi, Amherst, USA

37 Structural Dynamics
H. Ronald Riggs, Honolulu, USA
Solomon Yim, Corvallis, USA

38 Cable Dynamics for Marine Applications
Ioannis K. Chatjigeorgiou, Athens, Greece
Spyros A. Mavrakos, Athens, Greece

39 Offshore Geotechnics
Dong-Sheng Jeng, Griffith, Australia

40 Buoy Technology
Andrew Hamilton, Moss Landing, USA

41 Liquefied Natural Gas Carriers
Krish P. Thiagarajan, Orono, USA
Robert Seah, Houston, USA

42 Salvage Operations
Michael S. Dean, Washington D.C., USA

43 Oil Spills and Response
Merv Fingas, Edmonton, Canada

Chapter 33. This introductory chapter provides an overview of the state-of-the-art of offshore platforms. The chapter highlights the various structural concepts (such as fixed, gravity, compliant, floating, and subsea platforms) developed during the past 75 years of active offshore exploration and exploitation. The chapter focuses on the various factors that influence the choice of structure during project development and design, and gives an overview of the strength and weaknesses of various offshore development scenarios. The chapter concludes by discussing future trends and developments.

Chapter 34. The *Stability of Offshore Systems* chapter provides the theoretical background to analyze the buoyancy and stability of offshore platforms and discusses the regulatory framework, including IMO regulations, which sets standards of stability. Such standards apply criteria on the righting moment of a floating structure with appropriate safety factors. They aim at preventing incidents related to excessive heel or trim, capsizing, catastrophic loss of buoyancy, or even sinking. Intact stability and damaged stability of floating structures are considered in this chapter. Methods of calculating the righting moment are examined. The spirit and practical application of regulations are discussed in detail.

Chapter 35. The *Wave, Current, and Wind Loads* chapter describes various loads on fixed or floating offshore platforms. Both linear and nonlinear waves are discussed in deterministic and irregular seas. Wave loads on large structures are formulated using linear potential theory in the frequency domain. Morison's equation and the associated drag and inertia coefficients are introduced for slender structures. Also discussed are wave-current interaction and current-induced forces, including vortex-induced vibrations. Both steady and unsteady wind profiles are used to describe wind forces. Model tests and similarity laws are also discussed.

Chapter 36. The *Vortex-Induced Vibrations (VIV) of Slender Structures* chapter reviews the vibrations of flexible structures placed in a cross-flow, caused by vortices forming due to flow instability in the wake. The phenomenon is important for ocean and offshore applications because of possible fatigue damage and increased drag forces, and spans Reynolds numbers from as low as 50, reaching the highest Reynolds numbers recorded. The complex flow-structure interaction mechanisms result in a very rich dynamic behavior, which has made this a canonical problem of flow-structure interaction studies. The focus is on the response of flexible risers and cables, and the review of the effectiveness of vortex cancelation devices.

Chapter 37. The *Structural Dynamics* chapter presents the theory that is especially relevant to dynamics of flexible ocean structures. The primary focus of this chapter is on linear response to regular waves and random seas, but the extension to nonlinear systems is also discussed. Equations of motion in the time and frequency domains for a single-degree-of-freedom system with both viscous and hysteretic damping are presented. Generalization of these concepts to multi-degree-of-freedom systems is then covered. Hydroelasticity, which couples the structural and fluid models, is presented. Numerical solutions including spatial and time discretization and integration are presented. Nonlinear hydroelasticity, viz. fluid-structure interaction, is discussed briefly.

Chapter 38. The *Cable Dynamics* chapter provides information associated with the dynamic behavior of slender cables (e.g., mooring lines) with particular emphasis on marine applications. The scope of the chapter is to associate the aspect of dynamics, linear and nonlinear, with the most typical practical applications of cables in the marine and, especially, in the offshore sector. The chapter first derives a global formulation for the governing dynamical systems, and then numerical methodologies required for their solutions are presented.

Chapter 39. The *Wave-Seabed-Structure Interactions* chapter provides an overview of theoretical models for wave-induced seabed response in marine sediments. This includes the introduction of the existing three well-known models, a new analytical approximation for the wave-induced pore pressure accumulations, a new concept for linking oscillatory and residual mechanisms, and the inclusion of currents in the existing models for wave-seabed interactions. A case of caisson-type breakwater is considered as an application of the proposed numerical models. With the new model, the liquefaction zones around breakwater with different approximations are compared, and the differences identified. Finally, three topics are suggested for future studies in this area.

Chapter 40. The *Buoy Technology* chapter presents an introduction to small buoy systems used for oceanographic or other ocean-monitoring purposes. This chapter outlines typical buoy and mooring types and their applications, buoy and mooring material choices, specialty strength members, and expected dynamic behaviors and specific analysis techniques used to analyze a coupled buoy/mooring system. Tables of material and component properties useful to the designer, as well as a summary of hydrodynamic coefficients for small cylindrical buoys are included. The chapter concludes

with detailed examples of three mooring systems of different types that have been successfully fielded.

Chapter 41. The *Liquid Natural Gas Carrier* chapter discusses the challenges related to the transport of LNG. This chapter first addresses the type of vessels used, both historically and currently. It focuses next on the challenges of transporting such a cold and dangerous cargo and describes the recent advances in solving both environmental and engineering problems, inclusive of sloshing and cargo transfer. The chapter concludes by discussing design methodologies for the various containment systems.

Chapter 42. The *Salvage Operation* chapter presents an introduction to salvage engineering. This chapter outlines the various principles of naval architecture to be able to assess the strength and stability of a damaged vessel. It focuses on the various facets of salvage engineering, focusing on key steps, from planning to survey to reporting, and disseminate the critical information necessary for successful operations. The chapter concludes on types of specific salvages and highlights the differences between them.

Chapter 43. The *Oil Spills and Response* chapter is treats the random phenomenon of oil spills that occur in many sizes and forms. This chapter summarizes the response and cleanup of oil spills. The behavior of oil when spilled dictates the environmental impact it will have. Once spilled, oil on the sea is typically contained using booms and is recovered using skimmers. Dealing with recovered oil and disposing of it are important steps. Spill treating agents are occasionally used. In-situ burning of oil is now being used more frequently. Finally, spilled oil often contacts shorelines, therefore assessment and careful removal techniques for shorelines are essential.

33. Offshore Platforms

Arisi S.J. Swamidas, Dronnadula V. Reddy

An overview of the state-of-the-art on offshore platforms is provided, highlighting the various structural concepts (such as fixed, gravity, and compliant, floating, and subsea platforms) developed during the past 75 years of active offshore exploration and exploitation. It starts with a birds-eye-view of the various platforms that dot the ocean space. Some of the advantages and disadvantages that influence the choice of the structural concepts to be used are discussed. The holistic and integral structural behavior depends on the environmental influences exerted on the important components of the platform in terms of permissible stresses, allowable rotations/deformations, fatigue endurance, vortex-induced-vibration, unobstructed flow of hydrocarbons and overload on the total structural performance; the overview also mentions some effects of the above on marine risers, conductors, wells and pipelines. Finally, possible offshore structural developments that are likely to take place in future are discussed.

33.1 **Relevance** 745
33.2 **Types of Offshore Platforms** 747
33.2.1 Fixed and Compliant Offshore Platforms 748
33.2.2 Floating Offshore Platforms 750
33.2.3 Subsea Systems and Pipelines 750
33.3 **Future Trends and Developments in Offshore Platforms** 751
References 752

Offshore drilling and storage platforms play an important role in oil and gas developments around the world. The ever-increasing demand for augmenting our energy needs by offshore exploration and production activities across the globe, and especially by the drilling activities occurring at ultra-deep-water depths, is driving the market. Depending on the water depths at which the oil and gas reserves are available, different types of fixed, floating, and subsea platforms are used for the purpose. In addition, the transportation of oil and gas through subsea pipelines has facilitated the transfer of offshore resources to onshore facilities. For ultra-deep-water depths, semi-submersibles have the highest growth rate and are the most preferred rigs deployed for drilling, in conjunction with other FPSO (floating production storage and offloading) units. Other floating platforms, such as drill-ships, also are being utilized at a decent growth rate for ultra-deep waters, in conjunction with subsea structures. Latin America dominates the global offshore floating drilling platform market since it has high offshore potential reserves of oil and gas in deep and ultra-deep waters. The Asia-Pacific region has the largest increase in shallow-water exploration and production in the waters of Indonesia, Malaysia, and Australia, with the highest growth rate for shallow-water framed and gravity platforms. In spite of the *fracking* of shale (onshore and offshore) decreasing the need for new energy sources in USA and worldwide, dependence on crude oil and natural gas will continue to increase as the global energy demand continues to increase worldwide. In order to satisfy this growing demand for energy, drilling for exploration and exploitation will continue to move offshore from deeper to ultra-deep and colder Arctic waters. In addition, improvements in platform technology in the form of stronger, modular and light-weight construction will impact the structural innovations of existing offshore platforms for many years to come.

33.1 Relevance

A number of platforms, installed in the ocean, exploit the available natural resources, as shown in Fig. 33.1 [33.1]. In addition to the offshore platforms mentioned above, ship types and modes seen in the field of ocean transport include containerization, catamarans, hovercraft, floating terminals, and very large

Fig. 33.1 Opportunities for developing marine structures (after [33.1])

crude oil carriers, along with large numbers of boats and barges that ply the inland rivers. The development of transportation has increased the capability to exploit more of the oceans' alternate energy resources, such as ocean thermal energy conversion (OTEC) and wave and current energy conversion, through direct drive wave energy converters and ocean energy turbines, which have a strong potential for producing energy. Ocean sediments are recognized as a major source of mineral wealth, both hard minerals and petroleum. Offshore oil and gas now supply almost one third of the world's energy needs. While manganese nodules have been recovered from the deep ocean floors of the tropical and subtropical areas, coastal sediments have also been exploited using dredging structures, since they are rich in *industrial minerals* of gold, tin, chromium, platinum, and even sand and gravel [33.2].

Offshore hydrocarbons, a much sought after energy resource, are associated primarily with the continental margins of the world. These margins surround the continents and constitute 25% of the total ocean area. Extending from the shore to the deep ocean floor, the continental margin consists of three units – shelf, slope, and rise. Sediments generally accumulate on the coastal plains and continental shelves. These sediments periodically flow down the continental slopes as turbidity currents to form giant channels of submarine fans and debris flows at their base [33.3].

Due to their thick sedimentary nature, continental margins contain an estimated 99% of the oceans recoverable hydrocarbons, and the deep-ocean sediments only 1% thereof. Furthermore, it has been provisionally estimated that 65% of these hydrocarbons will be found at water depths less than 200 m, 30% from 200 to 2500 m, and only 5% at greater water depths [33.4]; these data are nearly 30 years old. This scenario is likely to change as more of the currently acquired offshore data are consolidated and made available [33.5]. According to the data given in Indexmundi's 2015 crude oil forecast [33.6], the total world consumption of oil per day (for 2013) is given as 90.35 m b/day, and for a year as 32.98 billion barrels.

The world's largest proven oil reserves(in million barrels) are given as (within brackets) [33.7]:

i) Venezuela (297 600)
ii) Saudi Arabia (267 910)
iii) Canada (173 105)
iv) Iran (154 560)
v) Iraq (141 035)
vi) Kuwait (104 000)
vii) UAE (97 800)
viii) Russia (80 000)
ix) Libya (48 010)
x) Nigeria (37 200).

In addition, the top ten oil-consuming countries of the world are listed as [33.8]:

i) USA (19.15 m b/day)
ii) China (9.40 m b/day)
iii) Japan (4.45 m b/day)
iv) India (3.18 m b/day)
v) Saudi Arabia (2.64 m b/day)
vi) Germany (2.50 m b/day)
vii) Canada (2.22 m b/day)
viii) Russia (2.21 m b/day)
ix) South Korea (2.20 m b/day)
x) Mexico (2.07 m b/day).

As a result, the development and use of the older and newly innovated offshore platforms will depend on the investments made in developing the offshore fields by most of the countries mentioned above.

Despite the challenges of offshore ocean exploration, exploitation, and maintenance, nearly 35% of global oil production and 27% of gas production comes from offshore areas. Although it is expected that continental shelves will satisfy a significant part of the future oil and gas requirements, and also that dependence on fossil hydrocarbon resources will last for at least another 50 years, the oil industry is already looking to the ultra-deep-water zones on the continental slopes for future oil supplies. The industry has continued to invest in infrastructure and new techniques for exploitation in deep waters, leading to enhanced underwater facilities. It appears that the huge fixed steel platforms used today will in future be increasingly replaced by floating production platforms and smaller-scale subsea production technology deployed directly on the seabed, similar to those already in existence on the North and Norwegian Sea basin margins,

Brazil's Campos Basin, and the Gulf of Mexico. Exploiting and transporting oil and gas reserves in the harsh and inhospitable Arctic regions, containing nearly 30% of world's undiscovered hydrocarbons, would require innovative structural development and management strategies [33.9].

33.2 Types of Offshore Platforms

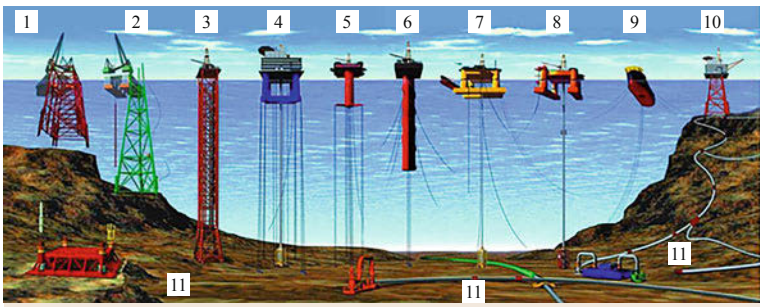
At the present juncture, more than 9500 offshore platforms are being used worldwide, operating in water depths ranging from 10 ft to nearly 10 000 ft. Topside weights range from 5 to 50 000 t, producing oil, gas, or both. Some of the major offshore oil and gas platforms, used in the context of ocean oil and gas exploitation, are shown in Fig. 33.2 [33.10, 11]:

i) Conventional fixed platforms (including gravity platforms) (Shell's *Bullwinkle* at 412 m deep water, in the Gulf of Mexico)
ii) Compliant tower (ChevronTexaco's *Petronius* at 534 m deep water in the Gulf of Mexico)
iii) Vertically moored tension leg and mini-tension leg platform (Conoco-Phillips' *Magnolia* in 1425 m deep water, in the Gulf of Mexico)
iv) Spar (Dominion's *Devils Tower* in 1710 m deep water, in the Gulf of Mexico)
v) Semi-submersibles (Shell's *NaKika* field in 1920 m deep water, in the Gulf of Mexico)
vi) Floating production, storage, and offloading facility (in 1600 m deep water, in the Campos Basin offshore Brazil)
vii) Subsea cluster completions and tie-back to host subsea facility (Shell's *Coulomb* field at 2307 m deep water tied to the *NaKika* field in the Gulf of Mexico)
viii) Subsea system with flow lines to a host fixed framed platform (Shell's *Mensa* field in 1620 m deep water, tied to a shallow-water jacket platform, in the Gulf of Mexico).

The platform concepts range from fixed and compliant, floating and subsea systems.

Referring to Fig. 33.2, platforms listed as categories 1–6 (along with 10) can be classified as fixed and compliant systems. Thus fixed and compliant systems include: steel jackets/towers (with or without tie-backs), concrete gravity platforms, jack-up, compliant towers, tension leg platforms (TLP), and spar concepts. The floating systems include: semi-submersibles, floating production storage, and off-loading units (FPSO), moored ship platforms and barges (listed under categories 7–9 in Fig. 33.2). The subsea systems include: (i) drilling and production units located on the seabed, listed under category 11 in Fig. 33.2 (economical for marginal fields and others in ultra-deep waters – tied to existing installations for access); and (ii) pipeline transportation systems for oil and gas.

Generally speaking an offshore platform has the primary function of allowing the extraction of hydrocarbons from below the sea-bottom, with the minimum treatment and maximum safety. The produced hydrocarbons are then transported safely to hydrocarbon refining plants on the nearby coasts, for treatment prior to commercialization. Due to the high costs of offshore construction, requisite offshore facilities are thus kept



1 & 2: Conventional fixed production platforms
3: Compliant tower production platform
4 & 5: Moored tension leg production platform
6: Spar production platform
7 & 8: Semi-submersible drilling rigs
9: Floating production storage and offloading facility
10: Subsea completion and tie-back to host facility
11: Subsea systems and flow lines

Fig. 33.2 Offshore structural concepts used offshore (after [33.10], courtesy of NOAA)

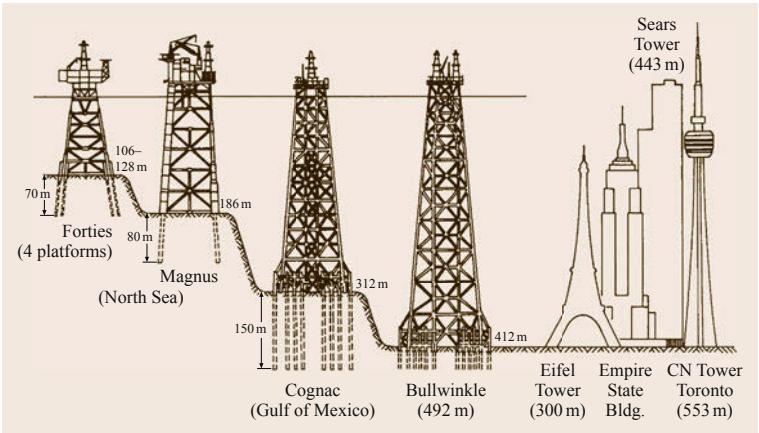


Fig. 33.3 Fixed platforms compared with known landmarks (after [33.12])

to a bare minimum by maximizing onshore prefabrication and modular assembly.

33.2.1 Fixed and Compliant Offshore Platforms

As stated earlier, this group of platforms is governed by fixity to the ocean bed. Depending on the fixity (or bottom supported) conditions the platforms can be classified as:

i) Fixed (rigidly or elastically)
ii) Bottom supported (with or without skirts and piles, providing transverse resistance)
iii) Compliantly restrained.

Use of these platforms is governed mainly by the maximum seawater depths at which they are to be located. Figure 33.3 [33.12] gives a typical comparison of various fixed platforms used in the offshore contexts, comparing them with some noted onshore landmarks. When water depths become deeper than 450 to 500 m, the fixed jacket and bottom supported gravity platform costs become exorbitant; hence the compliantly restrained platforms (at sea bottom) are needed to increase the depths (up to 1300 m), at which these platforms can be located economically. Even though the tension leg platforms and spars are moored to the ocean bottom through elastically restraining tension legs, mooring cables, and marine risers (similar to some floaters like semi-submersibles and FPSOs), they are still listed under fixed or compliant platforms, since their transverse motion is limited by the depths under which the marine riser systems can function properly due to:

i) Large bending stresses exerted on the conductors (from loads transmitted from the riser system)
ii) The need for uninterrupted hydrocarbon flow through the risers due to limited rotation (2°) permitted at its bottom.

In some of the recent TLP and spar systems, these restrictions have been removed (by using new tension riser systems) and they are now listed under floaters where there is no limit on the transverse surge motions [33.13]. These floaters can operate in up to 3000 m water depths.

Fixed platforms have a number of advantages over compliant and floating systems. These platforms can support very large deck loads (with facilities for refining the extracted oil and gas), can be prefabricated (in

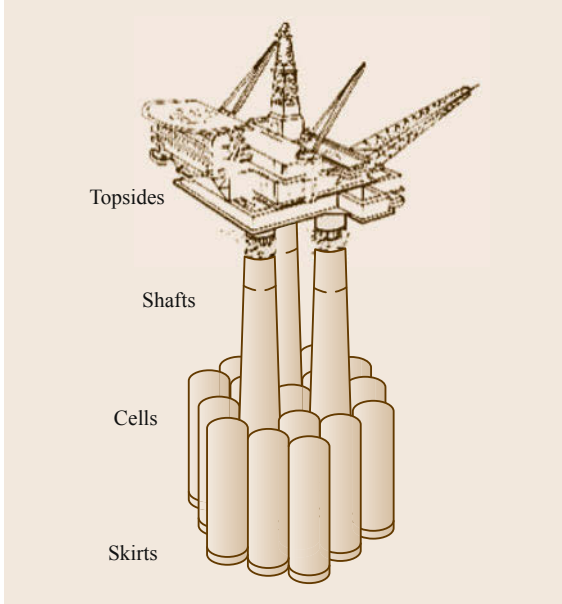


Fig. 33.4 Components of a gravity platform (after [33.14])

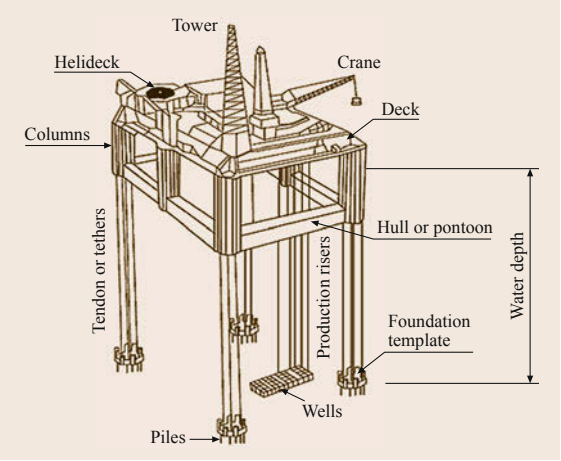


Fig. 33.5 Components of a tension leg platform (after [33.15])

on-land areas or dry docks) off their installation site in modular sections, assembled, and transported to the installation site, provide a stable support for a long-term usage, and are affected very little by sea-bottom scour around the piles. Some of their disadvantages are that the platform costs are very high initially, they have large maintenance costs due to fatigue and corrosion, and they are not reusable. The gravity platform (maximum permissible wear depth is 350 m) (Fig. 33.4 [33.14]), also enjoys similar advantages like the fixed jacket platforms with lower maintenance costs; they are also more resistant to fatigue and corrosion degradation. The main disadvantages are that they are more costly than framed steel jacket platforms and experience much larger foundation settlement/scour during their lifetimes. The jack-up platforms have retractable legs and are used for exploratory drilling and for servicing offshore wind farms. They are easily removable from one site to another. The depth of operation is limited to a maximum of 165 m.

Compliant offshore platforms (listed under this category) have a maximum water depth limitation of 1300 m. These platforms withstand and dissipate large transverse wave forces by moving laterally under the wave excitation. Due to their lower rigidity at the supporting base (with hinged end conditions at its bottom), the surge natural frequencies of these structures are much lower than the wave excitation frequencies. Hence, the phenomenon of resonance is avoided and the structure moves back and forth, executing a slow surge oscillatory motion. The tension leg platform (TLP), shown in Fig. 33.5 [33.15], is similar to a four-legged semi-submersible (with transverse submerged pontoons) held in place (with very small lateral motions) by vertically tensioned tendons, connected to the seafloor through templates and piles driven into the soil through them. TLPs have a water-depth limit of 1300 m; recently TLPs were also used at 2000 m.

The spar platform structure consists of a relatively long cylindrical tower of 200–250 m length (other types of spars, called trussed spars, are square in plan form), with watertight vertical cylindrical tanks (located inside the long cylindrical tower, with ballasting and de-ballasting capabilities). The spar is anchored to the seafloor by tensioned long cylindrical cables, connected to the seabed by suction anchors. Conventional spars have a limit imposed on their transverse motions by the permissible bending stress on the conductor, due to tensioned drilling risers and vortex-induced vibrations. In addition, the rotation at the base of the riser string is limited to 2° to permit proper flow of drilled

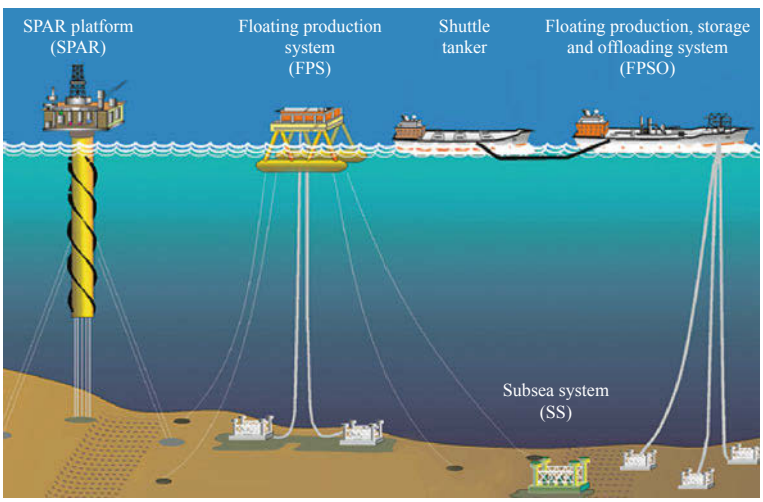


Fig. 33.6 Floating offshore structures (courtesy of The Bureau of Ocean Energy Management, after [33.16] BOEM)

hydrocarbons [33.13]. Initial spars were depth-limited to 1300 m.

33.2.2 Floating Offshore Platforms

As was mentioned before, floating offshore platforms, shown in Fig. 33.6 [33.16], consist of improved spars and TLPs, semi-submersibles, floating production storage and off-loading units (FPSO), moored ship-type platforms, and barges. In using these structures for water depths deeper than 1300 m, the previous limitation on the bending stresses in marine conductors is removed by using large diameter conductor pipes with increased wall thicknesses; the use of large diameter pipes increases their fatigue lives 5 to 10 times [33.13]. The blow-out-preventer (BOP) stacks provided at the top of the conductors have reduced lengths, along with the stick-up length (viz., the distance between the ocean bottom and the bottom of BOP stacks) of the conductors, to optimize the permitted stress levels. In addition, buoyancy provided at riser joints is increased by using syntactic foam elements with embedded micro glass or ceramic spheres. Along with an increase of the tension provided for riser strings, the risers are also provided with strakes to minimize or suppress vortex-induced vibration stresses; moreover the riser joints are provided with choke and kill lines made of titanium to minimize corrosion problems. These improvements have enabled spar platforms and other floating structures to operate efficiently at depths greater than 3000 m.

Floating production storage and off-loading (FPSO) units may be semi-submersibles or ship-shaped vessels, which float by their buoyancy. They are generally moored in place with flexible catenary-shaped wire ropes and/or polyester ropes, although they can also be kept in place by the use of dynamic positioning, if necessary. They are provided with subsea production systems, in which a group of wells is pre-drilled on a template installed on the seabed. Then the production unit is installed over the drilled wells and completed with subsea well heads. The FPSO unit is moored in a central position, with respect to the drilled subsea wells; these subsea wells are then connected by rigid (or flexible) pipelines to carry the reservoir fluids from the wellheads to surface storage vessels (with loading buoys) from which they can be offloaded to oil carriers. Other small-diameter flexible tubes, called umbilicals, connect the subsea production units to the surface units located on the floating vessel; these contain the electrical cables and hydraulic fluids required to control the wellheads from the surface. Drill-ships are vessels that are fitted with drilling equipment. Most often they are used for exploratory drilling for oil or gas wells in deep waters. Most drill-ships are fitted with dynamic positioning systems to maintain position over the well. They can drill in water depths up to 3660 m.

33.2.3 Subsea Systems and Pipelines

A subsea well consists of a wellhead template assembly and a wet Christmas tree, which is the underwater assembly of valves, spools, and fittings, used to control the flow of oil and gas out of the drilled well; since the system functions on the seabed, it is designed

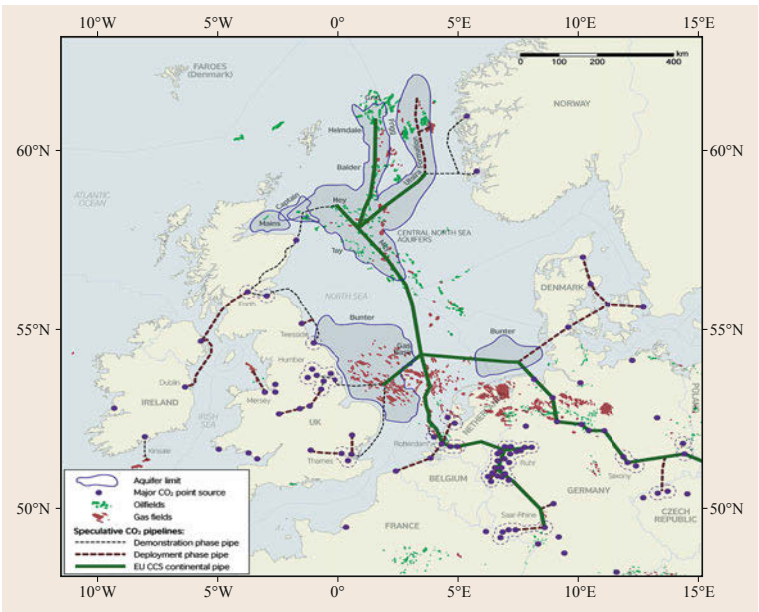


Fig. 33.7 Integrated pipelines in the North Sea Region (after [33.17])

such that it will always operate reliably. Generally, independent subsea wells are used for fixed or floating platforms for recovering reserves located beyond the reach of the horizontal drill-strings. Large multi-well subsea systems have also been installed with remote operating vehicle (ROV) intervention. In some of these subsea system operations they are tied to some existing deep-water or shallow-water platforms. A notable tieback (to a fixed platform located at a depth of 113 m) is the Mensa subsea development connected by flow lines of 101 km length, at a water depth of 1645 m in the Gulf of Mexico. Recently, much longer and deeper water-depth (≈ 1830 m) subsea systems have been developed with tiebacks to onshore systems in Brazil [33.18].

Submarine pipelines, which are laid on the seabed or buried inside a trench on the seabed, are used to carry produced oil and gas from the ocean depths to onshore systems. Figure 33.7 [33.17] shows one of the best and optimally used interconnections of pipelines around the North Sea offshore developments, transporting oil and gas to countries around its shores, viz., Ireland, Scotland, UK, Norway, Denmark, Sweden, Belgium, Holland, France, Germany, and the Czech Republic. The pipeline design and construction took into account the existing seabed ecology to preserve its biodiversity along its route to the shore [33.19, 20], international boundaries, geo-hazards, and the environmental loadings that will be exerted on the system. The pipeline diameters generally vary from 76 mm (for gas lines) to 1800 mm (for high capacity flow), and the wall thicknesses vary between 10–75 mm. The pipelines are protected against corrosion with epoxy or bituminous coatings and are often weighted down with concrete lining. The maximum internal pressure for design is taken as 10 MPa.

33.3 Future Trends and Developments in Offshore Platforms

Global energy consumption is forecast to expand 1.8 times by 2035 compared to that of 2010, and during the same period consumption of crude oil is expected to increase just over 30%, mainly in developing countries. Consumption of natural gas is forecast to rise greatly, by 50% or more in 2035, as compared with 2010. The depletion of existing fields will call for the discovery of new offshore drilling sites to maintain secure oil and gas supplies.

To give a context in terms of possible offshore development around USA, Fig. 33.8 [33.21] gives the undiscovered technically recoverable oil and gas resources available offshore. Hence, there is still a lot of scope for innovative structural developments to take place when the above-mentioned federal offshore acreages become available for development. Figure 33.8 gives the scope for possible offshore development for oil and gas resources around the federal offshore acreage in USA; the area shown consists of 89.9 billion barrels of oil and 404.6 trillion cubic feet of gas. Offshore industries are predicting a large investment (by a factor of three to four times current levels) on offshore structures and related facilities during the coming years [33.22]. It is believed that the Arctic re-

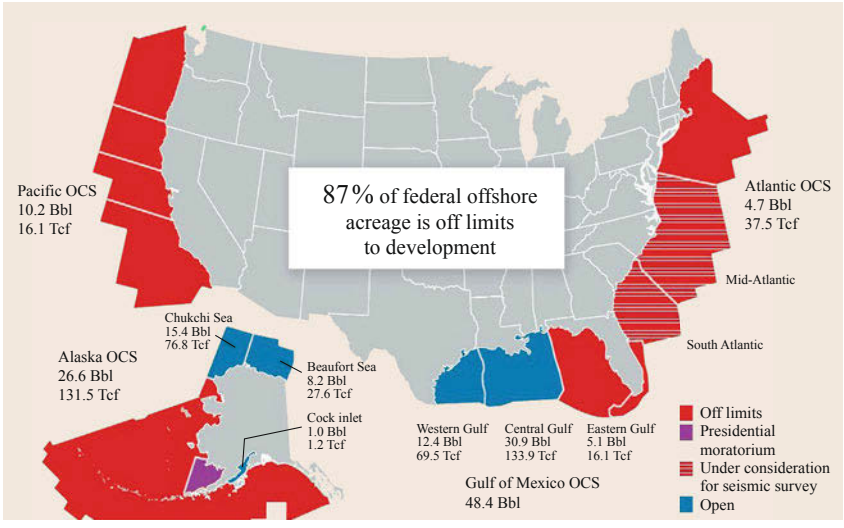


Fig. 33.8 Offshore technically undiscovered recoverable oil and gas resources around USA in federal acreage (Bbl and Tcf) (after [33.21])

gion holds significant volumes of oil and gas; until now investment in Arctic development and production has been rather small and most of it has been in areas with light to moderate ice conditions and relatively shallow waters, using artificial islands or gravity base structures [33.9].

A number of technology development targets are being set and will continue to be pursued. Some of these targets include [33.5, 9, 11, 13, 23, 24]:

- Lightweight, strong, and modular jackets, FPSO, and subsea concepts with storage capabilities for requisite water depths (varying from 1000 m to 4000 m), and with efficient long-distance tie-backs, would require more investment in offshore projects. Multi-phase pumping through pipelines would minimize costs considerably.
- Innovative developments in the float-over and self-installation of complete topsides of jackets and FPSOs (and their mooring systems) in a reliable and optimal manner (to avoid heavy offshore lift, intricate hook-up/commissioning and consequent accidents) is required. Reliable and automatically controlled seabed processing of oil and gas resources, with minimum intervention from above will make the operations economically more attractive.
- Advances in large-diameter, top-tensioned dry-Christmas-tree risers (the connecting valves, spools, and fittings are located on top of riser arrays), and arrays with catenary profiles that permit contact, optimal clashing, flow assurance, and long-term structural integrity at ultra-deep-water depths are very important [33.25].
- Prevention of the specific risks of a blowout in the ultra-deep waters by smart down-the-hole remote monitoring of conductors and risers, and provision of extra horizontal drilling conductors, that will become operative when reservoir pressures build up.

References

33.1 T. Moan: *Marine Structures for the Future – A Sea of Opportunities* (National University of Singapore, Singapore 2003)

33.2 D.V. Reddy, A.S.J. Swamidas: *Essentials of Offshore Structures – Framed and Gravity Platforms* (CRC, Boca Raton 2014) pp. 1–2

33.3 J. Morelock: Terrigenous Sediments. http://geology.uprm.edu/Morelock/dpseaterrig.htm (2015)

33.4 G.A. Lock: Technological factors in offshore hydrocarbon exploration, The Future of Offshore Petroleum. United Nations, Natural Resources and Energy Division, Ocean Economics and Technology Branch, Expert Meeting on the Future of Offshore Petroleum (McGraw-Hill, New York 1981) pp. 87–146

33.5 Offshore Operations Subgroup, Operations and Environment Task Group: Working Document of the NPC North American Resource Development Study. Subsea Drilling, Well Operations and Completions, Paper No. 2/11, 27–45, available at https://www.npc.org/Prudent_Development-Topic_Papers/2-11_Subsea_Drilling-Well_Ops-Completions_Paper.pdf

33.6 IndexMundi: Oil production country by country. In: *The World Factbook 2013–2014* (Central Intelligence Agency, Washington 2013), available at http://www.indexmundi.com/map/?v=88

33.7 A. Nagraj: Top Ten Countries With the World's Biggest Reserves, Gulf Business in Home/Energy/Features/Insights at http://gulfbusiness.com/2013/04/top-10-countries-with-the-worlds-biggest-oilreserves/#.Vi5kb4QnWFI (2013)

33.8 Central Intelligence Agency: The World Factbook, https://www.cia.gov/library/publications/resources/the-world-factbook (2014)

33.9 Eurasia Group: *Opportunities and Challenges for Arctic Oil and Gas Development*, http://www.wilsoncenter.org/sites/default/files/Artic%20Report_F2.pdf (Wilson Center, Washington 2013)

33.10 NOAA Ocean Explorer Gallery: Types of Offshore Oil and Gas Structures, http://oceanexplorer.noaa.gov/explorations/06mexico/background/oil/media/types_600.html (2010)

33.11 P. O'Connor, J. Buckness, M. Lalani: Offshore and subsea facilities. In: *Petroleum Engineering Handbook*, Vol. 3, ed. by L.W. Lake (Society of Petroleum Engineers, Richardson 2014)

33.12 Offshore Structures, Course Material Presented for OCE 3016 on Introduction to Coastal and Oceanographic Engineering, available at http://www.essie.ufl.edu/~sheppard/OCE3016/Offshore%20Structures.pdf (2015).

33.13 B. Middleditch: *Deepwater drilling riser technical challenges*, Technical paper (2H Offshore, Aberdeen 2011) pp. 10–35, http://www.2hoffshore.com

33.14 N. Terdre: Refloating Norway's Concrete Giants – Is it practical?, Offshore **71**(8), 88–91 (2011)

33.15 H.M. Refat, A.R. El-gamal: Influence of the density of water on the dynamic behavior of square tension leg platform, Amer. J. Civil Eng. Architecture **2**(4), 122–129 (2014)

33.16 American Petroleum Institute: Offshore Access to Oil and Natural Gas Resources, Unlocking America's Offshore Energy Opportunity, available at http://www.api.org/oil-and-natural-gas-overview/exploration-andproduction/offshore/~/media/Files/Oil-and-Natural-Gas/Offshore/OffshoreAccess-primer-highres.pdf, p. 15, (2015)

33.17 R.S. Haszeldine: Carbon capture and storage – How green can black be?, Science **325**(5948), 1647–1653 (2009)
33.18 J.M. Masset: Deep offshore, Proc. ASPO Conf. (2011)
33.19 D.V. Reddy: Offshore pipelines. In: *Offshore Structures*, Vol. 2, ed. by D.V. Reddy, M. Arockiasamy (Krieger Publications, Malabar 1991)
33.20 N. Huddleston: *Ocean Exploration, Highlights of National Academies Reports*, Ocean Science (National Academics, Washington 2008), available at http://dels.nas.edu/resources/staticassets/osb/miscellaneous/exploration_final.pdf
33.21 American Petroleum Institute: Offshore Access to Oil and Natural Gas Resources, Producing Offshore, available at http://www.api.org/oiland-natural-gas-overview/exploration-and-production/offshore/~/media/Files/Oiland-Natural-Gas/Offshore/OffshoreAccess-primer-highres.pdf, p. 1, (2015)
33.22 Kobelco Welding Worldwide, Trends in Global Demand for Offshore Structures, http://www.kobelco-welding.jp/education-center/technical-highlight/vol04.html (2014)
33.23 petrofed.winwinhosting.net: Offshore Oil and Gas Production Systems, available at http://petrofed.winwinhosting.net/upload/OffshoreoilandProd.pdf (2014)
33.24 M. Efthymiou: In deep water, http://www.offshore-technology.com/_features/feature52918/ (2009)
33.25 Det Norske Veritas: Riser Interference, Recommended Practice DNV-RP-F203 (Det Norskie Veritas, Høvik 2009)https://rules.dnvgl.com/docs/pdf/DNV/codes/docs/2009-04/RP-F203.pdf

34. Stability of Offshore Systems

Alexia Aubault, R. Cengiz Ertekin

Stability analysis is a major aspect of the design of offshore platforms. It focuses on averting any failures associated with excessive heel or trim, capsizing, catastrophic loss of buoyancy, or even sinking. In this chapter, we cover the buoyancy and stability of offshore platforms from both viewpoints of theoretical and regulatory issues. Intact and damaged stability of floating structures are discussed and essential formulations of the associated problems are introduced. In doing so, the stability of a platform (its ability to be at or to recover equilibrium at acceptable heel and/or trim angles) is assessed by comparison of the righting moment to overturning moments from external loads. Methods of calculating the righting moment that varies with the heel angle are discussed. At small heel and trim angles, the metacentric height and displacement may be used directly to calculate the righting arm. Overturning moments that account for all static loads, including environmental, operational, and live loads are listed as well. Assumptions are needed to calculate the applied loads, and stability analysis does not usually include the upsetting dynamic forces. To address these issues, regulators set criteria, with embedded safety factors, and they require sufficient reserve stability beyond the equilibrium of the platform. These and the risks associated with flooding are addressed through hull subdivision and damaged analysis in this chapter.

34.1 **Stability Criteria** 756

34.2 **Fundamentals** 757
- 34.2.1 Static Movements and Hull Position Float 757
- 34.2.2 Dynamic Movements 758
- 34.2.3 Some Geometric Definitions 758
- 34.2.4 Center of Buoyancy and Gravity 758
- 34.2.5 Irregular Shapes and Numerical Integration 759
- 34.2.6 Unit Systems 761

34.3 **Hydrostatic Forces and Moments** 761
- 34.3.1 Buoyancy and Displacement 761
- 34.3.2 Equilibrium of Forces and Moments 762
- 34.3.3 Shifting of Weight and Volume 763

34.4 **Stability** 763
- 34.4.1 Righting Arm, Righting Moment, and Metacentric Height for Small Angles 763
- 34.4.2 Transverse Metacentric Radius 764
- 34.4.3 Trim and Longitudinal Initial Stability 764
- 34.4.4 Weight Addition, Removal, and Shift 765
- 34.4.5 Effects of Liquid Free-Surface in Internal Tanks 766
- 34.4.6 Scribanti's Formula 767
- 34.4.7 Stability at Large Angles of Inclination 767
- 34.4.8 Dynamical Stability: Energy to Incline 768
- 34.4.9 Hydrostatic Stiffness Coefficients 769
- 34.4.10 Stability of Submersibles 770
- 34.4.11 Purpose 770
- 34.4.12 Industry Practice 773

34.5 **Loads** 776
- 34.5.1 Environmental Loads 776
- 34.5.2 Dead Loads 778
- 34.5.3 Live Loads 778
- 34.5.4 Operational Loads 778
- 34.5.5 Accidental Loads 779

34.6 **Lightship Parameters** 779
- 34.6.1 Technical and Commercial Significance 779
- 34.6.2 Regulatory Approach 779
- 34.6.3 Prediction 780
- 34.6.4 Determination 780
- 34.6.5 Control 781

34.7 **Subdivision** 781
- 34.7.1 Purpose and Criteria 781
- 34.7.2 Regulatory Requirements 781
- 34.7.3 Damaged Stability and Residual Stability 782

34.8 **Analysis** 784
34.8.1 Static- and Quasi-Static Analysis 784
34.8.2 Dynamic-Response-Based Analysis 784
References 785

Stability is one of the most critical aspects of offshore engineering, and the cornerstone of naval architecture. It aims at verifying that a vessel will remain afloat at equilibrium, and will not heel so much as to breach watertightness. For commercially built vessels, stability analysis follows internationally recognized rules. Rules rely on the calculation of a number of variables, which are common to the stability analysis of any floating body, and are based on classical rigid body mechanics.

The analytical framework for stability, much like many of the concepts of modern mathematics, was established in the mid-eighteenth century. French mathematician *Pierre Bouguer* [34.1] first introduced the concept of metacenter in his 1746 treatise of naval architecture, while, around the same time, Euler derived the notion of righting moment to assess the stability of a ship. As the first classification societies were created, in the nineteenth century, these tools were incorporated in newly devised rules and are still in use today.

The body of regulation has continued expanding to integrate lessons learned from past incidents. It has contributed to reducing casualties at sea. At first, regulations were aimed solely at ship-shaped vessels. However, as other types of floating vessels were built and serviced, new rules were developed. Although the mathematical tools remain the same, floating structures are subjected to varied loads from the environment. They also derive their stability in different ways. A deep draft platform, for instance, relies on a low center of gravity for stability, thus compensating a relatively small waterplane area. Traditional ship-shape vessels by contrast typically derive most of their stability from their large waterplane area. The function of a vessel further justifies variations in stability criteria. While a bulk carrier may be most susceptible to loss of buoyancy due to shifting of cargo, stability criteria must take into account the effect of umbilical and mooring on permanently moored offshore platforms. The variety of vessel types, functions, and exposure to the environment has resulted in a multiplicity of rules for stability. The present chapter focuses on stability criteria for offshore platforms and systems.

34.1 Stability Criteria

Offshore platforms and vessels are generally used for oil storage, exploration, or production. There are many types of offshore platforms; the five basic types are shown in Fig. 34.1.

Jacket-type platforms are the oldest platforms in existence since the late 1940s. The first jacket platform was installed in water depth of only 6 m! Today, jacket platforms can be installed in much deeper water depths. The drilling of oil is achieved by using risers that run from the deck to the ocean floor and then to great depths below the ocean floor depending on the location of the oil reservoir. These platforms are generally carried from the shipyard to the installation site by barges and in a number of pieces (or modules). Then they are upended from the barge(s) at the installation site. Upending and setdown of the modules involve very careful analysis of the buoyancy and stability characteristics of the platform and the barge combination before the installation takes place. This event is simulated through computer runs to make sure that upending will be successful and that modules would be able to mate the other modules when upended [34.2].

Gravity-type platforms are made of either concrete or steel. Although they are supported by their weights, sitting on the ocean floor, they are secured by piles, as are jacket platforms.

Semisubmersibles are the more common offshore platforms. Buoyancy is provided by the submerged pontoons and columns that are below the water surface. As a result, the wave forces that act on the platform are less than they would otherwise be; this in turn results in smaller platform motions, something that is very desirable. The other advantage of a semisubmersible is that it is mobile, that is, it can move on its own power to a new location when necessary, or sometimes transport vessels with caissons can piggyback a semisubmersible to move it to a new location. It can also be positioned dynamically through trusters.

TLPs have been more commonly used in recent years in oil production because of their very favorable motion characteristics. The very little dynamic movement of the platform is achieved by tensioning the tendons (which basically are pipes) attached to the legs (or columns of the platform) and the ocean floor. The vertical (static) equilibrium of the platform must consider the weight, buoyancy, and the pretension in the tendons of the platform (Sect. 34.3.2). There are a number of TLPs currently

operating also in the Gulf of Mexico and in the North Sea.

And finally, there are oil storage and exploration/production platforms that are of gravity-type. They are generally equipped with mud skirts that prevent them from sliding under the action of waves. Their conical shapes are mainly to break the ice sheets and thus prevent the collapse of the structure under the tremendous stresses caused by them. An advantage of such a platform is that it can be transported to a different location if necessary. A number of these platforms currently operate around the world. Installation of such platforms is also a challenging buoyancy and stability problem [34.2].

Basic concepts of naval architecture and mechanics of bodies are reviewed in this chapter. The quantities typical of stability analysis, such as metacenter, and righting moment, are introduced. Their formulas are derived and interpreted. These theoretical calculations apply to any vessel; therefore reference to a ship may be used in some cases for simplicity. In this regard, it is noted that many monographs [34.3–7] on the buoyancy and stability of ships can be used in the intact or damaged stability calculations of offshore structures. Stability analysis is inherently concerned with statics. Yet, dynamic forces may affect stability of a vessel temporarily, and are described as well. Understanding the loads that apply to a vessel is essential to set and apply appropriate stability criteria. Therefore, potential effects on offshore platforms from environmental, operational, as well as dead and live loads are outlined.

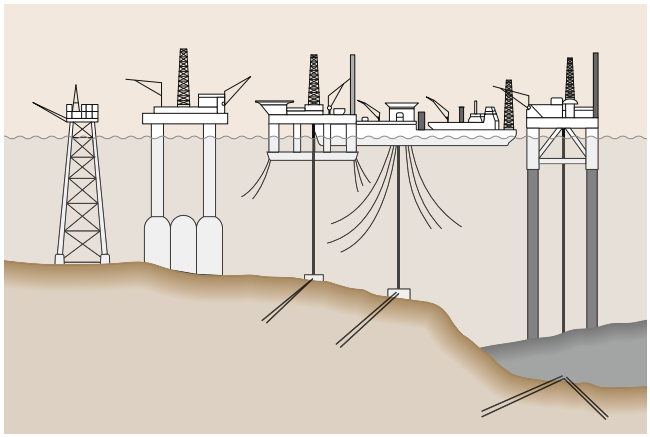


Fig. 34.1 Some offshore platform types. From left to right, they are: jacket-type platform (or *jacket* in short), gravity-type platform, semisubmersible (or *semi* in short), floating production ship or floating production, storage, offloading (FPSO) ship, and tension-leg platform (TLP) (after [34.8])

This chapter discusses performance standards for offshore platforms. A distinction is made between intact and damaged stability. Regulatory bodies enforce rules of stability for offshore platforms, using a set of tools throughout the lifetime of a vessel. The intent and possible implementation of the regulations are outlined herein as well.

34.2 Fundamentals

34.2.1 Static Movements and Hull Position Float

These movements include the three rotations of an offshore platform with respect to the three basic planes:

1. Transverse plane: A platform may heel to either the starboard or port side due to static forces, such as lifting a weight. Statically (when it is time-independent), this movement is called list angle or heel angle, as illustrated in Fig. 34.2. Dynamically (when it is time dependent), it is called rolling.
2. Middle-line plane: A platform may trim to either bow or stern due to a static force, such as cargo movement. Statically, this movement is called trim angle (Fig. 34.3). Dynamically, it is called pitching.
3. Water plane: A platform may drift to either port or starboard side due to a static force such as steady wind or current. Statically, this movement is called drift angle (Fig. 34.4). Dynamically, the rotation is called yawing. We will show that stability is not affected by movement in this plane, whether statically or dynamically.

Superposition of these movements can be done linearly as long as the angles are not very large.

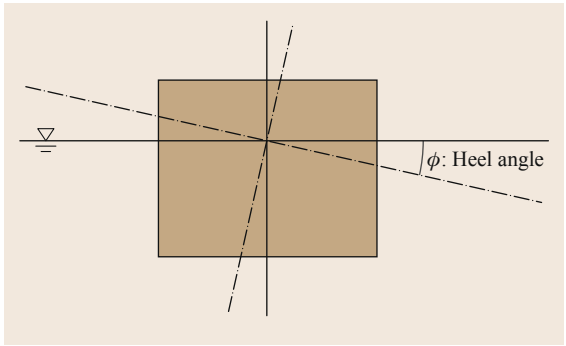


Fig. 34.2 List (heel) of a platform

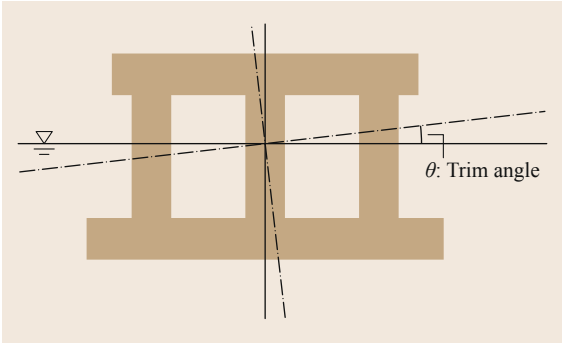


Fig. 34.3 Trim of a platform

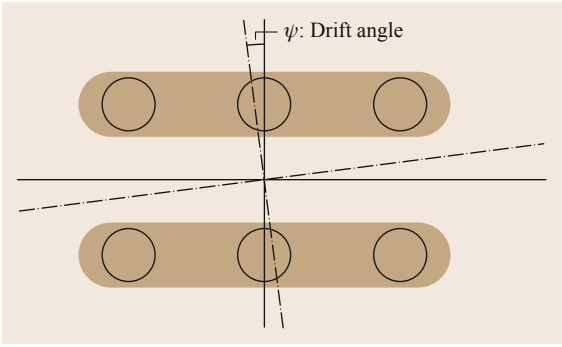


Fig. 34.4 Drift of a platform

34.2.2 Dynamic Movements

Like any three-dimensional rigid object, a ship has six degrees-of-freedom in its motions. These dynamic movements or motions are shown in Fig. 34.5. The x-axis runs through the centerline and positive x-axis points toward the bow. The positive y-axis points toward the port side of the ship and the positive z-axis points up, in the negative direction of the gravitational force.

The time-varying motion in the x-direction is called surge, in the y-direction it is called, sway and in the z-direction it is called heave. These are translational motions. The rotational or angular motion about the x-axis is called roll, about the y-axis pitch, and about the z-axis yaw. The dynamic motions are due to the presence of water waves (that are periodic) and so the six rigid-body motions are also periodic with their periods equal to the wave period. These dynamic translational and rotational motions are beyond the scope of this chapter. They will seldom be discussed further except for heave, roll, and pitch when we refer to the hydrostatic stiffness of offshore platforms in Sect. 34.4.9.

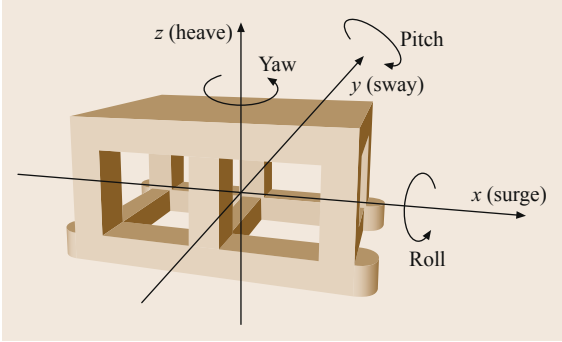


Fig. 34.5 Dynamic motions of a platform

34.2.3 Some Geometric Definitions

The three planes discussed in Sect. 34.2.1 are used both for ships and offshore platforms. Additionally, some essential geometric definitions are provided as follows:

- Keel (K): The baseline that runs at the lowermost of the platform, generally along the lowest surface of the pontoon or column, and generally parallel to the waterline.
- Draft (d): The vertical distance between the keel of the platform and the waterline.
- Freeboard: The vertical distance from the waterline to the uppermost watertight deck.
- Starboard: The right side of the platform when facing the bow.
- Port: The left side of the platform when facing the bow.
- Longitudinal center of flotation (LCF): The center of gravity of the waterplane area of the structure at the desired waterline (say the load or design waterline).

34.2.4 Center of Buoyancy and Gravity

When a structure immerses in water, the centroid of the submerged volume of the structure is called the center of buoyancy B. This centroid of the submerged volume is used in hydrostatics to do a number of calculations that will be discussed in later sections. Similarly, the location of the center of gravity of the structure is an important quantity and this location is used in stability calculations. We assume that the entire weight of the structure acts at the center of gravity G, of the structure. The first moments of both the center of buoyancy and center of gravity are zero with respect to themselves.

The longitudinal position of B, in the middle-line plane, is denoted by LCB; the transverse position, in a transverse plane, is TCB; and the vertical position, measured from the keel, is denoted by VCB (or $\overline{\text{KB}}$ sometimes). Similar notations apply to the position of the center of gravity (LCG, TCG, VCG).

34.2.5 Irregular Shapes and Numerical Integration

In buoyancy and stability calculations, we will see later that we need to determine various integral quantities, such as areas, volumes, and their moments. Most of the time, however, it is not possible to calculate these quantities analytically, but rather numerically. Therefore, we need to discuss first some basic integrals that will be used to do the integrations for some two-dimensional (2-D) and three-dimensional (3-D) shapes.

Areas and First and Second Moments

Most of the time, offshore platforms are made up of several different smaller shapes geometrically, such as rectangles, triangles, circles, and so forth. Then we can write composite equations for areas and their first moments as well as the center of gravity of areas

$$A_\mathrm{T} = \sum_{i=1}^{N} A_i\,, \quad M_x = \sum_{i=1}^{N} y_i A_i\,, \quad M_y = \sum_{i=1}^{N} x_i A_i$$

$$\overline{x}_\mathrm{c} = \frac{\sum_{i=1}^{N} x_i A_i}{A_\mathrm{T}}\,, \quad \overline{y}_\mathrm{c} = \frac{\sum_{i=1}^{N} y_i A_i}{A_\mathrm{T}}\,, \tag{34.1}$$

where N is the number of each smaller area, A_i is the area of each small shape, A_T is the total area, M_x and M_y are the first moments of the total area with respect to the x and y axes, and x_c and y_c are the coordinates of the center of gravity of the area (Fig. 34.6).

The moment of inertia of area (also known as the second moment of area) is used in the calculation of stability due to the presence of the waterplane area, as we will see later. Keeping Fig. 34.6 in mind, we can write, in general,

$$\mathrm{d}I_x = y^2\,\mathrm{d}A = y^2\,\mathrm{d}x\,\mathrm{d}y\,,$$

$$I_x = \sum \mathrm{d}I_x = \sum y^2\,\mathrm{d}x\,\mathrm{d}y\,, \tag{34.2}$$

and

$$\mathrm{d}I_y = x^2\,\mathrm{d}A = x^2\,\mathrm{d}x\,\mathrm{d}y\,,$$

$$I_y = \sum \mathrm{d}I_y = \sum x^2\,\mathrm{d}x\,\mathrm{d}y\,. \tag{34.3}$$

Parallel-Axis Theorem

Sometimes it is necessary to calculate easily the moment of inertia of an area with respect to an axis parallel to another axis (with respect to which we already know the second moment). If we know the moment of inertia of an area about an axis passing through the center of gravity of the area, then we can determine the moment of inertia of the same area with respect to any other axis which is parallel to the original one. Consider Fig. 34.7 for illustration. We write

$$I_x = I_{x'} + \overline{y}^2 A, \quad I_y = I_{y'} + \overline{x}^2 A\,, \tag{34.4}$$

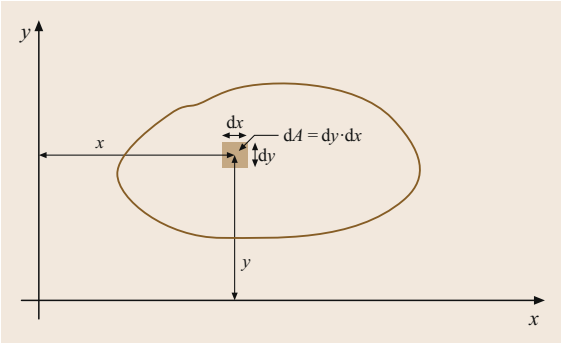


Fig. 34.6 Differential area

where $\overline{x}$ and $\overline{y}$ are the coordinates of the origin of the transformed coordinates measured from the original coordinates as shown in Fig. 34.7. Note that $I_{x'}$ and $I_{y'}$ should be the lesser of the inertia in the equation as both x' and y' pass through the neutral axes of the area (defined as the axes that pass through the center of gravity of the area), and hence the plus signs in front of the second terms on the right sides of the equations in (34.4).

Three-Dimensional Shapes

Three-dimensional shapes involve volumes and their moments. Such calculations are necessary to determine the displacement and the centroid of the submerged volume of the structure. In the case of platforms which have a number of buoyant components (such as columns and pontoons) connected to each other, one can calculate the volume and the center of buoyancy in a straight forward manner. Consider Fig. 34.8 and let $i = 1, 2, \ldots, N$ denote the label (used to identify a structural member) of the buoyancy component. Then, the volume of the platform is calculated by

$$\nabla = \sum_{i=1}^{N} V_i\,, \tag{34.5}$$

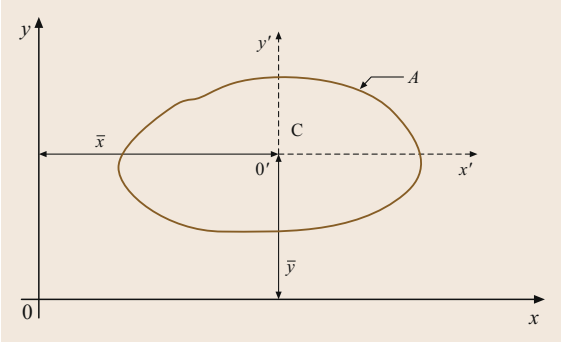


Fig. 34.7 Origin of the transformed coordinates for the parallel-axis theorem calculations

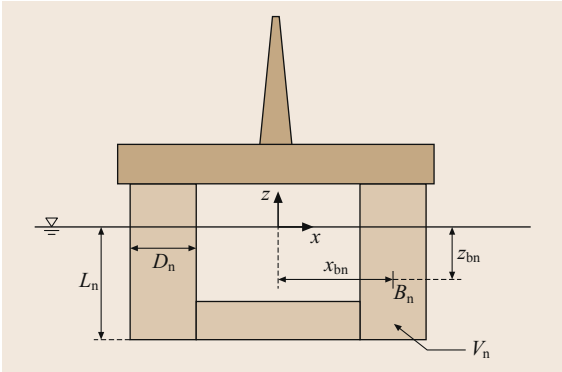


Fig. 34.8 Main dimensions of a floating platform

where V_i is the submerged volume of the i-th structural member.

The center of buoyancy of the platform can be calculated by taking the volume moment of each member, summing them up, and then dividing the sum by the total submerged volume of the platform,

$$x_B = \frac{1}{\nabla}\sum_{i=1}^{N} x_{bi} V_i \,,$$

$$y_B = \frac{1}{\nabla}\sum_{i=1}^{N} y_{bi} V_i \,,$$

$$z_B = \frac{1}{\nabla}\sum_{i=1}^{N} z_{bi} V_i \,, \qquad (34.6)$$

where $(x_{bi}, y_{bi}, z_{bi}) \equiv B_i$ denote the coordinates of the centroid (of the underwater volume) of the structural member i whose underwater volume is V_i.

Numerical Integration

There are many numerical methods of integration; however, the following are the most commonly used methods:

- Trapezoidal rule: Compound trapezoidal rule (Fig. 34.9) can be obtain by summing up the trapezoidal rule of integration for adjacent areas and by use of a first-order polynomial as the approximating polynomial for the ordinates. The area can be calculated, for example, by

$$A = \frac{h}{2}\left(y_0 + 2y_1 + 2y_2 + \cdots + 2y_{n-1} + y_n\right) \,, \qquad (34.7)$$

where n is equal to the number of nodes minus 1, and h is the constant *grid length* or *grid size* or *interval length*. Note that the truncation error in this method is $O(h)$.

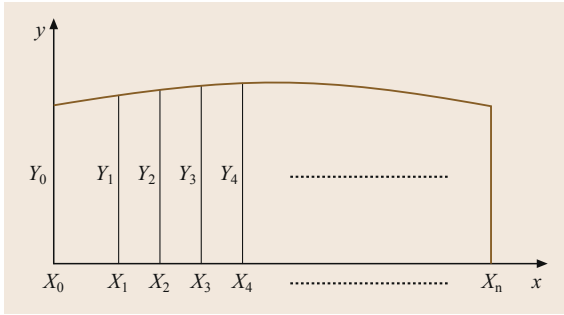


Fig. 34.9 Compound rule of integration

The static moment with respect to the y-axis can similarly be calculated as follows,

$$M_y = \int_0^L xy\,\mathrm{d}x = \frac{h}{2}(0 \times y_0 + h \times 2y_1 + 2h \times 2y_2 + 3h \times 2y_3 + \cdots \cdots + (n-1)h \times 2y_{n-1} + nh \times y_n) \,, \qquad (34.8)$$

where the origin is assumed to be at $x = 0$. The origin can be placed at any desired location, and the distance products, that is, $0, h, 2h, 3h, \ldots, nh$, in

Table 34.1 Unit system conversion

Physical quantity	Imperial unit	SI unit
Length	1.0 ft	0.3048 m
Length	1.0 in	0.0254 m
Length	1.0 yd	0.9144 m
Length	1.0 nautical mile (international)	1852 m
Length	1.0 nautical mile (UK)	1853.18 m
Length	1.0 mile	1609.344 m
Area	1.0 ft^2	0.092903 m^2
Area	1.0 in^2	645.16×10^{-6} m^2
Area	1.0 yd^2	0.836127 m
Volume	1.0 in^3	16.3871×10^{-6} m^3
Volume	1.0 ft^3	0.0283168 m^3
Velocity	1.0 ft/s	0.3048 m/s
Velocity	1.0 knot (international)	0.51444 m/s or 1.852 km/hr
Velocity	1.0 knot (UK)	0.51477 m/s or 1.85318 km/hr
Acceleration (gravitational)	32.174 (32.2) ft/s^2	9.80665 (9.81) m/s^2
Force	1.0 lb f	4.44822 N
Force	1.0 (long)ton f (2240 lb f)	9964.0128 N
Force	1.0 (short)ton f (2000 lb f)	8896.44 N
Mass	1.0 slug	14.59 kg
Mass	68.54 slug	1 ton = 1000 kg (metric)

(34.8), can be adjusted accordingly. Similarly, the equation for the second moment can be obtained.

- Simpson's first rule: This rule can be obtained by fitting a second-order polynomial to the ordinates. Compound Simpson's (first) rule is based on summing the adjacent integrals,

$$A = \frac{h}{3}(y_0 + 4y_1 + 2y_2 + 4y_3 + 2y_4 + \cdots + 4y_{n-1} + y_n) \,. \tag{34.9}$$

Note that the number of points (or nodes) must be an odd number ($n+1 = 3, 5, 7$, etc.) or the number of intervals should be an even number, $n = 2, 4, 6$, etc., to apply this method. Moreover, the interval size is constant. The coefficients in front of the ordinates (or offsets in the case of sectional area calculations), in (34.9) for example, are called *Simpson Multipliers* or in short *S.M.*, that is, $1, 4, 2, 4, \ldots, 4, 1$. Note that Simpson's formula is exact even for a third-degree polynomial. The truncation error in writing (34.9) is $O(h^3)$. Clearly, this is a much smaller error than in (34.7) when dimensionless $h \ll 1$. Similarly, formulas for first and second moments can be obtained.

34.2.6 Unit Systems

Most of the countries today use the SI unit system, and only a handful use the imperial unit system. Table 34.1 shows the unit conversion factors for some physical quantities that may be used in hydrostatics and stability calculations.

34.3 Hydrostatic Forces and Moments

Buoyancy is the ability of a structure in a fluid to be in vertical equilibrium. The subject of buoyancy is typically part of the field of hydromechanics which covers hydrostatics and hydrodynamics. Most problems of buoyancy indeed arise for bodies, including ships and platforms, floating in water, whether it is fresh water or salt water. All theoretical principles discussed herein may, however, be applied to problems in an alternative fluid, provided the value of fluid density is adjusted. When we discuss buoyancy herein, we are referring to the equilibrium of bodies in water, that is, hydrostatics.

34.3.1 Buoyancy and Displacement

Let us first discuss the hydrostatic pressure, and the forces and moments that a fluid element experiences under static conditions.

If a fluid element is in equilibrium, the static forces in the z (vertical) direction must sum up to zero, where

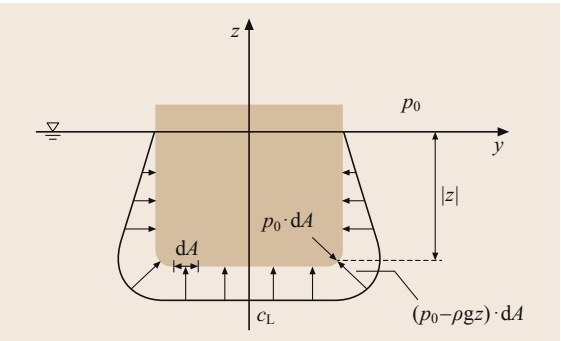


Fig. 34.10 Pressure force around the cross-section of a floating structure

z is positive above the still-water level (SWL), also referred to as the mean-water level (MWL), for a free-floating body. This leads to $p = p_0 - \rho g z$, where p_0 is the atmospheric pressure on the surface ($2116\,\mathrm{lb/ft^2}$ or $304\,704\,\mathrm{lb/in^2}$ (psi) or $101\,325\,\mathrm{Pa}$), and is small compared with p in general. And also, because of the continuity requirement for the pressure on the SWL, p_0 right above and right below the SWL must be the same. The addition of atmospheric pressure to p does not change the net pressure on the structure since inside the structure there is the atmospheric pressure as well; the outer and inner atmospheric pressures cancel each other. Therefore, we ignore p_0 and write

$$p = -\rho g z, \quad z \le 0 \,, \tag{34.10}$$

where p is called the hydrostatic pressure. The typical distribution of the force due to this pressure around a cross-section of a structure is shown in Fig. 34.10.

Archimedes' principle can be stated as *A structure immersed in a fluid is acted upon by a pressure force which is equal to the weight of the displaced fluid*. This force is in the vertical direction only. This principle then leads to hydrostatic forces that act on a floating or submerged structure be equal to

$$F_x = 0, \quad F_y = 0,$$
$$F_z = -w \int_A -\mathrm{d}\nabla = +w\nabla = \rho g \nabla \equiv \Delta \,, \tag{34.11}$$

where ∇ is the underwater (submerged) volume of the structure. In other words, $F_z = w\nabla = \rho g \nabla \equiv \Delta$, where Δ is the displacement (weight or force) or buoyancy

force. Since F_x and F_y are zero, we know that there is no hydrostatic restoration force in the x and y directions. Equations (34.11) can be obtained by integrating the distributed hydrostatic pressure on the wetted surface of the structure or it can be obtained by use of Gauss' divergence theorem.

The hydrostatic moments can be obtained by integration of the distributed pressure in the direction normal to the wetted surface of the structure

$$M_x = wy_B\nabla, \quad y_B = \frac{\int_A y\,d\nabla}{\nabla}, \quad M_y = -wx_B\nabla,$$
$$x_B = \frac{\int_A x\,d\nabla}{\nabla}, \quad M_z = 0, \tag{34.12}$$

where x_B and y_B are the coordinates of the center of buoyancy (centroid of the underwater volume)

$$x_B = \frac{-M_y}{\Delta}, \quad y_B = \frac{+M_x}{\Delta}. \tag{34.13}$$

In writing (34.12) or (34.13), we assumed of course that the first moments of the volumes were calculated with respect to the same coordinate system that x_B and y_B are measured in. Note also that M_x is the transverse moment of the displacement about the longitudinal axis, x, and M_y is the longitudinal moment of the displacement about the transverse axis, y, and that $M_z = 0$ indicates that there is no hydrostatic moment on the still-water level, that is, no static drifting moment (Sect. 34.2.1).

In summary then, there is no hydrostatic force or moment (and therefore no hydrostatic restoration) on the horizontal plane, something that is known intuitively; only the vertical force (displacement) and the two hydrostatic moments about the x-axis (list or heel moment) and the y-axis (trim moment).

34.3.2 Equilibrium of Forces and Moments

Here we refer to the equilibrium of a structure when we discuss the stabilizing forces and moments. Since there is no hydrostatic force or moment on the horizontal plane, we do not discuss equilibrium on that plane. A structure is said to be in vertical and rotational equilibrium if

1. Weight of the structure (W) is equal to the buoyancy force (or displacement) (Δ), i. e., $\sum F = 0$ (force equilibrium), or

$$W = \Delta = \rho g\nabla, \tag{34.14}$$

and

2. The center of buoyancy and the center of gravity must lie on the same vertical line so that $\sum M = 0$ (moment equilibrium). In other words,

$$\sum M_y = 0: \; +Wx_g - \rho g x_B\nabla = 0,$$
$$\sum M_x = 0: \; -Wy_g + \rho g y_B\nabla = 0,$$

after we used (34.14). And therefore,

$$x_B = x_g, \quad y_B = y_g, \tag{34.15}$$

where x_g and y_g are the coordinates of the center of gravity and, as before, x_B and y_B are the coordinates of the center of buoyancy. This means that the longitudinal center of buoyancy (LCB) must be the same as the longitudinal center of gravity (LCG); they are on the same vertical to the keel when the structure is in equilibrium. Similarly, the transverse centers of buoyancy (TCB) and gravity (TCG) must be on the same vertical.

If there are any cables, risers, or mooring lines that are under tension, the equilibrium of forces must take these into account. In the case of a tension-leg platform, for example, the pretension in the tendons must be added to the weight of the platform, and this will be equal to the displacement.

We must next mention the reserve buoyancy which is the volume between the design (or load) waterline and the main (watertight) deck as shown in Fig. 34.11. If we denote the reserve buoyancy by ∇', then

1. If $W > \rho g(\nabla + \nabla')$ the structure will sink, i. e., it is vertically unstable.
2. If $W < \rho g(\nabla + \nabla')$ the structure is vertically stable; it will come back to an equilibrium state if a small weight is added or removed. This is because of the hydrostatic restoration force the water provides in the vertical z direction.

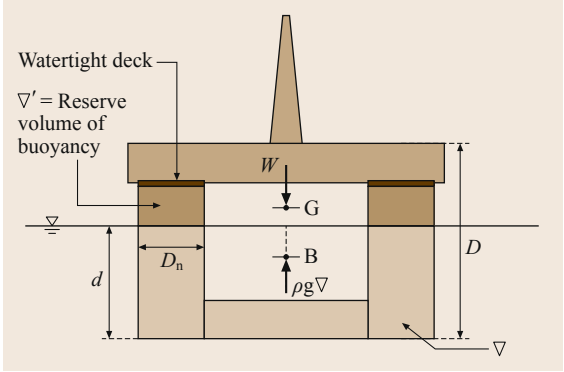


Fig. 34.11 Reserve buoyancy

3. If $W = \rho g(\nabla + \nabla')$ the structure is in neutral equilibrium (after the addition of a small weight, for example, it will not come back to its equilibrium state).

The equilibrium conditions can be thought of as the potential energy being, maximum (unstable), minimum (stable), and constant (neutral) as in an equilibrium model. For example, the equilibrium on the horizontal plane can be considered as neutral.

34.3.3 Shifting of Weight and Volume

Frequently, it is necessary to shift, add, remove weights, masses, areas, volumes, etc. on a structure. There is an elegant theorem called the law of moving (removing/adding) centroids (masses/weights/areas) that allows one to easily do this.

The upsetting forces that act on a floating or submerged structure may be due to, for example, the shift of some weight. It is of interest then to know how the center of gravity will shift as a result of weight movement on board. Similarly, when a platform heels, part of it immerses and part of it emerges, thus causing the centroid of underwater volume to shift. To see how these movements occur, let us consider the following theorem (Fig. 34.12):

Theorem 34.1

Let $\overline{\mathrm{M}}$ denote a structure of mass M with its center of gravity at G_0, and let some part $\overline{\mathrm{m}}$ of mass m (which is part of M) with its center of gravity at g be moved to a new position. Let the new center of gravity of $\overline{\mathrm{m}}$ be g_1. Then the new center of gravity G_1 of the altered structure has the property that the line $\mathrm{G}_0\mathrm{G}_1$ is parallel to gg_1. And furthermore, if $l = \overline{\mathrm{gg}_1}$ and $L = \overline{\mathrm{G}_0\mathrm{G}_1}$, then $ML = ml$.

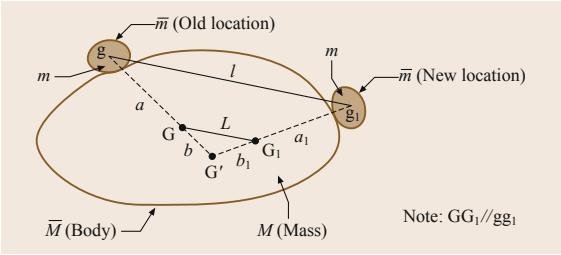


Fig. 34.12 The shifting of part of a structure mass to a new location

Without giving the proof, we give the final result

$$\frac{M}{m} = \frac{a+b}{b} = \frac{a_1+b_1}{b_1} = \frac{l}{L}\,. \tag{34.16}$$

Thus, $ML = ml$. Note that (34.16) of moving (or shifting) mass is applicable to other physical quantities such as volumes or areas or weights. It can also be used for removing and adding such physical quantities and therefore it can be used in hydrostatic and stability calculations. See, for example, Sect. 34.4.4 for some applications of this theorem.

34.4 Stability

The stability of a structure in water can be determined in terms of the existence of an equilibrium position, either at rest or subject to static movements such as sinkage, list or trim, and whether the structure has the ability to recover an equilibrium position after such static movement. This can be achieved by calculating the righting moment of the structure and comparing it with any external moment that will try to incline the structure. Let us start with the concept of righting (static stability) arm and moment.

34.4.1 Righting Arm, Righting Moment, and Metacentric Height for Small Angles

Consider a cross-section of a floating structure as shown in Fig. 34.13. Because of the effect of an externally applied moment (wind moment, moment due to the shifting of a weight, etc.), the structure heels or lists a small inclination angle of $\delta\phi$ and reaches a new equilibrium position. It is noted that *small angle* (less than

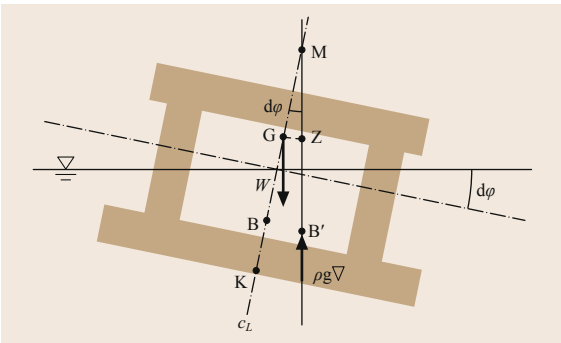


Fig. 34.13 The cross-section of a structure in inclined position (the center of gravity is below the metacenter)

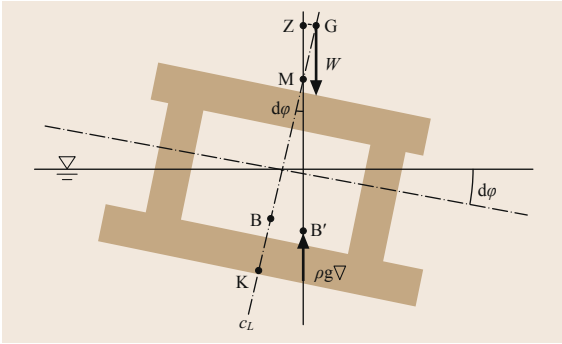


Fig. 34.14 The case of the center of gravity being above the metacenter

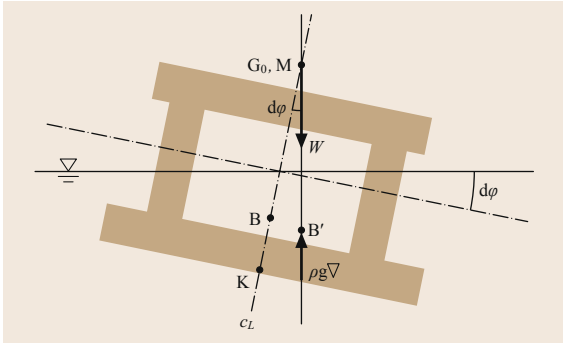


Fig. 34.15 The case of the center of gravity being on the metacenter

about 10°) is the key word here to have a fixed metacenter point M. The metacenter point is the intersection of the ship centerline with the line perpendicular to SWL and originating from the new center of buoyancy B′. The calculation of a fixed M point relies on the condition that as the platform lists, the volume that comes out of the water on one side is symmetrical to that which sinks under the waterline on the other side. It is valid only for small angles, when $\sin(\delta\phi) \approx \tan(\delta\phi) \approx \delta\phi$. The buoyancy force (or the displacement), $\Delta = \rho g \nabla$, remains the same (since no weight is added or removed and thus $\Delta = W$), but the center of buoyancy moves to a new position B′ because we lost some underwater volume on one side and gained the same on the other side.

The reaction (or righting) moment is $W\overline{GZ}$, and it can be determined from the following equation,

$$W\overline{GZ} = \Delta\overline{GZ} = \rho g \nabla \overline{GZ} = w\nabla\overline{GM}\sin\delta\phi \approx w\nabla\overline{GM}\delta\phi, \quad \delta\phi \ll 1\,\text{rad}\,, \tag{34.17}$$

where $\overline{GM}$ is called the transverse metacentric height (similarly, there also exists the longitudinal metacentric height). In short, these are referred to as the metacentric height or initial metacentric height. The term *initial* refers to the fact that small inclination angle assumption ($\delta\phi < 10^\circ = 0.175\,\text{rad}$) is used in the derivation of (34.17), and this is stressed. $\overline{GZ}$ is known as the righting-moment arm, or simply the *righting-arm*.

The metacentric height, $\overline{GM}$, is actually a measure of stability. To see this, consider the case in which G is above M, $\overline{GM} < 0$, as seen in Fig. 34.14. The weight causes the reaction moment in the same direction of list. This will result in the capsizing of the structure. If $\overline{GM} = 0$ then there is no reaction moment and the ship will be in neutral equilibrium as shown in Fig. 34.15. If $\overline{GM} > 0$, the righting moment will oppose heeling as was shown in Fig. 34.13. Therefore, $\overline{GM}$ can be used to determine the initial stability of the structure. Note that the metacentric height is given by

$$\overline{GM} = \overline{KB} + \overline{BM} - \overline{KG} = \overline{KM} - \overline{KG}\,, \tag{34.18}$$

where $\overline{KB}$ = centroid of the underwater volume (vertical component)(calculated), $\overline{BM} = \overline{BM}_T$ = transverse metacentric radius (calculated), and $\overline{KG}$ = vertical component of the center of gravity (either calculated/estimated or determined through inclining experiments).

34.4.2 Transverse Metacentric Radius

The transverse metacentric radius included in (34.18) need to be calculated for small list angles. This can be done by considering a longitudinal slice of a wedge of water at the waterline and calculating the amount of displacement shifted as a result of some external moment that inclines the structure. For small list angles, it can be shown that

$$\overline{BM} \equiv \overline{BM}_T = \frac{1}{\nabla}\frac{2}{3}\int_L y^3\,dx = \frac{I_T}{\nabla} = \frac{I_x}{\nabla}\,, \tag{34.19}$$

where I_x or I_T (the subscript T stands for transverse) is the second moment of the entire waterplane area with respect to the x-axis and ∇ is the underwater volume of the structure. To indicate that $\overline{BM}$ refers to the transverse rotation or rotation about the x-axis, the notation $\overline{BM}_T$ is generally used to denote the transverse metacentric radius.

34.4.3 Trim and Longitudinal Initial Stability

Trim by the stern means that excess trim is at A.P. (aft perpendicular) and trim by the bow (stem) means that

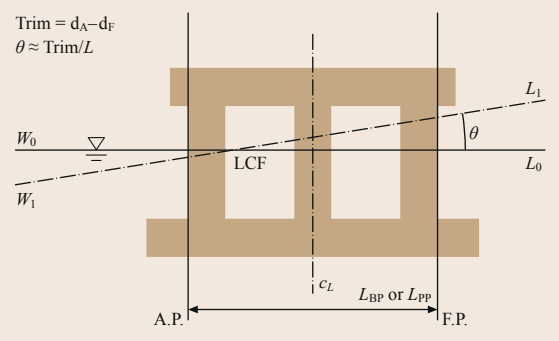


Fig. 34.16 Trim by the bow

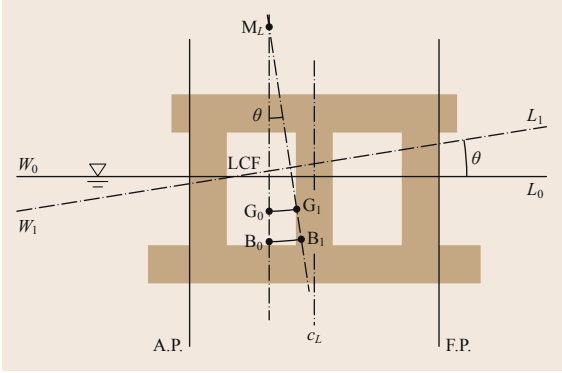


Fig. 34.17 The longitudinal metacenter

excess trim is at F.P. (forward perpendicular). For example, in Fig. 34.16, the trim is by the bow. We write

$$\text{trim} = t = d_A - d_F \,,$$
$$\tan\theta = \frac{t}{L} \Rightarrow \theta = \frac{t}{L} \text{ for small } \theta(< 10^\circ) \,, \quad (34.20)$$

where θ is the trim angle. With the (rather arbitrary) definition of trim given by (34.20), negative trim would mean that trim is by the bow.

If the *displacement* of the structure remains the *same* during trim, it can be proven that the structure must rotate about the LCF.

Trim can be calculated once the moment that causes the trim is known as

$$t = \frac{\text{Moment to change trim}}{\text{MCTC}} \,,$$
$$\text{MCTC} = \frac{\Delta\,\overline{G_0M_L}}{100\,L}, \text{ (ton m/cm)} \,, \quad (34.21)$$

where t is in centimeter, Δ is in ton, and $\overline{G_0M_L}$ and L in m, and the division by 100 is due to the definition of MCTC, that is *per centimeter*. Moment to change trim by one centimeter (MCTC) is used to calculate the trim easily as it is given in *hydrostatic curves*. In imperial unit system, the trim is written as

$$t = \frac{\text{Moment to change trim}}{\text{MCT1}''} \,,$$
$$\text{MCT1}'' = \frac{\Delta\overline{G_0M_L}}{12L}, \text{ (tonf ft/in)} \,, \quad (34.22)$$

where t is in inch, Δ is in longtonf (1 longtonf is 2240 lbf) and $\overline{G_0M_L}$ and L are in feet.

The longitudinal initial stability can be calculated very similar to the transverse one. The initial metacentric height for longitudinal inclination (trim) is written as

$$\overline{GM}_L = \overline{KB} + \overline{BM}_L - \overline{KG} \,, \quad (34.23)$$

where $\overline{GM}_L$ is the longitudinal metacentric height (initial) and $\overline{BM}_L$ is the longitudinal metacentric radius (Fig. 34.17). The vertical locations of the center of buoyancy, $\overline{KB}$, and center of gravity, $\overline{KG}$, are clearly the same as the ones used in the transverse stability calculations.

The longitudinal metacentric radius $\overline{BM}_L$ in (34.23) is obtained from

$$\overline{BM}_L = \frac{I_L}{\nabla} \,, \quad (34.24)$$

where I_L is the second moment of the waterplane area about the transverse axis, y', passing through the LCF.

The longitudinal metacentric height is much larger than the transverse metacentric height for ship-shaped vessels, and thus, the longitudinal intact stability of a vessel is not of concern most of the time, unless, of course, the displacement is very small. Nevertheless, it should be calculated in any case, and moreover, the knowledge of the longitudinal metacentric height is necessary in trim calculations as well as in damaged stability calculations.

34.4.4 Weight Addition, Removal, and Shift

When weights are added or removed or shifted, the centers of gravity and buoyancy will move. These movements can be calculated rather easily by use of the law of moving centroids, discussed in Sect. 34.3.3.

For example, assume that a weight, already onboard, is shifted to one side a distance of $\overline{g_0g_1}$ (Fig. 34.18). Both the center of gravity and the center of buoyancy will move to the same side until they are on the same vertical line which passes through the metacenter M. Note that, from the law of moving centroids, we must have G_0G_1 parallel to g_0g_1. Hence,

$$\tan\varphi = \frac{\overline{G_0G_1}}{\overline{G_0M}} \,. \quad (34.25)$$

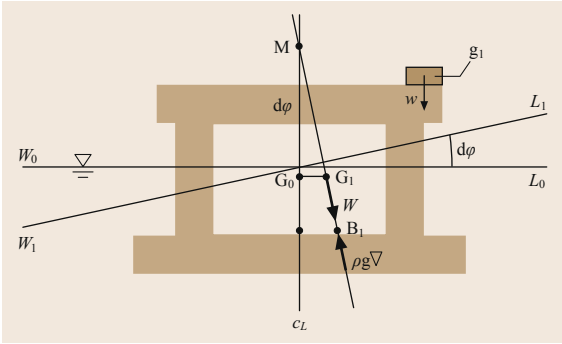


Fig. 34.18 A structure with a weight already onboard shifted to one side

This equation can be used to determine the new location G_1, of the center of gravity given by φ, G_0 and M. It can also be shown (or using the law of moving centroids when $L = \overline{G_0G_1}$, $M = W$, $m = w$, $l = \overline{g_0g_1}$ in (34.16)) that

$$\overline{G_0G_1} = \frac{w\overline{g_0g_1}}{W} , \tag{34.26}$$

so that the list angle is calculated from $\tan\varphi = \overline{G_0G_1}/\overline{G_0M}$, or

$$\varphi = \arctan\left\{\frac{w\overline{g_0g_1}}{W\overline{G_0M}}\right\} . \tag{34.27}$$

The list can then be calculated easily since

$$\text{List} = d_{\text{port}} - d_{\text{starboard}} ,$$
$$\tan\varphi = \frac{\text{List}}{\text{B}} \Rightarrow \text{List} \approx \text{B}\varphi$$

for small φ, where d_{port} and $d_{\text{starboard}}$ are the drafts on the port and starboard sides, respectively.

In the case of weight addition, there is a very useful quantity, called TPC (or TPI, depending on the unit system) that can be used. In deriving the new center of buoyancy of a floating structure when a weight, w, is added, we considered a slice of volume added to the ship. Can we find the approximate thickness of this slice? To answer this question, consider the weight, w, that sinks the structure by 1.0 cm. And let it float in water of mass density ρ. Then, for not *very large* weights, it is true that $A_{\text{WP}} \times 1.0\,\text{cm}$ is approximately equal to the added volume, and of course, this is exact for a wall-sided structure. The mass of the weight added must then be equal to $m_{\text{w}} = w/g = A_{\text{WP}} \times \rho \times 1.0\,\text{cm}$.

We can now define TPC as the ton (mass) per centimeter immersion (when the waterplane area A_{WP} is given in m^2 in SI units)

$$\text{TPC} = \frac{m_{\text{w}}}{1.0\,\text{cm}} = \frac{\rho A_{\text{WP}}}{100} = 0.01025 A_{\text{WP}} , \tag{34.28}$$

where 100 on the denominator comes from the *meter* to *centimeter* conversion, and the last equality above uses $\rho = 1.025\,\text{ton/m}^3$ (salt water). Note again that *ton* refers to mass, and not to weight (1000 kg = 1 ton), and A_{WP} is in m^2.

In the imperial unit system, the tonf-per-inch immersion of a waterplane is defined as the weight required to effect a parallel sinkage of a structure, at that waterline, of 1.0 in. If A_{WP} is given in ft^2, then the weight required to immerse the structure by 1.0 in is given by TPI (tonf per inch immersion)

$$\text{TPI} = \frac{w}{1.0\,\text{in}} = \frac{\rho g A_{\text{WP}}}{12} , \tag{34.29}$$

which, for $\rho = 2.0\,\text{slug/ft}^3$ (salt water) and $g = 32.0\,\text{ft/s}^2$, becomes

$$\text{TPI} = \frac{A_{\text{WP}}}{12}\frac{64\,\text{lb/ft}^3}{2240\,\text{lb/tonf}} = \frac{A_{\text{WP}}}{420} ,$$

where $1/(\rho g) = 2240/64 \approx 35\,\text{ft}^3/\text{tonf}$ is called the reciprocal weight density of salt water, and 1.0 longtonf = 2240 lb, and A_{WP} is given in ft^2.

We can now define parallel sinkage as

$$s = \frac{w(\text{tonf})}{\text{TPI}}\,(\text{in}) \quad \text{or} \quad s = \frac{w(\text{ton})}{\text{TPC}}\,(\text{cm}) , \tag{34.30}$$

where w is the added weight (similarly, r = rise can be defined if w is the removed (or negative) weight). Equation (34.30) is exact for wall-sided structures, and approximate for others.

When a weight is removed along the centerline, the same arguments used for parallel sinkage apply, except the weight becomes $-w$, and the structure will rise (this is called parallel rise) and is denoted by r.

34.4.5 Effects of Liquid Free-Surface in Internal Tanks

There may be many internal tanks in a floating structure that is partially or fully filled with a liquid. These can be ballast tanks, oil tanks, septic tanks, etc. When the structure heels due to some external force, the free surface of the liquid in these tanks will incline and, therefore, there will be a shift of liquid weight due to the movement of the liquid free-surface. This shift of weight will cause the metacentric height, $\overline{GM}$, to reduce, and thus, the stability of the ship will suffer as a result.

One can calculate the moment of weight of the shifted liquid due to the free surface of the liquid inside the tank and determine the amount the center of gravity

of the structure is shifted. This way, one can calculate the reduction in the righting (static stability) arm of the structure since the metacentric height is now different, and obtain

$$\overline{\mathrm{GM}_{\mathrm{Eff}}} = \left\{\overline{\mathrm{G_0M}} - \frac{\rho_\mathrm{l}}{\rho_\mathrm{s}}\frac{i_\mathrm{T}}{\nabla}\right\}, \qquad (34.31)$$

where ρ_l is the density of the liquid in the tank, ρ_s is the density of water that the structure floats in, ∇ is the submerged volume of the structure, and i_T is the second moment of the free-surface area of the tank with respect to its own longitudinal axis; $\overline{\mathrm{GM}_{\mathrm{Eff}}}$ is called the virtual or effective metacentric-height. The second term on the right-side, $(\rho_\mathrm{l}/\rho_\mathrm{s})(i_\mathrm{T}/\nabla)$, is called the virtual loss in metacentric height. If there is more than one tank, then the losses due to each tank's liquid free-surface must be summed. Note that this result is obtained by assuming that the list angle is *small*.

The loss of stability due to the presence of liquid free-surface in internal tanks can be reduced in some cases by creating subdivisions in each tank so that the virtual loss in metacentric height becomes $(\rho_\mathrm{l}/\rho_\mathrm{s})(i_\mathrm{T}/\nabla)(1/n^2)$, where n is the number of subdivisions.

34.4.6 Scribanti's Formula

There is a formula that can be used to calculate the righting arm, $\overline{\mathrm{GZ}}$, of a wall-sided (or nearly wall-sided around the still-water level) structure or vessel for inclining angles up to about 25°. This formula can be derived by calculating the vertical and horizontal moment of the sectional wedge volume and integrating along the length of the structure, and then determining the new value of the righting arm, see [34.3] Chapt. 4, Sect. The Wall-Sided Formula

$$\overline{\mathrm{GZ}} = \left(\overline{\mathrm{GM}} + \frac{1}{2}\overline{\mathrm{BM}}\tan^2\phi\right)\sin\phi\,. \qquad (34.32)$$

It is noted that for small ϕ, we had $\overline{\mathrm{GZ}} = \overline{\mathrm{GM}}\,\phi$ previously. The righting-arm, $\overline{\mathrm{GZ}}$, formula given by (34.32) is known as Scribanti's formula. The righting moment follows from $M_\mathrm{st} = \Delta\overline{\mathrm{GZ}}$.

34.4.7 Stability at Large Angles of Inclination

Recall that the static moment or righting moment of a floating structure or vessel for small inclination angles is given by (34.17). This assumes that the metacenter is a fixed point. In the case of large list/trim angles, however, the metacenter M is no longer fixed but moves along a curve called the metacenter curve or M-curve. A vessel rotates about the point where the centerline meets the waterline when the list/trim angle ϕ is small. For large inclination angles, this is no longer true in general unless the vessel is wall-sided.

To calculate the righting arm, $\overline{\mathrm{GZ}}$, for any list/trim angle, one needs to determine the location of the center of buoyancy as well as the center of gravity. The latter is generally assumed initially and then a correction due to it can easily be made in calculating the righting arm, once the exact location of the center of gravity is determined.

We can summarize the procedure for these calculations as follows:

1. Assume that the vessel is in upright position, meaning G and B are on the same vertical line (and no trim) as shown in Fig. 34.19.
2. Rotate it without trim, about the x-axis as shown in Fig. 34.20. The center of buoyancy may move horizontally and, therefore, B and G will not be on the same vertical line.
3. Allow it to trim so that the weight-buoyancy and trim equations are satisfied; i. e., B and G are on the same vertical line as shown in Fig. 34.21. The righting arm, $\overline{\mathrm{GZ}}$, shown on the right, is now the correct one which must be determined in this position.
4. To determine $\overline{\mathrm{GZ}}$, calculate the static/righting moment M_st of the structure/vessel. The righting arm is then obtained from $\overline{\mathrm{GZ}} = M_\mathrm{st}/\Delta$.

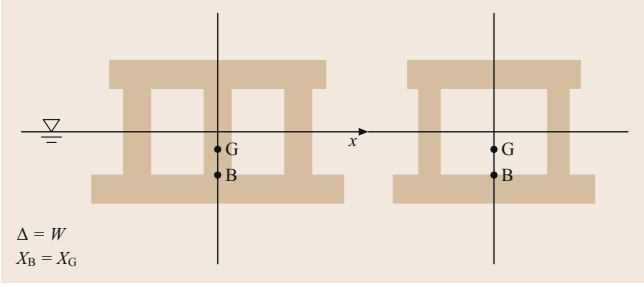


Fig. 34.19 The vessel is in upright position

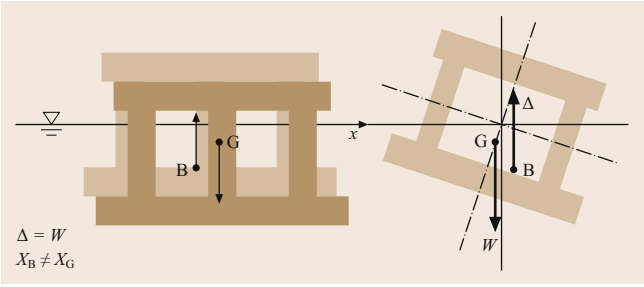


Fig. 34.20 The vessel lists with no trim

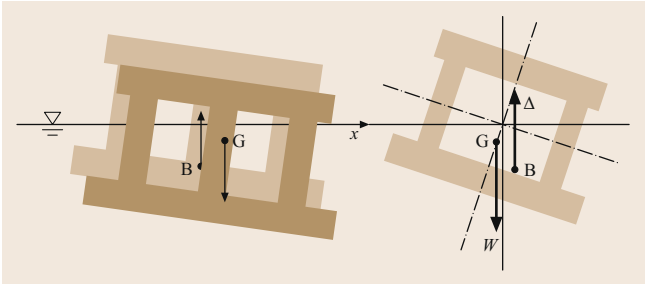


Fig. 34.21 The vessel lists and trims

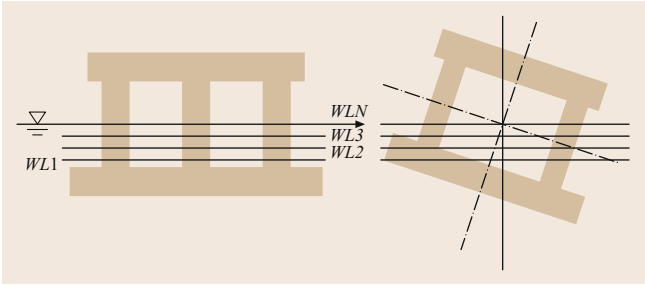


Fig. 34.22 The water lines used in righting-arm calculations

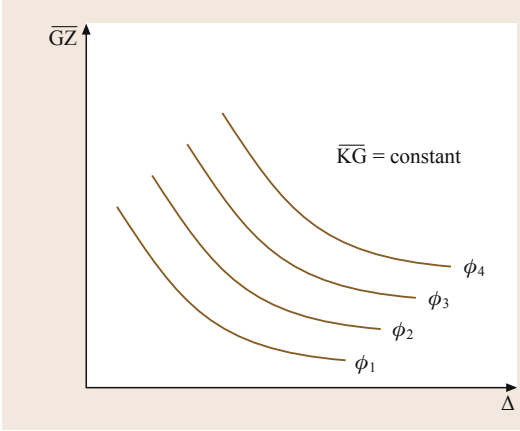


Fig. 34.23 Cross-curves of stability for different list angles

5. The above procedure is repeated for a series of waterlines as shown in Fig. 34.22, and for a constant list angle ϕ, and similarly for a constant trim angle θ. This way, one can obtain $\overline{\mathrm{GZ}}$ as a function of the displacement Δ for a given set of ϕ (or θ) and $\overline{\mathrm{KG}}$ as shown in Fig. 34.23, called the cross-curves of stability.
6. The procedure in the previous step should be repeated for a series of inclination angles to obtain the cross-curves of stability, for a constant $\overline{\mathrm{KG}}$, as shown in Fig. 34.23. Or, one can plot the constant-displacement curves that are called the static-stability curves shown in Fig. 34.24.

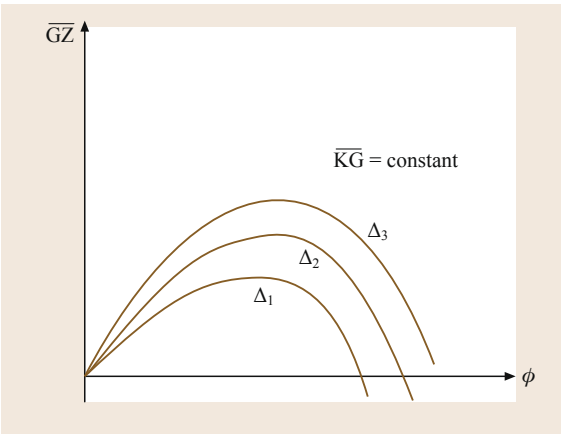


Fig. 34.24 Static-stability curves

34.4.8 Dynamical Stability: Energy to Incline

A suddenly applied heeling moment (drop of heavy cargo, gusty wind, etc.) can cause dynamical stability problems. Let us discuss this next by considering the following heeling moment M,

$$M = 0 \quad \text{for} \quad t < 0\,,$$
$$M = \text{constant} \quad \text{for} \quad t \geq 0\,, \tag{34.33}$$

where $t = 0$ denotes the initial time. This moment heels the structure to a new angle.

We now define dynamical stability as the work done in inclining the structure to an angle ϕ. To see the relation between the work done and the list of a ship, consider the stability of an object that is discussed in mechanics (Fig. 34.25). The weight, which is equal to mg, where m is the mass acting at G which is at a distance of h_0 above the floor. R is the reaction force exerted by the floor. Now, if the object is raised from one side, its center of gravity G will move up and will be h_1 away from the floor. Therefore, a moment couple is formed and the object will return to its initial position when the object is released if $h_1 > h_0$, that is, potential energy is increased by moving the object up. Before the object was raised, the potential energy was minimum (stable equilibrium). The work done on the object is equal to the change in potential energy. In other words, the vertical separation distance $(h_1 - h_0)$ between the center of gravity G and the location of the reaction force R determines the change in potential energy, which we

define by dynamical stability, that is,

$$E_{\mathrm{st}} = mg\,(h_1 - h_0)\ , \tag{34.34}$$

in the example above. E_{st} has the unit of work. Note that the moment equilibrium is reached when

$$Pz = mgy\,. \tag{34.35}$$

In the case of a floating body, Pz is the heeling (or overturning) moment and mgy is the righting (or restoring) moment. It can be shown that

$$E_{\mathrm{st}} = \Delta \int_0^{\phi} \overline{\mathrm{GZ}}\,\mathrm{d}\phi'\,, \quad \Delta = W\,. \tag{34.36}$$

This is the dynamical stability of the floating structure when it heels to an angle ϕ, and it shows that the area under the $\overline{\mathrm{GZ}}(\phi)$ function is equal to the change in potential energy. The unit of E_{st} is ton m, therefore, ϕ must be in radians in these calculations, and not degrees. The work done by the heeling (or overturning) moment M is

$$W_{\mathrm{h}} = \int_0^{\phi} M\,\mathrm{d}\phi'\,. \tag{34.37}$$

Equilibrium is reached when $W_{\mathrm{h}} = E_{\mathrm{st}}$.

34.4.9 Hydrostatic Stiffness Coefficients

A rigid body has six degrees-of-freedom (surge, sway, heave, roll, pitch, yaw) in motions. These modes of motions can be labeled 1 through 6, respectively. We already have seen that hydrostatic restoration in modes 1, 2, or 6 is zero (Sect. 34.3.1).

The forces and moments that act on a floating structure are governed by the equations of motion so that a floating body can be in static and dynamic equilibrium through Newton's second law of motion, that is, that forces must be equal to mass times translational accelerations, $F_i = m_{ij}a_j$, and moments must be equal to mass moment of inertia times rotational accelerations. The arranged (linear) equations of motion of a floating body can be written as

$$\left(M_{ij} + \mu_{ij}\right)\ddot{x}_j + \lambda_{ij}\dot{x}_j + C_{ij}x_j + k_{ij}x_j = F_i + \cdots\,, \tag{34.38}$$

where there is summation from 1 to 6 over the repeated index j Einstein's summation convention, M_{ij} are the elements of the mass matrix, x_j are the time-dependent motions (or displacements) in mode j, $\dot{x}_j$ are the body velocities, $\ddot{x}_j$ are the accelerations, μ_{ij} are the elements of the added-mass matrix, λ_{ij} are the elements of the damping matrix, C_{ij} are the elements of the hydrostatic restoration matrix, k_{ij} are the elements of the mooring stiffness matrix, F_i are the elements of the wave exciting force/moment vector, and $i, j = 1, 2, \ldots, 6$, i = force or moment, j = direction of motion. The elements of the hydrostatic restoration forces and moments are such that

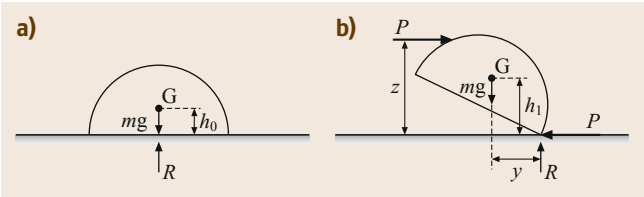


Fig. 34.25a,b Stability of an object

$$F_{\mathrm{H}i} = -C_{ij}x_j \quad i,j = 1, 2, \ldots, 6\,. \tag{34.39}$$

Since hydrostatic restoring coefficients for the movements on the horizontal plane are zero, $C_{ij} = 0$, $i, j = 1, 2, \ldots, 6$. Also, C_{ij} can be shown to be the elements of a symmetric matrix, that is, $C_{ij} = C_{ji}$.

The heave restoring force is written as $F_{\mathrm{H3}} = -(C_{33}x_3 + C_{34}x_4 + C_{35}x_5)$, where the heave restoring coefficient is given by

$$C_{33} = \rho g A_{\mathrm{WP}}, \quad A_{\mathrm{WP}} = \sum_{m=1}^{M} A_{\mathrm{WP}m}\,. \tag{34.40}$$

Clearly, if there is no waterplane area, there is no vertical restoration hydrostatically, provided, of course, that the body is in equilibrium. The index $m = 1, 2, \ldots, M$ refer to the surface-piercing members only. In other words, if a member is not surface piercing, then it is not included in (34.40) since a motion in the x_3 direction is not going to result in a change in volume (as this would be the case for a completely submerged structure).

The heave-roll restoring coefficient is given by (Fig. 34.26)

$$C_{34} = C_{43} = \rho g \iint_{A_{\mathrm{WP}}} y\,\mathrm{d}S = \rho g \sum_{m=1}^{M} y_m A_{\mathrm{WP}m}\,. \tag{34.41}$$

Note that $C_{34} = C_{43} = 0$ if there is symmetry with respect to the x–z plane of the platform, and this is common.

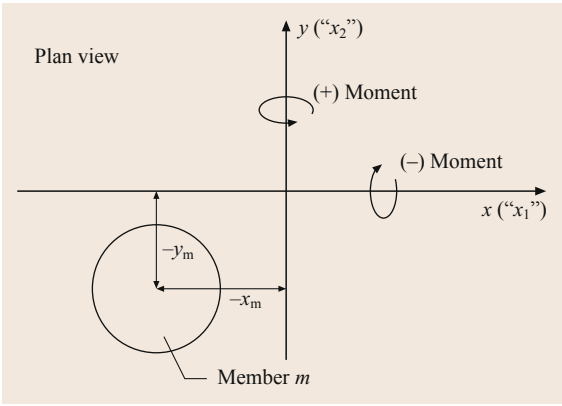


Fig. 34.26 Calculations for the heave-roll and heave-pitch restoring coefficients (z is positive up)

On the other hand, the heave-pitch restoring coefficient is given by (Fig. 34.26 again)

$$C_{53} = C_{35} = -\rho g \iint_{A_{WP}} x \, dS = -\rho g \sum_{m=1}^{M} x_m A_{WPm} \, . \tag{34.42}$$

Note that $C_{53} = C_{35} = 0$ if there is symmetry with respect to the y–z plane of the platform. This can be seen directly from (34.42). This also is a common occurrence.

The roll-restoring moment is written as $F_{H4} = -(C_{43}x_3 + C_{44}x_4 + C_{45}x_5)$, where the roll-restoring coefficient C_{44} is given by (Fig. 34.27)

$$\begin{aligned} C_{44} &= \rho g \nabla (z_B - z_G) + \rho g \iint_{A_{WP}} y^2 \, dS \\ &= -\rho g \nabla \overline{BG} + \rho g \nabla \overline{BM}_T = \Delta \, \overline{GM}_T \, . \end{aligned} \tag{34.43}$$

Note that this assumes that the angular displacements (roll and pitch) are *small* (definitely less than about 10°, linear theory). Also *note that* $C_{44}\,x_4$ is the transverse righting moment, where $x_4 = \phi$ is the list angle.

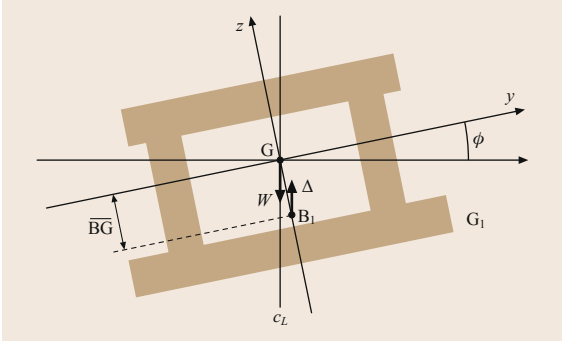


Fig. 34.27 Calculations of the roll-restoring coefficient

The roll-pitch restoring coefficient, on the other hand, is given by

$$C_{45} = C_{54} = -\rho g \iint_{A_{WP}} xy \, dS = -\rho g \sum_{m=1}^{M} x_m y_m A_{WPm} \, . \tag{34.44}$$

Note that if there is symmetry with respect to both the x–z and y–z planes of symmetry, then $C_{45} = C_{54} = 0$.

The pitch-restoring moment is written as $F_{H5} = -(C_{53}x_3 + C_{54}x_4 + C_{55}x_5)$, where the pitch-restoring coefficient is given by

$$\begin{aligned} C_{55} &= \rho g \nabla (z_B - z_G) + \rho g \iint_{A_{WP}} x^2 \, dS \\ &= -\rho g \nabla \overline{BG} + \rho g \nabla \overline{BM}_L = \Delta \, \overline{GM}_L \, , \end{aligned}$$

and $C_{55}\,x_5$ is the longitudinal righting moment and $x_5 = \theta$ is the trim angle. Note that all these hydrostatic restoring coefficients are valid when the motions are *small* or linear.

34.4.10 Stability of Submersibles

When a submersible is completely submerged, it has no water-plane area, and therefore the metacentric radius, $\overline{BM}_T$ or $\overline{BM}_L$, tends to zero. This means that B → M and would rise over G. This is just the opposite of what we have in the case of a floating structure. As a result, the stability moment becomes very simple,

$$M_{St} = \Delta \, \overline{GZ} = W \overline{BG} \sin\phi = W \overline{GM} \sin\phi \, . \tag{34.45}$$

The period of submerging and surfacing need to be treated very carefully because of the presence of ballast tank free-surfaces that do reduce the metacentric height further. The stability criteria for submersibles must at least require positive values of $\overline{GM}$.

34.4.11 Purpose

Service Requirements

Stability analysis applies to any floating structure. It aims at determining if the structure will be able to remain afloat within an allowable range of list angles when subjected to overturning loads. The extent of stability of a hull is typically assessed by comparison of the righting-moment curve with the extreme overturning curve, under all predicted loading conditions. This is exemplified in a typical righting-moment curve in Fig. 34.28. The overturning-moment curve meets the righting-moment curve at equilibrium, and must be lower than it at higher heel angles to provide reserve stability.

Insufficient stability of a floating structure may lead to:

1. Capsizing, the loss of righting moment resulting in the hull heeling beyond acceptable limits, and in some cases turning upside down, or
2. Loss of buoyancy of the hull, which is due to ingress of water through loss of watertight integrity or progressive flooding, and may lead to sinking.

The durability of stability is ensured by preserving watertightness of the hull within the range of operational drafts and heel angles of the hull. The minimum extent of watertightness is defined as an envelope of draft and heel conditions as illustrated on heel angle in Fig. 34.28. If the mean waterline under any load condition extends beyond this envelope, the hull may start taking in water from outside through nonwatertight openings. Some hull openings may be within the extent of watertightness so long as they are designed to be watertight when fully submerged with an appropriate pressure head. This may be the case of bolted hatches, or some capped sounding pipes for instance.

Additionally, we define the extent of weathertightness as the acceptable limit of reach of *green water*, or wave dynamics, above the mean waterline. The waterlines to which weathertightness extend on a typical semisubmersible hull are shown in Fig. 34.29. Openings within the extent of weathertightness shall be designed to withstand a low water pressure. Vents with nonreturn valves or blind flanges, some quick-open manways, and similar openings are designed for weathertightness for instance.

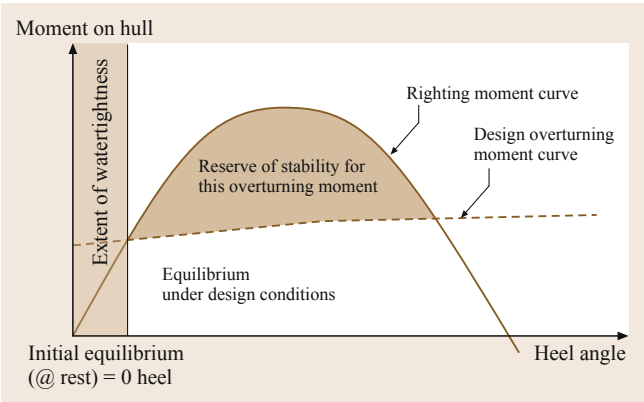


Fig. 34.28 A typical righting-moment curve

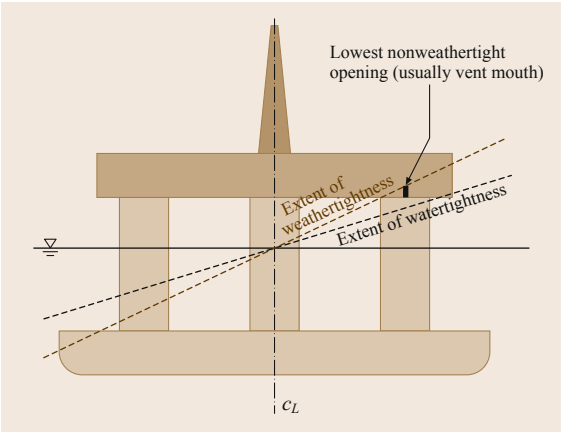


Fig. 34.29 Example of the extent of watertightness and weathertightness on a semisubmersible

Capsizing

Capsizing may be defined as the loss of static righting moment which leads to increasing heel of the vessel past the range of acceptable angles. Some hulls, such as monohull sailboats, may be designed to minimize the extent of the equilibrium once the boat is capsized upside down so that it will not remain in a capsized position, but right itself back up.

For a free-floating structure, capsizing typically results from one of two phenomena:

1. If the hull is intact, the center of gravity and center of buoyancy are unchanged from their original envelope. However, overturning loads in excess of predicted loads may result in loss of equilibrium and heel of the platform. Excessive loads may be static or dynamic. If the righting moment is comparatively too small at any given heel angle, the hull will keep heeling until a position of equilibrium is found. Unless the energy to incline increases again as the hull heels or the overturning energy starts decreasing, it will capsize and may end up lying on its side or upside down. This is illustrated on a righting-moment curve in Fig. 34.30.
2. Capsizing can also be due to a change in mass or buoyancy characteristics of the hull. These affect the metacentric height of the structure, and consequently modify the righting-moment curve. The righting moment may become negative or too small compared to expected overturning moment. This is illustrated in Fig. 34.31.
   Loss of buoyancy, resulting in a shift in the center of buoyancy location, thus in a shift of the righting-arm curve, typically occurs in case of loss of watertight integrity of the hull. This may be due to flooding of a compartment due to loss of structural integrity on an external bulkhead. It can also be due to accidental water ingress, such as downflooding or progressive flooding.
   Changes in mass properties may be linked to events such as accidental loss of cargo, internal leaking,

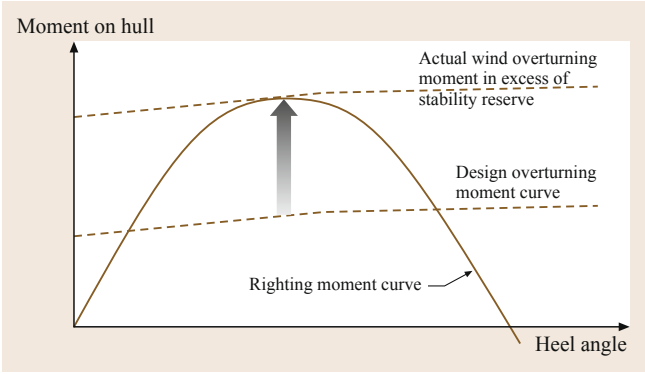


Fig. 34.30 Capsizing due to excess of overturning moment: the effect on the moment curves

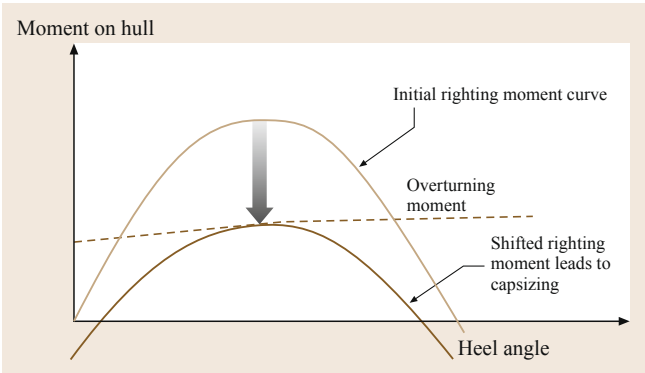


Fig. 34.31 Capsizing due to changes in stability characteristics: the effect on the moment curves

or accidental transfer of cargo. These may result in draft changes as well as a shift of the center of gravity. The latter will affect the metacentric height and may lead to capsizing.

Other mechanical failures may lead to capsizing in the case of a structure with additional static constraints. This is the case of moored structures with high mooring pretension, such as a TLP. The pretension on the mooring contributes to stability. This may be modeled as a correction on the center of gravity and weight, thus affecting the displacement

$$W_{\text{corr}} = W + T\,,$$
$$\text{VCG}_{\text{corr}} = \frac{\text{VCG} \times W + z_{\text{fairlead}} \times T}{W_{\text{corr}}}\,, \qquad (34.46)$$

where W is the weight, T the pretension of the mooring, VCG the vertical position of the center of gravity, and z_{fairlead} the vertical position of the fairleads. Similar formulas may be used for the horizontal position of the center of gravity, TCG (transverse) and LCG (longitudinal).

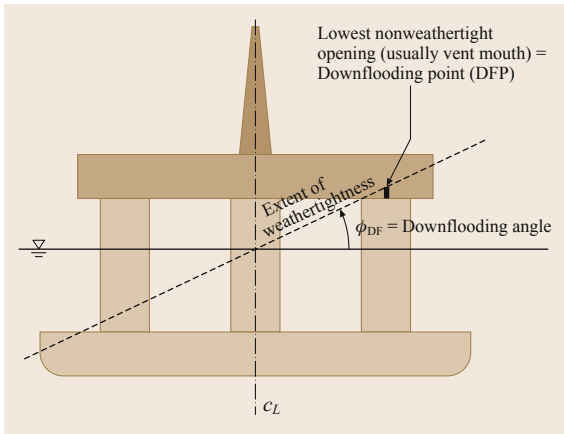


Fig. 34.32 Example of downflooding angle on a semisubmersible hull

If loss of pretension in the tendons occurs, the corrected mass properties are affected, and the platform may capsize.

Downflooding

The downflooding point of a hull marks the limit of the extent of weathertightness. If green water from wave action reaches this point, there is water ingress in the hull. A loss of weathertightness may result in flooding of the hull.

The downflooding point is usually a vent, a door, or any similar nonweathertight opening on deck or in external bulkheads. For each operational draft condition, a downflooding angle may be defined. It is the heel angle at which the waterline reaches the downflooding point. Figure 34.32 shows an example of a downflooding angle on a semisubmersible. For stability of the hull to be preserved, the downflooding angle must be greater than the angle of equilibrium of the hull, under all predicted extreme overturning loads.

If the hull heels beyond the downflooding angle, water ingress in the hull may lead simultaneously to sinking, due to loss of buoyancy (or increase in weight), and to heeling, due to change of metacentric height and weight distribution. Depending on the hull design and on the location of water ingress, the hull may sink or it may capsize first. Even if the mean heel angle of a platform remains below the downflooding angle, consideration should be given to the effect of dynamic heel and wave elevation, which would contribute to a loss of weathertightness and flooding. For this reason, regulations typically require that a reserve of stability exists between the maximum mean heel angle under static loads and the downflooding angle. The downflooding angle of a hull is an important parameter to add to the righting-moment curve (Fig. 34.33).

Progressive Flooding

Progressive flooding refers to the accidental ingress of water due to system failure. For instance, it may involve open-looped piping systems, which have the capability of pumping external sea water into tanks. Due to human errors or control failures, such a system may pump sea water from the ocean beyond acceptable limits. Progressive flooding may also occur if marine systems or bulkhead openings are operated in a way that spreads flooding between compartments. This results in a change of mass properties of the hull, with risks of sinking or capsizing.

A variety of solutions may be implemented to mitigate the risks of progressive flooding:

- Redundant alarms, to alert operators of unusual levels of water in tank, unusual draft or heeling conditions, internal leaking, etc.
- Fail-safe systems, where tanks may be limited in capacity to avoid catastrophic failure in case of progressive flooding.
- A bilge system, with controls integrating the flood detection systems. This is especially useful in case progressive flooding occurs due to internal leaking of a compartment into other tanks or voids, thus triggering the pumping of external water to replace the loss.

Human Factors

Floating platforms are equipped with complex systems and watertight openings (manways, hatches), and these help to manage compartmentalization. These systems are implemented, maintained, and operated by designers and mariners. Errors in design, fabrication, or operation can lead to lack of sufficient stability, loss of watertight integrity, or progressive flooding among other things.

Human errors in design often stem from overlooking or underpredicting critical overturning loads, as well as errors in calculating metacentric height and righting-arm curves. Such errors are mitigated by quality control and quality assurance. Regulatory bodies routinely require third-party reviews, which often entail independent calculation of stability quantities to verify design.

Conversely, safe design practices may enhance stability in operation. Better control of stability conditions may be achieved when appropriate fail-safe approaches are implemented in design. These include the redundancy of control systems or the design of passive control systems. Closed-loop marine systems are passive systems, since they do not allow the intake or egress of ocean water. The use of redundancy or nonreturn valves to close pipes and the use of watertight tank subdivisions are also passive solutions. Additionally, operational contingency plans should be incorporated

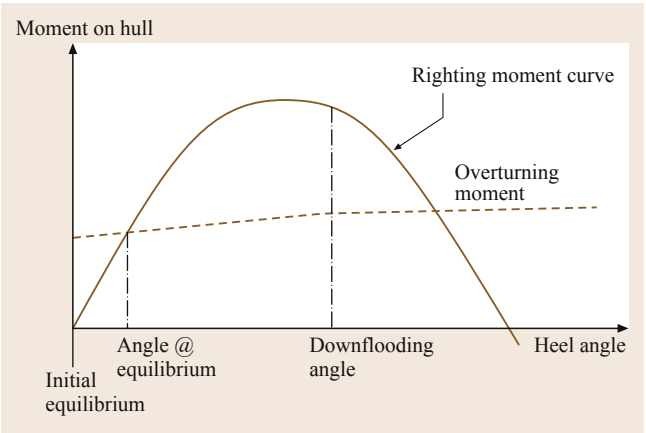


Fig. 34.33 Downflooding angle on a righting-moment curve

in design. In terms of stability, the design of a bilge system allows corrective action to be taken in case of loss of watertight integrity. Control systems shall be designed to identify all failure modes and react appropriately, through alarm systems and direct action.

Fabrication and installation are also critical phases for hull stability. Human errors in bulkhead construction, assembly of piping systems, or even marking of the hull may lead to rapid loss of stability of the hull, once it is put in service. Improperly mounted check valves on an open-loop system is not an uncommon encounter, and may result in progressive flooding. These issues are addressed through system testing, watertight boundary testing, and surveying throughout fabrication.

In operations, the risks to stability are addressed procedurally and manuals detail proper operation of piping systems, including ballast systems, and procedures for opening and closing watertight compartments. Operation manuals also mention acceptable draft limits and live load patterns, and their impact on stability. Inspection and maintenance of watertight boundary, corrosion systems, bilge systems, and flood detection systems should also be part of the overall mitigation plan, to minimize risks of and from external flooding.

34.4.12 Industry Practice

Regulatory

Industry practice is guided by international recommendations, such as the International Code on Intact Stability of the International Maritime Organization (IMO) [34.9]. These IMO guidelines are typically adopted by local regulatory agencies such as the US Coast Guard and by classification societies which adapt them into class rules.

The intent of the rules is to prevent vessels from losing stability leading to the point of capsizing or sink-

ing in any operational condition. They lay a common ground for all vessels at sea. Regulations also enforce a certification process aimed at validating that intended stability properties are met by a given vessel. The process is applied throughout the lifecycle of the vessel. Classification societies typically require third-party design reviews and both class rules and flag authorities enforce a regular cycle of surveys to record actual stability characteristics and ensure conformity with plans. Stability rules are updated regularly to take into account technological changes and learnings from vessel operations and recorded casualties.

Rules are adjusted for different types of vessels, according to their shapes, their operational conditions, static loads and specific sensitivity to dynamic loadings, such as wave loads. They are stated in general terms for ship-shape hulls. Additional considerations are typically made for special applications: passenger vessels, oil tankers, cargo ships, high-speed crafts, fishing vessels, offshore supply vessels, tug boats, and mobile offshore drilling units (MODU) all are subject to special rules. Classification societies and national regulatory bodies may define additional subcategories to refine the criteria further according to the type and application of vessels. MODU rules, from ABS (American Bureau of Shipping) [34.10] or Det Norske Veritas (DNV) [34.11] for instance, both distinguish between self-elevating (jack-up rigs), surface-type units and column-stabilized units (semisubmersible for instance). The variables that require calculation and analysis are the same for all these types. However, the exact criteria are slightly different accounting for differences in shape and typical response to the environment.

Requirements are also adjusted according to the risk to life at sea and to the environment. This is reflected in the difference between stability criteria for otherwise similar vessels. Passenger ships, for instance, must meet additional criteria compared to other ships, which reflect both the specific loading of such vessels and the increased risk to human life. Although the shape and configuration of hulls used for renewable energy applications is often very similar to those in the oil and gas extraction industry, stability guidelines differ significantly, acknowledging the lower risk associated with the former.

Regulations use a few variables to define criteria that apply to a given class of vessel. These variables consist of a number of aforementioned characteristic values for a given hull. In particular, metacentric height, equilibrium heel angle, and righting-arm curves are commonly used to assess stability. For hull types with constant displacement over given operational conditions, the use of righting-moment curves instead of righting-arm curves is widespread. Limits are applied to these basic parameters; they form the set of criteria for a given hull type. The most typical criteria are reviewed in the following section. Although criteria vary slightly from one classification society or national regulator to the other, they are based on similar principles and rely on the same set of variables. Also, international organizations, such as IMO and the International Association of Classification Societies (IACS), have helped significantly to standardize criteria imposed by regulators throughout the world.

First Principles

Stability criteria are designed to ensure that a vessel stays within an acceptable range of heel angles under all operating conditions. Criteria are set by industry guidelines and rely heavily on a few calculated parameters to assess overall stability of a vessel. In most standards, a one size fits all approach is assumed for a given category of vessel. Minimum values are set for all relevant parameters. This is motivated by the difficulty to predict reliably at the design level the destabilizing combinations of loads the vessel will encounter. Standards rely on statistical analysis of similar vessels to define an acceptable envelope of static loading, given the risk associated with the type of vessel and its applications.

The following variables are typically considered to define stability criteria:

- Metacentric height, $\overline{\mathrm{GM}}$, on an upright vessel shall be positive in all operating conditions. This is the minimum stability requirement, since a negative $\overline{\mathrm{GM}}$ means that the vessel righting moment is negative at small heel angles, resulting in excessive list or trim and potentially in capsizing. It is good practice to ensure a minimum positive $\overline{\mathrm{GM}}$ to provide sufficient righting moment to counteract any overturning moment. IMO regulations recommend a minimum $\overline{\mathrm{GM}}$ of 0.15 m for all vessels. This number is derived from statistical analysis of casualties and includes a minimum safety factor to account for errors in calculating the metacentric height. Industry standards may fix larger values depending on the type of vessel.
- Criteria are also applied to the righting-arm curve (since it represents the reserve righting capacity of a heeling vessel). The righting curve of an intact vessel is expected to increase at low heel angles, and up to an acceptable value (called α herein). IMO recommends that vessels have an increasing righting arm up to $\alpha = 25°$ of heeling; this is adjusted for special vessels. This is ensured by verifying that the point at which the righting arm starts decreasing is larger than α, as illustrated in Fig. 34.34. The

righting arm needs to remain positive over an extensive range of heel angles.

- To guarantee that there is sufficient reserve stability to right the vessel, a criterion should also be applied to energy. It combines the amplitude of the righting curve with the extent of acceptable heel angles to assess the stability of the vessel. This is typically achieved by requiring a minimum value λ be attained for the integral under the righting arm or righting-moment curve as a function of heel, from 0° heel to a limiting heel angle (Fig. 34.35). The limiting angle is defined according to the characteristics of the vessel. It is typically the lesser of the downflooding angle and the angle at which the righting arm becomes negative. Where weather and operating conditions are considered, it may also involve an angle of equilibrium between the righting moment and the overturning moment due to external loading. This is explained further below.
- The effect of weather is typically considered by calculating the overturning moment due to external loads, especially wind. Industry rules provide guidance on the minimum wind speed to consider, as well as on the method to calculate the overturning arm. These methods vary with the type and applications of the vessel. In general the overturning moment can be calculated by

$$M_O = \boldsymbol{OP} \times \boldsymbol{F} \,, \tag{34.47}$$

where $\boldsymbol{P}$ is the center of application of the horizontal force $\boldsymbol{F}$ and $\boldsymbol{O}$ is the center of rotation.

It is common to require that the overturning moment arm $\boldsymbol{OP}$ be taken as the vertical distance between the center of wind pressure and the center of lateral pressure on the submerged hull since there are uncertainties on the exact position of the center of rotation of the vessel. The applied force $\boldsymbol{F}$ should represent extreme static overturning loads acting on the vessel. As a minimum, it should include the effect of wind, calculated as the product of the lateral wind pressure by the projected wind-exposed surface area, or windage area. The windage area may vary with the heel angle.

Stability analysis of floating offshore platforms is usually performed by comparison of the moment curves. Most ship-shaped vessels will experience the largest overturning moment with a wind direction associated with the lowest righting moment. Indeed, the longitudinal windage area is larger than the transverse windage area on ship-shaped vessels leading to larger rolling moment, while these vessels are most sensitive to rolling, due to lower transverse metacentric height. However, in some cases, alternative directions may be associated with lower stability, due to different loading patterns, involving current for instance, or when the vessel is not ship-shaped. On semisubmersible and spar platforms, all directions of the heel angle must be studied, to determine the worse one.

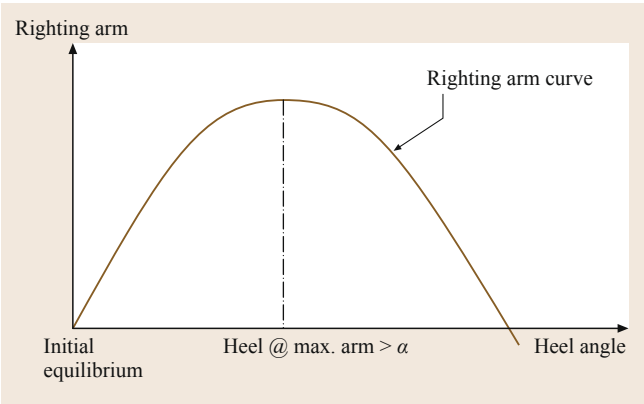


Fig. 34.34 Illustration of a righting arm criterion for stability

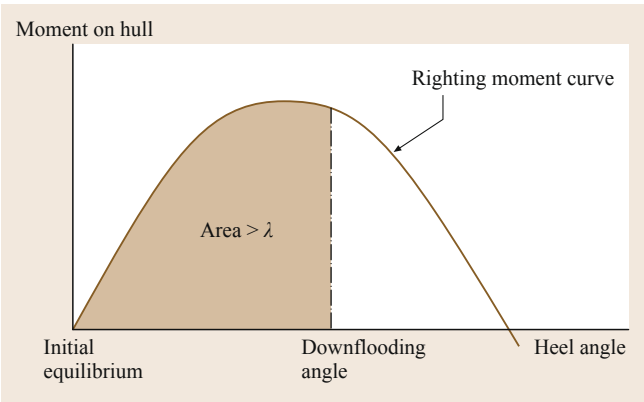


Fig. 34.35 Illustration of an energy criterion for stability

There exists no criteria on the overturning moment alone. Its adequacy with stability criteria is determined by comparison with the righting moment. This comparison must be made for all relevant directions of the loading and heel angle.

In comparing the overturning moment and the righting moment, two sets of variables are of interest:

- The angle of equilibrium of the vessel subject to the overturning moment. There are two such angles of particular interest, referred to as the first intercept and the second intercept. The first intercept is the first angle of equilibrium, where the righting-moment curve crosses the overturning moment curve. It is also the limit of watertightness. The downflooding angle may be smaller than the angle at the second intercept. However it must be sufficiently larger than

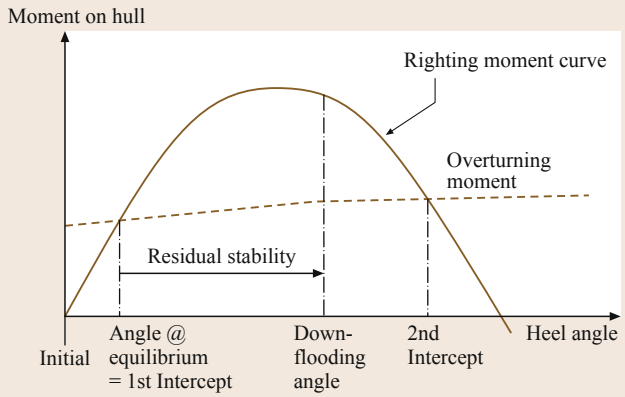


Fig. 34.36 Righting and overturning moment intercepts

the angle at the first intercept to provide reserve stability for dynamic loads. These variables are illustrated in Fig. 34.36.

- The ratio of righting energy to overturning energy can also be used to assess the extent of stability. This criterion usually considers the energy up to the downflooding angle – meaning within the extent of weathertightness.

Usually, wave dynamics is not directly accounted for in the basic stability criteria. The overturning moment is meant to capture the effect of extreme static loads on the vessel. Rather, waves and other dynamic effects, such as wind turbulence, are addressed by applying safety factors on the variables above. Ratios of righting energy versus overturning energy within the extent of weathertightness must be large enough (1.4 for general MODUs for instance) to allow for a temporary increase in overturning moment.

Alternative Performance Standards

Stability criteria cover very broad ranges of vessels. In some cases, meeting the regulatory requirements may be overly conservative or inadequate. This is especially true of novel concepts, which scarcely fall into any predefined category.

In such cases, alternative criteria may be derived. The criteria must consider all the variables defined above. However, rather than using standard formulas to account for dynamic effects and assess stability, criteria may be derived from vessel-specific data. Alternative variables may be defined from stability parameters, if deemed more appropriate. For instance, criteria on the ratio of righting to overturning energy may be replaced by criteria on a ratio of reserve righting energy to initial righting energy.

When alternative criteria are derived, the uncertainties on loadings as well as adequate margins of error on the vessel characteristics and righting capacity must be considered.

A common alternative criterion considers the specific dynamic response of a vessel rather than using the standard dynamic heel angle. Local metocean data, rather than broader wind speed, may also be accepted by local authorities. This is especially useful for vessels with restricted operational envelopes. Numerical analysis may be used to demonstrate the validity of the predicted alternative response of the platform. Regulators will often require that calculations are backed up by model tests and wind tunnel tests, to justify drag coefficients and numerical parameters.

34.5 Loads

We have mentioned earlier that the righting moment needs to be compared to any external moments that may try to overturn a floating structure. Some kind of an upsetting external force or moment may act on the structure, and this may change the equilibrium position either temporarily of permanently. These forces may be due to environmental factors, like wind, or to activities on board the vessel (lifting of a weight overside), or to accidental or operational weight shifts. Common loads and their effect on the stability of offshore platforms are briefly discussed in this section.

34.5.1 Environmental Loads

Environmental loads need to be considered in the derivation of righting- and overturning-moment curves. These loads may be adjusted according to the applications and areas of operation of the vessel. In particular, the following effects should be considered.

Wind Loads

Wind pressure exerts a force which is mostly concentrated on the superstructure since wind velocity grows exponentially as we are farther above the still-water level, and thus causes a moment which tries to heel or trim the structure from its upright position. Wind force is proportional to wind speed squared times the drag coefficient times the projected surface area of the ship exposed to the wind action, that is, above the SWL, and including the superstructure. The only opposing force is the viscous drag-force or water-resistance force for a nonconstrained body (a body which is not moored).

The equilibrium is reached when the moment due to the weight and buoyancy of the ship equals the moment developed by the wind force. The destabilizing effect of wind on the vessel is considered by calculating the overturning moment on the topsides.

The general formula for the wind force applied on the vessel is given by

$$F = \int_{\substack{\text{Windage}\\ \text{Area}}} p \, \mathrm{d}A \,, \tag{34.48}$$

where p is the wind pressure. Depending on the type of vessel, p may be fixed by stability standards, or may be recalculated. IMO provides a minimum pressure of 756 Pa (of which 504 Pa corresponds to steady winds) for general applications. It is typically calculated based on a drag formula

$$p = \frac{1}{2} C_\mathrm{d} \, \rho_\mathrm{air} \, U^2 \,, \tag{34.49}$$

where C_d is the drag coefficient and U is the extreme wind velocity. The drag coefficient C_d may vary for different types of structures on deck, according to their shapes, and their orientation to the wind direction. For MODUs, a set of accepted values of drag coefficients is commonly used in the pressure calculations and generally accepted by regulatory bodies. The C_d of extraneous shapes must be determined by model tests.

The wind velocity may be fixed by the standard. For intact stability, it is typically higher than 50 kn. This wind velocity is commonly increased to 100 kn for intact stability of permanently moored structures.

Alternative wind speeds may be considered acceptable in special cases, based on local meteorological data and specific applications such as renewable energy applications. Adequate statistical analysis must be run to determine the extreme wind speed. The effect of wind gust must also be included in the local wind assessment.

Also, extreme wind speed is often expressed in terms of the wind speed at 10 m height above the mean waterline. For a vessel with a windage center of area higher than 10 m, the effect of shear in the wind should be considered. Formulas for wind shear calculations are provided in many standards.

Current Loads

Current loads may be detrimental to stability, especially if it is foreseen that they will act in opposite direction to wind effects. They should be taken into account where the extreme wind speed is based on meteorological data. They can be calculated based on drag force calculations on the wetted surface of the vessel. When deemed necessary, the effect of current loads should be added to the overturning moment to assess stability.

Wave Loads

Wave loads are dynamic. They are addressed by enforcing appropriate safety factors that provide reserve stability to withstand these dynamic loads and induced motions (rolling and pitching). In some cases, higher order wave loads, such as slow drift, combined with station-keeping of a vessel, may indirectly result in trimming or listing. This may happen in the case of a moored vessel, for instance. The mooring resisting the drift motion may pull the vessel down on one side. These detrimental mooring effects should be taken into account in the stability analysis. They are addressed in the description of operational loads below.

Ice Loads

The effect of icing on stability is associated with ice accretion on topsides of a vessel. The built-up of an ice layer on deck results in a change in the weight properties of the vessel. In particular, the formation of an ice sheet on deck may affect the following:

- The overall displacement of the vessel (change of mass). This should translate in an adjustment of the righting-moment curve.
- The location of the center of gravity. Ice on deck will result in a higher center of gravity, which has a destabilizing effect on the vessel. But the ice sheet may not be uniformly distributed, which could result in a horizontal shift of the center of gravity as well, leading to the list and/or trim of the hull. This effect should again be associated with a correction on the righting-arm curve.
- In some cases, ice accretion could increase the windage area of the vessel, thus affecting the overturning-moment curves.

The effect of ice should be included as a variation on the stability conditions. However, it will have a larger effect on small vessels, such as small passenger vessels, and on vessels with large deck area to displacement ratios (ship-shaped platforms).

When ice is considered, the maximum ice thickness and distribution patterns are generally defined according to local standards for the area of operation. The effect of ice accretion should be considered for all vessels operating in the Arctic ocean, in the Bering Sea, and close to the Antarctica in the Southern Ocean. The IMO stability code defines minimum icing layer thickness to be considered, as well as the exact location where icing occurs. However, local authorities may apply different standards, based on local meteorological knowledge.

Vessels which are foreseen to operate in icing regions should be designed to minimize the accumulation of ice on topsides.

Water Retention from Wave Splashing and Rain

If water accumulates on deck, subject to heavy rain or wave splashing or for some other reason, stability of a vessel may be affected. The effect is similar to that of ice loads: the mass properties of the vessel should be adjusted to account for the accumulation of water. The center of gravity can be shifted upward or sideways, and the displacement may increase, thus leading to a loss of righting arm. Moreover, the effective $\overline{GM}$ may reduce due to a partial loss in the waterplane area.

The topsides of a vessel should be designed with adequate draining system to minimize the collection of water on deck.

34.5.2 Dead Loads

Dead loads correspond to the weight of permanently installed structures and equipment on the vessel. The following elements shall be included in the inventory of dead loads:

- Lightship weight from structure and permanently installed systems, as of the end of construction.
- Any permanently installed topsides.
- The weight of any permanent ballast, and its effect on the location of the center of gravity.
- The equivalent weight from mooring pretension when applicable. This is especially important in cases where mooring may have a destabilizing effect on the vessel. When mooring has a stabilizing effect, a mooring damage case may be recommended to ensure that the vessel would remain stable in the event of a broken line.

34.5.3 Live Loads

Live loads are the weight of the structure hull and its components that may vary and move throughout the life of it. If a weight which is already on board is moved, for example, transversely, to a new location, the center of gravity of the structure will also move and the platform will heel. As a result, the center of buoyancy will shift and a new equilibrium will be reached when the two centers are on the same vertical line. The longitudinal movement of a weight which is already on board will also cause the structure's center of gravity to move longitudinally, resulting in the trim of it. Live loads include the following:

- Ballast: During a vessel's operation, ballast water may be pumped from seawater, moved between tanks and ejected to the ocean. These changes of configuration of ballast affect the mass properties of the vessel. The displacement may change as ballast water is pumped in and out. The center of gravity of the vessel shifts. All these effects must be taken into account in the calculation of the righting-moment curves. Ballast tanks should be clearly identified and extreme loading situations should be described and used for stability analysis. The stability of intermediate ballast configurations should also be investigated.
  Additionally, particular attention needs to be paid to the effect of the freely moving water surface in partially filled tanks. In such tanks, heeling of the vessel will displace liquid toward the heeling side, thus shifting the center of gravity further to that side, and potentially increasing the heel. This is referred to as the free-surface effect. For deep tanks (tanks with large depth to breadth ratios), it can be accounted for in the stability analysis by calculating the drop of metacentric height due to the presence of the free surface (Sect. 34.4.5). For shallower tanks, or tanks where the moving water may reach the top of the tank, the righting-arm correction may have to be recalculated at each heel angle. For rectangular tanks, and at typical filling ratios (50%, 95%, and 98%), parametric tables are usually available. Such tables may be found, for example, in Lewis' principles of naval architecture [34.5].
- Cargo and passenger loads. These loads encompass all types of cargos, from liquid oil and gas in tanks, to bulk cargo, and including containers, cars, and people. For liquid cargo, the detrimental effect of free surfaces shall be included in the stability analysis, by correction of the righting arm, as described for ballast tank above. The destabilizing effect of freely moving loads, such as grains and gravel in bulk carrier, or suspended loads (fish nets on fishing vessels, or load on a crane) shall be considered as well. For cargo and passengers, an envelope of allowable live loads must be derived, with associated displacement and center of gravity. Stability must be ensured within the envelope of loads. International and national standards typically set minimum requirements of magnitude and distribution on cargo and passenger loads, according to the type and application of a vessel. General guidelines may be found in IMO stability code [34.9].

34.5.4 Operational Loads

Operational loads correspond to additional loads related to the operation of the vessel. They stem from passive and active control systems, as well as from the use of equipment on the vessels. These loads must be included in the analysis if they have a detrimental ef-

fect on stability. Stabilizing effects may be included as well, provided that there is enough redundancy in the controls and systems involved in starting and maintaining the load on the vessel. Operational loads include the following:

- Propulsion and dynamic positioning loads from propellers or thrusters, as well as loads from antirolling device. Using the stabilizing effect of the latter in analysis is only allowed in cases where the antirolling system is equipped with sufficient levels of redundancy.
- Detrimental mooring loads due to drift should be considered in the stability analysis as well.
- Overside weight lifting: When a weight is lifted over the side of a structure, the center of gravity will move up and toward the lifted weight. As a result, the structure will heel/trim until the new center of buoyancy and the new center of gravity are on the same vertical line.
- Additionally, any load associated with the functionality of the vessel should be investigated. Offshore wind turbines, for instance, have special overturning wind loads due to the thrust on the turbine. Similarly, wave energy converters may experience increased static and dynamic loads from electricity generators. MODUs and oil production platforms, under operation are under loading from well-head activities or riser connection. The effect of umbilical and riser connection to a platform is of special interest to stability because it may affect the platform displacement and effective center of gravity, much as mooring pretension does.

34.5.5 Accidental Loads

Accidental loads result from incidents which extend beyond the operational envelope of the platforms. They include loads from collision, grounding, loss of cargo, etc. The static forces resulting from such incidents may also be considered in stability analysis:

- Grounding: If a vessel runs aground, there will be a reaction force applied to the ship at the point of grounding. This force produces a list and/or trim moment. Recalling the vertical equilibrium of forces due to weight and buoyancy in the floating structure case, we must then include this new reaction force due to grounding in the vertical equilibrium equation such that the weight of the vessel is now equal to the reaction force plus the buoyancy force. Since the weight of the vessel remains the same, the buoyancy force must be less, the vessel rises (parallel) as a result of grounding as expected.

34.6 Lightship Parameters

34.6.1 Technical and Commercial Significance

As discussed in previous sections, the stability assessment of a vessel depends heavily on the ability of the designers and operators to predict and control the displacement and position of the center of gravity accurately. This is possible if the mass properties of the vessel are known. This means, in particular, the following:

- The cargo, passenger, and operating loads on the vessel are well understood and may be calculated with acceptable precision. Typically, designers make assumptions, which are stated in the operational manual of the vessel. Operators are responsible to control their cargo and passenger loads to remain within this envelope.
- The weight and weight distribution of the hull structure, with all permanent systems, must be calculated adequately. These parameters are used throughout design to assess stability.

Operators later have access to lightship parameters stated in the operational manual. These parameters can be used to recalculate the exact stability characteristics of a vessel when combined with known cargo conditions.

To ensure a consistently accurate projection of lightship parameters, regulatory bodies have set procedures and milestones.

34.6.2 Regulatory Approach

Regulatory bodies rely on reviews and surveys during the lifecycle of the vessel to ascertain that stated lightship parameters are accurate for a given vessel. Three phases are of particular importance:

- Reviews at the design level focus on regulatory conformity of plans and design calculations. In particular, stability calculations and plans of watertight and weathertight integrity as well as compartment plans are reviewed. Lightship parameters are predicted at this stage.

- During fabrication, surveys and inspections ensure the consistency between the plans and the erected hull. They also witness the testing of the hull to verify the lightship parameters. This is the determination phase.
- The operating vessel is surveyed on a regular basis throughout its operating lifetime to verify continued adequacy with the plans. This is the control phase.

Completion of each phase results in the issuance of a certificate by the regulatory body, to certify the adequacy of the vessel design, construction, and operation with national, international, or class regulations.

34.6.3 Prediction

During design and construction, lightship weight cannot be directly measured. Instead they must be calculated from projected weight distribution on the hull.

This is achieved during careful weight management. All components of the ship, their weight, and location are recorded and the lightship weight and center of gravity are calculated by summation of all parts. Construction plans must be updated constantly with any design modification or construction adjustment, and the effect on weight recorded.

Modern software, using 3-D modeling and extensive database, typically provide very accurate calculations of weight and center of gravity.

34.6.4 Determination

Lightship parameters are also measured directly using surveys of the vessel throughout construction and at the time of loadout. In particular, the following tests and inspections are carried out:

- A survey of draft marks to confirm their conformity with design drawings and regulatory requirements. Draft marks are positioned around the hull to allow reading of the draft of the vessel at any load line. They may also be used to determine the trim and list of the hull, if placed at extremities of the hull.
- Other elements of stability are surveyed during construction. In particular, watertightness is typically ensured by airtightness testing on watertight bulkheads. Hull openings are inspected for watertightness and weathertightness depending on their location. Tank sizes and bulkhead dimensions are also surveyed to ensure that the buoyancy and tank dimensions are in accordance with engineering design.
- At the time of load out, the weight of the hull, free of any cargo or operational load, may be verified by reading the draft using the marks. This reading is compared to the predicted draft of the lightship, which confirms the weight predictions of the hull free of cargo. This is the lightship weight.
- Finally, an inclining experiment is often requested to confirm the weight distribution of the lightship hull and the exact location of the center of gravity of the structure. It is easy to see that taking moments of every bit of weight in the *light* structure to determine the position of the center of gravity is a very tedious task (but of course it can be done). On the other hand, $\overline{\mathrm{KG}}$ can be found quite accurately by conducting inclining experiments on the *light* structure (after it is launched or in service). An inclining experiment consists of loading the vessel with a known mass (typically weights on deck, or if allowed ballast water in tanks) offset from the center. The heel angle of the hull is recorded during the test, allowing accurate calculation of the position of the center of gravity. A pendulum (plumb line) is used during the inclining experiments as shown in Fig. 34.37. Setting the list moment equal to the moment due to the weight shifted, we must have (so that the structure is upright)

$$w\,\overline{\mathrm{g_0g_1}} = \varphi\, W\left(\overline{\mathrm{KB}} + \overline{\mathrm{BM}} - \overline{\mathrm{KG}}\right) \quad \text{or}$$

$$\overline{\mathrm{KG}} = \overline{\mathrm{KB}} + \overline{\mathrm{BM}} - \frac{w\,\overline{\mathrm{g_0g_1}}}{W\,\varphi}\,. \tag{34.50}$$

The quantities $\overline{\mathrm{KB}}$ and $\overline{\mathrm{BM}}$ can be determined from the hydrostatic curves for the structure, as well as the weight from $W = \Delta$. Then, once φ is measured (since the length of the pendulum and the deflection are known by measurements), one can calculate $\overline{\mathrm{KG}}$ from (34.50) quite accurately. The initial metacentric height can also be obtained from these experiments. One would normally use several weights shifted transversely for greater accuracy.

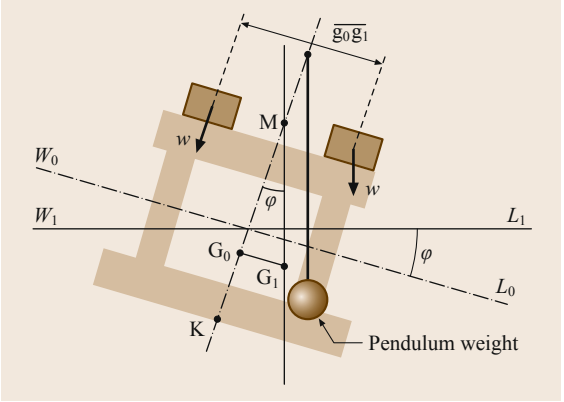


Fig. 34.37 Inclining experiments done by use of a pendulum and shifting a weight

There are many important conditions that must be satisfied during the inclining experiments, that is, all weights onboard must be secured or no (or very small) wind present, and these conditions are specified by the certification agencies and must be strictly followed. The calculated center of gravity is compared with predicted values from design and construction drawings. For an inclining experiment to be successful, the difference between the predicted and measured location is expected to be within a few percent. These experiments may also be conducted during marine surveys requested by the owner, regulatory agencies or insurance companies, after for example, the structure is overhauled.

Local authorities and classification agencies witness these surveys and ensure that they follow approved procedures. Such procedures are typically set by international standards, and adjusted for each type of vessel. Standards determine the conditions of the vessel before the tests, acceptable means of measurements of lightship weight, and inclination and extent of the inclining experiment. The results of the lightship weight measurement and inclining experiment are submitted to relevant authorities and to the class surveyor where applicable. They are recorded in the vessel's manuals, and may be used for further stability calculations.

In cases where identical vessels have been built and the lightship weight is confirmed within an acceptable margin, local authorities may waive the need for an inclining experiment. Upon successful completion of the lightship weight and inclining experiment, a certificate is provided to the vessel's owner.

34.6.5 Control

Lightship parameters are controlled throughout the life of the ship. Most types of vessels undergo regular inspection, at least every 5 years. At a minimum, the lightship weight may be verified at such inspection by survey of the draft. The measured draft should correspond to the predicted load line according to the given cargo conditions.

Retrofit or modifications of the hull usually triggers recertification of the lightship parameters.

34.7 Subdivision

34.7.1 Purpose and Criteria

Offshore platforms are susceptible to flooding due to collision with boats, such as supply vessels, and due to cracks in the structure from corrosion or environmental loads. The latter may be adverted by regular inspection. Procedures mitigate the risks of collision. Yet, for maximum reliability at sea, vessels are designed to retain sufficient stability in the event of flooding of a compartment. This is typically ensured by subdivision of the vessel.

Offshore platforms, like most commercial ships, are divided internally in compartments. These compartments, whether they are voids, tanks for ballast or cargo, machinery compartments or living quarters, are arranged to prevent the spreading of flooding throughout the hull. They are insulated from each other by watertight boundaries (bulkheads, decks, or flats). Openings between compartments are provided with means of closure to prevent progressive flooding, which must develop the same watertightness characteristics as the boundary they cross.

If damaged stability is considered, the extent of watertightness is defined as the worse of the extreme waterline defined in intact stability and the final damaged waterline. Watertight boundaries are designed to withstand water pressure due to a head to the damaged waterline.

The purpose of damaged stability analysis is to ensure that an arrangement of compartments within a vessel provides sufficient subdivision, that is, that the vessel does not capsize or sink with one compartment flooded, and that progressive flooding is avoided.

34.7.2 Regulatory Requirements

Most classes of offshore vessels are subject to regulatory requirements of damaged stability by local authorities and classification companies. The requirements depend, however, on the risk to life at sea and to the environment posed by the potential loss of the vessel.

Within a vessel, a distinction is often made between those compartments wholly or partially adjacent to sea at operating draft and in operating trim conditions and other compartments.

The most stringent requirements apply to compartments which are associated with collision damage. Most rules require that any internal bulkhead or flat within a defined extent be assumed damaged so that consecutive compartments separated by these bulkheads and flats be flooded simultaneously. The extent

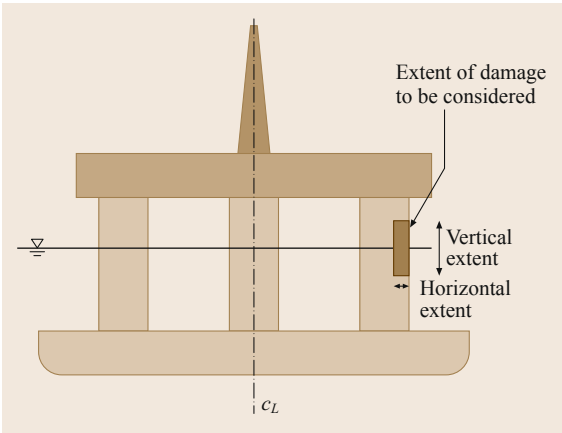


Fig. 34.38 Illustration of the extent of damage on a semisubmersible hull

of damage is defined horizontally by a depth of penetration and vertically by a vertical extent of damage. These variables are identified in Fig. 34.38. The values of depth of penetration and vertical extent vary by type of platform and function. Most offshore platforms are designed with a 1.5 −m depth horizontal penetration (see ABS MODU rules [34.10], and DNV Offshore Standard C301 [34.11]). Consequently, watertight bulkheads are usually located farther than 1.5 m from the external hull for maximum efficiency in design.

On column-stabilized units and deep-draft platforms, such as spars, considerations of damage may be limited to compartments within a predefined vertical extent, in the region of the waterline. The collision damage may also be limited to outer portions of the hull, disregarding those parts of columns and hulls which are unlikely to be hit by approaching vessels. However, all possible transit and operating waterlines are considered. For other types of platforms, such as self-elevated platforms, or ship-shape hulls, all compartments adjacent to sea from the keel upward may be considered damaged.

Additionally, depending on the type of units, a damage stability criterion may also apply to all compartments, regardless of their locations, or, in the case of column-stabilized units, to all compartments wholly or partially below the operating waterline. Flooding of machinery compartments, regardless of their locations, may also be considered.

Ultimately, the extent of damage to be considered in the analysis is determined by class regulation and local and international regulations based on the risk associated with the vessel. For instance, column-stabilized units designed as offshore wind turbine installations are not required to meet damage stability requirements for collision damage, unlike MODUs, due to the low risk of collision in absence of large supply vessels.

A designer should determine which compartments may be flooded for damaged stability purposes according to the following:

- The type of vessel to be considered (ship-shape, self-elevated unit, deep-draft platform, etc.)
- The function of the vessel (MODU, accommodations, renewable energy production, etc.)
- The body of regulations which apply to the design. If the unit is expected to be classed, class rules by a given classification society should apply. Local authorities, such as US Coast Guard, also may have specific requirements.

34.7.3 Damaged Stability and Residual Stability

Damaged stability criteria depend on the calculation of stability characteristics in a damaged condition. These calculations are very similar to stability calculations in intact conditions. As for the intact stability, righting-moment and overturning-moment curves are calculated to assess damaged stability. However, the inclination angle is usually defined with reference to the equilibrium at rest of the intact hull. Two methods are used to determine the righting-moment curves:

- Lost-buoyancy method: In the first method, called the lost-buoyancy method, we consider the total weight (or displacement) and center of gravity of the ship to remain constant before and after it is damaged. The buoyancy is lost in that part of the flooded compartment below W_0L_0, and equal amount of buoyancy is added between W_0L_0 and W_1L_1. This method is also called the constant-buoyancy method (which is a more appropriate terminology). In this case, if the flooded compartment was free of liquid cargo or ballast, the mass properties of the vessel are unchanged.
  The waterline and center of buoyancy will be modified to account for the lost buoyancy at operating draft. Due to flooding, the hull will sink to recover the displacement of the platform as determined by the mass and additional static forces (mooring pretension, etc.). The adjusted waterline and center of buoyancy may be used to calculate the damaged draft and damaged inclination of the platform. The inclination is calculated with respect to the intact even-keel position. The waterplane area may also be affected if the flooded compartment crosses the waterplane area at any relevant heel angle, causing a loss of waterplane area of the hull.
  These losses of buoyancy in the compartment and waterplane area result in a loss of metacentric height. Righting arm and righting-moment curves

Table 34.2 Comparison of the lost- (or constant-) buoyancy method (LBM; CBM) versus added-weight method (AWM) of damaged-stability calculations showing when the physical quantities remain the same or are different, before and after bilging occurs

Quantity	CBM (LBM)	AWM
Δ, ∇	Same	Different
Center of Flotation	Different	Same
I_L , I_T	Different[a]	Same
$\overline{BM}_T$, $\overline{BM}_L$	Different	Different
$\overline{GM}_T$, $\overline{GM}_L$	Different	Different
$\overline{KG}$	Same	Different
$\overline{KB}$	Different	Different
MCTC, MCT1″	Different	Same
TPC, TPI	Different	Same

[a]Unless there are horizontal water tight bulkheads (or flats) above the damaged compartment (but below the intact waterline) that would keep the waterplane area intact after bilging occurred.

are calculated similar to the intact stability case. At zero-degree trim, which may correspond to the intact inclination angle, the damaged righting arm is negative. The hull recovers stability at an equilibrium point which identifies the damage inclination angle.

- Added-weight method: In the second method, called the added-weight method, we assume that the structure remains completely intact after flooding, although its equilibrium position changes. We do this by assuming that a *weight* is added to the structure. The weight is the weight of water that enters one or more of the compartments. The problem here is that the weight is not known since it depends on the final waterline that must be determined as part of the solution. This then suggests an iterative solution procedure when we consider the flooding of a compartment. In this case, the hull volume, center of buoyancy, and waterplane area are the same as in intact stability.
  The displacement and center of gravity are adjusted to reflect the new mass properties. The damaged waterline is calculated and this results in an adjusted center of buoyancy location. The damaged inclination angle is also calculated based on the equilibrium between adjusted buoyancy and adjusted weight.

Table 34.2 lists which quantities are different or remain the same in the use of the two different methods before and after damage occurs

The righting-moment curves can be calculated for any damaged compartment as they were for any intact load line following one of the approaches described above. It may be corrected for internal free surfaces. The righting moment may be compared to an overturning moment from static external forces. The overturning moment is typically associated with a mean wind speed corresponding to a storm of lesser intensity than those considered in intact stability.

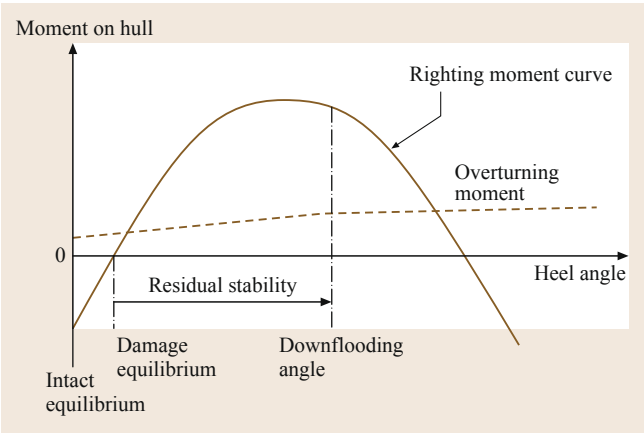


Fig. 34.39 Righting-moment curve after damage has occurred (0° heel corresponds to the equilibrium position of the intact hull)

For damaged stability to be deemed sufficient, a number of criteria may be applied. The major variables at stake are shown in Fig. 34.39:

- Equilibrium in damaged conditions: An analysis of the righting-moment curve provides basic information. Although the righting moment is negative at 0° inclination, the damaged inclination, where the moment becomes positive again, should be significantly lower than the downflooding angle. Some

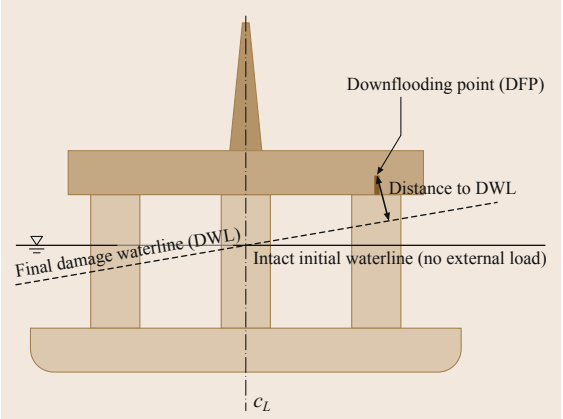


Fig. 34.40 Stability criteria associated with residual stability of column-stabilized vessels: minimum distance between final waterline (including overturning moment) and downflooding point

design codes require that the righting moment remain positive for a predefined range of trim angle beyond the damaged inclination angle.

- Residual stability: Additional criteria are usually applied to verify that the hull has sufficient residual stability. Residual stability may be defined as the range of heel angles at which the hull righting moment remains sufficient (positive or greater than applied overturning moment, depending on the regulatory authorities and type of vessels). For column-stabilized units for instance, a first and second intercept between the righting-moment curve and the overturning moment curve is defined. The residual stability in this case is defined as the difference of the heel angles between the first intercept and the downflooding angle or the second intercept. It is sometimes enforced by verifying that the downflooding point is positioned sufficiently far from the final damaged waterline at the first intercept. The final damage waterline, shown on a semisubmersible in Fig. 34.40, is defined as the maximum angle at equilibrium of the damaged hull under static-load conditions. More generally, residual stability is often defined as the range of heel angles over which the righting-moment curve of the vessel is positive.

34.8 Analysis

34.8.1 Static- and Quasi-Static Analysis

Hydrostatic analysis relies on the definition of the external boundaries of the hull combined with either an even-keel displacement or draft, and the position of the center of gravity of the hull. Designers may use commercial tools or do hand calculations to determine the corresponding center of buoyancy, initial metacentric height, and righting-arm curves based on this data, and using the methods presented herein.

For complex geometries, designers often rely on commercial software, which can update stability properties for each inclination angle, accounting for geometric discontinuities in the submerged hull.

Similarly, simple models are used to calculate the overturning moment. The wind-overturning moments may be calculated based on numerical determination of the windage coefficient according to the wind direction. These may be calculated by hand from the projection of the wind expose-area in any direction. Alternatively, for complex topside geometries, numerical programs and commercial software may be used.

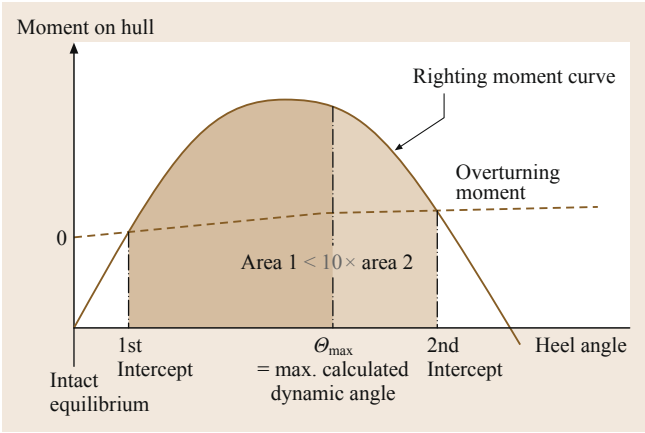


Fig. 34.41 Dynamic angle stability criteria

34.8.2 Dynamic-Response-Based Analysis

A common alternative stability criterion involves the response-based analysis. With the rising accuracy and availability of numerical analysis, it has become easier to demonstrate the divergence between the one-size-fits-all rules applied to entire categories of vessels, and the behavior of specific designs. Some categories of vessels are more design-sensitive than others. This is especially true of semisubmersibles within the MODU category. Column-stabilized semisubmersibles are less constrained than other MODUs, and design strategy influences significantly the response in waves of the platform. Traditional stability criteria, however, would not capture the better dynamic performance of a platform, penalizing more dynamically stable semisubmersible designs.

As an alternative, stability criteria which take into account the actual dynamic response of a vessel, rather than projected averages, have been derived. Some codes published by classification companies provide detailed methods to carry out the dynamic-response-based analysis. This is the case in [34.10] for column-stabilized mobile offshore drilling units. When a dynamic-response-based approach is adopted, the constant ratio of the righting moment to overturning moment is abandoned and replaced by the following considerations:

- A dynamic-angle criteria, to verify that there is sufficient reserve stability between the maximum predicted dynamic inclination angle and the second intercept with the overturning moment. This is typically achieved by ensuring that the energy to incline

between the maximum dynamic angle and the second intercept is at least 10% of the energy to incline between the first intercept and the maximum dynamic angle, as illustrated in Fig. 34.41. It ensures that design accounts for uncertainties in loadings and dynamic response calculations.

- A downflooding criteria, which ensures sufficient clearance between the maximum wave crest and the downflooding point, in reduced wind speed. The initial downflooding distance between the mean waterline and the downflooding point is compared to the predicted reduction in downflooding distance. The latter is computed according to preset formulas provided in the IMO stability code, and in many classification societies' guidelines. The formula includes safety factors and takes into account the relative motions of the platform.

These criteria rely on the determination of dynamic motions of the vessels. Two approaches are typically acceptable for this. Designers may use empirical formulas based on the vessel's geometric characteristics and mass properties. Alternatively, a full-fledged numerical analysis may be used. If so, the numerical model must account for linear and nonlinear loadings on the vessel, and, if calibration is necessary, model test results comparing the numerical model to measured data may be required.

References

34.1 P. Bouguer: *Traité du Navire* (Research Publications, New Haven 1746), Microfilm

34.2 B. Padmanabhan, R.C. Ertekin: Setdown of a catenary-moored gravity platform, Mar. Struct. **9**(7), 721–742 (1996)

34.3 K.J. Rawson, E.C. Tupper: *Basic Ship Theory*, Vol. 1, 3rd edn. (Longman, London 1983)

34.4 W. Muckle: *Muckle's Naval Architecture*, Marine Engineering Series, 2nd edn. (Butterworths, London 1987)

34.5 E.V. Lewis: *Principles of Naval Architecture: Stability and Strength*, Vol. 1 (Society of Naval Architects and Marine Engineers, Alexandria 1988)

34.6 D.R. Derrett: *Ship Stability for Masters and Mates*, 5th edn. (Butterworth-Heinemann, Oxford 1999)

34.7 A. Biran: *Ship Hydrostatics and Stability*, 1st edn. (Butterworth-Heinemann, Amsterdam 2003)

34.8 O.M. Faltinsen: *Sea Loads on Ships and Offshore Structures*, Ocean Technology Series (Cambridge Univ. Press, Cambridge 1990)

34.9 I.M. Organization: *The International Code on Intact Stability* (International Maritime Organization, London 2009)

34.10 A.Bo. Shipping: *Rules for Building and Classing Mobile Offshore Drilling Units – Part 3: Hull Construction and Equipment* (American Bureau of Shipping, Houston 2014)

34.11 D.N. Veritas: *Offshore Standard: Stability and Watertight Integrity* (Det Norske Veritas AS, Nøvik 2013)

35. Wave, Current and Wind Loads

R. Cengiz Ertekin, George Rodenbusch

This chapter describes wave, current, and wind loads on fixed or floating offshore platforms. Both linear and nonlinear waves are discussed in deterministic and irregular seas. Linear waves are written as a subset of the more general wave theory based on the perturbation method. Nonlinear waves include Stokes waves in deep waters and cnoidal and solitary waves in shallow waters. Wave loads on both large and slender structures are formulated, and solution methods, such as the Green function method, are introduced. For large structures, linear potential theory is formulated in the frequency domain. However, time-domain methods and drift loads are also discussed. For slender structures, Morison's equation and the associated drag and inertia coefficients are introduced.

These are followed by wave–current interaction, many types of uniform and nonuniform currents, wave–current kinematics, and current-induced forces, as well as vortex-induced vibrations. A number of important quantities, such as the Doppler shift, velocity estimation through the power law, lift and drag coefficients are also introduced.

Wind forces on offshore structures are discussed through both the steady and unsteady wind profiles and forces, and through spectral analysis. Other considerations include sections on model tests and similarity laws and how various physical quantities can be scaled to prototype, both commercial and open-source computational fluid dynamics (CFD) tools, and extreme response estimation.

35.1 **Wave Loads** 787
35.1.1 Linear Waves 787
35.1.2 Nonlinear Waves 790
35.1.3 Shallow-Water Waves 791
35.1.4 Random Waves 791
35.1.5 Large Bodies 794
35.1.6 Slender-Member Bodies 798

35.2 **Current Loads** 802
35.2.1 Nonuniform Currents 802
35.2.2 Wave–Current Interaction 803
35.2.3 Wave Current Kinematics 804
35.2.4 Current-Induced Forces 804
35.2.5 Vortex-Induced Vibrations 805

35.3 **Wind Loads** 806
35.3.1 Wind-Speed Profile 807
35.3.2 Wind Spectra and Gusts 808
35.3.3 Steady-State Forces 809
35.3.4 Unsteady Forces 809

35.4 **Model Tests** 810
35.4.1 Principles and Similarity Laws 810
35.4.2 Scaling of Loads 812
35.4.3 Elastic Structures 812

35.5 **CFD Tools** 814

35.6 **Extreme Response Estimation** 814

References 816

35.1 Wave Loads

35.1.1 Linear Waves

Linear waves are characterized by the ratio of the wave amplitude to its wavelength as a *small* quantity. The fluid is assumed to be incompressible and inviscid, and the flow is irrotational, so that the particle velocity vector is given by

$$\boldsymbol{u} = \nabla\phi \,, \tag{35.1}$$

where ϕ is the velocity potential. Because of the incompressible fluid assumption, the continuity (or the con-

servation of mass) equation is given by $\nabla \boldsymbol{u} = 0$. And as a result, the continuity equation becomes Laplace's equation

$$\Delta\phi = 0\,. \tag{35.2}$$

We also have Euler's integral given by

$$\frac{\partial\phi}{\partial t} + \frac{1}{2} \parallel \nabla\phi \parallel^2 + \frac{p}{\rho} + \mathrm{g}x_2 = \frac{p_\mathrm{A}}{\rho}\,, \tag{35.3}$$

where p_A is the atmospheric pressure, ρ is the mass density of the fluid, g is the gravitational acceleration, and x_2 is the vertical coordinate (Fig. 35.1). Equation (35.3) is to be used in the determination of the pressure; it is called unsteady Bernoulli's equation by some. This equation is a result of the conservation of momentum equation which need not be simultaneously solved with the conservation of mass equation since that equation only involves a single unknown, the velocity potential.

On any material surface, whether free or not, we have the general kinematic boundary condition

$$\boldsymbol{u}\cdot\boldsymbol{n} = \frac{\partial\phi}{\partial n} = \boldsymbol{q}\cdot\boldsymbol{n}\,, \tag{35.4}$$

where $\boldsymbol{u}$ is the fluid particle velocity vector, $\boldsymbol{q}$ is the solid boundary velocity vector, and $\boldsymbol{n}$ is the unit normal vector on the boundary, pointing out of the fluid. Clearly, (35.4) is also the body-boundary condition and the seafloor boundary condition since it represents no flux through the surface. By defining the boundary surface by the equation, $F(x_1, x_2, x_3, t) = x_2 - \eta(x_1, x_3, t) = 0$, where η is the free-surface elevation, and requiring that the material derivative of F vanishes, we obtain the kinematic free-surface condition

$$\frac{\partial\phi}{\partial x_2} - \frac{\partial\eta}{\partial t} - \frac{\partial\phi}{\partial x_1}\frac{\partial\eta}{\partial x_1} - \frac{\partial\phi}{\partial x_3}\frac{\partial\eta}{\partial x_3} = 0 \quad \text{on } x_2 = \eta\,. \tag{35.5}$$

The dynamic condition on the free surface is that the pressure is continuous, that is, $p = p_\mathrm{A} \cong 0$ on $x_2 = \eta$, where we have taken the atmospheric pressure equal to zero without loss in generality. The surface tension is ignored, which means that the water waves we deal with here exclude the capillary waves (whose lengths are less than about 1.5 cm). Therefore, from Euler's integral, (35.3), we obtain the dynamic free-surface condition

$$\frac{\partial\phi}{\partial t} + \frac{1}{2}\|\nabla\phi\|^2 + \mathrm{g}\eta = 0 \quad \text{on } x_2 = \eta\,. \tag{35.6}$$

The two difficulties associated with these boundary conditions are: (i) they must be imposed on a boundary which is unknown, and (ii) they are nonlinear. Note, however, that the governing equation (35.2) is linear.

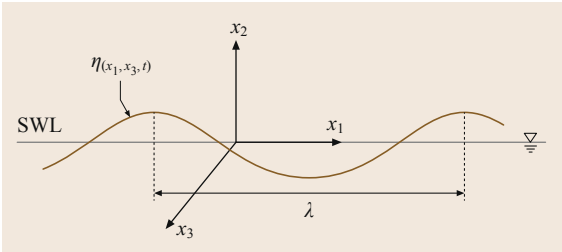


Fig. 35.1 Free surface of a wave of length λ

Perturbation Expansion

To overcome the difficulties associated with solving the nonlinear free-surface boundary conditions, one generally resorts to the use of the perturbation expansion of the quantities involved, and then linearizing the problem. To do this, one usually assumes that the wave motion is *small* and, therefore, the nonlinear terms can be discarded as a result of the argument that their magnitudes will be smaller than those of the linear terms. In this respect, one can assume that the velocity potential, as well as the surface elevation, can be expanded in a perturbation series for which a perturbation parameter ϵ is taken to be equal to, for example, Ak, where A is the wave amplitude and k is the wave number, that is, $k = 2\pi/\lambda$, where λ is the wavelength. In shallow-water wave problems, however, one would rather deal with two small parameters, one representing the nonlinearity and the other representing the dispersion of long waves.

A perturbation series is a series expansion of an unknown function about a known function, provided that the deviation of the unknown function from the known function is *small* (say the known function is the potential which can, for example, be taken as constant everywhere; this corresponds to a quiescent fluid). Then, we can write

$$\begin{aligned}
\phi &= \epsilon\phi^{(1)} + \epsilon^2\phi^{(2)} + \epsilon^3\phi^{(3)} + \cdots\,,\\
\eta &= \epsilon\eta^{(1)} + \epsilon^2\eta^{(2)} + \epsilon^3\eta^{(3)} + \cdots\,,\\
\epsilon &= Ak\,,
\end{aligned} \tag{35.7}$$

where $\phi^{(1)}$ is called the first-order potential, $\phi^{(2)}$ is the second-order potential, etc., and similarly for η, that is, $\eta^{(1)}$ is called the first-order surface elevation, $\eta^{(2)}$ is called the second-order surface elevation, etc. The expansion in (35.7) is such that when $\epsilon = 0$, there is no fluid motion; therefore, ϕ and η vanish.

Now, if the wave motion is small, meaning $\epsilon \ll 1$, we may, without giving any necessary justification, discard all the higher order terms after we substitute these

expansions in the boundary conditions. Moreover, we can expand each of the terms of the boundary conditions in a Taylor series about the still-water surface, $x_2 = 0$. For example, the time derivative of the potential is written as

$$\frac{\partial \phi}{\partial t}(x_1, x_2 = \eta, x_3, t) = \frac{\partial \phi}{\partial t}(x_1, 0, x_3, t) + \eta \frac{\partial^2 \phi(x_1, 0, x_3, t)}{\partial t \partial x_2} + \cdots . \tag{35.8}$$

Since η and ϕ are small, the higher order terms can sometimes be ignored. This means that only the linear terms involving ϕ and η have to be evaluated on the still-water level ($x_2 = 0$) instead of on the exact boundary surface, $x_2 = \eta(x_1, x_3, t)$. This is required to be consistent with the perturbation expansion. Therefore, we have the linearized versions of the boundary conditions, given by (35.5) and (35.6), as follows

$$\frac{\partial \eta^{(1)}}{\partial t}(x_1, x_3, t) = \frac{\partial \phi^{(1)}}{\partial x_2}(x_1, 0, x_3, t) \text{ (kinematic)} , \tag{35.9}$$

$$\eta^{(1)}(x_1, x_3, t) = -\frac{1}{\mathrm{g}} \frac{\partial \phi^{(1)}}{\partial t}(x_1, 0, x_3, t) \text{ (dynamic)} . \tag{35.10}$$

In summary, the first-order problem, $O(\epsilon)$ problem, becomes

$$\begin{aligned} &\Delta \phi^{(1)}(x_1, x_2, x_3, t) = 0 , \\ &\phi^{(1)}_{x_2}(x_1, -h, x_3, t) = 0 , \\ &\phi^{(1)}_{x_2}(x_1, 0, x_3, t) - \eta^{(1)}_t(x_1, x_3, t) = 0 , \\ &\phi^{(1)}_t(x_1, 0, x_3, t) + \mathrm{g}\eta^{(1)}(x_1, x_3, t) = 0 , \end{aligned} \tag{35.11}$$

where subscripts indicate differentiation with respect to the indicated variable. And for this $O(\epsilon)$ problem, the dynamic pressure is given by the linearized Euler's integral

$$p^{(1)}(x_1, x_2, x_3, t) = -\rho \phi^{(1)}_t(x_1, x_2, x_3, t), \quad x_2 < 0 , \tag{35.12}$$

anywhere in the fluid.

We can now assume that we have two-dimensional or long-crested, linear water waves so that the associated functions do not depend on the x_3-coordinate. Of course, in the case of short-crested waves, which represents the real situation in the oceans, we cannot rule out the x_3-dependence.

Let us now assume that a monochromatic wave, propagating in the positive x_1-direction, is given by

$$\eta(x_1, t) = A \cos(kx_1 - \omega t) . \tag{35.13}$$

Here, A is the wave amplitude. Equation (35.13) does not depend on time in a moving coordinate system, whose constant (phase) speed is given by $c = \frac{\omega}{k}$. In other words, the motion is steady in the moving coordinates. In a fixed coordinate system, η is a time-harmonic function. Because η is periodic, ϕ must also be periodic, so that we can write

$$\phi(x_1, x_2, t) = \mathrm{Re}\left\{ Y(x_2) \mathrm{e}^{\mathrm{i}[kx_1 - \omega t]} \right\} . \tag{35.14}$$

Equation (35.14) is a result of the separation-of-variables technique used in solving linear partial differential equations.

By enforcing the dynamic free-surface boundary condition and the no-flux sea-floor condition, the linear solution for the velocity potential can be obtained

$$\phi(x_1, x_2, t) = \frac{\mathrm{g}A}{\omega} \frac{\cosh[k(x_2 + h)]}{\cosh(kh)} \sin(kx_1 - \omega t) . \tag{35.15}$$

However, we have not yet used the kinematic free-surface condition given by the third equation in (35.11). When we enforce this condition by using (35.15), we obtain the dispersion relation

$$\omega^2 = \mathrm{g}k \tanh(kh) , \tag{35.16}$$

where h is the water depth and k is the wave number. In deep water, $kh \to \infty$, so that we have $\omega^2 = \mathrm{g}k$, and in shallow water, $kh \ll 1$, so that $\omega^2 = \mathrm{g}hk^2$.

In the deep-water case, the real part of the velocity potential of the incoming wave (or incident wave potential) becomes

$$\phi(x_1, x_2, t) = \frac{\mathrm{g}A}{\omega} \mathrm{e}^{kx_2} \sin(kx_1 - \omega t) . \tag{35.17}$$

It is useful to give particle velocity components (linear) for finite water depth by using (35.15)

$$\begin{aligned} u_1 &= \frac{\partial \phi}{\partial x_1} = \frac{\mathrm{g}Ak}{\omega} \frac{\cosh[k(x_2 + h)]}{\cosh(kh)} \cos(kx_1 - \omega t) , \\ u_2 &= \frac{\partial \phi}{\partial x_2} = \frac{\mathrm{g}Ak}{\omega} \frac{\sinh[k(x_2 + h)]}{\cosh(kh)} \sin(kx_1 - \omega t) . \end{aligned} \tag{35.18}$$

Table 35.1 Some physical quantities resulting from linear theory

Vertical particle displacement	$\zeta = A\dfrac{\sinh[k(x_2+h)]}{\sinh(kh)}\cos(kx_1-\omega t)$
Horizontal particle displacement	$\xi = A\dfrac{\cosh[k(x_2+h)]}{\sinh(kh)}\sin(kx_1-\omega t)$
Group velocity	$c_\mathrm{g} = \dfrac{1}{2}\left[1+\dfrac{2kh}{\sinh(2kh)}\right]c$
Mean energy density	$E_\mathrm{m} = \dfrac{1}{2}\rho g A^2$
Energy flux	$P = E_\mathrm{m}c_\mathrm{g}$

The total pressure (linear) can then be obtained from Euler's integral as

$$p(x_1,x_2,t) = \rho g A\frac{\cosh k(x_2+h)}{\cosh(kh)}\cos(kx_1-\omega t) - \rho g x_2\,, \tag{35.19}$$

where the first term on the right-hand side represents the dynamic and the second term represents the hydrostatic pressure.

Water particle accelerations (linear) are given by

$$\begin{aligned}\frac{Du_1}{Dt} &\cong \frac{\partial u_1}{\partial t}\\ &= \mathrm{g}Ak\frac{\cosh[k(x_2+h)]}{\cosh(kh)}\sin(kx_1-\omega t)\,, \qquad (35.20)\\ \frac{Du_2}{Dt} &\cong \frac{\partial u_2}{\partial t}\\ &= -\mathrm{g}Ak\frac{\sinh[k(x_2+h)]}{\cosh(kh)}\cos(kx_1-\omega t)\,,\end{aligned} \tag{35.21}$$

where D/Dt is the material derivative, approximated here as the local time derivative only, due to the linearity of the problem. Some other physical quantities for linear water waves are listed in Table 35.1.

35.1.2 Nonlinear Waves

To obtain the linear wave solution, the perturbation expansion introduced in the last section was truncated at $O(\epsilon)$. Clearly, this expansion can be carried out to higher orders, and this is generally done in offshore engineering and in deep-water applications, up to the fifth order. The higher order infinitesimal wave theory based on the systematic power series expansion in $\epsilon = Ak$ is due to [35.1]. The proof of convergence can be found in [35.2]. *Schwartz* [35.3] obtained the infinite-depth expansion up to $O(\epsilon^{117})$ by using a computer algorithm. However, because of the complexity of algebra and the rapid convergence of the asymptotic series, it is mostly unnecessary to consider problems of $O(\epsilon^6)$ and higher, unless, perhaps, if the water depth is very shallow. But then the Stokes expansion in shallow water gives inaccurate results in general, and thus should not be used if the water depth is shallow. Instead, a cnoidal wave theory can be used in shallow waters [35.4, 5]. The fifth-order Stokes waves, commonly used in offshore engineering, were calculated by [35.6].

After using the Taylor series expansion of the functions and its derivatives, and substituting them in the boundary conditions for each of the perturbation terms seen in (35.7), one can obtain, for example, the kinematic free-surface boundary condition at the first and second orders in two dimensions

$$\begin{aligned}&O(\epsilon): \phi^{(1)}_{x_2} - \eta^{(1)}_t = 0\,,\\ &O(\epsilon^2): \phi^{(2)}_{x_2} - \eta^{(2)}_t = \phi^{(1)}_{x_1}\eta^{(1)}_{x_1} - \phi^{(1)}_{x_2x_2}\eta^{(1)}\,.\end{aligned} \tag{35.22}$$

Note that once the first-order problem is solved, the right-hand side of the second-order boundary condition in (35.22) is known, and thus it can be treated as an applied or external pressure on the free surface located on the still-water level.

Let us consider the dynamic free-surface boundary condition given by (35.6) in two dimensions. Following the same procedure, that is, by using the Taylor series expansion of the functions in each term of the perturbation expansion, one can obtain

$$\begin{aligned}\phi_t &= \epsilon\phi^{(1)}_t + \epsilon^2(\phi^{(2)}_t + \phi^{(1)}_{x_2t}\eta^{(1)})\\ &\quad + \epsilon^3(\phi^{(3)}_t + \phi^{(1)}_{x_2t}\eta^{(2)} + \cdots) + O(\epsilon^4)\,,\\ \frac{1}{2}\{\phi^2_{x_1} + \phi^2_{x_2}\} &=\\ &\frac{1}{2}\{[\epsilon\phi^{(1)}_{x_1} + \epsilon^2(\phi^{(1)}_{x_2x_1}\eta^{(1)} + \phi^{(2)}_{x_1}) + \cdots]^2\\ &+ [\epsilon\phi^{(1)}_{x_2} + \epsilon^2(\phi^{(1)}_{x_2x_2}\eta^{(1)} + \phi^{(2)}_{x_2}) + \cdots]^2\}\\ &+ O(\epsilon^5)\,.\end{aligned} \tag{35.23}$$

Then, we have

$$\begin{aligned}&O(\epsilon): \phi^{(1)}_t + \mathrm{g}\eta^{(1)} = 0\,,\\ &O(\epsilon^2): \phi^{(2)}_t + \mathrm{g}\eta^{(2)} =\\ &\qquad -\phi^{(1)}_{x_2t}\eta^{(1)} - \frac{1}{2}(\phi^{(1)^2}_{x_1} + \phi^{(1)^2}_{x_2})\,.\end{aligned} \tag{35.24}$$

The dynamic and kinematic free-surface conditions can be combined into one equation for each $O(\epsilon)$ and $O(\epsilon^2)$

as follows

$$
\begin{aligned}
&O(\epsilon): \phi_{tt}^{(1)}(x_1,0,t) + g\phi_{x_2}^{(1)}(x_1,0,t) = 0\,,\\
&O(\epsilon^2): \phi_{tt}^{(2)}(x_1,0,t) + g\phi_{x_2}^{(2)}(x_1,0,t) =\\
&\qquad -\eta^{(1)}\left[\phi_{ttx_2}^{(1)} + g\phi_{x_2x_2}^{(1)}\right]\\
&\qquad -2(\phi_{x_1}^{(1)}\phi_{tx_1}^{(1)} + \phi_{x_2}^{(1)}\phi_{tx_2}^{(1)})\,. \qquad (35.25)
\end{aligned}
$$

The first equation in (35.25) is the combined form of the third and fourth equations, respectively, in (35.11).

It is important to note that the first-order potential, which will be given explicitly later, also satisfies the second-order problem if the water depth is infinite. In other words,

$$
\phi(x_1,x_2,t) = \frac{gA}{\omega}e^{kx_2}\sin(kx_1-\omega t) + O(\epsilon^3)\,. \qquad (35.26)
$$

However, the second-order surface elevation, that is,

$$
\eta = \epsilon\eta^{(1)} + \epsilon^2\eta^{(2)} + O(\epsilon^3) \qquad (35.27)
$$

is not the same as the first-order surface elevation given by (35.13) [35.4]. The second- and fifth-order solutions of Stokes waves are given, for example, in [35.5, Tables 4.3–4.4].

Next, we briefly discuss another type of nonlinear waves that occur when the water depth is relatively shallow.

35.1.3 Shallow-Water Waves

In relatively shallow waters, when the wavelength is greater than about eight water depths, the Stokes expansion no longer works, and an alternative wave theory must be used. One such theory is called the cnoidal wave theory as established by [35.7]. Subsequently, other cnoidal wave theories were developed by [35.4, 8–10].

The infinite length limit of cnoidal waves is known as solitary waves. These waves are generally used to model tsunami propagation and arrival times in the oceans. A number of solitary wave solution are available [35.8, 11, 12]. In more recent years, solitary waves based on the Green–Naghdi theory were also developed [35.13, 14].

Some of the equations that can be used in engineering calculations of cnoidal and solitary waves are listed in [35.5].

35.1.4 Random Waves

Irregular sea waves can be thought of as the sum of an infinite number of sinusoidal waves (each of which has a different amplitude and frequency) whose phase angles are random. In general then, the amplitude of each component wave may be represented by $A(\omega,\gamma)$, which is a random variable itself. Here, ω is the angular wave frequency and γ is the heading angle of incoming waves. Because of this randomness of waves, a probabilistic approach is necessary to describe various parameters associated with a *confused sea*. First, *Denis* and *Pierson* [35.15] introduced the probabilistic description of confused seas in marine hydrodynamics involving ship motions. As an example of the superposition of regular waves of different (however, infinitesimal) heights and frequencies, consider Fig. 35.2. Even this limited number of regular waves gives an irregular wave pattern when they are superposed. Furthermore, the resulting irregular shape is totally random, that is, a slight change in wave amplitude, frequency, or phase of the waves will result in a different pattern for irregular waves (Fig. 35.3, which also shows how the frequency domain and time domain representations of waves are related to each other in long-crested seas). Therefore, irregular waves cannot be identified by their shapes (surface elevation).

Because we cannot characterize an irregular sea by its shape, we need another criterion to base our approach on. This criterion is that the total (potential and kinetic) energy E of an irregular wave train is the sum of the energies of all components of individual waves, that is,

$$
E = \frac{\rho g}{2}\{A_1^2 + A_2^2 + \cdots\} = \frac{1}{2}\rho g\sum_{n=1}^{N}A_n^2\,. \qquad (35.28)
$$

This concept leads us to the energy spectrum in which waves of many frequencies are present. In the limit,

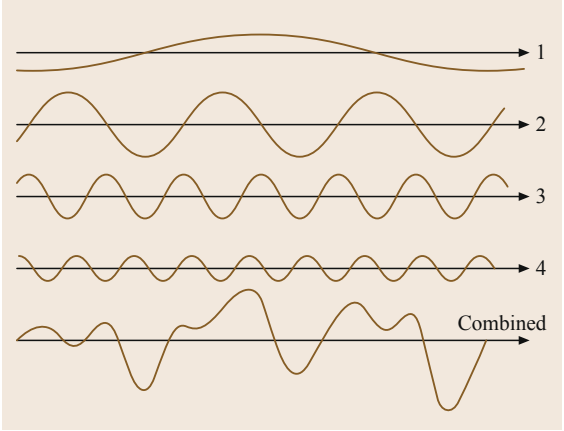


Fig. 35.2 Superposition of four regular waves with different amplitudes and lengths, shifted randomly (after [35.16])

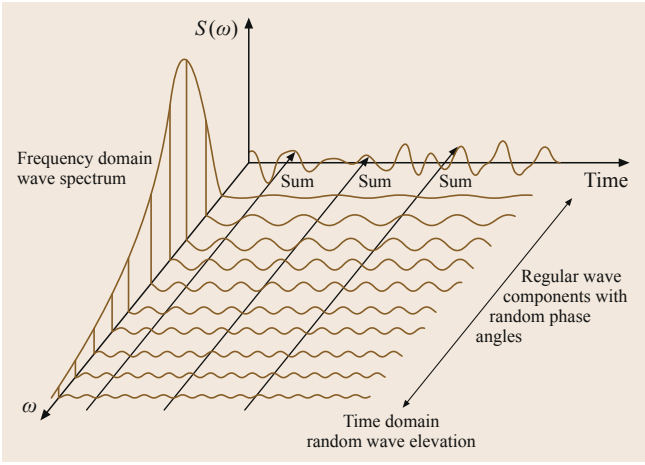


Fig. 35.3 The relation between the frequency-domain and time-domain representations of waves in long-crested seas (after [35.17])

the number of individual wave components N tends to infinity, and the summations become integrals, for example

$$\eta(x_1, x_3, t) = \mathrm{Re}\left\{\int_0^{\infty}\int_0^{2\pi} A(\omega,\gamma)\exp\{\mathrm{i}k(\omega)(x_1\cos(\gamma) + x_3\sin(\gamma)) - \mathrm{i}\omega t\}\,\mathrm{d}\gamma\,\mathrm{d}\omega\right\}, \tag{35.29}$$

where the wave amplitude $A(\omega,\gamma)$ is a random quantity. Equation (35.29) admits all possible wave directions and frequencies.

The wave spectrum can also be calculated by considering the auto-correlation function (of the surface elevation) defined by

$$R_\mathrm{T}(\tau) = \frac{1}{\mathrm{T}}\int_{-\frac{\mathrm{T}}{2}}^{\frac{\mathrm{T}}{2}} \eta_\mathrm{T}(t)\,\eta_\mathrm{T}(t+\tau)\,\mathrm{d}t\,. \tag{35.30}$$

The one-sided wave spectrum for a long-crested irregular wave system (unidirectional) is then given by

$$S(\omega) = \frac{1}{2\pi}\int_0^{\infty} R_\mathrm{T}(\tau)\cos(\omega\tau)\mathrm{d}\tau\,. \tag{35.31}$$

The total mean energy of the wave system is

$$\bar{E} = \frac{1}{2}\rho g\int_0^{\infty}\int_0^{2\pi} S(\omega,\gamma)\mathrm{d}\gamma\mathrm{d}\omega\,. \tag{35.32}$$

Since the mean square value of a monochromatic wave is $A^2/2$, we can represent a random process by

$$\eta(t) = \lim_{\Delta\omega\to 0}\left\{\sum_{n=1}^{N}\sqrt{2S(\omega_n)\Delta\omega}\cos(\omega_n t + \epsilon_n)\right\}, \quad N\to\infty\,, \tag{35.33}$$

where ϵ_n is a random phase angle uniformly distributed in the interval 0–2π, and $\omega_n = n\Delta\omega$. Even though different choices of ϵ_n would produce different time histories, the resulting wave spectrum would be the same. The mean square value of $\eta(t)$ is the area under the spectral density versus the frequency curve.

Let us consider the mean value of $\eta(t)$

$$\langle\eta(t)\rangle = \frac{1}{\mathrm{T}}\int_0^{\mathrm{T}}\eta(t)\mathrm{d}t\,, \tag{35.34}$$

and the standard deviation σ

$$\sigma = \sqrt{\frac{1}{\mathrm{T}}\int_0^{\mathrm{T}}(\eta(t) - \langle\eta(t)\rangle)^2\,\mathrm{d}t}\,. \tag{35.35}$$

The standard deviation σ is a measure of how η deviates from the mean. Water-surface elevation is assumed to have a Gaussian probability distribution, with zero mean. Thus, the variance σ^2 becomes

$$\sigma^2 = \langle\eta^2(t)\rangle = \frac{1}{\mathrm{T}}\int_0^{\mathrm{T}}\eta(t)^2\mathrm{d}t = \int_0^{\infty} S(\omega)\mathrm{d}\omega = m_0\,, \tag{35.36}$$

that is, the area under the spectral density function is the variance of $\eta(t)$ and m_0 is known as the zeroth-moment of the spectrum. In other words, the root mean square (RMS) of $\eta(t)$ is equal to the standard deviation given by (35.35) if the probability distribution is Gaussian, that is

$$\mathrm{RMS} = \sqrt{\langle\eta^2(t)\rangle} = \sqrt{m_0} = \sigma\,. \tag{35.37}$$

The spectrum of any quantity of interest, such as the motion of a platform or forces acting on it, can be obtained in a random seaway by

$$S_y(\omega) = |H(\omega)|^2 S_x(\omega) \,, \tag{35.38}$$

where the modulus (or magnitude) of $H(\omega)$ is indicated since the transfer function, $H(\omega)$, can also be a complex function of the frequency. In other words, the output spectrum is linearly proportional to the square of the transfer function. A typical transfer function could be equal to the force divided by the wave amplitude. The force clearly varies with the angular frequency ω(rad/s) (the cyclic frequency, $f = \omega/2\pi$ is used sometimes).

The transfer function, $H(\omega)$, can represent the wave force, wave run-up, surface-elevation amplitude, etc., as long as the system is linear. In other words, $L[\,]$ is an operator such that for the input $x(t)$ and the output $y(t)$, and the relation $y(t) = L[x(t)]$ between them, we must have $L[x_1(t) + x_2(t)] = L[x_1(t)] + L[x_2(t)]$. Also, for any constant α, $L[\alpha x(t)] = \alpha L[x(t)]$ needs to be satisfied. This is just the definition of a linear operator. Here, $x_1(t)$ and $x_2(t)$ are the two inputs corresponding to the two outputs $y_1(t)$ and $y_2(t)$, respectively.

For a narrow-banded spectrum (most of the energy present in waves is concentrated around a rather small interval of wave frequencies), it can be shown for any physical quantity of interest by using the Rayleigh probability distribution that

$$\bar{y}^{1/3} \cong 4.0\sqrt{m_0} = 4.0\,\sigma \,, \tag{35.39}$$

where m_0 is the area under the spectrum curve and σ is the root mean square (RMS) as given by (35.37). If, for instance, $y(t)$ is equal to the wave height, $H(t)$, we have

$$H_{1/3} = 4.0\sqrt{m_0} = 4.0\sqrt{\int_0^\infty S(\omega)\,\mathrm{d}\omega} \,, \tag{35.40}$$

where $H_{1/3}$ is called the significant wave height and $S(\omega)$ is the wave spectrum. Note that if m_0 is taken as the area under the response spectrum, $y_{1/3}$ would give the significant height (double amplitude) of the force, moment, motion, etc., whatever the response (output) corresponds to.

If one is interested in the significant amplitude response, then, for example

$$\begin{aligned} F_{1/3} &= 2.0\sqrt{m_\mathrm{R}} = 2.0\sqrt{\int_0^\infty S_\mathrm{R}(\omega)\,\mathrm{d}\omega} \\ &= 2.0\,\mathrm{RMS\ (Force)} \end{aligned} \tag{35.41}$$

will be the significant force amplitude, while $S_\mathrm{R}(\omega)$ is the force amplitude response spectrum,

$$S_\mathrm{R}(\omega) = |H(\omega)|^2\, S_W(\omega) \,, \tag{35.42}$$

where $H(\omega)$ is the force amplitude transfer function, $F(\omega)/A$, and $S_W(\omega)$ is the wave amplitude spectrum.

Once the significant response is known, it is possible to predict (if the Rayleigh distribution is valid) the short-term design extreme by the following formula (for a derivation of this equation, see for example [35.18, p. 319]),

$$y_\mathrm{extreme} = y_{1/3}\left(\frac{1}{2}\log\frac{N}{0.01}\right)^{1/2} \,, \tag{35.43}$$

where N is the number of waves expected to be encountered during a storm. For example, if the storm lasts for 3 hours and the average wave period is 15 s, then $N = 3 \times 60 \times 60/15 = 720$, and, therefore, $y_\mathrm{extreme} = 1.5584\, y_{1/3}$.

Sometimes we may know the spectrum given as a function of one quantity, say cyclic wave frequency or period or encounter frequency, and we may need to convert it to a wave spectrum as a function of another quantity. To do this, one has to keep in mind that the energy content of the waves must remain the same no matter what coordinate system is used (this is called *Galilean Invariance*), including the steadily moving one. It is common, for example, to see that the cyclic wave frequency, $f = 1/T$ (Hz), is used in calculations. In such cases, one can convert the angular-frequency spectrum into the cyclic-frequency spectrum by writing

$$S(f)\mathrm{d}f = S(\omega)\mathrm{d}\omega,\ \omega = 2\pi f \ \Rightarrow\ S(f) = 2\pi S(\omega) \,. \tag{35.44}$$

So far, we have concentrated on the long-crested wave spectrum. The waves in the ocean are, in fact, short crested, meaning that they move in different directions in general. If waves are multidirectional, and, therefore, are short crested, with a dominant wave heading angle γ, the directional wave-energy spectral density can be written as

$$\begin{aligned} &\bar{S}_\mathrm{W}(\omega,\theta) = S_\mathrm{W}(\omega)G(\theta) \,, \\ &G(\theta) = \frac{2}{\pi}\cos^2\theta \,, \\ &-\frac{\pi}{2} \le \theta \le \frac{\pi}{2} \,, \end{aligned} \tag{35.45}$$

where θ, the heading angle of each component wave, is measured from the axis of the dominant wave heading,

and $G(\theta)$ is called the spreading function, which, by assumption, is set to zero if $|\theta| > \pi/2$. Note that there exist spreading functions that are different from (35.45). This is because of a better fit of a particular spreading function to observational data at a specific ocean site.

The wave spectrum for a given location is, in general, not available from observational data. As a result, we must use one or more of a number of formulas developed for estimating the wave spectrum. Here, we summarize some of them, and the reader is referred to other works for more detailed analysis of and references on the subject [35.5] or [35.19].

The Bretschneider spectrum is based on the significant wave height H_s and peak wave (angular) frequency, $\omega_p = 2\pi/T_p$, where T_p is the peak period of the wave spectrum, that is, it is a two-parameter spectrum. For fully developed seas, the wave spectrum is

$$S(\omega) = \frac{5H_s^2}{16\omega_p}\frac{1}{(\omega/\omega_p)^5}\exp\left\{-\frac{5}{4}\left(\frac{\omega}{\omega_p}\right)^{-4}\right\} . \tag{35.46}$$

The dimension of the wave spectrum is L^2T. An example of a Bretschneider spectrum is shown in Fig. 35.4.

The Pierson–Moskowitz (P–M) spectrum is based on the wind speed U_w(m/s) alone, it is a one-parameter spectrum, and is given by

$$S(\omega) = \frac{\alpha g^2}{\omega^5}\exp\left(-\frac{B}{\omega^4}\right) , \tag{35.47}$$

where $\alpha = 8.1\times10^{-3}$ is Phillips' constant, $B = 0.74(g/U_w)^4$, and g is the gravitational acceleration. It is possible to express the P–M spectrum in terms of the significant wave height, H_s, that is the average height of the highest one-third of the waves. Recall that H_s is given by (35.40)

$$H_s = 4.0\sqrt{m_0} = 4.0\sqrt{\int_0^\infty S(\omega)\mathrm{d}\omega}$$
$$= \frac{2{U_w}^2}{g}\sqrt{\frac{\alpha}{0.74}} \quad \text{or} \quad {U_w}^{-4} = \frac{0.044}{g^2(H_s)^2} , \tag{35.48}$$

where we used (35.47). Therefore, we can now write the one-parameter P–M spectrum in terms of H_s

$$S(\omega) = \frac{8.1-3g^2}{\omega^5}\exp\left(-0.032\frac{g^2}{\omega^4(H_s)^2}\right) . \tag{35.49}$$

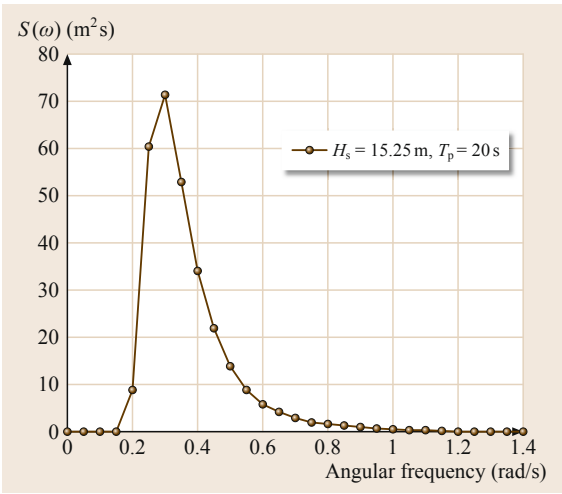


Fig. 35.4 Bretschneider spectrum for significant wave height of 15.25 m and peak period of 20 s

35.1.5 Large Bodies

When an offshore structure is *large* (generally the dimensions of the structure and perhaps even of each member is not small compared with the wavelength), one would assume that the viscous effects are much smaller than the inertial effects as far as the wave loads are concerned. In this section, we discuss wave loads on large structures by using linear potential theory valid for rather small wave slopes and under the assumptions of inviscid and incompressible fluid and irrotational flow.

Potential Theory

Since we have a linear system, meaning that the governing equation (Laplace's equation) and the boundary conditions do not contain any nonlinear terms, all physical quantities related to the response should be linearly proportional to wave amplitude. As a result, the complicated problem of a freely floating structure impacted by linear waves can be decomposed into multiple problems, each of which are easier to solve. The sum of all the solutions will then be the solution of the complicated problem, as long as the linearity assumption made holds true, that is, that the wave slope is small.

Consider a freely floating body in the *absence* of forward motion. We assume that the body makes *small* motions in six degrees of freedom, and that they can be written as

$$x_j = x_j^0 e^{-i\omega t} , \quad i = \sqrt{-1}, \quad j = 1, 2, \ldots, 6 . \tag{35.50}$$

Each x_j, $j = 1, 2, 3$, refers to the translational displacements of surge, heave, and sway, respectively, and each x_j, $j = 4, 5, 6$, refers to the angular displacements (or rotations) of roll, yaw, and pitch, respectively, and x_j^0 denotes the complex amplitude of the motion.

The complex total potential due to the interaction of waves with the body can be written in a compact form as

$$\Phi_{\mathrm{T}} = \sum_{j=0}^{7} \phi_j(x_1, x_2, x_3) \mathrm{e}^{-\mathrm{i}\omega t} \,, \tag{35.51}$$

where $\phi_0 \equiv \phi_I$ is the incoming wave potential, ϕ_j, $j = 1, 2, \ldots, 6$, are the radiation potentials, and $\phi_7 \equiv \phi_D$ is the diffraction potential, all being complex functions of the independent spatial variables. The incoming wave potential is only due to the periodic linear waves propagating in the absence of the body, the diffraction potential is due to a fixed body impacted by the incoming waves, and the radiation potentials are due to a body oscillating in a prescribed mode of motion, one at a time, and in the absence of any incoming waves.

Each potential in (35.51) must satisfy

$$\begin{aligned} \nabla^2 \phi_j(x_k) &= 0, \quad x_k \in D \,, \\ \frac{\partial \phi_j(x_k)}{\partial x_2} - \frac{\omega^2}{\mathrm{g}} \phi_j(x_k) &= 0, \quad x_k \in S_{\mathrm{f}} \,, \\ \frac{\partial \phi_j(x_k)}{\partial x_2} &= 0, \quad x_k \in S_{\mathrm{s}} \,, \end{aligned} \tag{35.52}$$

for $j = 0, 1, 2, \ldots, 7$ and $x_k \equiv (x_1, x_2, x_3)$ (Fig. 35.5). The sea-floor condition (as used here) implies that the water depth is constant (although this is not a necessary assumption in general).

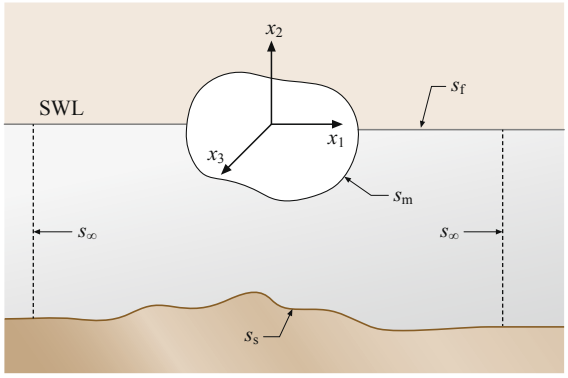


Fig. 35.5 A control volume in the fluid, bounded by the body, still-water surface, sea floor, and a large control cylinder of radius R

In addition, we must have the following body-boundary conditions to be satisfied

$$\begin{aligned} \frac{\partial \phi_j(x_k)}{\partial n} &= n_j \,, \; j = 1, 2, \ldots, 6, \; x_k \in S_{\mathrm{m}}; \\ \frac{\partial \phi_7(x_k)}{\partial n} &= -\frac{\partial \phi_0(x_k)}{\partial n} \,, \; x_k \in S_{\mathrm{m}} \,. \end{aligned} \tag{35.53}$$

And all ϕ_j, $j = 1, 2, 3, \ldots, 7$, except the incident wave-potential, must also satisfy the Sommerfeld radiation condition:

$$\lim_{r \to \infty} \left[\sqrt{r} \left(\frac{\partial}{\partial r} - ik \right) \phi_D \right] = 0 \,. \tag{35.54}$$

This equation basically states that the particle velocities $\to 0$ as $r \to \infty$ and that the waves are outgoing. Equation (35.54) is called the radiation condition (or the Sommerfeld condition) for a three-dimensional (3-D) body of bounded extent in a fluid (with a free surface) of unbounded extent on the horizontal plane. Note that

$$r = \sqrt{x_1^2 + x_3^2}$$

on the horizontal plane. This equation must be satisfied by the diffraction and radiation potentials, but not by the incoming potential.

The total pressure can be obtained from linearized Euler's integral

$$p_{\mathrm{T}} = -\rho \mathrm{g} x_2 - \rho \frac{\partial \Phi_{\mathrm{T}}}{\partial t} \,; \tag{35.55}$$

and the total force ($j = 1, 2, 3$) and moment ($j = 4, 5, 6$) are obtained from

$$F_{\mathrm{T}j} = -\int_{S_{\mathrm{m}}} p_{\mathrm{T}} n_j \, \mathrm{d}S \,, \quad j = 1, 2, \ldots, 6 \,. \tag{35.56}$$

The hydrostatic forces and moments due to the first term on the right hand side of (35.55) can be written as

$$F_{\mathrm{S}i} = -k_{ij}^{(\mathrm{S})} x_j \,, \quad i, j = 1, 2, \ldots, 6 \,, \tag{35.57}$$

where $k_{ij}^{(\mathrm{S})}$ are the hydrostatic stiffness (or restoring) coefficients, $k_{22}^{(\mathrm{s})} = \rho \mathrm{g} A_{\mathrm{WP}}$ is the restoring coefficient in heave, where A_{WP} denotes the water-plane area of the body. Because there is no restoration in the horizontal plane, it is clear that $k_{ij}^{(\mathrm{s})} = 0$ if $i, j = 1, 3, 5$. Also, note that, in a linear system and for a rigid body, $k_{ij}^{(\mathrm{S})} = k_{ji}^{(\mathrm{S})}$, that is, the hydrostatic stiffness matrix (or tensor) is

symmetric. For elastic (or deformable) bodies, this can also be shown [35.20].

The total forces and moments given by (35.56) includes the hydrostatic, wave-exciting and radiation forces, and moments. The wave-exciting forces can be written as

$$F_{Wj} = -\rho e^{-i\omega t} \int_{S_m} \left(\frac{\partial \phi_0}{\partial t} + \frac{\partial \phi_7}{\partial t} \right) n_j dS$$
$$= AE_j e^{-i\omega t}, \quad j = 1, 2, \ldots, 6, \tag{35.58}$$

where E_j is the complex amplitude of the exciting force divided by the wave amplitude A. Note that (35.58) includes both the Froude–Krylov force ($\partial\phi_0/\partial t$ term) and the scattering force ($\partial\phi_7/\partial t$ term).

The hydrodynamic (or radiation) forces due to the motion of the body in each mode can be determined, once the radiation potentials are solved for. They are given by

$$F_{Ri} = -\mu_{ij}\ddot{x}_j - \lambda_{ij}\dot{x}_j, \quad i, j = 1, 2, \ldots, 6, \tag{35.59}$$

where

$$\mu_{ij} = \rho \int_{S_m} \phi_j^{(R)} \frac{\partial \phi_i^{(R)}}{\partial n} \, dS,$$
$$\lambda_{ij} = \rho\omega \int_{S_m} \phi_j^{(I)} \frac{\partial \phi_i^{(R)}}{\partial n} \, dS, \tag{35.60}$$

where $\phi_j^{(R)}$ is the real and $\phi_j^{(I)}$ is the imaginary part of j-th radiation potential. The components of the second-order tensor μ_{ij} are called the added-mass coefficients, and λ_{ij} are called the wave-damping or, simply, damping coefficients. Note that, in the unbounded-fluid case, there is no free surface and, as a result, there is no wave generation. And, therefore, the damping coefficients do not exist since no energy is carried by waves toward infinity. See, for example, [35.21] for some added-mass and damping coefficients for a floating structure of a semisubmersible type.

By considering a control volume enclosed by the free surface, the body surface, sea floor, and an imaginary cylinder at $r = \infty$, one can show that $\mu_{ij} = \mu_{ji}$ and $\lambda_{ij} = \lambda_{ji}$ by using Green's second identity. Note that these results are valid if there is no forward motion. It can also be shown that the average rate at which work is done by the body upon the fluid is directly proportional to λ_{ij}, and, therefore, that the matrix λ_{ij} must be positive definite for all ω. In particular, $\lambda_{ii} > 0$ for all $i = 1, 2, \ldots, 6$ and ω, but in some cases $\lambda_{ij}, i \neq j$ may be negative [35.22].

There may be mooring lines which are attached to the body to keep it in location. These mooring lines tend to restore the motion of the body, and, thus, can be treated as restoring coefficients

$$F_{Mi} = -k_{ij}^{(M)} x_j, \quad i, j = 1, 2, \ldots, 6. \tag{35.61}$$

And they can provide restoring (however small) in all modes of motion unlike the hydrostatic restoring.

We can now assemble the forces and moments to obtain

$$F_{Ti} = -(k_{ij}^{(S)} + k_{ij}^{(M)})x_j - \mu_{ij}\ddot{x}_j - \lambda_{ij}\dot{x}_j + AE_i e^{-i\omega t}, \tag{35.62}$$

where we have also included the mooring line loads in (35.62) (they were not included in (35.56)).

We are now ready to consider Newton's equations which govern the motions of a body. These equations can be written as

$$F_{Ti} = m_{ij}\ddot{x}_j, \quad i, j = 1, 2, \ldots, 6, \tag{35.63}$$

where

$$m_{11} = m_{22} = m_{33} = m, \; m_{44} = I_1, \; m_{55} = I_2,$$
$$m_{66} = I_3, \; m_{54} = m_{45} = -I_{12}, \; m_{46} = m_{64} = -I_{13},$$
$$m_{56} = m_{65} = -I_{23},$$
$$I_{jk} = \int_V \bar{x}_j \bar{x}_k \rho_B \, dV, \; I_1 = I_{22} + I_{33}, \; I_2 = I_{11} + I_{33},$$
$$I_3 = I_{11} + I_{22}, \tag{35.64}$$

where any unspecified mass or mass moment, m_{ij}, is zero, provided that F_{Tj} refers to a coordinate system which is fixed in the mean position, S_m, of the body, and whose origin coincides with the center of gravity of the body. Here, ρ_B is the mass density of the body and I_{ij} are called the moment of inertia coefficients when $i = j$, and products of inertia when $i \neq j$ [35.23].

We can now set (35.62) equal to (35.63) and arrange the equation by moving some terms around to obtain

$$(m_{ij} + \mu_{ij})\ddot{x}_j + \lambda_{ij}\dot{x}_j + (k_{ij}^{(S)} + k_{ij}^{(M)})x_j = AE_i e^{-i\omega t}. \tag{35.65}$$

Considering (35.50), we can write the equations of motion given by (35.65) as

$$\frac{x_j^0}{A} = \left[-\omega^2(m_{ij} + \mu_{ij}) - i\omega\lambda_{ij} + \left(k_{ij}^{(S)} + k_{ij}^{(M)}\right) \right]^{-1} E_i. \tag{35.66}$$

Recalling that x_j^0 is complex, we can write it as

$$x_j^0 = x_j^{0R} + \mathrm{i}x_j^{0I} = |x_j|\mathrm{e}^{-\mathrm{i}\delta} ,$$

where δ is the phase angle of the motion relative to the incoming-wave crest. As a result, (35.66) becomes two sets of simultaneous 6×6 linear equations to be solved for the motion response, $|x_j^0|/A$. These motion responses are commonly called transfer functions. The square of a transfer function is sometimes known as the response amplitude operator (RAO). However, it is not uncommon that a transfer function, itself, is called an RAO, rather than its square. RAOs are used in irregular sea analysis to determine the random or stochastic response of a floating body in random waves. It is also noted that the term *transfer function* is also used for wave forces or moments per unit wave amplitude A.

Solution Methods

The boundary-value problems above can be solved to obtain the diffraction and radiation potentials so that the hydrodynamic coefficients and wave loads can be calculated. These potentials can be evaluated using the 3-D source-distribution method [35.24–26]. The pulsating source potential, which is a complex function, in infinite water depth can be found in [35.27, Art. 13]. This Green function can be written in an alternate form [35.28]. A different series expansion of the Green function is also given in [35.27]. Although this series expansion is more accurate and more computationally efficient to evaluate, compared with the integral form given in [35.27], it is limited to cases where kr is not very small. On the other hand, the evaluation of the Green function given by [35.28] is much more efficient, and no difficulty with regard to the size of kr is expected.

The solution of the integral equation for the unknown strengths of the distributed sources requires the discretization of the body surface by panel elements. Once the unknown strengths of the sources are determined on the body boundary (discretized by constant panels) by enforcing the body boundary condition, the diffraction and radiation potentials can be determined. These calculations can be performed by a number of commercially available computer programs.

A Green function is also called a source function, and it is one of the bases for finding solutions to diffraction and radiation problems in coastal and offshore engineering. A numerical method, called the boundary-element method (BEM), uses this approach. However, the BEM uses the simple Rankine source rather than the complicated Green function, and, as a result, all boundaries of the flow field need to be discretized. The Green function (GFM), requires the discretization of the boundaries on which there are distributed singularities with unknown strengths. For example, in the case of the linear problem of a floating body, only the body boundary need to be discretized. The BEM or the GFM is in contrast with the finite-element method (FEM) which requires that the entire fluid domain and its boundaries be discretized to solve for the velocity potential, and, therefore, the velocities. For more information on the BEM, the readers are referred to [35.29].

One of the alternative methods for solving hydrodynamic problems is the FEM. The disadvantage of this method compared to the GFM, provided that a particular Green function exists and is available, is that ϕ has to be calculated everywhere inside the domain. On the other hand, i) the FEM results in a banded matrix which can be solved efficiently, ii) the function calculated in the FEM is simpler, and iii) the FEM requires less knowledge of mathematics and fluid mechanics than the GFM does. Another numerical method that can be used is the finite-difference method, although it is used infrequently in fluid–structure interaction problems involving floating bodies.

Time-Domain Methods

Sometimes, the frequency-domain method discussed before is inadequate to incorporate certain nonlinearities and/or interactions in the system, and, as a result, one needs to resort to a time-domain method of calculating the loads and/or motions of an offshore structure. Typical examples could include problems involving the current and wave interaction, and low-frequency motions of platforms. These time-domain methods are based, in most cases, on the hydrodynamic coefficients and loads that have been previously calculated through a frequency-domain method and utilize the fact that the frequency and time contents of a system are related through the Fourier transforms. It is noted that some of the loads and/or the equations of motion could be nonlinear in these calculations.

Memory effects are included through the velocity-based convolution integrals. The memory effects considered are largely based on the early works of [35.30, 31] and [35.22] in the application of linear potential theory in the time domain, and on the work of [35.32]. Viscous effects may be included through the nonlinear drag term of Morison's equation. An earlier review of linear and nonlinear methods of time-domain calculations of motions can be found in [35.33]. This is an important reference as it is not always possible to use a time-domain method based on the linear hydrodynamic coefficients or exciting forces obtained through a frequency-domain method, by a Green function panel method, as some physical events are nonlinear and even transient.

There are many applications of the mentioned time-domain methods applied to various ocean engineering

problems, including to wave-energy conversion (WEC) problems [35.34]. There also are software packages, for example, ANSYS AQWA, that use some of these methods.

Drift Loads

The first-order wave loads discussed so far are time harmonic and, therefore, do not produce a steady component. This is mainly because the second-order component of the pressure on the body was assumed to be zero in the linear problem, see (35.3). Clearly, if this term was included, the time-mean force would have been of second order even though the first-order (linear) potential was used. Moreover, if a body is surface-piercing, there would be an additional term related to the runup on the body. In regular waves, this mean force is called the drift force in brief. In irregular waves, there is the mean drift force, but in addition there is also the slowly varying drift force. All these forces, although smaller in magnitude compared with the first-order forces, may become very important especially for bodies that are moored, as they may amplify motions due to a possible resonance in the system.

Maruo [35.35] showed that the mean drift force can be shown to be proportional to the square of the reflected wave amplitude, and since the reflected wave can be thought of as the reflection coefficient times the incoming wave amplitude, the mean drift force can be seen to be proportional to the square of the incoming wave amplitude. Reference [35.36] used the far-field radiation potentials (momentum) approach to obtain the drift forces on a floating body, and these were used later by [35.25] to calculate the drift force and moment on a floating body. The momentum approach results in expressions that involve the Kochin function. *Pinkster* [35.37], on the other hand, used the near-field potentials to calculate the mean drift forces and moments in all six degrees of freedom, and this approach was used later by [35.38] in conjunction with the panel method based on the Green-function method for a 3-D floating structure. These studies are for single bodies; multi-body calculations of drift forces became important later due to plans to build very large floating stuctures (VLFS) [35.39, 40].

The slowly varying component of the drift force on rigid bodies in irregular waves appears to have been first studied experimentally by [35.41] and later theoretically by [35.42].

35.1.6 Slender-Member Bodies

A slender member of a body is defined as a structural member whose characteristic dimension, for example, diameter, is *small* compared to the wavelength. With this in mind, we discuss Morison's equation which is sometimes used in offshore engineering during the preliminary design stage.

Morison's Equation

Wave forces on cylindrical structural members, such as the pontoons or columns of a platform or a pile that extends to the sea floor, when both the inertia and viscous forces are important, are discussed here. Due to the nonlinear nature of the Navier–Stokes equations and the boundary conditions, many have attempted to simplify the computation of wave forces on cylindrical piles. Most of the studies have been based on the experimental determination of the inertia and drag coefficients that appear in Morison's equation [35.43]. Morison's equation was developed as an ad hoc approach to a limited set of experimental data. However, because of the importance of cylindrical piles both in offshore engineering (jack-up platforms) and coastal engineering (pier piles, bridge columns, etc.), there have been many investigations on the proper coefficients to be used in almost every different case that one can imagine. A few of the different cases include inclined cylinders, group of cylinders, roughened cylinders, and horizontal cylinders [35.5].

The original form of Morison's equation is given by

$$\delta F = (1 + C_a)\rho\pi \left(\frac{D}{2}\right)^2 \dot{u}_1 + \frac{1}{2}\rho D C_d u_1 |u_1| \,, \tag{35.67}$$

where δF is the sectional (in the x_2 (vertical) direction) horizontal force, D is the diameter of the vertical pile, u_1 is the wave horizontal velocity, $\dot{u}_1$ is the wave horizontal acceleration, C_a is called the added-mass or virtual-mass coefficient, and C_d is the form-drag coefficient. C_a and C_d are dimensionless. $C_m = (1 + C_a)$ is called the inertia coefficient. Note that u_1 is the particle velocity only due to the incoming wave, there is no diffraction effects accounted for, unlike in the theory of *MacCamy* and *Fuchs* [35.44] based on the linear potential theory. The justification for this is that the diffraction effects are small because the cylinder is *slender*; therefore, u_1 and $\dot{u}_1$ are evaluated along the cylinder axis, since the errors made would be *small*.

Consider an oscillatory flow, in an *unbounded fluid* (no free surface or sea floor), with a period $T = 2\pi/\omega$, and velocity $u_1(t) = U\cos(\omega t)$, where U is the amplitude of velocity. The functional dependence of the force (per unit length) acting on the body, whose characteristic length is D, can be written as

$$\delta F = f(U, D, \rho, \nu, T) \,. \tag{35.68}$$

By applying dimensional analysis to (35.68), in which the set (ρ, U, D) is chosen as dimensionally independent, one can obtain

$$\frac{\delta F}{\frac{1}{2}\rho D U^2} = C_\mathrm{d}\left(\frac{UD}{\nu}, \frac{UT}{D}\right) , \tag{35.69}$$

or for the total force,

$$\frac{F}{\frac{1}{2}\rho D^2 U^2} = C_\mathrm{d}(\mathrm{Re}, \mathrm{KC}) , \tag{35.70}$$

where $\mathrm{Re} = UD/\nu$ is the usual diameter-based Reynolds number, and $\mathrm{KC} = UT/D$ is called the Keulegan–Carpenter(KC) number [35.45]. For large values of the KC number, $\mathrm{KC} \gg 1$, we expect the force coefficient, C_d, in (35.70) to approach the value of the steady (form-)drag coefficient, since large KC numbers correspond to long periods. In other words

$$C_\mathrm{d}(\mathrm{Re}, \mathrm{KC}) \to C_\mathrm{d}(\mathrm{Re}) \quad \text{if} \quad \mathrm{KC} \gg 1 . \tag{35.71}$$

On the other hand, when $\mathrm{KC} \ll 1$, the inertia effects will dominate viscous effects since the end of the time duration necessary to develop the boundary layer and flow separation could not have been reached. Therefore, we have basically an inviscid fluid since $\mathrm{Re} \to \infty$, or

$$C_\mathrm{d}(\mathrm{Re}, \mathrm{KC}) \to C_\mathrm{d}(\mathrm{KC}), \quad \mathrm{Re} \gg 1, \quad \mathrm{KC} \ll 1 . \tag{35.72}$$

If we now consider the presence of a *periodic free surface*, we can then anticipate that the force depends on

$$F = f(h, H, \lambda, D, \rho, \mathrm{g}, \nu) , \tag{35.73}$$

where h is the water depth, λ is the wavelength, and H is the wave height. In fact, it should not be difficult to obtain

$$\frac{F}{\rho \mathrm{g} H D^2} = f\left(\frac{h}{\lambda}, \frac{H}{\lambda}, \frac{D}{\lambda}, \mathrm{Re}\right) \tag{35.74}$$

by using dimensional analysis. If the problem is linear and the fluid is inviscid, we can write (35.74) as

$$\frac{F}{\rho \mathrm{g} H D^2} = f_1\left(\frac{D}{\lambda}, \frac{D}{h}\right) . \tag{35.75}$$

Therefore, if the fluid is inviscid, the force coefficient given in (35.74) does depend only on the ratio D/λ for a fixed-body geometry and water depth, that is, D/h is constant. So that $F = F(\omega)$ only, where ω is the angular wave frequency related to the wavelength through the linear dispersion relation. If we had explicitly included the frequency (or period) in the equations above, the KC number would also appear.

To see the effect of body dimension, wave height, and wavelength on the forces on a vertical, circular cylinder, consider Fig. 35.6. We see that as D/λ becomes small, the viscous effects (i. e., separation) become important. For a circular cylinder, and if $D/\lambda > 0.15$, the diffraction effects are more important. The KC number given in Fig. 35.6 is

$$K \equiv \mathrm{KC} = \frac{\pi H/\lambda}{\frac{D}{\lambda}\tanh(kh)} , \tag{35.76}$$

and it is evaluated at the still-water level and is for finite water depth. In infinite water, (35.76) becomes $\mathrm{KC} = \pi H/D$, since $\tanh kh \to 1.0$. Also, one can deduce from Fig. 35.6 that if $H/D > 1$, the viscous effects become important for a fixed ratio of D/λ. Therefore, in engineering calculations, it is recommended that, for a vertical circular cylinder that extends to the sea floor:

- If $\frac{H}{D} < 1.0$ and $\frac{D}{\lambda} \geq 0.15$, diffraction theory be used
- If $\frac{H}{D} \geq 1.0$ and $\frac{D}{\lambda} < 0.15$, Morison's equation be used.

It is noted that these should be cautiously used for other body geometries. In Fig. 35.6, $(H/\lambda)_{\max}$ is the maximum wave steepness (before wave breaking occurs) given by (for finite but large meaning not shallow-water depth, h)

$$\frac{H}{\lambda} = 0.14\tanh(kh) . \tag{35.77}$$

In deep water, this clearly becomes $H/\lambda = 0.14$.

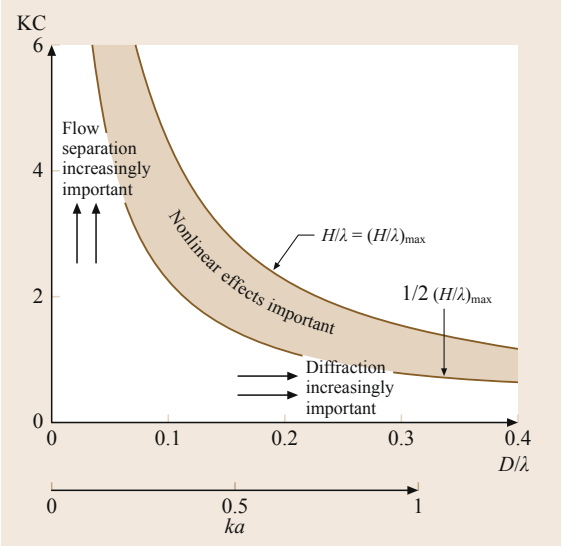


Fig. 35.6 Different wave force regimes (after [35.46])

Morison's equation is obtained by summing the inertia and drag terms. There can be no rational basis for this summation, especially since the coefficients C_a and C_d are intended to be constants and it can be shown easily that the $(1 + C_a)$ term is nothing but the sum of the Froude–Krylov force and the virtual mass force, that both can be obtained through the inviscid fluid and irrotational flow assumptions. In reality, the inertia and drag coefficients are also functions of the wave frequency. However, this formula is widely used in offshore and coastal engineering to determine wave and current forces on slender cylinders, such as drilling risers or jetty piles, and rather successfully.

Clearly, the drag term is nonlinear. This causes a slight problem in using Morison's equation, especially when the body is freely floating since the location where u_1 is calculated is unknown, and one must use the total *relative velocity* rather than u_1 alone if the body is freely floating. To overcome this problem, and also, to be consistent with the assumption of linearity, the drag term is linearized by defining a linear drag coefficient. There is one more reason for linearizing the drag force, and it is related to the use of spectral analysis in irregular waves which require that the system be linear.

To turn to the question of the linear drag coefficient, we write the horizontal component of the particle velocity as

$$u_1 = u_0 \cos(\omega t) \quad \text{at } x_1 = 0\,, \tag{35.78}$$

where u_0 is the amplitude of the velocity. The drag force in the direction of wave motion becomes

$$\delta F_d = \frac{1}{2}\rho D C_d u_0^2 \cos(\omega t)|\cos(\omega t)|\,. \tag{35.79}$$

The linear drag force can now be defined by equating the net work done (on a differential cylinder element) by the drag force δF_d, given by (35.79), and the linearized drag force, given by

$$\delta F_{dL} = \frac{1}{2}\rho D C_{dL} u_0 \cos(\omega t)\,. \tag{35.80}$$

This means that the viscous energy dissipated (per wave cycle) is the same whether we use (35.79) or (35.80). Therefore, by requiring equal energy dissipation per wave cycle, we must have $E_d = E_{dL}$

$$E_d = \int_0^\lambda \frac{1}{2}\rho C_d D u_0^2 \cos(\omega t)|\cos(\omega t)|\,\mathrm{d}x_1$$

$$= E_{dL} = \int_0^\lambda \frac{1}{2}\rho C_{dL} D u_0 \cos(\omega t)\,\mathrm{d}x_1\,. \tag{35.81}$$

Recalling also that $u_1 = \mathrm{d}x_1/\mathrm{d}t$ or $\mathrm{d}x_1 = u_0\cos(\omega t)\mathrm{d}t$, and $|\cos(\omega t)|\cos(\omega t) = \cos^2(\omega t)$ if $0 \le t \le T/4$, and that the total energy is equal to four times the energy per quarter of a wavelength for constants C_d and C_{dL}, we obtain, from (35.81)

$$C_{dL} = C_d \frac{\int_0^{T/4} u_0^3 \cos^3(\omega t)\mathrm{d}t}{\int_0^{T/4} u_0^2 \cos^2(\omega t)\mathrm{d}t} = \frac{8}{3\pi} C_d\, u_0\,. \tag{35.82}$$

Note that C_{dL} has the dimension of velocity, unlike C_d (which is dimensionless).

The amplitude of the horizontal velocity, u_0, is given, in linear potential theory, by

$$u_0 = \frac{gAk}{\omega}\frac{\cosh[k(x_2 + h)]}{\cosh(kh)}\,. \tag{35.83}$$

Substituting (35.82) into (35.80), we obtain

$$\delta F_{dL} = \frac{4}{3\pi}\rho D C_d u_0^2 \cos(\omega t)\,. \tag{35.84}$$

We need to note that there are other methods of linearization of the drag term, different from the method discussed here. Also, see [35.47–49] on the linearization of the drag force when a current is present, and when the waves are random.

There may also be current in the vicinity of the cylinder. Let us denote the steady shear-current velocity by $u_c = u_c(x_2)$. Also, it is possible that the cylinder is moving. In the case of a cylinder that extends to the sea floor, the only possible mode of motion is in the horizontal plane, $x_1 - x_3$, and this is called surging. However, Morison's equation is used not only to determine the forces on fixed platforms, such as

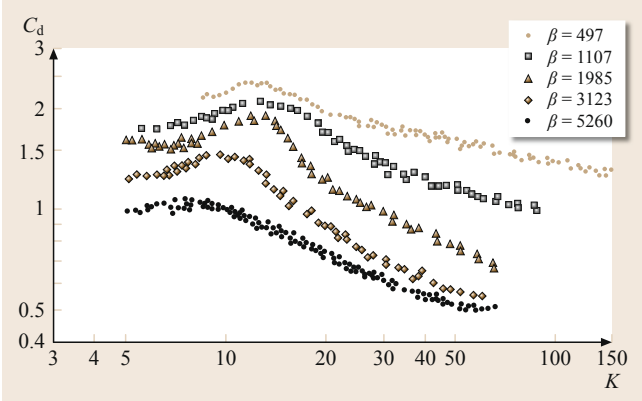


Fig. 35.7 Drag coefficient for a smooth circular cylinder as a function of the Reynolds and Keulegan–Carpenter numbers (after [35.51])

jack-up rigs, but also on floating platforms, such as semi-submersibles, which may have tubular structural members, such as columns, pontoons, or braces. These members may also be inclined, rather than just be vertical or horizontal. Therefore, in general, one can have three translational and three rotational (or angular) velocity components for the structure motion. In such cases, the velocity calculated in (35.67) must be the relative velocity

$$u_r = u_p + u_c - u_b \,, \tag{35.85}$$

where u_r is the relative velocity, u_p is the particle velocity, u_c is the current velocity, and u_b is the body velocity.

Drag and Inertia Coefficients

There have been many experimental studies on the drag and inertia coefficients used in Morison's equation. The most comprehensive references on the subject are the two monographs, [35.5] and [35.50]; they cite almost all previous experimental works on the inertia and drag coefficients. For example, [35.51] gives Figs. 35.7 and 35.8 for the drag and inertia coefficients, respectively, for smooth cylinders that were tested in a U-tube water tunnel as functions of the KC number and $\beta = \mathrm{Re}/\mathrm{KC}$, where KC is the Keulegan–Carpenter number and Re is the Reynolds number (Sect. 35.1.6).

A number of organizations also publish their recommended coefficients, For example, [35.52] recommends the following drag and inertia coefficients ($C_A = C_m - 1$) shown in Figs. 35.9 and 35.10

Some other organizations that recommend the use of certain coefficients are American Petroleum Institute, American Bureau of Shipping and Society of Naval Architects and Marine Engineers, and many other classified organizations whose publications are frequently referenced by regulatory agencies.

Viscous Drift Loads

Viscous drift forces originate mainly due to the existence of the drag force over the instantaneous submerged length of the members of platforms, due to the presence of current, and due to wave–current interaction effects. Traditionally, these forces have been computed using the drag force term of Morison's equation [35.43] to determine the mean forces and moments. Most of the studies on viscous drift forces prior to 1995 concentrated on analytical and/or experimental results for a single circular cylinder. For example, *Chakrabarti* [35.53] presented analytical expressions for viscous and potential drift forces on a vertical cylinder and compared the predictions with experimental data. The relative importance of the viscous and potential drift contributions has also been discussed.

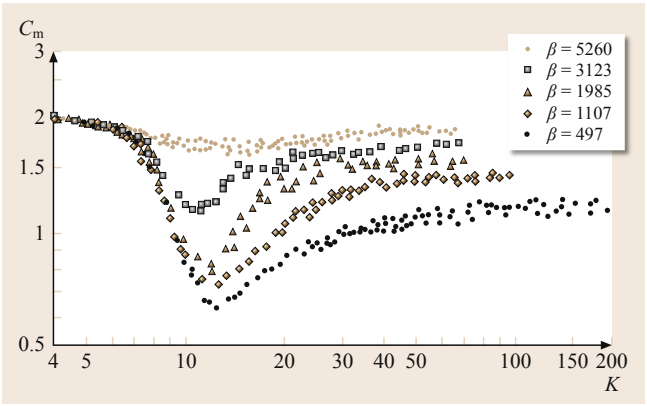


Fig. 35.8 Inertia coefficient for a smooth circular cylinder as a function of the Reynolds and Keulegan–Carpenter numbers (after [35.51])

Burns [35.54] used the relative velocity model of Morison's equation where the platform motion in surge is considered in determining the total relative velocity between the body and fluid. A method to obtain nonlinear viscous drift force transfer functions that can be used to determine the mean and slowly varying surge drift forces in irregular seas has also been presented. Other studies of viscous drift forces on the tension leg platform (TLP) [35.55] were also limited to the analysis of surge drift forces and motions. *Ertekin* and *Chitrapu* [35.56] computed the wave- and current-induced viscous drift forces and moments in all six degrees of freedom of a TLP. In that study, as well as in [35.57], first-order motions of the platform in all six degrees of freedom, given by the

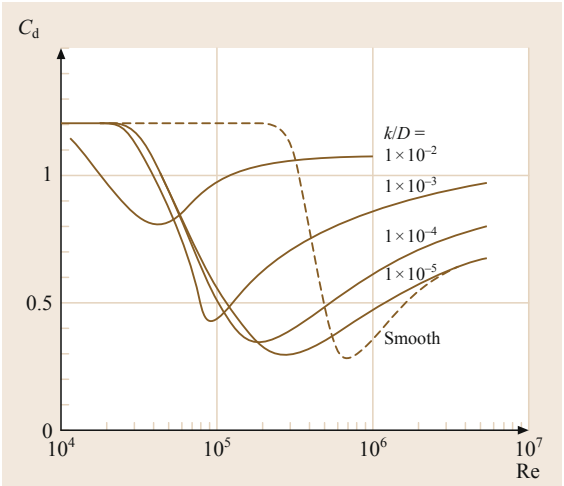


Fig. 35.9 Drag coefficient for a circular cylinder as a function of the roughness coefficient and Reynolds number (after [35.52])

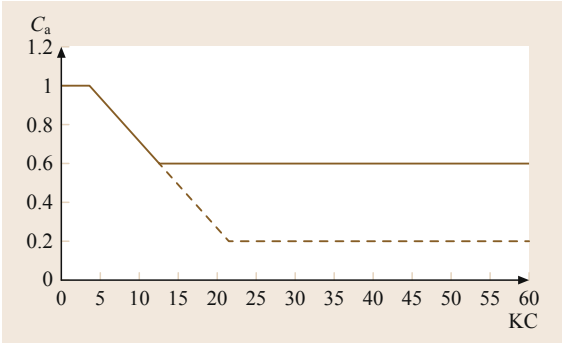


Fig. 35.10 Inertia coefficient for a circular cylinder as a function of the Keulegan–Carpenter number; *solid line* smooth cylinder and *dotted line* rough cylinder (after [35.52])

frequency-domain motion transfer functions are used to calculate the body velocity, and the resulting relative velocity is used to compute the drift forces and moments.

The evaluation of viscous drift forces in irregular waves is complicated due to the fact that the drift force is a nonlinear function of the wave height. From the results obtained in regular wave analysis, it appears that the viscous drift force, in the presence of current, is proportional to the third power of the wave amplitude as compared to the potential drift force which is proportional to the wave amplitude squared. Suitable methods have been developed to obtain potential drift forces in irregular waves using the regular wave results in conjunction with spectral analysis methods. The alternative is to use a time-domain method in which all nonlinearities can be included. However, frequency-domain methods for the computation of viscous drift forces in irregular waves are preferred in the preliminary design stage since they are computationally more efficient, although time-domain analysis is indispensable for the final design purpose [35.17, 57]. For this reason, several investigations are being carried out to develop frequency-domain methods for computing viscous drift forces in irregular waves [35.58, 59].

35.2 Current Loads

For the purpose of offshore engineering, it is the horizontal current that is of interest. Currents are important for vessel anchoring, installation work, riser interference, and vortex-induced vibration (VIV). Surface and mid-depth currents cause in-line and transverse forces on fixed and floating structures. Sea-floor currents may cause scour around structures and pipelines. This scour may compromise structural foundations or create unsupported spans in pipelines. The bottom current may also induce VIV in the unsupported spans of pipelines.

There are many components of the current in the ocean. Some of these currents extend to depths of several thousand meters and are poorly understood. Current types include:

1. Tidal currents
2. Wind driven
3. Ocean circulation
4. Boundary currents including loop and eddy currents
5. Internal waves and solitons.

The superposition of these current components generates the total current which can be represented by a current profile which gives speed and direction as a function of depth.

Currents are generally considered time invariant for purposes of offshore structure design, though in most cases, they comprise turbulent flow and vary in speed and direction with time. Currents are usually characterized by averages of the horizontal current vector over several minutes. Figure 35.11 illustrates how currents may vary in direction and magnitude at different depths and with time.

A description of currents should include the general circulation pattern in the area, tidal currents, and wind-driven currents. In most parts of the world, there is scarce current data because of the expense and the need for long-term observations required to capture a reasonable number of severe events. Yet design, planning for installation, and operation require information regarding the frequency of occurrence and seasonal variations of the current speed and direction. Whenever possible, site-specific measurements should be obtained throughout the water column and over sufficient time to capture several major events. When current models are used in lieu of site-specific measurements, the model should be validated against nearby measured data.

For design purposes, sufficient current information should be gathered to permit an estimate of the 100 year (1% annual probability of exceedance) event.

35.2.1 Nonuniform Currents

Current is rarely uniform with depth. Current profiles are frequently modeled as piecewise linear functions with the profile described in a table of depth, speed, and direction. Simple profiles, such as a uniform cur-

rent near the surface and zero elsewhere in the water column, are often used for design. Frequently, directional variation with depth is ignored or simplified for the purpose of design calculations. These simplifications are used when the simplification is conservative and yet does not appear to increase cost unreasonably.

The currents in a region will be a function of the local topography and oceanography, including density distribution and the flow into or out of the area (Fig. 35.12). Shallow-water currents are frequently driven by tides, and simple profiles of speed versus depth provide an adequate description. Deepwater currents as illustrated in Fig. 35.11 can vary greatly in the speed and direction with depth and may require a more complex description.

Shallow-water and near-bottom currents will often exhibit a power-law profile, such as given by (35.86).

$$U_c(z) = U_{c0}\left(\frac{z+h}{h}\right)^{\alpha}, \tag{35.86}$$

where $U_c(z)$ is the current speed at elevation z, where $z = 0$ at the surface; U_{c0} is the surface current speed; h is the water depth; α is an exponent (typically $1/7$).

A method for developing design current profiles from long-term measured current profile data sets is described in [35.62]. The data are parameterized using empirical orthogonal functions, and then the design current profile with the required return period is selected through a process involving an inverse first-order reliability method (FORM).

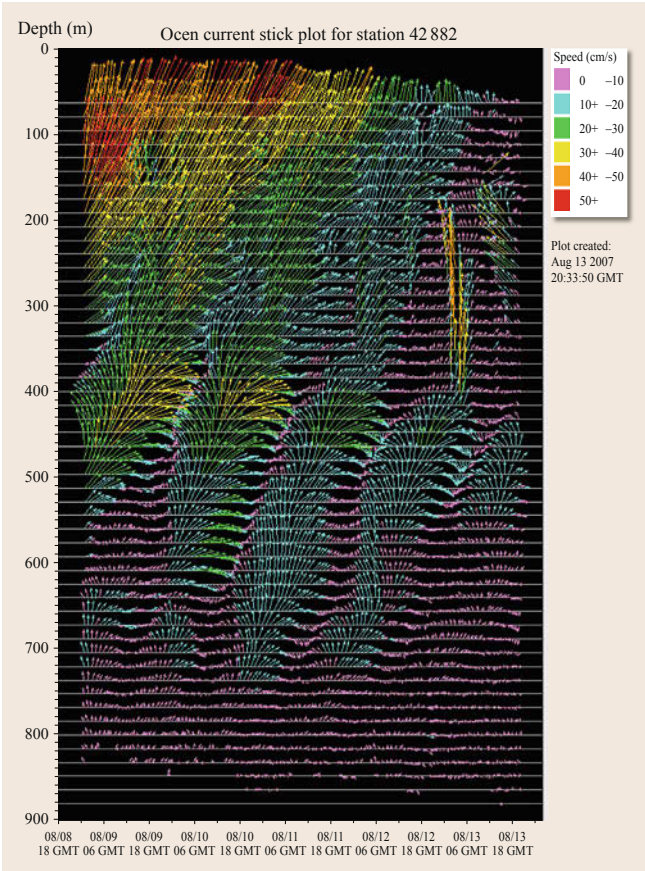


Fig. 35.11 National Data Buoy Center (NDBC) stick plot of currents acquired by oil and gas companies in the northern Gulf of Mexico (after [35.60])

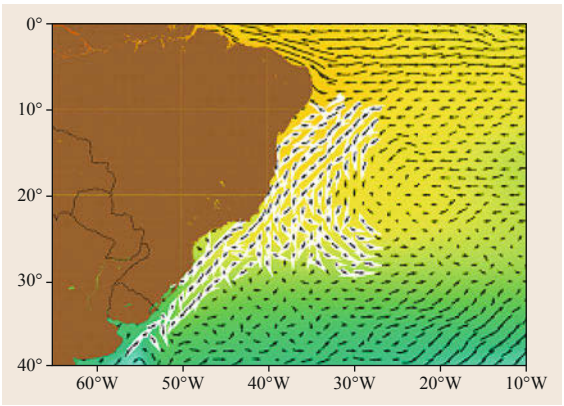


Fig. 35.12 The Brazil current as represented by the Mariano Global Surface Velocity Analysis (MGSVA) (after [35.61])

35.2.2 Wave–Current Interaction

When waves propagate on a current, the encounter frequency of the wave encountering a fixed body is different from the intrinsic frequency of the waves. It is the intrinsic frequency that determines the wavelength and the wave kinematics. Much wave data are collected from fixed platforms that measure the encounter frequency. This fixed platform data must be corrected to the intrinsic frequency since wave spectra for criteria are usually represented in terms of the intrinsic frequency.

For an effective current speed U, moving in the same direction as the waves, the encounter wave frequency is higher than the wave frequency alone. The encounter frequency is $\omega_A = 2\pi/T_A$, where T_A is the encounter period seen by a fixed observer. The two frequencies are related by the Doppler shift as

$$\omega = \omega_A - kU\,, \tag{35.87}$$

where k is the wave number and ω is the intrinsic frequency. The last term in (35.87) is called the convective frequency. The Doppler effect is computed based on the component of the ambient current in the direction of

waves U. If the current profile is not uniform, *Kirby* and *Chen* [35.63] showed that the encounter wave period is computed from (35.87) and the following equations

$$\omega^2 = gk \tanh(kh) \,. \tag{35.88}$$

$$U = \frac{2k}{\sinh 2kh} \int_{-h}^{0} U_c(z) \cosh[2k\,(z+h)]\,\mathrm{d}z \,, \tag{35.89}$$

where h is the water depth and $U_c(z)$ is the current profile.

The frequency parameter in wave spectra formulations (Sect. 35.1.4) is the intrinsic frequency. A fixed or floating stationary structure in a wave field with current responds to the encounter frequency rather than to the intrinsic frequency. For wave-response calculations, the wave-frequency spectrum should be transformed into the encounter frequency. The wave energy per frequency band is independent of the reference frame so that $S(\omega)\,\mathrm{d}\omega = S(\omega_A)\,\mathrm{d}\omega_A$. Equation (35.87) is used to make the frequency transformation, where the wave number k is a function of the intrinsic wave frequency ω through (35.88).

35.2.3 Wave Current Kinematics

The kinematics of waves propagating through a uniform current can be modeled simply by adding the current vector to the wave velocity vector. When the current profile varies significantly through the wave zone, this simple model does not apply. Several approximations have been put forward to model the combined regular wave and current velocity. These approximations are extensions of the same approximations used to estimate wave kinematics near the free surface when no current is present.

A current profile is defined from the mean water line to the bottom. Waves change the location of the free surface and transform the current profile as they pass [35.64, 65].

A linear transformation of the current profile incorporates a stretching factor F_s

$$F_s = \frac{\eta + h}{h} \,, \tag{35.90}$$

where η is the water-surface elevation measured upward from the still-water level and h is the still-water depth.

The transformed current beneath the instantaneous water surface is $U_c\left(\frac{h+z_s}{F_s} - h\right)$, where z_s is the elevation measured upward from the still-water level and $U_c(z)$ is the current profile absent waves. A linearly stretched current profile is exactly analogous to the stretching of linear wave kinematics as applied by [35.66].

The nonlinear transformed current at an elevation z_s is $U_c(z)$, where z is determined by solving

$$z_s = z + \eta \frac{\sinh(k_{nl}(z+h))}{\sinh(k_{nl}h)} \,, \tag{35.91}$$

where k_{nl} is the wave number for the regular wave under consideration for water depth h, crest elevation η, and the intrinsic wave frequency ω. k_{nl} is calculated using the dispersion relationship of the nonlinear wave theory being used.

A simple modification of the nonlinear regular wave method described above can be used for random waves. In this case, the wave period and length correspond to the period and length of the spectral peak frequency.

Once the current profile is transformed to correspond to the instantaneous wave, the total horizontal water velocity at an elevation is the sum of the current vector and the wave velocity vector at that elevation.

35.2.4 Current-Induced Forces

Forces induced by currents on offshore structures are usually modeled as constant. These forces are nonlinear and, thus, the interaction with waves and with structure motions must be considered. The members of a structure in a steady flow experience a drag force proportional to the square of the flow velocity

$$f = \frac{1}{2}\rho U^2 A C_d \,, \tag{35.92}$$

where ρ is the mass density of the fluid, U is the flow velocity, A is the area of the member, and C_d is the drag coefficient. The drag coefficient is a function of the Reynolds number (Re) and the relative roughness of the surface (e) (Sect. 35.1.6).

This same drag force is a component of the force on a member in the oscillating flow, in waves calculated using the Morison equation (Sect. 35.1.3). In oscillating flows the drag coefficient is also a function of the Keulegan–Carpenter(KC) number already discussed in Sect. 35.1.3. Note also that the combination of the steady-current velocity and the oscillating wave-induced velocity causes an increase in the average drag force on a body due to the quadratic nonlinearity of the drag term in the Morison equation.

A current also induces a nonlinear time-varying force transverse to the direction of the current known as the lift force

$$f_L = \frac{1}{2}\rho U^2 A C_L \,, \tag{35.93}$$

where C_L is the lift coefficient. The lift force even in a steady flow is time varying. Thus, the lift coefficient reflects the RMS lift force or the maximum lift force over a period encompassing a number of oscillations of the lift force. These transverse forces are also referred to as vortex-induced forces and can cause large deflections of slender members free to vibrate known as VIV or large deflections of flexible structures known as vortex-induced motion (VIM) when the shedding period of vortices in the wake of the structure is close to natural periods of structure vibration. The shedding of vortices in the wake of a body also induces unsteady forces in line with the current direction. These unsteady inline forces are generally smaller than the transverse forces and much smaller than the average inline current force. The unsteady inline forces usually occur at twice the frequency of the transverse forces but can induce significant motions when they occur at a frequency close to the natural frequency of vibration of the structure.

Structures which appear transparent to currents can induce significant current blockage. The structure causes some of the current to go around the structure rather than through it. Current flows through these structures at a reduced velocity. Accounting for the reduced currents may be of interest for the design structures, especially when they accommodate a large number of conductors or risers. Current speed reduction factors ranging between 7 and 9 for jacket-type structures are provided in [35.67]. These factors are applied to the undisturbed current profile to obtain the current to use in force calculations.

The wake of an object located within a few diameters upstream of another object will influence the forces on the downstream object. Shielding will reduce the drag force on the downstream object and shear in the wake will induce lift forces on the downstream object. The drag and lift forces will depend on the spacing of the objects. The average current speed U_w in the turbulent wake of a cylinder at a point x downstream and y transverse to the current per [35.68] is

$$U_w = U - k_2 U \sqrt{\frac{C_d D}{x_s}} e^{-0.693\left(\frac{y}{b}\right)^2}, \tag{35.94}$$

$$x_s = x + \frac{4D}{C_d}, \tag{35.95}$$

$$b = k_1 \sqrt{C_d D x_s}, \tag{35.96}$$

where the empirical coefficient $k_1 = 0.25$ and coefficient $k_2 = 1.0$, and D is the upstream cylinder diameter and C_d is the upstream cylinder drag coefficient. This current in the wake is used to calculate the drag force on the downstream object.

The shielding may be beneficial in that it reduces forces on the downstream object, or it may be detrimental if the reduced forces lead to reduced distance between the objects which leads to clashing. This clashing is of particular concern for closely spaced risers in deep water.

For configurations comprising multiple closely spaced cylinders, experimental data possibly supported by computational fluid dynamics (CFD) should be used. Otherwise, interference effects should be ignored where beneficial and experiments should be considered where clashing is a concern.

Forces due to currents are calculated using the Morison equation (Sect. 35.1.6). The superposition of current and wave velocities should be considered when the local amplitude of the wave-induced water motion is larger than the radius of the member under consideration. The current vector and wave particle velocity vector should be added before the force is computed. Due to the nonlinear nature of the drag force in the Morison equation, currents can affect dynamic forces and should thus be modeled when dynamics is important.

35.2.5 Vortex-Induced Vibrations

VIVs can be caused by the flow of any fluid past a structure. Vortices are shed in the wake of the structure causing force transverse to the direction of the flow which may cause motions transverse to the flow which in turn may reinforce the vortex shedding. This may lead to large oscillations particularly in long slender elements normal to their long axis.

Parameters that influence these VIVs include: slenderness (L/D), mass ratio ($m^* = m/\left(\frac{1}{4}\pi\rho D^2\right)$), damping ration ($\zeta$), Reynolds number ($\mathrm{Re} = UD/\nu$), reduced velocity ($V_R = U/f_n D$), and flow properties, such as oscillations, turbulence (σ_U/U), and profile, where L is the member length, D is the member diameter, m is the mass per unit length, ζ is the ratio between damping and critical damping, ρ is the fluid density, ν is the fluid kinematic viscosity, U is the mean fluid speed, f_n is the natural frequency of the member, and σ_U is the standard deviation of the flow speed.

For the steady flow, the vortex shedding frequency is

$$f_s = \mathrm{St}\frac{U}{D}, \tag{35.97}$$

where St is the Strouhal number. For a smooth fixed circular cylinder in the steady flow, the Strouhal number is a function of the Reynolds number as shown in Fig. 35.13.

Table 35.2 Efficiency of helical strakes and helical wires (after [35.52])

	Number of windings	Height of strakes	Pitch	Lift coefficient C_L	Drag coefficient C_d
Helical strakes	3	.11D	4.5D	0.238	1.6
	3	.11D	15D	0.124	1.7
Helical wires	3–4	.118D	5D	0.2	1.17
	3–4	.118D	10D	0.2	1.38
No spoilers				0.9	0.7

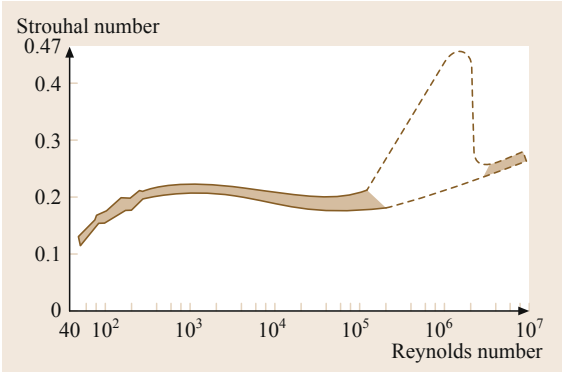


Fig. 35.13 Strouhal number for a circular cylinder as a function of the Reynolds number (after [35.69])

When the vortex shedding frequency coincides with natural frequencies of vibration, large-amplitude motions may arise. This phenomenon is called lock-in because the motions and vortex shedding will occur at the natural frequency of the body over a range of flow speed. The natural frequency of the body may be different from that observed in still water due to flow-induced variation in added mass. Lock-in can occur in line with the flow as well as transverse to the flow. Lock-in occurs when the reduced velocity V_R is close to the inverse of the Strouhal number.

The mass ratio m^* is an indication of the relative importance of the body mass and the fluid forces. A high mass ratio indicates that the fluid force is relatively small compared to the body inertia. For situations with a low mass ratio, such as a pipe in water, the lock-in range is $3 < V_R < 16$ and for a high mass ratio, such as wire in air, the lock-in range is $4 < V_R < 8$.

VIV on a slender member will cause fatigue damage, may increase the drag coefficient of the member causing additional load and possibly clashing, and may excite vibration and VIV of downstream objects.

The responses of structures to flow-induced vibrations can be estimated via model tests, response-based methods, force-based methods, or flow-based methods. Response-based methods [35.70, 71] estimate steady-state responses of systems as a function of structural and hydrodynamic parameters making use of conservative envelopes of experimental data. Force-based methods make use of a structural model excited by inertia and damping forces determined from empirical data. They attempt to model the flow explicitly and include CFD which solve the Navier–Stokes equations coupled with a structural model.

Vortex shedding may also induce motions in the hull of floating systems. These motions are important because these motions can influence mooring and riser design. Cross-flow oscillations are generally larger than inline oscillations and are, thus, of most interest. For large circular cylinders, like Spars, the amplitude of VIMs can be up to 80% of the diameter. This can be reduced to 40% or less with suppression devices. Model tests are generally used to estimate the motions of these structures with and without suppression devices.

The suppression of VIV is generally done with devices which spoil the wake and reduce the coherence of vortex shedding along the structure. This can be done with fins which weather a vane and help keep the wake symmetric and reduce the size of shed vortices. The wake can also be spoiled by making the member irregular along its length by wrapping it with helical strakes or wires (Table 35.2).

35.3 Wind Loads

Offshore platforms are exposed to wind fields which pose a serious overload threat to the overall platform structural system as well as to topside components, such as helidecks, drilling derricks, cranes, and living quarters. Fluctuating wind loads as well as VIVs can cause fatigue damage and lead to failures. Reliable design to resist these failure mechanisms requires a good understanding of the wind environment and as well as the mechanisms of wind loading.

The wind speed and direction vary in space and time. Sufficient data to describe the spatial and time variations in a great detail are rarely available and for most applications are unnecessary. Wind field descriptions are based on statistical parameters, such as the

mean and the standard deviation of the speed, as well as the mean direction. Both length scales and time scales influence the definition of these statistical parameters.

Both local and global wind effects on platforms must be considered. The local effects influence the design of the deck structure and equipment. Global effects, such as overturning moments and total lateral loads, drive the design of foundations and mooring systems. Both mean and fluctuating forces must be considered. Compliant structures such as TLPs have natural frequencies of motion in the horizontal plane which can be excited by wind forces and make them sensitive to the slowly varying fluctuations of the wind.

Extreme and normal wind conditions should be specified when designing offshore structures. The response of the structure to wind will influence which conditions are of interest. Three-dimensional spatial scales of wind are related to the durations of turbulent gusts. Thus, gusts of a few seconds duration will have less influence on a large structure than will gusts of a minute duration. The time variation of the wind, usually characterized by a spectrum, should be considered for structures with appreciable dynamic response.

A description of the wind environment for a location should include an estimate of extreme wind speeds in specified directions and specified averaging times as a function of the recurrence interval. Information about the measurement sites and a description of the measured data used to generate the environment description should be available. The frequency of exceedance of specified thresholds from specified directions during the service life should be estimated. The types of storms causing high winds should be described. Similar information should be provided for normal or short-term conditions providing descriptions by month or season. Figure 35.14 is an example of such a monthly description.

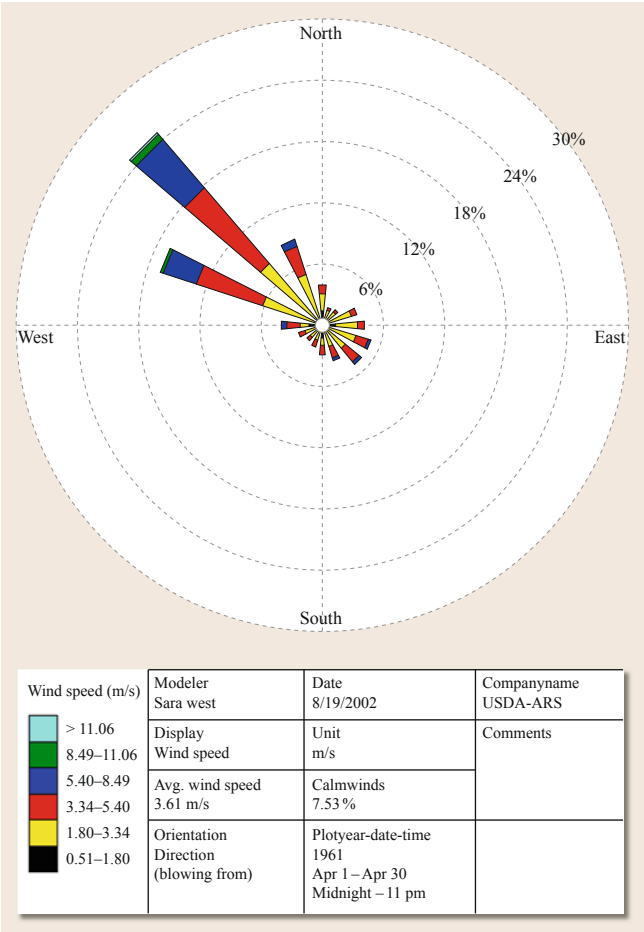


Fig. 35.14 Wind rose showing the statistical variation of wind direction and speed during the month. The length of each *line* indicates the percentage of time the wind blew from that direction during the month. The *colors* indicate different wind-speed bins (after [35.72])

35.3.1 Wind-Speed Profile

On length scales typical of even the largest offshore structures, the mean and standard deviation of the wind speed, averaged over durations of the order of an hour, do not vary horizontally, but they do change with elevation. For averaging durations shorter than an hour, there will be periods with higher mean speeds and the spatial variations will increase. To be meaningful, a wind-speed value must be qualified by an elevation and a duration over which it is averaged. An elevation of 10 m above the mean sea level is used as a standard reference height.

According to *Planning* [35.67], the mean wind-speed profile $U_w(z)$ under storm conditions can be more accurately described by a logarithmic profile as given in (35.98) than by the traditional power-law profile

$$U_{w,1h}(z) = U_{w0}\,[1 + C \ln (z/z_r)\,]\,, \qquad (35.98)$$

where $U_{w,1h}(z)$ is the 1 hour sustained wind speed at a height z above the mean sea level; U_{w0} is the 1 hour sustained wind speed at the reference elevation z_r and is the standard reference speed for sustained winds; C is a dimensionless coefficient, the value of which is dependent on the reference elevation and the wind speed, U_{w0}. For $z_r = 10\,\text{m}$, $C = 0.0573\sqrt{1 + 0.15U_{w0}}$, where U_{w0} is in units of meters per second (m/s); z is the height above the mean sea level; and z_r is the reference elevation above the mean sea level ($z_r = 10\,\text{m}$).

For the same conditions, the mean wind speed for averaging times shorter than 1 hour may be expressed

by (35.99) using the 1 hour sustained wind speed $U_{w,1h}(z)$ of (35.98)

$$U_{w,T}(z) = U_{w,1h}(z)\left[1 - 0.41 I_u(z) \ln\left(\frac{T}{T_0}\right)\right], \tag{35.99}$$

where $U_{w,T}(z)$ is the sustained wind speed at height z above mean sea level, averaged over a time interval $T < 3600\,\mathrm{s}$; $U_{w,1h}(z)$ is the 1 hour sustained wind speed at height z above mean sea level, see (35.98); T is the time-averaging interval with $T < T_0 = 3600\,\mathrm{s}$; T_0 is the standard reference time-averaging interval for the wind speed of $1\,\mathrm{h} = 3600\,\mathrm{s}$; $I_u(z)$ is the dimensionless wind turbulence intensity at a height z above the mean sea level, given by (35.100), where U_{w0} is in m/s

$$I_u(z) = 0.06\,(1 + 0.043 U_{w0})\left(\frac{z}{z_r}\right)^{-0.22}. \tag{35.100}$$

Note that the equations in this section are derived from curve fitting through available data [35.73] and contain numerical constants that are only valid in the SI units of meters and seconds. The above equations are not valid for the description of winds in short-lived events, such as squalls and tornados, since the duration is often less than 1 hour. Adjustments to the wind profile at a particular location or under certain conditions can be made when specific appropriate measured data from an offshore location are available (i. e., measured data for the kind of event used in design).

35.3.2 Wind Spectra and Gusts

The wind spectrum characterizes the time-varying properties of the wind. These variations are due to boundary-layer turbulence which depends on the wind speed and the thermal stability of the air. The following wind spectrum formulation is based on measurements under conditions of nearly neutral thermal stability as documented in [35.73].

Equation (35.101) describes the spectrum of wind speed at a point in space and is analogous to the spectra used to describe the time-varying kinematics of waves in Sect. 35.1.4.

$$S(f,z) = \frac{320\left(\frac{U_{w0}}{U_{ref}}\right)^2\left(\frac{z}{z_r}\right)^{.45}}{\left(1+\tilde{f}^n\right)^{5/(3n)}}, \tag{35.101}$$

where $S(f,z)$ is the wind spectrum (spectral or energy density function) at frequency f and elevation z in m^2/s; U_{w0} is the 1 h sustained wind speed at the reference elevation z_r (the standard reference speed for sustained winds); U_{ref} is the reference wind speed, $U_{ref} = 10\,\mathrm{m/s}$; f is the frequency in cycles per second (hertz) over the range $0.00167\,\mathrm{Hz} < f < 0.5\,\mathrm{Hz}$; z is the height above the mean sea level; z_r is the reference elevation above the mean sea level ($z_r = 10\,\mathrm{m}$); $\tilde{f}$ is a non dimensional frequency defined by (35.102) where the numerical factor 172 has the unit of second (s)

$$\tilde{f} = 172 f\left(\frac{z}{z_r}\right)^{2/3}\left(\frac{U_{w0}}{U_{ref}}\right)^{-.75}. \tag{35.102}$$

n is a coefficient equal to 0.468.

Integrating the spectrum over the applicable frequency range yields the standard deviation of the wind speed. When comparing the spectrum to data, comparable frequency ranges should be used.

The spatial and temporal wind fields are correlated. Thus, the wind spectrum should be complemented with a description of the spatial coherence. It is generally conservative to assume that the wind speed is fully correlated over a complete structure. But it is reasonable to take advantage of the reduced correlation when estimating loads on structures. Equation (35.103) describes the coherence between two points P_1 and P_2, with positions in the wind directions of x_1 and x_2, and position transverse to the wind directions of y_1 and y_2, and elevations above the mean water of z_1 and z_2

$$F_{Coh}(f, P_1, P_2) = \exp\left[-\frac{1}{U_{w0}}\left(\sum_{i=1}^{3} {A_i}^2\right)^{1/2}\right], \tag{35.103}$$

where $F_{Coh}(f, P_1, P_2)$ is the coherence function between turbulence fluctuations at P_1 and at P_2; U_{w0} is the 1 hour sustained wind speed at 10 m above the mean sea level in meters per second (m/s); A_i is a function of frequency and the position with units of m/s calculated from (35.104)

$$A_i = \alpha_i f^{r_i} D_i^{q_i}\left(\frac{z_g}{z_r}\right)^{-p_i}, \tag{35.104}$$

where f is the frequency in Hertz (Hz), D_i is the distance, measured in meters (m), between points P_1 and P_2 in the x, y, and z directions for $i = 1, 2$, and 3, respectively (Table 35.3); z_g is the geometrical mean height of the two points, $z_g = \sqrt{z_1 z_2}$; z_r is the reference elevation above the mean sea level, $z_r = 10\,\mathrm{m}$; α_i, p_i, q_i, r_i are coefficients given in Table 35.3.

The concept of a wind spectrum is only applicable to steady wind conditions. The time and spatial variation of the wind speed in a squall or tornado cannot be described by a wind spectrum. The analysis of forces

Table 35.3 Coefficients in (35.103) for points P_1 and P_2

i	D_i	a_i	p_i	q_i	r_i
1	$\lvert x_1 - x_2 \rvert$	2.9	0.4	1.00	0.92
2	$\lvert y_1 - y_2 \rvert$	45.0	0.4	1.00	0.92
3	$\lvert z_1 - z_2 \rvert$	13.0	0.5	1.25	0.85

and responses caused by squalls and tornados requires the specification of a time series of wind velocity.

Wind speeds are frequently classified as sustained winds or gusts. Sustained winds are usually hourly averages and gusts are usually the hourly maxima of averages of a minute or less. In either case, the averaging duration should always be stated as well as the elevation. The duration of interest depends on the dimensions and natural period of the structure being analyzed. Small structures should be designed for a shorter gust duration (and hence a higher gust wind speed) than a larger structure. Gusts naturally arise due to the turbulence in wind and can be considered local maxima rather than transient wind conditions. Gust wind speeds are generally derived from the wind spectrum.

Equation (35.99) may be used to calculate the gust speed for various gust durations. Squalls, thunderstorms, downbursts, tornados, and water spouts are relatively short-lived phenomena which can induce extreme winds. The ratio of the maximum gust wind speed to hourly mean wind speed at any one location in these examples can be large. Turbulence alone generates gusts during periods of high mean wind speed, but in this case the ratio of the maximum gust wind speed to hourly mean wind speed over the sea is typically less than about 1.5.

35.3.3 Steady-State Forces

For fixed structures, global wind forces are generally much less important than those caused by waves and currents. Wind forces on individual components of these structures can be significant, however. Global forces on structures are determined using a time-averaged wind speed in the form of a sustained wind speed. For the design of individual structural components, a time-averaged wind speed can also be adequate, but the averaging duration should be reduced to allow for the smaller turbulence scales that can affect individual components. The wind in a 3-s gust is appropriate for determining the maximum static wind load on individual members; 5-s gusts are appropriate for maximum total loads on structures whose maximum horizontal dimension is less than 50 m; and 15-s gusts are appropriate for the maximum total static wind load on larger structures.

Wind acts on a structure above the water, as well as on any equipment, deck houses, bridges, flare booms, and derricks that are located on the topsides. The height of the component above the sea level should be taken into account when estimating the wind speed.

Table 35.4 Wind shape coefficients

Area	Shape coefficient C_S
Beams	1.5
Sides of buildings	1.5
Cylindrical sections	0.5
Overall projected area of platform	1.0

Equation (35.99) can be used to calculate the gust speed for various gust durations for a given elevation z.

The steady wind pressure q is calculated as

$$q = \frac{1}{2}\rho U^2 \,. \tag{35.105}$$

And the force F, acting normal to the body axis or surface, is calculated as

$$F = qC_S A \sin\alpha \;\;, \tag{35.106}$$

where ρ is the density of air (1.22 kg/m^3 for standard temperature and pressure), U is the wind speed (m/s), A is the object area (m^2), α is the angle between the direction of the wind and the axis of the exposed member or surface, and C_S is the shape coefficient.

For smooth circular tubular structures, the shape coefficient is $C_S = 0.65$ for the Reynolds number $> 5 \times 10^5$, and $C_S = 1.2$ for the Reynolds number $< 5 \times 10^5$. A good collection of shape coefficients for long bodies and bodies of finite length is presented in [35.52]. Table 35.4 provides shape coefficients recommended in [35.67] for several bodies for the perpendicular wind approach angle.

On most structures, objects exposed to wind loads are closely spaced and shield each other from the wind depending on the wind direction. If a detailed model of the wind-loaded objects is used, shielding coefficients should be included to account for this interaction. Wake models similar to that presented in (35.94) can be used to estimate shielding.

Wind tunnel tests should be considered to determine pressures and resulting loads on complex structures. Testing should include the variation of wind speed with elevation as well as turbulence.

35.3.4 Unsteady Forces

Modeling the time and spatial variation of the wind should be considered for structures and components which respond dynamically to wind loads. A dynamic analysis of a structure is generally necessary when the wind field contains energy at frequencies near the natural vibration frequencies of the structure; this is

generally the case for compliant bottom-founded platforms as well as for floating structures. Time-varying wind forces can cause resonant surge, sway, and yaw motion of floating anchored structures.

Good load and response estimates can be obtained from boundary-layer wind tunnels and from CFD. Reasonable results can be obtained by simulating forces and responses using a time-domain synthesis of the wind spectrum, (35.101), and adding it to the mean wind. The spatial variation of the wind speed can also be modeled by making use of the coherence function (35.103).

The instantaneous wind force can be calculated by the summation of the forces on individual members exposed to wind. The pressure q can be estimated by

$$q = \frac{1}{2}\rho_a \left|U_{w,T} + u - \dot{x}\right| \left(U_{w,T} + u - \dot{x}\right) , \quad (35.107)$$

where $U_{w,T}$ is the mean wind velocity, u is the gust velocity which may vary spatially as well as with elevation, $\dot{x}$ is the velocity of the body, and ρ_a is the mass density of air ($1.226\,\mathrm{kg/m^3}$ for dry air at 15 °C).

When the structural velocity $\dot{x}$ is negligible compared to the wind speed, the wind pressure can be linearized to

$$q = \frac{1}{2}\rho_a {U_{w,T}}^2 + \rho_a U_{w,T} u . \quad (35.108)$$

These time-varying pressures can be used with (35.106) to calculate time-varying forces.

A special case of dynamic response is VIV of relatively slender structures subjected to steady winds in which alternate vortex shedding excites components. Components of fixed steel offshore structures can be exposed to wind VIV during construction and transportation. Flare structures and telecommunication towers can also be susceptible to wind VIV throughout their lives. Vortex shedding and VIV in wind are much the same as for currents. See Sec. 35.2.5 for a discussion of these phenomena.

35.4 Model Tests

Regardless of how good we think the theoretical predictions of a physical quantity is, there will always be errors, however small, due to the particular modeling we use, or errors in the numerical analysis we perform. We may also have truncation and round-off errors, not to mention any human errors that may be present. All these potential errors can be assessed in a comprehensive way by conducting model tests. However, one has to realize that model tests themselves are not necessarily free of potential errors either, mainly due to the fact that the tests are conducted in a rather finite domain that may be the cause of reflections from tank walls, the wavemaker and absorption beach(es) that in turn may contaminate the measured data.

Physical modeling is the most important area that dimensional analysis can be applied to. By physical modeling, we refer to the technique of reproducing a physical phenomenon on a greater or smaller scale. The motions of or the wave loads on offshore platforms that can be measured by means of model tests in a test basin are the examples to model testing at a smaller scale. There are two basic issues that need to be addressed in any model testing: 1) how do we conduct the experiments so that the data obtained at the model scale are accurate, and 2) how do we extrapolate the data to the prototype scale?

In most offshore engineering problems, the three dimensions, length (L), mass (M), and time (T), are the fundamental dimensions. All physical quantities can then be measured in terms of these three fundamental dimensions. However, some physical quantities, such as temperature and angle, may occasionally be used as one of the fundamental dimensions. If a measurement results in a real number, which can directly be compared to one of these fundamental dimensions, then the measurement is called a direct measurement. For example, a distance of 5 m is a direct measurement since its unit is one of the fundamental units, namely the length. On the other hand, if a measurement is a result of various comparisons that give a real number, whose dimension is a combination of two or more fundamental units, then the measurement is called a derived measurement. A typical example is the velocity, since both length and time have to be measured to obtain the velocity. In derived measurements, there is always a function that expresses the relation between the direct measurements. In the case of velocity measurements, for example, this function is $f(x,t) = \mathrm{d}x/\mathrm{d}t$.

35.4.1 Principles and Similarity Laws

The scale model of any prototype system has to satisfy certain conditions called the laws of similarity or similitude so that the behavior of the prototype phenomenon can accurately be reproduced. The similarity can refer to one or more of:

1. Geometric similarity (refers to length)
2. Kinematic similarity (refers to velocity)
3. Dynamic similarity (refers to force).

The wave loads or motions of offshore platforms or any other physical quantity of interest have to obey one or more of these similarity laws in order that one can properly conduct experiments [35.74].

It is necessary at this point to introduce various force mechanisms that are present in the flow of real fluids. This necessity arises because, under certain conditions of the specific problem being investigated, some force mechanisms may be dominant over the other. As a result, we may isolate, or even neglect, the smaller forces by choosing a proper similarity law. The three principal types of internal force mechanisms in a real fluid are:

1. Inertial force
2. Gravitational force, and
3. Viscous force.

The inertial force is due to the fluid particle acceleration and is proportional to $\rho u(\partial u/\partial x)$, where ρ is the fluid mass density, and u is the particle velocity component in the x direction (other terms are left out without loss of generality). The gravitational force is due to the weight of the fluid itself and is proportional to ρg, where g is the gravitational acceleration. The viscous force is due to the difference between the shear forces acting on a fluid element, and is proportional to $\mu\partial^2 u/\partial x^2$, where μ is the dynamic viscosity coefficient (again other terms are left out). All these forces are included in the Navier–Stokes equations.

The ratios of the internal forces (per unit volume) with one another give us the relative importance or dominance of these forces. For example, we can write the ratio of

$$\frac{\text{Inertial force}}{\text{Gravitational force}} \propto \frac{\rho u\frac{\partial u}{\partial x}}{\rho g} = \frac{u}{g}\frac{\partial u}{\partial x} \propto \frac{U^2}{g\ell} \,. \tag{35.109}$$

In (35.109), U is a characteristic velocity, ℓ is a characteristic length of the problem, and ρ and g are the fluid mass density and gravitational acceleration, respectively. It is seen that the ratio is proportional to $U^2/g\ell$. Similarly, we can write the ratio of

$$\frac{\text{Inertial force}}{\text{Viscous force}} \propto \frac{\rho u\frac{\partial u}{\partial x}}{\mu\frac{\partial^2 u}{\partial x^2}} \propto \frac{U\ell}{\upsilon} \,, \tag{35.110}$$

where $\upsilon = \mu/\rho$ is the kinematic viscosity coefficient. The third ratio, the ratio of the gravitational force to viscous force, can be obtained by a combination of (35.109) and (35.111), and, therefore, it need not be considered. The square root of (35.109) is the Froude number Fr and (35.110) itself is the Reynolds number Re. So, for small Fr, for instance, we can say that the flow is gravity dominated, and for large Re, for instance, we can say that the flow is inertia dominated or viscosity is negligible (or the inviscid fluid assumption is a good one).

If we require that the flow about a body be similar fully both in the model and prototype scales, then it is necessary that both the Fr and Re numbers be the same for the model and prototype. In other words, the Froude number of the model scale must be the same as the Froude number of the prototype scale, that is, $\mathrm{Fr_m} = \mathrm{Fr_p}$, and similarly for the Reynolds number, that is, $\mathrm{Re_m} = \mathrm{Re_p}$, where the subscripts m and p refer to the model and prototype, respectively. Unless under very special circumstances, it is not practically possible to scale a prototype by holding both the Fr and Re numbers the same. To see this, consider the Froude number and the Reynolds number, keeping both constant for the model and prototype scales, that is

$$\begin{aligned} \mathrm{Fr_m} &= \frac{U_\mathrm{m}}{\sqrt{g_\mathrm{m}\ell_\mathrm{m}}} = \mathrm{Fr_p} = \frac{U_\mathrm{p}}{\sqrt{g_\mathrm{p}\ell_\mathrm{p}}} \,, \\ \mathrm{Re_m} &= \frac{U_\mathrm{m}\ell_\mathrm{m}}{\upsilon_\mathrm{m}} = \mathrm{Re_p} = \frac{U_\mathrm{p}\ell_\mathrm{p}}{\upsilon_\mathrm{p}} \,. \end{aligned} \tag{35.111}$$

The first equation in (35.111) requires that if $\ell_\mathrm{m} < \ell_\mathrm{p}$, then $U_\mathrm{m} < U_\mathrm{p}$, assuming that $g_\mathrm{m} = g_\mathrm{p}$, whereas the second requires that if $\ell_\mathrm{m} < \ell_\mathrm{p}$, then $U_\mathrm{m} > U_\mathrm{p}$, assuming that $\upsilon_\mathrm{m} = \upsilon_\mathrm{p}$. Therefore, unless the gravitational acceleration is considerably increased during the experiments (this is not uncommon considering, for example, some soil-mechanics experiments conducted in a centrifugal apparatus) or ν_m is considerably decreased or some combination of the two, the ratio of Fr to Re can not be held constant simultaneously.

One of the two similarity laws, that is, Froude's law, given by the first equation in (35.111), or Reynolds' law, given by the second, has to be used depending on the particular application in mind. Froude's law (or scale) is generally used in conjunction with the experiments on platform motions and surface waves. Reynolds' law or scale, on the other hand, is used in experiments related to deeply submerged bodies, pipe flow, etc.

Another force mechanism is the surface tension, although it is not a principle one. For example, if we consider the ratio of the inertial force to surface tension force (per unit volume), we obtain

$$\frac{\text{Inertial force}}{\text{Surface tension force}} \propto \frac{\rho u\frac{\partial u}{\partial x}}{\frac{\sigma}{L^2}} \propto \frac{\rho U^2}{\frac{L\sigma}{L^2}} = \frac{\rho L U^2}{\sigma} \,, \tag{35.112}$$

Table 35.5 Some of the frequently used dimensionless numbers

Dimensionless number	Symbol	Definition
Froude number	Fr	$U/\sqrt{gL}$
Reynolds number	Re	UL/υ
Strouhal number	St	fD/U
Keulegan–Carpenter number	KC	UT/D
Euler number	Eu	$p/\rho U^2$
Cauchy number	Cy	$\rho U^2/E$
Ursell number	Ur	HL^2/h^3
Weber number	We	$\rho U^2 L/\sigma$

where σ is the surface tension (lb/ft or N/m). The last equality is called the Weber number, that is

$$\mathrm{We} = \frac{\rho L U^2}{\sigma} \,. \tag{35.113}$$

There are a number of other important numbers used in offshore engineering: the Strouhal number, $\mathrm{St} = fD/U$, where f is the vortex shedding frequency and D is the characteristic length, for example, diameter, and KC number, $\mathrm{KC} = UT/D$, where T is the wave period, are two examples. Some of the dimensionless numbers used are shown in Table 35.5, where U is the velocity, g is the gravitational acceleration, L or D is the characteristic length, f is the cyclic frequency, υ is the kinematic viscosity of water, T is the wave period, H is the wave height, h is the water depth, E is Young's modulus, p is the pressure, and σ is the surface tension.

35.4.2 Scaling of Loads

In planning model tests, one has to decide which scaling law to be used. In most offshore engineering model tests, the Froude scaling law is used as offshore platforms mostly encounter gravity waves in nature. This means that during such experiments, one must make sure to satisfy that $\mathrm{Fr_m} = \mathrm{Fr_p}$. Having also decided what length scale, $S_\mathrm{L} = L_\mathrm{m}/L_\mathrm{p}$, to use, one can scale the other physical quantities, accordingly. For example, if we want to scale the wave forces by using Froude's law, so that $U_\mathrm{m} = U_\mathrm{p} S_\mathrm{L}^{1/2}$, we can first determine the scaling of time and acceleration

$$\begin{aligned} U_\mathrm{m} &= \sqrt{S_\mathrm{L}} U_\mathrm{p} = \frac{L_\mathrm{m}}{t_\mathrm{m}} = \sqrt{S_\mathrm{L}} \frac{L_\mathrm{p}}{t_\mathrm{p}} \Rightarrow t_\mathrm{m} = \sqrt{S_\mathrm{L}} t_\mathrm{p} \,, \\ a_\mathrm{m} &= \frac{U_\mathrm{m}}{t_\mathrm{m}} = \frac{\sqrt{S_\mathrm{L}} U_\mathrm{p}}{\sqrt{S_\mathrm{L}} t_\mathrm{p}} = a_\mathrm{p} \,, \end{aligned} \tag{35.114}$$

where t is the time. That the scales for accelerations are the same in the model and prototype scales is not surprising since the gravitational acceleration in the model and prototype scales is the same, $g_\mathrm{m} = g_\mathrm{p}$.

We also need to scale the masses of the model and the prototype. It is

$$S_\mathrm{m} = \frac{m_\mathrm{m}}{m_\mathrm{p}} = \frac{\rho_\mathrm{m} L_\mathrm{m}^3}{\rho_\mathrm{p} L_\mathrm{p}^3} = S_\rho S_\mathrm{L}^3 \,. \tag{35.115}$$

Therefore, the force scaling can be written as

$$S_\mathrm{F} = \frac{m_\mathrm{m} \dot{U}_\mathrm{m}}{m_\mathrm{p} \dot{U}_\mathrm{p}} = \frac{m_\mathrm{m}}{m_\mathrm{p}} = S_\mathrm{m} = S_\rho S_\mathrm{L}^3 \,, \tag{35.116}$$

since the acceleration scale is 1.0.

This method of obtaining the scaling for forces can be used on any other physical quantity to determine how it is scaled to the model. However, there is another way of obtaining the same scaling result. Let us show this by way of an example. Consider the dimension of force and write it as $[F] = (L, M, T^{-2}) = (1, 1, -2)$, that is, as a vector in the 3-D (L, M, T) space. Writing the functional form of force as $F = f(\rho, L, \mathrm{g})$, and taking the set (ρ, L, g) as a dimensionally independent set of quantities, one can use the Pi theorem [35.75] to obtain the single dimensionless π

$$\begin{aligned} [F] &= (L, M, T^{-2}) = (1, 1, -2) = [\rho]^p \, [L]^q \, [\mathrm{g}]^r \\ &= (-3, 1, 0)^p \, (1, 0, 0)^q \, (1, 0, 2)^r \\ &\text{or} \\ p = 1, q = 3, r = 1 &\Rightarrow \frac{F}{\rho L}^3 \mathrm{g} = \pi \Rightarrow \frac{F_\mathrm{m}}{\rho_\mathrm{m} L_\mathrm{m}^3} \\ &= \frac{F_\mathrm{p}}{\rho_\mathrm{p} L_\mathrm{p}^3} \Rightarrow S_\mathrm{F} = S_\rho S_\mathrm{L}^3 \,. \end{aligned} \tag{35.117}$$

This is the same result given by (35.116).

In Table 35.6, we show the model scales for some of the physical quantities that would be of interest during an offshore engineering experiment (L: length, M: mass, T: time, $S_\mathrm{L} = L_\mathrm{m}/L_\mathrm{p}$ is the length ratio, and $S_\rho = \rho_\mathrm{m}/\rho_\mathrm{p}$ is the specific gravity of salt water). Many others are given in [35.74], but for the special case of $S_\rho = 1.0$.

35.4.3 Elastic Structures

Model tests of elastic structures bring additional complexities to the problem as it is not practical in many cases to correctly scale the stiffness of the structure at the model scale. Such elastic structures could be TLP tendons, oil-production risers, catenary-mooring lines, etc., but the structure itself may be elastic as well, especially very large floating structures (VLFS) [35.76]. To see this, consider a tubular, beam-like object, and write the scalings, from Table 35.6, of the axial and bending

Table 35.6 Some model scaling obtained by using Froude's scaling law

Quantity	Dimension	Scale
Length	L	S_L
Mass	M	$S_\rho S_L^3$
Mass moment of inertia	L^2M	$S_\rho S_L^5$
Moment of inertia of area	L^4	S_L^4
Time	T	$S_L^{1/2}$
Acceleration	LT^{-2}	1
Velocity	LT^{-1}	$S_L^{1/2}$
Linear spring constant	MT^{-2}	$S_\rho S_L^2$
Axial stiffness	LMT^{-2}	$S_\rho S_L^3$
Bending stiffness	L^3MT^{-2}	$S_\rho S_L^5$
Work	L^2MT^{-2}	$S_\rho S_L^4$
Power	L^2MT^{-3}	$S_\rho S_L^{7/2}$
Energy	L^2MT^{-2}	$S_\rho S_L^4$
Force	LMT^{-2}	$S_\rho S_L^3$
Moment	L^2MT^{-2}	$S_\rho S_L^4$
Stress	$L^{-1}MT^{-2}$	$S_\rho S_L$
Pressure	$L^{-1}MT^{-2}$	$S_\rho S_L$
Modulus of elasticity	$L^{-1}MT^{-2}$	$S_\rho S_L$

stiffness, and its diameter as

$$S_{EA} = \frac{(EA)_m}{(EA)_p} = \frac{F_m}{F_p} = S_F = S_m = S_\rho S_L^3 ,$$
$$S_{EI} = \frac{(EI)_m}{(EI)_p} = \frac{F_m L_m^4}{L_m^2}\frac{L_p^2}{F_p L_p^4} = S_F S_L^2 = S_\rho S_L^5 ,$$
$$\frac{D_m}{D_p} = \frac{L_m}{L_p} = S_L . \qquad (35.118)$$

Either the bending or shear stress, τ, is scaled as

$$S_\tau = \frac{F_m}{L_m^2}\frac{L_p^2}{F_p} = \frac{S_F}{S_L^2} = \frac{S_\rho S_L^3}{S_L^2} = S_\rho S_L . \qquad (35.119)$$

And the area moment of inertia is scaled as

$$S_I = \frac{I_m}{I_p} = \frac{L_m^4}{L_p^4} = S_L^4 . \qquad (35.120)$$

Therefore, Young's modulus is scaled as

$$S_E = \frac{E_m}{E_p} = \frac{S_{EI}}{S_I} = \frac{S_\rho S_L^5}{S_L^4} = S_\rho S_L . \qquad (35.121)$$

There are basically two problems one encounters as a result of these scalings. One is that it is very difficult to scale the geometry because D_m is typically 50–100 times smaller than D_p. The other is that the modulus of elasticity of the material used in the experiments is also 50–100 times smaller than what is used in the prototype, for example, steel. Some engineering solutions to these kinds of problems is necessary to conduct the experiments. For example, *Dillingham* [35.77] suggested in modeling the tendons of a TLP that all parameters are correctly scaled for the tendon except the axial stiffness that is modeled by a spring placed at the top or bottom of the tendon to provide the correct stiffness. Even this approximation involves some errors that must be carefully assessed.

The structural rigidity can sometimes be reduced by distorting the structure as discussed, for example, by [35.78]. Another type of a distorted model can be achieved in offshore model tests in rather shallow water where the horizontal length dimensions are much larger than the vertical length dimension. In such cases, two different model scale ratios are used in the experiments [35.74, 79]. Let us again set the horizontal length scale to S_L but set the vertical length scale to $S_V = h_m/h_p$, where h_m and h_p are the water depths in the model and prototype scales, respectively, and S_V, in general, is different from S_L. Next, let us consider the linear shallow-water phase speed, $c_p = (gh)^{1/2}$, and use S_V to write the Froude scaling law to obtain

$$\frac{c_m}{\sqrt{gh_m}} = \frac{c_p}{\sqrt{gh_p}} \Rightarrow c_m = \sqrt{S_V} c_p , \qquad (35.122)$$

and since $c = \lambda/T$, where λ is the wavelength, scaled by S_L, $S_L = \lambda_m/\lambda_p$, we have the following scaling for the wave period

$$\frac{T_m}{T_p} = \frac{\lambda_m}{C_m}\frac{C_p}{\lambda_p} = \frac{S_L}{\sqrt{S_V}} \Rightarrow T_m = \frac{S_L T_P}{\sqrt{S_V}} . \qquad (35.123)$$

If the water is not very shallow, of course, the full dispersion relation based on the linear wave theory can be used in deriving the corresponding scales that will now involve hyperbolic functions. Finally, recall that the accelerations scale as 1.0 if a single length scale is used according to the Froude scaling. In the distorted model used, however, that will not be the case. The accelerations in a distorted model would scale as

$$a_m = \frac{U_m}{T_m} = \frac{\sqrt{S_V} U_p \sqrt{S_V}}{S_L T_p} \Rightarrow \frac{a_m}{a_p} = \frac{S_V}{S_L} . \qquad (35.124)$$

35.5 CFD Tools

With the very rapid advancement in personal and server-based computer hardware and numerical methods, as well as the development of commercial software in recent years, CFD is becoming a viable tool in the computations of wave, current, and wind loads on and the resulting motions of offshore platforms. Here, we use the term CFD to mean the solution of the exact governing equations subject to the instantaneous boundary conditions. The governing equations could be the Reynolds-averaged Navier–Stokes (RANS) equations (steady or unsteady) or Euler's equations in which viscosity is assumed to be negligible, although the definition of CFD computations generally is reserved for viscous flows. We do this since there are many problems in offshore engineering in which the effect of viscosity is negligible.

Even though there is very rapid growth in computer hardware and software, the stumbling blocks in the use of CFD tools are concentrated [35.80] around the user-related problems (some of which are related to training and lack of user-friendly GUI) and theoretical problems (some of which are (1) continuation of computations after wave breaking, (2) radiation/reflection conditions for simulating wave tanks or open ocean conditions, (3) fluid–structure interaction including multiphase flow structural modeling, (4) higher order discretization methods, (5) viscous model selection, which one to use and when, among many other problems). Of course the speed at which the computations can be completed is another major issue even if parallel computing is heavily utilized. It is estimated that by the year 2030 the transistor count will be more than 10^{11} and the computing speed will be more than $1.0E^{+18}$ (Exascale). With the estimated increase in the computing speed, it is expected that the cost of running the same model-scale or full-scale case in the year 2030 would be 250 times less than what it was in the year 2013 for as many as 2500 cores.

With these shortcomings of the field at present, however, there are many commercially available or open-source software that are used to solve problems related to offshore platforms. Some of them are ANSYS AQWA, CD-ADAPCO, Reef3D, FLOW3D, and OpenFOAM. The last open-source software appears to be the most popular one currently, not only because it is free, but also because of its relative ease of use, flexibility, and speed. As examples to the use of OpenFOAM software, see [35.81] for the calculations of wave and current forces (viscous fluid) on a semisubmersible and [35.82] for the calculations of nonlinear (inviscid fluid) wave loads on a coastal bridge deck.

35.6 Extreme Response Estimation

Offshore structure design is based on proportioning a structure so that it resists actions (forces) which may occur during its design life with an appropriate reliability. For a random environment, such as that presented by offshore wind, wave, and current, the extreme loads or responses imposed by the environment are required to assess the fitness of the structure for its intended purpose. A target probability of exceedance in combination with a long-term distribution of the response is used to set these extreme responses.

Annual exceedance probabilities of 10^{-2} or 10^{-4} are used to characterize environmental forces for ultimate and accidental limit states. Scatter diagrams and joint probability density functions are used to describe the variability of forces from multiple sources or with multiple parameters of interest, such as frequency and direction. These diagrams or functions are used to generate combinations of environmental parameters which correspond to exceedance probabilities of interest.

The joint probability of occurrence is used to create combinations of wind, wave, and current conditions causing the extreme loads. For most fixed, tower, gravity, and caisson types of platforms, the design environmental load is predominantly due to waves, with current and wind playing secondary roles. The design conditions comprise the design waves and the currents and winds likely to coexist with the design waves. For compliant structures, response to waves is reduced, so that winds and currents become relatively more important.

For fixed structures, the collinear environment usually controls a design and intensities for various sources of environmental load can be selected from Table 35.7 per [35.83].

For floating installations, environmental conditions that involve large differences in direction need to be considered.

When design forces due to wind need to be combined with forces due to waves and current, the following is appropriate:

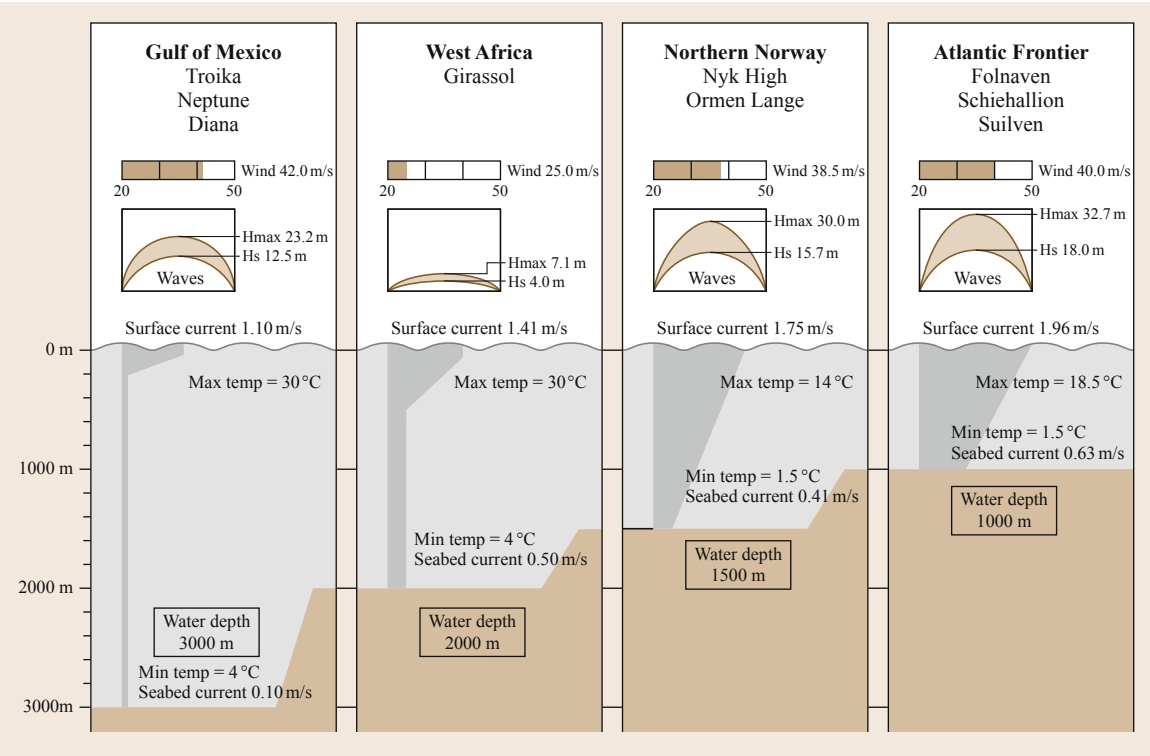


Fig. 35.15 Environmental conditions at a number of deep water sites (after [35.84])

Table 35.7 Combination of environmental forces with expected mean values and annual probability of exceedance

Limit state	Wind	Waves	Current
Ultimate limit state	10^{-2} 10^{-1} 10^{-1} – –	10^{-2} 10^{-1} 10^{-1} – –	10^{-1} 10^{-2} 10^{-1} – –
Accidental limit state	10^{-4} 10^{-2} 10^{-1} – –	10^{-2} 10^{-4} 10^{-1} – –	10^{-1} 10^{-1} 10^{-4} – –

1. For structures with negligible dynamic response, the 1 h sustained wind can be used to determine quasi-static global actions caused by wind in conjunction with ultimate or accidental loads due to waves and currents
2. For structures that are moderately dynamically sensitive, but do not require a full dynamic analysis, the 1 min mean wind can be used to determine quasi-static global loads caused by wind, again for wind in conjunction with ultimate or accidental quasi-static loads due to waves and currents
3. For structures with significant dynamic response to excitation with periods longer than 20 s, a full dynamic response analysis to fluctuating winds should be considered.

Reduced design requirements can be used for the design or relocation of structures that are unmanned or evacuated during the design event or where the loss of or severe damage to the structure would not result in a high consequence of failure. Risk analysis may justify either longer or shorter recurrence intervals for design criteria. Where sufficient information available about the variation under environmental conditions expected to occur from different directions can be considered.

Figure 35.15 illustrates some examples of combined wind, wave, and current environments in four different ocean basins.

References

35.1 G.G. Stokes: On the theory of oscillatory waves, Trans. Camb. Phil. Soc. **8**, 441–455 (1847)
35.2 T. Levi-Civita: Determination rigoureuse des ondes permanentes d'ampleur finie, Math. Ann. **93**, 264–314 (1925)
35.3 L.W. Schwartz: Computer extension and analytic continuation of Stokes' Expansion for gravity waves, J. Fluid Mech. **62**, 553–578 (1974)
35.4 R.L. Wiegel: *Oceanographical Engineering* (Prentice-Hall, New Jersey 1964)
35.5 T. Sarpkaya, M. Isaacson: *Mechanics of Wave Forces on Offshore Structures* (Van Nostrand Reinhold Co., New York 1981)
35.6 L. Skjelbreia, J.A. Hendrickson: Fifth order gravity wave theory, Coast. Eng. Proc. (1960) pp. 184–196
35.7 D.J. Korteweg, G. de Vries: On the change of form of long waves advancing in a rectangular canal, and on a new type of long stationary waves, Phil. Mag. **39**(5), 422–443 (1895)
35.8 E.V. Laitone: The second approximation to cnoidal and solitary waves, J. Fluid Mech. **9**, 430–444 (1960)
35.9 J.E. Chappelear: Shallow water waves, J. Geophys. Res. **67**, 4693–4704 (1962)
35.10 R.C. Ertekin, J.M. Becker: Nonlinear diffraction of waves by a submerged shelf in shallow water, J. Offshore Mech. Arct. Eng. **120**(4), 212–220 (1998)
35.11 J. Boussinesq: Théorie de l'intumescence liquide appelée onde solitaire ou de translation, Comptes Rendus Acad. Sci. Paris **72**, 755–759 (1871)
35.12 J.D. Fenton: A ninth-order solution for the solitary wave, J. Fluid Mech. **53**, 237–246 (1972)
35.13 A.E. Green, N. Laws, P.M. Naghdi: On the theory of water waves, Proc. R. Soc. Lond. A **338**, 43–55 (1974)
35.14 B.B. Zhao, R.C. Ertekin, W.Y. Duan, M. Hayatdavoodi: On the steady solitary-wave solution of the Green–Naghdi equations of different levels, Wave Motion **51**(8), 1382–1395 (2014)
35.15 M.S. Denis, W.J. Pierson: On the motions of ships in confused seas, Trans. Soc. Nav. Archit. Mar. Eng. **61**, 280–357 (1953)
35.16 W.H. Michel: Sea spectra simplified, Mar. Technol. **5**(1), 17–30 (1968)
35.17 O.M. Faltinsen: *Sea Loads on Ships and Offshore Structures*, Ocean Technology Series (Cambridge Univ. Press, Cambridge 1990)
35.18 J.N. Newman: *Marine Hydrodynamics* (MIT Press, Cambridge 1978)
35.19 S.K. Chakrabarti: *Hydrodynamics of Offshore Structures* (Comput. Mechanics, Southampton 1987)
35.20 L.L. Huang, H.R. Riggs: The hydrostatic stiffness of flexible floating structures for linear hydroelasticity, Mar. Struct. **13**, 91–106 (2000)
35.21 R.C. Ertekin, H.R. Riggs, X.L. Che, S.X. Du: Efficient methods for hydroelastic analysis of very large floating structures, J. Ship Res. **37**(1), 58–76 (1993)
35.22 J.V. Wehausen: The motion of floating bodies, Ann. Rev. Fluid Mech. **3**, 237–268 (1971)
35.23 H. Goldstein: *Classical Mechanics*, 2nd edn. (Addison-Wesley, Reading 1980)
35.24 R.W. Yeung: *A Singularity-Distribution Method for Free-Surface Flow Problems with an Oscillating Body*, Rep. No. 73-6 (University of California, Berkeley 1973)
35.25 O.M. Faltinsen, F.C. Michelsen: Motions of large structures in waves at zero Froude number, Int. Symp. Dyn. Mar. Veh. Struct. Waves (1974) pp. 91–106
35.26 C.J. Garrison: Hydrodynamic loading of large offshore structures: Three-dimensional source distribution methods. In: *Numerical Methods in Offshore Engineering*, ed. by O.C. Zienkiewicz, R.W. Lewis, K.G. Stagg (Wiley, New York 1978)
35.27 J.V. Wehausen, E.V. Laitone: Surface waves. In: *Handbuch der Physik*, Vol. 9, ed. by S. Flugge (Springer, Berlin, Heidelberg 1960)
35.28 J.N. Newman: Algorithms for the free-surface Green function, J. Eng. Math **19**, 57–67 (1985)
35.29 C.A. Brebbia: *The Boundary Element Method for Engineers* (Wiley, New York 1978)
35.30 A.E. Cummins: The impulse response function and ship motions, Schiffstechnik **9**(47), 101–109 (1962)
35.31 T.F. Ogilvie: Recent progress toward the understanding and prediction of ship motions, Proc. 5th Symp. Nav. Hydrodyn. (1964) pp. 3–79
35.32 J.O. de Kat, J.R. Paulling: The simulation of ship motions and capsizing in severe seas, Soc. Nav. Archit. Mar. Eng. **97**, 139–168 (1989)
35.33 R.F. Beck: Time-domain computations for floating bodies, Appl. Ocean Res. **16**, 267–282 (1994)
35.34 J. Nolte, R.C. Ertekin: Wave power calculations for a wave energy conversion device connected to a drogue, J. Renew. Sustain Energy (AIP) **6**(1), 013117-1–013117-21 (2014)
35.35 H. Maruo: Thel drift of a body floating on waves, J. Ship Res. **4**(1), 1–10 (1960)
35.36 J.N. Newman: The drift force and moment on ships in waves, J. Ship Res. **6**(1), 10–17 (1967)
35.37 J.A. Pinkster: Mean and low frequency wave drifting forces on floating structures, Ocean Eng. **6**, 593–615 (1979)
35.38 C.J. Garrison: *Hydrodynamic interaction of waves with a large displacement floating body*, Rep. No. NPS-69GM77091 (Naval Postgraduate School, Monterey 1977)
35.39 M. Takaki, Y. Tango: Wave drift forces on very huge floating structures, Int. J. Ofshore Polar Eng. **5**(3), 204–211 (1995)
35.40 X.Q. Liu, R.C. Ertekin, H.R. Riggs, D. Xia: Mean wave-drift loads on connected semisubmersible modules, Proc. 17th Int. Conf. Offshore Mech. Arct. Eng. (OMAE) (1998)
35.41 F.H. Hsu, K.A. Blenkarn: Analysis of peak mooring forces caused by slow vessel drift oscillations in random seas, Offshore Technol. Conf. (1970) pp. 135–146
35.42 J.N. Newman: Second-order slowly varying forces on vessels in irregular waves, Int. Symp. Dyn. Mar. Veh. Struct. Waves (1974) pp. 182–186

35.43 J.R. Morison, M.P. O'Brien, J.W. Johnson, S.A. Schaaf: The force exerted by surface piles, Petroleum Trans. **189**, 149–154 (1950)

35.44 R.C. MacCamy, R.A. Fuchs: *Wave Forces on Piles: A Diffraction Theory*, Tech. Memo. No. 69 Beach Erosion Board (Army Corps of Engineers, Washington 1954)

35.45 G.H. Keulegan, L.H. Carpenter: Forces on cylinders and plates in an oscillating fluid, J. Res. Nat. Bureau Stand. **60**(5), 423–440 (1958)

35.46 M. de S.Q. Isaacson: Wave Induced Forces in the Diffraction Regime. In: *Mechanics of Wave-Induced Forces on Cylinders*, ed. by T.L. Shaw (Pitman, London 1979) pp. 68–89

35.47 J.R. Paulling: Frequency-domain analysis of otec CW pipe and platform dynamics, Offshore Technol. Conf. (1979) pp. 1641–1651

35.48 J.R. Paulling: An equivalent linear representation of the forces exerted on the OTEC CW pipe by combined effects of waves and current, Ocean Eng. OTEC (1979) pp. 21–28

35.49 L.P. Krolikowski, T.A. Gay: An improved linearization technique of frequency domain riser analysis, Offshore Technol. Conf. (1980) pp. 341–353

35.50 T. Sarpkaya: *Wave Forces on Offshore Structures*, 1st edn. (Cambridge Univ. Press, Cambridge 2010)

35.51 T. Sarpkaya: *In-Line and Transverse Forces on Smooth and Sand-Roughned Cylinders in Oscillatory Flow at High Reynolds Numbers, Rep. No. NPS-69SL76062* (Naval Post Graduate School, Monterey 1976)

35.52 DNV: *Recommended Practice DNV-RP-025. Environmental Conditions and Environmental Loads* (Det Norske Veritas, Høvik 2010)

35.53 S.K. Chakrabarti: Steady drift force on vertical cylinder–viscous vs. potential, Appl. Ocean Res. **6**, 73–82 (1984)

35.54 G.E. Burns: Calculating viscous drift of a tension leg platform, Proc. 2nd Int. Offshore Mech. and Arct. Eng. Conf., ASME, Houston (1983) pp. 22–30

35.55 D.L.R. Botelho, T.D. Finnigan, C. Petrauskas, S.M. Lui: Model test evaluation of a frequency-domain procedure for extreme surge response prediction of tension leg platforms, Proc. 16th Annual Offshore Technol. Conf., Houston, Texas OTC 4658 (1984) pp. 105–112

35.56 R.C. Ertekin, A.S. Chitrapu: Wave- and current-induced viscous drift forces on floating platforms, Proc. 6th Int. Symp. Offshore Eng. (1987) pp. 625–629

35.57 A.S. Chitrapu, R.C. Ertekin, J.R. Paulling: Viscous drift forces in regular and irregular waves, Ocean Eng. **20**(1), 33–55 (1993)

35.58 Y. Li, A. Kareem: A description of hydrodynamic forces on tension leg platforms using a multivariate Hermite expansion, Proc. 9th Int. Offshore Mech. Arct. Eng. Conf. (1990) pp. 133–142

35.59 P.D. Spanos, M.G. Donley: Stochastic response of a tension leg platform to viscous drift forces, Proc. 9th Int. Offshore Mech. Arct. Eng. Conf. (1990) pp. 107–114

35.60 NDBC: Does NDBC measure ocean current velocities?, http://www.ndbc.noaa.gov/adcp.shtml (2013)

35.61 B. Bischof, E. Rowe, A.J. Mariano, E.H. Ryan: The Brazil current, http://oceancurrents.rsmas.miami.edu/atlantic/brazil.html (2004)

35.62 G.Z. Forristall, C.K. Cooper: Design current profiles using Empirical Orthogonal Functions (EOF) and Inverse FORM methods, Offshore Technol. Conf. (1997)

35.63 J.T. Kirby, T.M. Chen: Surface waves on vertically sheared flows, approximate dispersion relations, J. Geophys. Res. **94**(C1), 1013–1027 (1989)

35.64 R.A. Dalrymple, J.C. Heideman: Non-linear water waves on a vertically-sheared current, E&P Forum Workshop (1989)

35.65 J.W. Eastwood, C.J.H. Watson: Implications of wave-current interactions for offshore design, E&P Forum Workshop (1989)

35.66 J.D. Wheeler: Method for calculating force produced by irregular waves, J. Petroleum Tech. **22**, 473–486 (1970)

35.67 A.R.Pf. Planning: *Designing and Constructing Fixed Offshore Platforms: Working Stress Design*, 21st edn. (American Petroleum Institute, Washington 2000)

35.68 H. Schlichting: *Boundary Layer Theory*, 2nd edn. (McGraw-Hill, New York 1968)

35.69 R.D. Blevins: *Flow-Induced Vibration*, 2nd edn. (Van Nostrand Reinhold, New York 1990)

35.70 DNV: *Recommended Practice DNV-RP-F105. Free Spanning Pipelines* (Det Norske Veritas, Nøvik 2006)

35.71 DNV: *Recommended Practice DNV-RP-F204. Riser Fatigue* (Det Norske Veritas, Nøvik 2005)

35.72 NRCS: Wind Rose Data, http://www.wcc.nrcs.usda.gov/climate/windrose.html (2002)

35.73 O.J. Andersen, J. Løvseth: The Frøya database and maritime boundary layer wind description, Mar. Struct. **19**(2/3), 173–192 (2006)

35.74 S.K. Chakrabarti: *Offshore Structure Modeling* (World Scientific, Singapore 1994)

35.75 E. Buckingham: On physically similar systems; illustrations of the use of dimensional equations, Phys. Rev. **4**(4), 345–376 (1914)

35.76 R.C. Ertekin, J.W. Kim (Eds.): *The Proceedings of 3rd International Workshop on Very Large Floating Structures (VLFS '99)* (SOEST, Honolulu 1999)

35.77 J.T. Dillingham: Recent experience in model-scale simulation of tension-leg platforms, Mar. Technol. **21**(2), 186–200 (1984)

35.78 M.P. Tulin: Hydroelastic scaling, Proc. 3rd Int. Workshop Very Large Float. Struct. (VLFS) (1999) pp. 483–487

35.79 S.A. Hughes: *Physical Models and Laboratory Techniques in Coastal Engineering* (World Scientific, Singapore 1993)

35.80 W.Y. Duan: Report of the 1st workshop on numerical wave tank, Int. Theory Advis. Panel Numer. Tank (College of Shipbuilding Engineering, Harbin Engineering University, Harbin, China 2014)

35.81 M.A. Benitz, D.P. Schmidt, M.A. Lackner, G.M. Stewart, J. Jonkman, A. Robertson: Comparison of hydrodynamic load predictions between engineering models and computational fluid dynamics for the

Oc4-DeepCWind semi-submersible, Proc. 33rd Int. Offshore Mech. Arct. Eng. Conf. (OMAE) (2014)
35.82 B. Seiffert, M. Hayatdavoodi, R.C. Ertekin: Experiments and computations of solitary-wave forces on a coastal-bridge deck. Part I: Flat plate, Coast. Eng. **88**, 194–209 (2014)
35.83 S. Norway: *NORSOK Standard N-003: Actions and action effects* (Standards Norway, Lysaker 2007)
35.84 A. Moros, P. Fairhurst: Production riser design: Integrated approach to flow, mechanical issues, Offshore Mag. **59**(4), 82 (1999)

36. Vortex-Induced Vibrations

Michael S. Triantafyllou, Rémi Bourguet, Jason Dahl, Yahya Modarres-Sadeghi

Starting at a low Reynolds number of about 50, and reaching the highest Reynolds numbers recorded, bluff bodies placed within an external flow form an unstable wake that results in the formation of a regular pattern of vortices, the Karman street. If the structure is flexible or flexibly mounted, these vortices may cause vibrations, leading to stresses and fatigue damage. This motion of the body influences, in turn, the vortex formation process, establishing a feedback mechanism that may lead to stable or unstable dynamic equilibria. As a result, vortex-induced vibrations are controlled by complex physical mechanisms characterized by rich dynamic properties. When elongated, flexible structures are placed in a sheared cross-flow, the fluid–structure interaction process is distributed along their length, resulting in added complexity, as parts of the structure act to transfer energy from the flow to the structure, while other parts damp the response.

Vortex-induced vibrations can be a serious concern for operations and for the structural integrity of mooring cables, towing hawsers, marine risers, and moored bluff structures, such as spar buoys. We review some of the basic properties of vortex-induced responses and the effectiveness of vortex cancellation devices.

36.1 **VIV Prediction of Wide-Span Rigid Cylinders in Uniform Cross-Flow** ... 820
36.1.1 Flow Instability and the Formation of the Karman Street ... 820
36.1.2 VIV of Flexibly Mounted Cylinders in Cross-Flow ... 822
36.1.3 Forced Cylinder Vibrations ... 823
36.1.4 Lift Coefficient, Velocity and Amplitude ... 825
36.1.5 Correlation Length ... 828
36.2 **VIV Prediction of Flexible Structures in Nonuniform Flow** ... 831
36.2.1 Traveling Structural Waves and Multimodal Responses in Sheared Flows ... 831
36.2.2 Lock-In and Fluid-Structure Energy Transfer Within Shear Flow ... 835
36.2.3 Synchronization of In-Line and Cross-Flow VIV ... 838
36.3 **Experimental Studies and Fatigue Analysis** ... 838
36.3.1 Riser Orbital Motions and Excitation Region ... 839
36.3.2 Higher Harmonic Strain and Acceleration Components in Flexible Structures ... 839
36.3.3 Periodic (Type-I) and Chaotic (Type II) Signals ... 840
36.3.4 Response Reconstruction Using Experimental Data ... 841
36.3.5 Fatigue Calculations ... 841
36.3.6 VIV Prediction Tools ... 842
36.4 **Effectiveness of Vortex Canceling Devices** ... 843
36.4.1 Helical Strakes ... 843
36.4.2 Fairings ... 844
36.4.3 Other VIV Suppression Devices ... 844
36.5 **Multiple Interfering Bluff Bodies** ... 845
36.6 **Effect of Reynolds Number** ... 845
References ... 846

The literature on the subject of vortex-induced vibrations (VIV) is so vast that one cannot hope to be all-inclusive, or even to make justice to all significant contributions. Besides, several review articles and books have already covered several topics [36.1–10]. In this chapter, we will provide an outline of those concepts that are essential to understanding the physics of VIV, and then we will cover recent developments while identifying outstanding issues. Particular emphasis is placed on vibrations in water, where the added mass force plays a very significant role in shaping the properties of VIV.

The topic of VIV of flexibly mounted bluff bodies has become the canonical problem for studying the

flow–structure interaction of nonstreamlined objects, because of the richness of the dynamic behavior. At the same time, VIV is of great importance to applications, such as risers and floating structures used in offshore oil and gas production, towing and anchoring of ships and buoys, acoustic towed arrays, remotely operated vehicles, and cable laying. Also, structures with bluff shapes are moored within currents, such as spar buoys and tension leg platforms, which also exhibit vortex-induced motions (VIM). We will collectively refer to such responses as *vortex-induced vibrations* (VIV), while sometimes we will refer to the VIM of spar buoys.

In towed cables and hawsers, the drag coefficient of the cable can be amplified due to VIV from a value near 1.2 for stationary cylinders to a value of 3.0 or higher, depending on the amplitude of VIV. The drag coefficient affects the static shape, while unsteady loads can cause relatively high-frequency vibrations and hence failure by fatigue [36.11–16]. Similarly, in moored ships and structures, especially in open waters, the drag coefficient may be amplified due to VIV, causing the damping of the overall system, which depends on the drag force acting on the mooring lines, to rise substantially. Since moored systems at sea have long-period natural frequencies, of the order of 100 s, and which are underdamped, a good estimate of the total damping is crucially important. The damping due to the presence of mooring lines can be from 30 to 80% of the total damping [36.17–20]. Another application where drag coefficient amplification due to VIV has significant effect is in cable-laying, because the precise location of where the cable will eventually land, and the amount of tension needed to avoid buckling depends on the drag force. For oceanographic moorings in the open ocean, the most crucial loads are those due to VIV, which seriously limit their life at sea [36.15]. Offshore moorings and risers are used routinely today at depths of 2500 m and beyond. It is widely recognized that, particularly for risers, VIV is the most important problem leading to large static loads and fatigue due to dynamic loads [36.21–23]. Finally, very large structures in the shape of spar buoys are used in the ocean today, such as the Genesis platform with a diameter of 37 m and a vertical height of 198 m, fitted with strakes to reduce VIV. These platforms, operating at Reynolds number around $Re = 10^8$ exhibit VIV, which cannot be reliably predicted on the basis of low Reynolds number VIV data.

The hydrodynamic problem of VIV, which involves interaction of a flexible structure with vortex shedding in a turbulent wake, has been, and still is an area of intense research effort, as evidenced by a vast literature [36.6, 24–28]. The viscous flow behind vibrating cables and risers is dominated by complex flow mechanisms, including changing patterns of vortex formation, formation of secondary vortical structures, turbulence in the wake, and correlation length effects. A series of at-sea experiments has demonstrated that the behavior of cables at sea is strongly influenced by shear and can result in substantial changes in the drag coefficient [36.14, 15]. In the offshore industry, similar VIV problems apply for risers. At shallow depths, risers are dominated by their bending stiffness and behave principally like beams, hence very few modes may be excited and the prediction of VIV is somewhat easier because single-mode prediction procedures, based on empirical models [36.21, 22] or laboratory experiments, can lead to realistic predictions. As the water depth increases, however, risers tend to behave increasingly closer to tensioned cables, with the possibility of having multimodal response in a sheared current. Intense activity in the offshore industry has led to the development of experiment-based predictive programs for risers whose results are usually compared successfully against full-scale experiments, although predicting multimodal response and high Reynolds number behavior, remain outstanding problems [36.29–32].

One of the basic outstanding problems today is the sparsity of data in the open literature for VIV at high Reynolds numbers. While for a stationary cylinder in cross-flow it is well known that significant transitions occur with dramatic effects on the loads and frequency of vortex formation [36.33], far less information is available for vibrating cylinders [36.34, 35].

36.1 VIV Prediction of Wide-Span Rigid Cylinders in Uniform Cross-Flow

We start by outlining the basic concepts for VIV prediction as applied to a uniform, wide-span, flexibly mounted cylinder of circular diameter in a uniform flow, because they are well understood, while many of these concepts can be extended to shear flow as well. The literature on this topic is vast because it has become a canonical problem of flow-induced vibration, as it exhibits an impressive richness of behavior.

36.1.1 Flow Instability and the Formation of the Karman Street

The primary cause of all the phenomena we describe is, at least initially, the formation of the Karman street, a remarkable phenomenon because it occurs at practically every scale beyond a threshold value of the Reynolds number $Re \approx 50$; for example, formation of

vortex streets behind islands, observed in the clouds, occurs at $Re \approx 10^{11}$ (Fig. 36.1). The Karman street consists of a staggered array of vortices forming at a frequency f, which obeys the Strouhal law: based on the cylinder diameter d and current speed U, $fd/U = \mathrm{St}$, where the Strouhal number $\mathrm{St} \approx 0.20$ for subcritical flows above $Re = 150$, and takes somewhat higher values for supercritical flows.

The Karman street forms spontaneously due to an instability of the separated flow behind bluff bodies [36.36–38]. This instability is equivalent to an unstable pole in finite-dimensional systems, and it is always excited by even minute perturbations. The variation of the Strouhal number as a function of the Reynolds number parallels the variation of the drag coefficient, underlining the fact that the Strouhal number is basically a wake parameter, and hence it is related closer to the width of the wake, rather than the diameter of the cylinder. In fact, if the width of the wake h is used to define the Strouhal number, that is, $\mathrm{St}_h = fh/U$, a nearly universal value can be obtained for all Reynolds numbers.

We distinguish, then:

a) The regime below $Re = 50$ where no Karman street forms.
b) The regime above $Re = 50$ and below $Re = 150$, where the Strouhal number varies nearly linearly with Re from a value of $\mathrm{St} \approx 0.1$ to a value of $\mathrm{St} \approx 0.2$.
c) The regime above $Re = 150$ and below roughly $Re = 250\,000$, where the Strouhal number stays roughly at a value of $\mathrm{St} \approx 0.2$.
d) The regime above $Re = 250\,000$ and below roughly $Re = 500\,000$, where the Karman street disappears for perfectly smooth cylinders placed in a very low turbulence stream.
e) And the regime above roughly $Re = 500\,000$ when the Karman street re-appears eventually reaching a higher constant value of $\mathrm{St} \approx 0.26$.

While the value of the Strouhal number is interesting for stationary cylinders, for VIV, several other parameters are equally important, as we outline below.

Starting at Reynolds number roughly 250, a secondary instability sets in the flow. Regularly spaced vortices, whose axis of rotation is roughly parallel to the stream (streamwise vortices), appear on top of the Karman street. The distance between the vortices is between 1 and 2 cylinder diameters; these have high vorticity but small circulation, and they are called *braid* or *hairpin* vortices because of their shape (Fig. 36.2). As the Reynolds number increases, these hairpin vortices cause the flow to become increasingly turbulent; above the Reynolds number 1000 the wake of a circular cylinder is fully turbulent. The braid vortices reduce significantly the value of the lift coefficient and lead to a stabilization of the basic properties of the Karman street. Finally, there are also secondary instabilities at the critical Reynolds number, around $Re = 250\,000$.

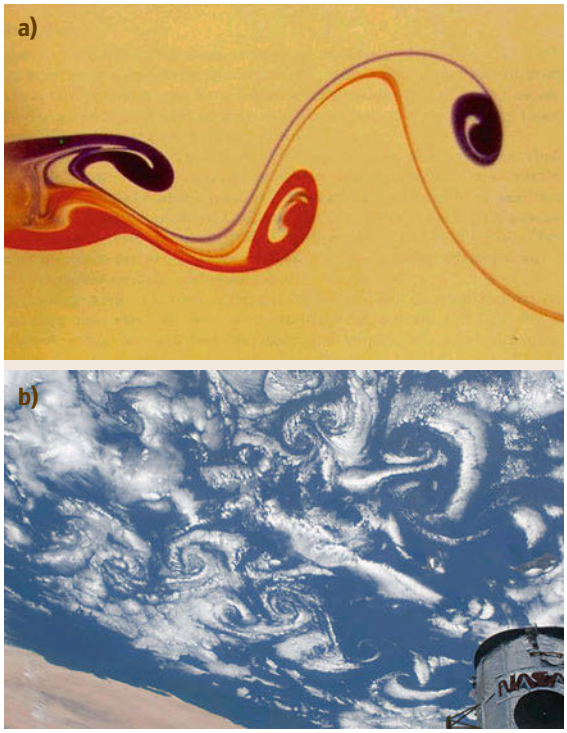


Fig. 36.1 **(a)** Karman street behind a circular cylinder at $Re = 250$ (after [36.39]); **(b)** Karman street behind islands at $Re \approx 10^{11}$ (courtesy of NASA)

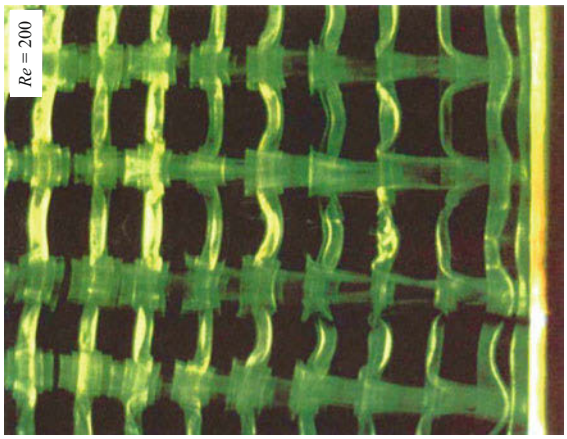


Fig. 36.2 A secondary instability causes the formation of *braid* vortices, that affect the lift force and the Karman street properties (after [36.5])

In the beginning of the critical regime, at $Re \approx 250\,000$, the drag crisis occurs, wherein the boundary layer becomes turbulent, forcing the flow separation points to move from an average position of 86°, with respect to the front stagnation point, to a value of 120°; the drag coefficient drops from a value of nearly 1.2 to a value as low as 0.20. Most importantly, asymmetric re-attachment of the flow occurs before the final separation point is reached, and hence the symmetry of the time-averaged flow is broken. As a result, the instability of the flow is altered and no Karman street forms. The roughness of the cylinder surface or turbulence in the external flow can trip the flow, which is extremely sensitive to such perturbations – in particular, roughness can cause the recovery of symmetric separation. As a result, the reported properties within the critical regime can vary substantially depending on the conditions of the experiment [36.33].

The richness of the Karman street formation properties behind stationary cylinders as a function of the Reynolds number and the conditions of the cylinder surface and the flow turbulence is remarkable [36.6, 10]. We limit the discussion herein to only an outline of these results, because we focus on the vibrations resulting from the Karman street, when the cylinder is flexibly mounted.

36.1.2 VIV of Flexibly Mounted Cylinders in Cross-Flow

Consider a circular cylinder allowed to oscillate transverse to the incoming flow. The cylinder is mounted on linear springs with constant k, a linear damper with constant b, and the cylinder has mass m. For this cylinder, VIV may be excited as periodically shed vortices exert unsteady loads on the cylinder. Under synchronization conditions, where vortex shedding occurs at the frequency of vibration of the structure, the system will oscillate at a frequency relatively close to the structural natural frequency in quiescent fluid, but the precise frequency will depend on the oscillation amplitude, frequency, and the structure of the wake.

Figure 36.3 shows an example of the amplitude of the transverse response, nondimensionalized by the diameter, A/d, as a function of the reduced velocity $V_{rn} = U/f_n D$, where f_n denotes the natural frequency of the system in quiescent water, that is, with added mass coefficient equal to 1. The response, depending on the ratio of the cylinder mass m to the displaced mass m_d, $m* = m/m_d$, where $m_d = \rho\pi sd^2/4$, with ρ the water density, and s the span, covers, typically, a wide range of nominal reduced velocities, in contrast with the classical underdamped mass–damping–spring system, which exhibits a high response only over a narrow parametric frequency range, close to the natural frequency. As we explain below, the reason for this difference is that the effective added mass of a vibrating cylinder within cross-flow is a strong function of the reduced velocity; in fact, it can take on negative values. This is one of the most significant concepts in VIV.

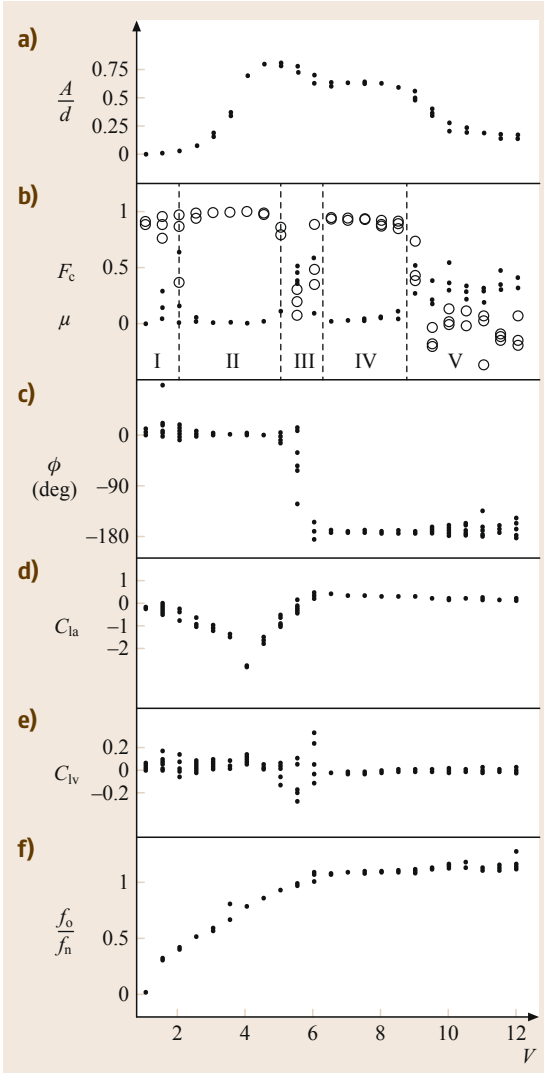


Fig. 36.3a–f Response of a flexibly mounted cylinder versus nominal reduced velocity. **(a)** A/d, **(b)** correlation coefficient between cylinder ends, **(c)** lift force phase angle with respect to the motion, **(d)** lift coefficient in phase with acceleration, **(e)** lift coefficient in phase with velocity, **(f)** response frequency to natural frequency, versus reduced velocity. Reynolds number is 30 000, aspect ratio 26, mass ratio 3, damping ratio 3.5% (after [36.40])

As shown in Fig. 36.3, there are different regimes that characterize the VIV response. As V_{rn} increases,

the response initiates at a value around 3 and increases, reaching a maximum value at about $V_{rn} = 5$, and then the amplitude of response reduces and reaches a plateau at around $V_{rn} = 6$, which lasts all the way to a value of 10. Beyond this value, response decreases rapidly and eventually vanishes.

36.1.3 Forced Cylinder Vibrations

In dynamics, forced vibrations of a system are caused by an external oscillatory input. Although a (steady) external flow is imposed in the case of VIV, the resulting vibrations are caused by an instability in the flow that also determines the frequency and magnitude of the forcing function. Hence, only in a loose way can these vibrations be viewed as forced vibrations. Likewise, in classical dynamics, free vibration refers to the transient oscillations of a dynamic system due to nonzero initial conditions, when there is no external forcing function. VIV, then, is not a free vibration since there is a fluid forcing that continuously drives such vibrations. Finally, it should be mentioned that in forced vibrations of a lightly damped extended structure, the natural frequencies of the structure largely determine the frequency and amplitude content of the response; however, in VIV the forcing function is affected significantly by the motion of the structure. This feedback mechanism renders the use of the original natural frequencies and, especially, the natural modes, in most cases of questionable value in VIV.

Moreover, the terms *forced* and *free* vibration when pertaining to VIV have very specific meanings, as established in the relevant literature. A forced vibration VIV experiment refers to an experiment where the body is fixed to an actuator and forced to move through the fluid with a prescribed trajectory. Typically, hydrodynamic forces exerted on the body are measured during these forced motions. In this case, the body does not respond to the fluid forces since the trajectory is prescribed. This differs from the classical definition of forced vibration.

Also, a VIV experiment where a cylinder is elastically mounted on springs and free to react to the fluid forces is referred to as a *free vibration*. The vibration is called *free* since there is no oscillatory external forcing to drive this oscillation; the energy input is at the expense of the kinetic energy of the oncoming fluid stream, extracted through the instability in the flow. Under steady-state oscillations, fluid forces cancel the structural damping forces, thus providing a vibration analogous to the response of an unforced system with zero damping. For the remaining sections, forced and free vibration will refer to the definitions in regards to VIV experimentation.

In order to explain the basic fluid forces in VIV, we will use results from forced vibrations rather than free vibrations. The cylinder is forced to vibrate transversely to an oncoming stream U with a harmonic motion of amplitude A and frequency f, or circular frequency $\omega = 2\pi f$. The transverse and in-line loads, $X(t)$ and $Y(t)$, respectively, are measured as a function of time; the average of the inline force provides the steady drag force. If the cross-flow motion and its derivatives are described by

$$\begin{aligned} y(t) &= A\cos(\omega t)\,, \\ \frac{\mathrm{d}y}{\mathrm{d}t}(t) &= -\omega A\sin(\omega t)\,, \\ \frac{\mathrm{d}^2y}{\mathrm{d}t^2}(t) &= -\omega^2 A\cos(\omega t)\,, \end{aligned} \tag{36.1}$$

then the inline force contains a constant drag term $\overline{X}$ and an unsteady term whose principal components are even harmonics of the frequency of vibration; the second-harmonic X_2 is typically the most significant. The cross-flow force is found to contain a large harmonic component, $Y_h(t)$, which has a phase ϕ with respect to displacement, and hence can be written as

$$\begin{aligned} Y_h(t) &= Y_o\cos(\omega t+\phi) \\ &= Y_o\cos(\phi)\cos(\omega t) - Y_o\sin(\phi)\sin(\omega t) \\ &= -(-Y_o\cos(\phi))\cos(\omega t) \\ &\quad -(Y_o\sin(\phi))\sin(\omega t) \\ &= -Y_a\cos(\omega t) - Y_v\sin(\omega t)\,. \end{aligned} \tag{36.2}$$

Hence, the harmonic transverse force is decomposed into a component in phase with acceleration Y_a and a component in phase with velocity Y_v, as shown in (36.2); the latter determines whether energy is transferred from the flow to the structure (when positive), or the other way around if negative.

The forces, properly nondimensionalized, provide the mean drag coefficient c_D and unsteady components of the drag coefficient, such as the second-harmonic component, at twice the frequency of vibration c_{d2}. Perpendicular to the direction of flow, the decomposition and normalization of the lift force gives the lift coefficient in phase with velocity c_{lv}, and the lift coefficient in phase with acceleration c_{la}, which can be reformulated to provide the added mass coefficient c_m assuming harmonic body motion [36.41]

$$\begin{aligned} c_D &= \frac{\overline{X}}{\rho dsU^2/2}\,, \\ c_{d2} &= \frac{X_2}{\rho dsU^2/2}\,, \end{aligned} \tag{36.3}$$

$$c_{lv} = \frac{Y_v}{\rho d s U^2/2}\,,$$
$$c_{la} = \frac{Y_a}{\rho d s U^2/2}\,,$$
$$c_m = \frac{Y_a}{A\omega^2 \rho \pi d^2 s/4}\,. \qquad (36.4)$$

Variable Effective Added Mass and Frequency of Response

The concept of an added mass force is well established for inviscid flow, where it is shown that the acceleration of a body, along a straight line for example, within an unbounded fluid causes a reactive force whose magnitude is equal to a constant times acceleration. This constant depends only on the shape of the body and hence is equivalent to a (constant) *added mass*. The general motion of a body in six degrees of freedom can be similarly formulated with a tensor of added masses and moments of inertia [36.42].

However, even within a potential flow, the added mass may not be constant if the body, for example, is vibrating near the free surface of a fluid. For the harmonic motion of a body piercing the surface, the added mass (and damping due to the radiated waves) is a strong function of the frequency. This is an *effective* added mass, in the sense that it is strictly valid for harmonic motion. The variability of the added mass is due to the different wave patterns forming at the fluid surface as a function of the frequency. If the motion is not harmonic, one must take the inverse Fourier transform of the effective added mass as a function of the frequency to find the impulse response and then perform convolution with the acceleration of the body to evaluate forces exerted on the body. What is remarkable is that there are cases when the added mass can be negative. For a submerged body very close to a free surface, the added mass in heave can become negative due to the flux of the fluid mass contained between the body and the free surface at rest.

Hence, for someone familiar with free-surface flows, it is not surprising to discover that the added mass of a vibrating cylinder in cross-flow is a function of the frequency and can even become negative, as there is a similarity with free-surface flows because the wake of a bluff body also supports the formation of waves. The additional complication for VIV is that nonlinearities play a strong role and hence the superposition principle, which is very useful for linearized free-surface flows, are not applicable. Reference to a negative mass without further explanation, may be confusing; however, it is important to note that this is an effective mass, resulting from the confluence of two mechanisms, on one hand, the inertia of the displaced fluid particles (classical added mass definition), and on the other hand, the forces in phase with acceleration due to the forming vortices, which are synchronized and have the same frequency. As a result, as shown in Table 36.1, we define the natural frequency in vacuo (no added mass), the nominal NF with added mass coefficient 1, and the actual NF in water, which varies with reduced velocity.

An alternative explanation of a negative effective added mass is provided in [36.7], where it is explained that the effective added mass force is not a true added mass force and is the result of a vortex force, that

Table 36.1 Definition of the natural frequency (NF) of a mass (m) – spring (k) – dashpot (b) system, as used in the text. m_a is the added mass in quiscent fluid, m_{ae} is the effective added mass

NF in vacuo	$\sqrt{(k/m)}$
Nominal NF in water	$\sqrt{(k/(m+m_a))}$
NF in water	$\sqrt{(k/(m+m_{ae}))}$

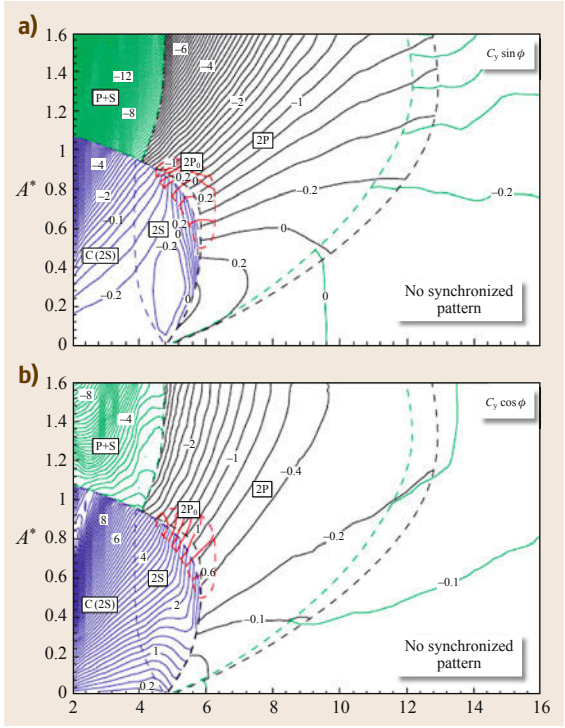


Fig. 36.4a,b Fluid forces from sinusoidally forced cylinder motions in a free stream. **(a)** Lift coefficient in phase with velocity c_{lv} and **(b)** lift coefficient in phase with acceleration c_{la} as a function of A/D and reduced velocity. Reynolds number 20 000. Superimposed are free vibration tests for three different values of the damping coefficient (after [36.43])

is, a force caused by the shed vorticity. This explanation provides further understanding of the physics of the problem and underlines the importance of vortex shedding synchronization with the motion of the cylinder.

When the added mass coefficient and the lift in phase with velocity are plotted as a function of the amplitude and frequency of response, as shown in Fig. 36.4, it is easy to see the strong dependence of these quantities on the frequency of vibration. The variation in the added mass coefficient is particularly striking, since it can be as low as −0.5 and can exceed a value of 2 – all compared to a nominal value of 1.

The cause of the variation in the added mass is the variation in the timing of vortex formation relative to the cylinder motion measured, for example, by the timing of shedding a clockwise vortex relative to the time when the oscillating cylinder reaches a maximum in the oscillatory motion. As shown in Fig. 36.5, as the frequency of oscillation varies, the formation of a clockwise vortex moves earlier in the cycle; note that the value of the nondimensional frequency where this occurs is exactly the value of rapid variation in the added mass coefficient (Fig. 36.5). The variation in the lift force is easy to explain, since when a vortex is close to the cylinder, it causes a suction force due to the low pressure associated with the vortex. Hence, the timing of vortex formation alters the suction force caused by the vortex on the cylinder. If the suction force is in the direction of cylinder acceleration, it will reduce the inertia force, resulting in a reduced added mass; if the suction force is in the opposite direction, the apparent added mass increases.

36.1.4 Lift Coefficient, Velocity and Amplitude

Just as the lift force in phase with acceleration provides for an effective added mass of the system, the lift force in phase with velocity provides for an effective overall damping. When the effective damping is negative (the lift force in phase with velocity minus the structural damping force is positive), it will drive the system to higher amplitudes and when the effective damping is positive, it will drive the system to smaller amplitudes. Hence, the balance of this fluid-forcing term with the structural damping of the system governs the amplitude of the oscillation.

In nondimensional terms, the lift coefficient in phase with velocity, c_{lv} provides the energy to drive VIV. The coefficient reaches peak positive values at around the reduced frequency

$$f^* = fd/U \approx 0.17 \,.$$

This frequency of peak excitation is of critical importance to the prediction of VIV and also coincides with the maximum variability of the effective added mass.

A critical property of c_{lv} can be established when it is plotted as a function of the amplitude of response, for constant frequency of vibration. It follows the curve shown in Fig. 36.6, namely, as the amplitude increases,

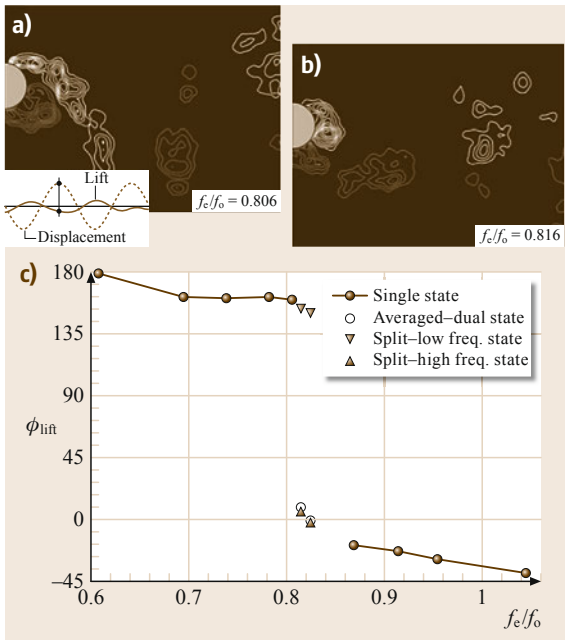


Fig. 36.5 Variation in the timing of vortex shedding (after [36.44])

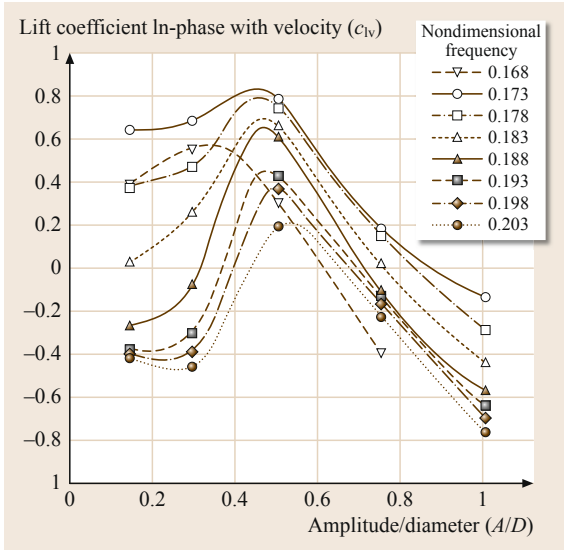


Fig. 36.6 Lift coefficient in phase with velocity as a function of amplitude of vibration (after [36.45])

it initially increases as the lift force becomes increasingly better correlated along the span, but then, beyond a value of typically $A/D = 0.5$, it decreases nearly linearly with A/D, reaching a value of zero at around one diameter (this value varies from 0.6 at low Reynolds numbers to a value of 1.2 or even larger at large subcritical Reynolds numbers, and then returns to a value of around 0.9 for supercritical Reynolds numbers). For large amplitudes, where c_{lv} becomes negative, the effective damping of the system will always be positive, driving the system to have a smaller amplitude. The eventual sign change of c_{lv} at high amplitudes explains the self-limiting nature of VIV, since vibrations beyond typically one diameter are unsustainable.

From an energy balance point of view, in a freely vibrating system, the ultimate value of c_{lv} must be such as to balance the structural damping. When damping is negligible, c_{lv} must be equal to zero. This implies that a condition of *dynamic equilibrium* must be reached.

Strouhal Frequency and Peak Excitation Frequency

As mentioned earlier, the instability in the flow is the cause of VIV, so one might expect that the cylinder is excited most strongly for frequencies close to the Strouhal frequency. One should note, however, that when the cylinder is allowed to vibrate, the process of vortex formation is affected; also, vibrations occur at those frequencies that the added mass coefficient dictates, as we will explain in the next section. As a result, the value of peak VIV response is close, but not equal to the Strouhal frequency, with a nondimensional value close to 0.17 for subcritical Reynolds numbers (as compared to a Strouhal number of around 0.20 for subcritical Reynolds numbers). It should also be noted that this is the value where the added mass coefficient undergoes its largest variation.

Free Versus Forced Vibrations

When free vibrations exhibit purely harmonic motion, there is direct correspondence between forces and motions in free-vibration and forced vibration experiments. Indeed, as shown in Fig. 36.4, the free-vibration results closely follow the iso-contours of constant C_{lv}.

As a result, hydrodynamic coefficients determined through the forced motion of a cylinder in a streaming flow act as fundamental building blocks for predicting VIV. One can write the equation of motion for the vibrations of an elastically mounted cylinder $y(t)$

$$m\frac{\mathrm{d}^2}{\mathrm{d}t^2}y(t) + b\frac{\mathrm{d}}{\mathrm{d}t}y(t) + ky(t) = f(t)\,, \tag{36.5}$$

where $f(t)$ denotes the fluid force, which we decompose into a part in phase with velocity and one in phase with acceleration, as explained earlier. By assuming harmonic motion with the complex amplitude Y and frequency $\omega = 2\pi f$, $y(t) = \Re[Y\mathrm{e}^{\mathrm{i}\omega t}]$, we write

$$f(t) = \Re\left[\left(-c_{\mathrm{m}}\frac{\pi}{4}d^2s\omega^2 Y + \mathrm{i}\frac{1}{2}\rho c_{\mathrm{lv}}dsU^2\frac{Y}{|Y|}\right)\mathrm{e}^{\mathrm{i}\omega t}\right], \tag{36.6}$$

where $\Re[x]$ stands for the real part of x. By substituting into (36.5), and resolving the real versus imaginary part, we find

$$M\omega^2 = k\,, \quad b\omega A = \frac{1}{2}\rho c_{\mathrm{lv}}dsU^2\,, \tag{36.7}$$

where $M = m + m_{\mathrm{a}}$ is the mass plus added mass, with $m_{\mathrm{a}} = c_{\mathrm{m}}\pi/4D^2s$, and A is the absolute value of Y. Equations (36.7) can be solved to provide the amplitude and frequency of response. Their solution must be iterative because the added mass and damping coefficients are functions of the amplitude and frequency.

The second equation expresses the fact that the energy from the fluid is balanced by the damper; hence, if the damping is negligible, free vibrations will take place at values of $c_{\mathrm{lv}} = 0$. The first of (36.7) simply expresses the fact that vibrations occur at resonance; except that the natural frequency varies because of the variability of the added mass coefficient. Indeed, this variability, which can be significant, is the root cause of the noted VIV response over a wide parametric range, and is a crucially important concept for accurate VIV prediction.

We can put (36.7) in a nondimensional form using the definitions of the mass ratio m^*, damping ratio ζ, natural frequency in a vacuum ω_n, added mass coefficient c_{m}, lift coefficient in phase with velocity c_{lv}, reduced velocity V_{r}, and nondimensional frequency f/f_{n}

$$m^* = \frac{m}{\rho\pi\frac{D^2}{4}S}\,, \tag{36.8}$$

$$\omega_n = \sqrt{\frac{k}{m}}\,, \tag{36.9}$$

$$\zeta = \frac{b}{2m\omega_n}\,, \tag{36.10}$$

$$V_{\mathrm{r}} = \frac{U}{fD}\,, \tag{36.11}$$

$$f^* = \frac{f}{f_{\mathrm{n}}}\,. \tag{36.12}$$

In a nondimensional form, we find that (36.5) gives two equations from the real and imaginary parts, where the nondimensional amplitude and frequency are deter-

mined directly from the fluid forcing terms

$$\frac{A}{D} = \frac{c_{\mathrm{lv}} V_{\mathrm{r}}^2 f^*}{4\pi^3 m^* \zeta}\,, \tag{36.13}$$

$$f^* = \sqrt{\frac{m^*}{m^* + C_{\mathrm{m}}}}\,. \tag{36.14}$$

We can see the importance of added mass variability with two simple examples:

- Let us assume that the natural frequency of a system with $m^* = 1$, using a nominal added mass coefficient $c_{\mathrm{m}} = 1, \hat{f}_{\mathrm{n}}$, provides $\hat{f}_n d/U = 0.10$, at the edge of excitation. When we account for the true value of the added mass coefficient, however, $c_{\mathrm{m}} = -0.5$, we realize that the true natural frequency changes to give $f_{\mathrm{n}} d/U = 0.16$, near peak excitation.
- Next we assume that the velocity changes, so that using again the nominal natural frequency, we find $\hat{f}_{\mathrm{n}} d/U = 0.20$, at the other side of the edge of excitation. When we account, again, for the true value of the added mass coefficient, however, $c_{\mathrm{m}} = 2.0$, we realize that the true natural frequency changes to give $f_{\mathrm{n}} d/U = 0.163$, back to near peak excitation.

These two examples show why the specific form of the added mass variation shown in Fig. 36.4 causes the excitation to spread to a wide range of the nondimensional parameter *nominal reduced frequency*, $f_{\mathrm{n}}^* = \hat{f}_{\mathrm{n}} d/U$, or, equivalently, in terms of the *nominal reduced velocity*, $V_{\mathrm{rn}} = U/(\hat{f}_{\mathrm{n}} d)$. If the added mass had the reverse trend, that is, negative values at large frequencies and large values at lower frequencies, VIV would be an insignificant phenomenon, because it would be virtually impossible to find a resonant peak.

Wake Capture and Excitation Cause Lock-In

Two basic concepts that result from forced vibrations are those of *wake capture* and *wake excitation* regions. Wake capture is the gray-shaded region in Fig. 36.7, wherein the wake of a vibrating cylinder forms vortices at exactly the frequency of cylinder vibration, rather than at the Strouhal frequency. Outside this region, the loads are found to contain two frequencies, one at the cylinder vibration frequency, and another close to the Strouhal frequency.

The second shown region, denoted as the positive energy region in the graph, is the region where the lift coefficient in phase with velocity is positive (Fig. 36.7), hence providing energy from the flow to the system. Free vibrations will take place only within this region, since the fluid provides the needed energy for vibration, which is dissipated through the structural damping. Vibrations in the overlap region between these two regions are monochromatic free vibrations; hence the overlap region is what is usually referred to as the *lock-in region*.

A defining feature of lock-in is effective resonance. In classical dynamics, resonance occurs when a system is excited at its natural frequency. Large motions occur at resonance as energy is stored in a balance between the kinetic energy of the moving system mass and potential energy in the system springs. In VIV, as described above, the frequency at which the system oscillates is dependent on the added mass, which in turn is a function of the oscillation frequency and amplitude. The dynamic equilibrium achieved results in the system's oscillation frequency acting as an effective natural frequency for the cylinder, such that

$$f_{\mathrm{n,eff}} = \sqrt{\frac{k}{m + m_{\mathrm{a}}}}\,. \tag{36.15}$$

Since the effective natural frequency changes as a function of the effective added mass and this effective natural frequency is equal to the frequency at which the system oscillates, the cylinder is always in a state of effective resonance when lock-in occurs.

Vortical Patterns of an Oscillating Body

Once the body is set in motion, the fluid forces are altered, and hence the vortical patterns change. If a steady-state vibration sets in, this is a dynamic equilibrium between the structural forces and the fluid forces, where one must account for the strong feedback mechanism that alters the fluid forces.

We start with bodies undergoing free cross-flow vibration only. As a function of the reduced velocity V_{rn}, initially the wake resembles the Karman street, forming two vortices per period (*2S* mode). After the peak response has been reached, however, four vortices per cycle form, appearing as two pairs of counter-rotating vortices (*2P* mode); initially, the two additional vortices

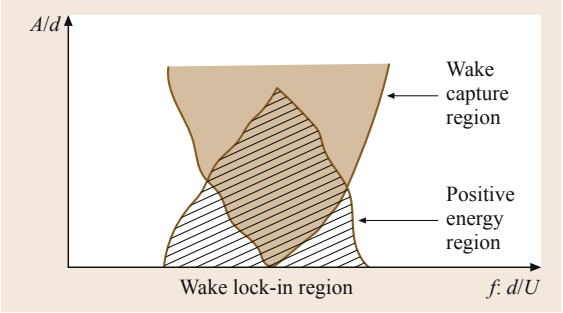


Fig. 36.7 Wake capture versus wake excitation (positive energy) regions

are weak and grow in strength with increasing reduced velocity. *Williamson* and *Roshko* [36.25] mapped the form of the wake for forced cross-flow vibrations as a function of the amplitude and frequency of vibration, discovering a variety of patterns. The basic patterns, as applicable to one degree of freedom, free vibrations are the *2S* and *2P* patterns.

More complex patterns can appear for large enough forced oscillatory motion of a cylinder [36.25], including asymmetric patterns. When in-line motions of the body are allowed in combination with transverse oscillations, such as in two degree of freedom oscillations or cantilevered cylinder oscillations, multiple vortex patterns may form [36.7, 46, 47], including a *2T* pattern, where triplets of vortices form per half-cycle.

The particular vortex shedding pattern combined with the motions of the cylinder govern the magnitude, phasing, and frequency content of the forces exerted on the body. The particular shedding patterns can be directly related to observations in the change of the phase between the lift force and transverse motion of the body along with a change in the sign of the effective added mass. In the *2S* mode of vortex shedding, the body moves toward the nearest shedding vortex in the wake. The motion slightly lags behind the force if there is positive damping in the system; however, the phase difference between the force and body motion will be very small. When the wake transitions to the *2P* mode, the second vortex of the pair forms as the body is moving away from the pair. Hence, the body is moving away from the nearest shed vortex, rather than toward it. This leads to a 180° phase shift, since the sign of the body's motion is switched with respect to the force exerted by the vortices. This sign change also corresponds to a change in the sign of the effective added mass, as the force exerted by the vortex now opposes the body motion (but is completely in phase with acceleration of the body).

36.1.5 Correlation Length

In order for significant VIV to occur, the vortex shedding, and hence the excitation force, must be well correlated with the length of the structure, that is, forces must have the same phase along the span. Studies show that the motion of the cylinder reinforces correlation; a three-dimensional (3-D) instability that causes the appearance of streamwise intense-vorticity, but of low circulation (compared to the Karman vortices) *rib* vortices, is the cause of an apparent drop in correlation. Nonetheless, by applying proper filtering, it is found that the correlation length is large in vibrating cylinders in uniform or mildly sheared flow.

It has been found, however, that for values of the reduced velocity near peak response, the correlation between the lift forces measured at the ends of a finite-span cylinder section drops to very low values [36.41]. For a tapered cylinder in a uniform flow, this lack of correlation has been associated with the appearance of a *hybrid mode* of vortex shedding [36.48]. Numerical simulations [36.49] have shown the details of such hybrid mode formation in uniform-section cylinders placed in linearly sheared flow.

The drop in correlation has been found for a wide range of Reynolds numbers, showing that the basic phenomenon is controlled by large-scale vortex dynamics. A comparison of the response and correlation length in a smooth cylinder to that of a cylinder fitted with tripping wires, used to force earlier transition to turbulence in the boundary layer of the cylinder (following the procedure of [36.50–52]), shows that there is a decrease in the reduced velocity region where the correlation drops for the cylinder with wires, but the qualitative features remain the same. Remarkably, the region of low correlation is associated with a region of very high amplitude response. Detailed studies in the wake of a cylinder [36.40], using hot wire anemometry, show that the lack of correlation manifests itself in the velocity fluctuations at various locations downstream and along the cylinder span.

The lack of correlation is resolved by considering the fact that, especially when the structural damping is small, the free vibrations are found to occur at the transition between *2S* and *2P* patterns. It is possible, therefore that part of the cylinder is under the *2S* regime and the rest under *2P*. Small variations in the incoming flow or the diameter can trigger such variations in the vortical patterns which are better viewed when the cylinder diameter is tapered [36.48]. The paradox, however remains, because one has to explain how an uncorrelated lift force causes maximum vibrations.

The fact that low correlation is associated with high amplitude vibrations turns, in fact, not to be a paradox. In the case of the formation of hybrid modes, these are structurally stable patterns, not randomly varying excitation. Also, it is important to note that the magnitude of the lift coefficient in phase with velocity is not what controls how high the response will be, at least for very low damping systems. The controlling parameter is the value of A/D at which the c_{lv} coefficient becomes zero. In Fig. 36.6, we show a typical c_{lv} versus A/D curve for constant reduced velocity, as well as the related lift coefficient in phase with acceleration. The slope at the zero down-crossing is equal to the hydrodynamic damping [36.53], while the maximum value of lift is not a crucial quantity as far as maximum amplitude is concerned.

Combined In-Line and Cross-Flow Motions

In experimental testing, the dynamics involved with VIV are typically simplified to study only cross-flow motion of the body since for a fully flexible structure, transverse oscillations of the body are typically larger compared with in-line oscillations. The constraints imposed by this simplification, however, lead to different forces exerted on the body, since the cylinder is not free to react to fluid forces in the in-line direction. Allowing combined in-line and cross-flow motions undergoing VIV leads to larger amplitude oscillations, larger magnitude forces, and larger harmonic components of the forces due to changes in the relative motion of the cylinder with respect to vortices shed in the wake.

As explained earlier, the peak frequency in the oscillatory drag force is typically at twice the frequency in lift. This unsteady drag force may also lead to a significant oscillation of a flexible body in the in-line direction. If the 2 : 1 frequency ratio is maintained in the resulting in-line and cross-flow motion, the structure oscillates with either a figure-eight or crescent shape motion, depending on the phase relation between the two motions. If we define a harmonic motion in the cross-flow direction $y(t)$ with frequency ω, and a harmonic motion in the in-line direction $x(t)$ with twice the cross-flow frequency, we can denote the phase difference between these motions, θ,

$$y(t) = A_y \sin(\omega t) \,, \tag{36.16}$$

$$x(t) = A_x \sin(\omega t + \theta) \,. \tag{36.17}$$

The phase angle θ alters the orbital motion of the cylinder, such that the cylinder may oscillate with a figure-eight shape, curved figure-eight, or crescent shape as illustrated in Fig. 36.8. Depending on the value of this phase, the curved figure eight or crescent shape may be oriented to curve toward the upstream or downstream current. Additionally, the direction of the path of the orbital shape is also dependent on the phase. For example, both phases of 0 and 180° will look like a regular figure-eight; however, the direction of the path relative to the free stream will be opposite in these cases. To define the direction of the figure-eight motion, we can use the direction of the motion when the cylinder is at the top of the cross-flow motion. For example, if the fluid is flowing left to right and the cylinder is oscillating such that it moves upstream at the two extremes of the cross-flow motion, then the cylinder is defined to be moving counterclockwise and the phase is between −90 and 90°. If the cylinder is moving in the opposite direction, such that its motion is downstream at the extremes of the cross-flow response, then it is said to be moving in a clockwise motion and the phase if between −180 to −90° and 90 to 180°.

Since the excitation frequency of the in-line motion will typically be at twice the excitation frequency of

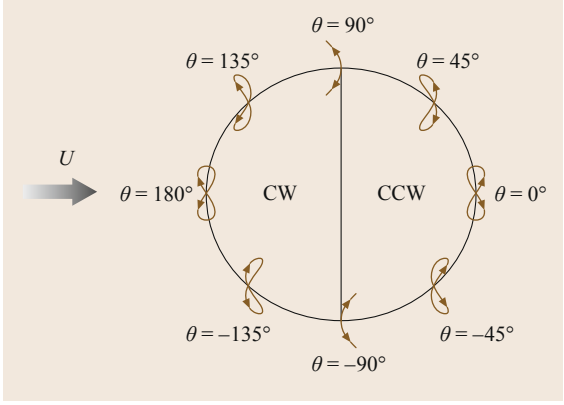


Fig. 36.8 Definition of phase angle between in-line and cross-flow motion θ based on sine motion. Flow is defined as left to right. Definition of counterclockwise (CCW) and clockwise (CW) orbits are noted. Cosine motion of the body will result in a 90° phase shift of θ

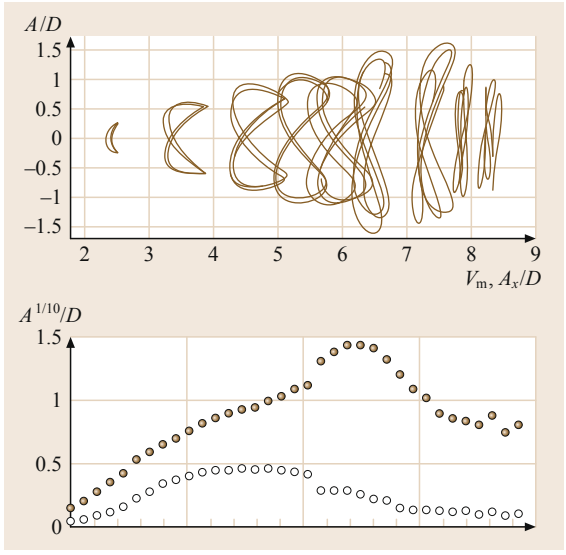


Fig. 36.9 **(a)** Horizontal axis V_{rn}, A_x/D; vertical axis A_y/D. (The horizontal axis plots both V_{rn} which is marked 2, 3, 4, etc., as well as the inline motion A_x for several subfigures shown) **(b)** Horizontal axis V_{rn}; vertical axis $A^{1/10}/D$. Definitions: D is the diameter of the cylinder, A_x is the amplitude of motion in the in-line direction, A_y is the motion in the cross-flow amplitude of motion, $A^{1/10}$ is the one-tenth highest average amplitude of cross-flow (*black circles*) and in-line (*open circles*) motion; V_{rn} is the reduced velocity, defined as $V_{rn} = U/(f_n D)$, with U the current velocity, and f_n the natural frequency of the flexible cylinder [36.54]

the cross-flow motion, a condition of dual resonance may be defined where the cylinder will oscillate with a 2 : 1 in-line to cross-flow frequency ratio, such that the effective natural frequencies of the system have this 2 : 1 ratio [36.35]. For long, flexible structures, the structure will behave as a tensioned beam or tensioned string, such that there are a large number of natural frequencies associated with various modal excitations of the structure. It is therefore possible for long, flexible structures to have varying natural frequency ratios along the length of the structure, depending on which modes of the structure are excited. This phenomenon may be modeled using an elastically mounted rigid cylinder by tuning the vacuum natural frequency of the cylinder separately in the in-line and cross-flow directions. A natural frequency ratio of 1 : 1 would denote equal natural frequencies in the in-line and cross-flow directions, which would result in the excitation of the same mode on a long, flexible structure. A frequency ratio of 2 : 1 would denote the in-line natural frequency to be twice the cross-flow natural frequency. For an infinite string, for example, this would result in exciting the cross-flow motion at the fundamental mode and the in-line motion at the second mode.

Figure 36.9 shows the in-line and cross-flow responses, as well as the trajectories at mid-length of a flexible beam towed at constant velocity as a function of the reduced velocity. Figure 36.9 counterclockwise trajectories dominate the peak vibratory response [36.54].

This experimental technique has been implemented to study the response of a freely vibrating rigid cylinder, elastically mounted, allowing combined cross-flow and in-line motions [36.35, 46, 55, 56]. In general, the magnitude of the observed response of a cylinder free to oscillate in two degrees of freedom can be significantly larger than for a cylinder constrained to cross-flow motion only. Whereas a cylinder with low mass ratio and damping, constrained to cross-flow motions will typically reach a maximum response near $A_y/D \approx 1$, a similar cylinder free to oscillate in both the in-line and cross-flow directions may exceed transverse amplitudes of $A_y/D = 1.5$. As a function of the increased response for these motions, additional vorticity may shed in the wake in the form of vortex triplets, leading to the *2T* wake mode, which consists of two triplets of vortices shedding per cycle of motion [36.46].

In addition to increased cylinder motion and slight alterations to the wake, the forces exerted on a cylinder in combined cross-flow and in-line motion may be significantly different than those observed for equivalent cross-flow motion only [36.57]. The largest difference is the presence of higher harmonic forces in the lift and drag forces. Large magnitude third-harmonic forces in lift have been observed for two degrees of freedom VIV [36.46, 47]. As vortices shed from the cylinder, the relative motion of the cylinder with respect to the shed vortices governs the magnitude and frequency of the forces exerted on the body. For a cylinder shedding a typical Karman street and oscillating with a figure-eight pattern, the relative motions of the body can

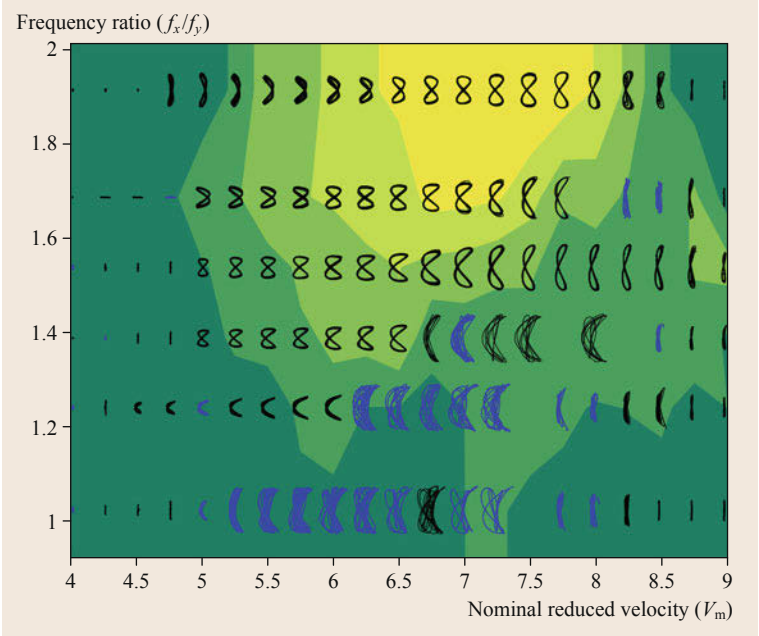


Fig. 36.10 Response trajectory of a flexibly mounted cylinder as a function of reduced velocity and ratio of natural frequencies in the in-line and cross-flow directions f_x/f_y (after [36.35]). *Black* denotes counterclockwise motion, *blue* clockwise motion. Color denotes intensity of high harmonic forces (*yellow* highest, *green* lowest)

lead to a dominant third-harmonic force, even though the dominant response of the cylinder occurs at the fundamental frequency [36.35]. Figure 36.9 illustrates how the relative magnitude of the third-harmonic force varies as a function of the natural frequency ratio and orbital shape.

Although the ratio between the natural frequencies in the in-line and cross-flow directions may vary, due to the significant variation in the effective added mass, the effective natural frequency ratio between in-line and cross-flow directions will be 2 : 1 when lock-in occurs in both directions. This condition is referred to as dual resonance [36.35], since the system resonates in both the in-line and cross-flow directions. Regardless of the nominal natural frequency ratio, in dual resonance, the values of the effective added mass in both the in-line and cross-flow directions vary until they provide a double resonance condition. This is sustained as long as the wake vortical patterns can be sustained to provide the appropriate values of added mass (Fig. 36.10).

In an analogous phenomenon, it has been shown that in cross-flow oscillations and for a mass ratio less than a threshold value of about 0.56, the effective added mass adjusts to a negative value that expands the effective natural frequency over an infinite range of reduced velocity [36.7].

36.2 VIV Prediction of Flexible Structures in Nonuniform Flow

Moving to the case of a flexible cylinder placed in steady cross-flow, one must reformulate the governing equations to account for the fact that the vibration is a function of length along the span as well as time. Equations (36.7) turn to integral relations (integrating across the span) rather than algebraic relations. The details including numerical implementation can be found in [36.58].

Slender flexible bodies undergoing VIV can be modeled as linear tensioned beams subjected to fluid forcing. If, as previously introduced in Sect. 36.1.3, x and y denote the in-line and cross-flow displacements, respectively, which are now functions of time t and spanwise position z, and X and Y denote the sectional forces exerted by the flow on the body, then the dynamics of such structure can be expressed as follows,

$$\begin{cases} m\dfrac{\partial^2 x}{\partial t^2} - T\dfrac{\partial^2 x}{\partial z^2} + EI\dfrac{\partial^4 x}{\partial z^4} + b\dfrac{\partial x}{\partial t} = X\,, \\ m\dfrac{\partial^2 y}{\partial t^2} - T\dfrac{\partial^2 y}{\partial z^2} + EI\dfrac{\partial^4 y}{\partial z^4} + b\dfrac{\partial y}{\partial t} = Y\,, \end{cases} \tag{36.18}$$

where m is the structure mass per unit length, T is its tension, EI is its bending stiffness, and b is the structural damping. This second-order system must be complemented by initial and boundary conditions such as, for example, $x = y = \partial^2 x/\partial z^2 = \partial^2 y/\partial z^2 = 0$ at the ends, for a simply supported structure. The expansion of the body displacements on a suitable modal basis and the Galerkin projection of expression (36.18) onto this basis lead to the following dispersion relation, which links in vacuum, the n-th structural mode of nondimensional spatial wavelength $\lambda_n = 2L/nD$, to the corresponding nondimensional natural frequency f_n

$$f_n = \frac{nD}{2L}\sqrt{\frac{1}{mU^2}\left(T + EI\frac{n^2\pi^2}{L^2}\right)}\,, \tag{36.19}$$

where U is the oncoming flow velocity. A slender flexible structure presents an infinite set of natural frequencies. Since VIV are characterized by synchronization of body oscillation and vortex formation over a range of frequencies in the vicinity of the Strouhal frequency, the above dispersion relation may provide a rough estimate of the possible excited modes. It also shows that when such a body is immersed in a sheared current, that is, with a frequency of vortex shedding that varies along the structure span, the vibrations may involve multiple structural modes.

36.2.1 Traveling Structural Waves and Multi-Modal Responses in Sheared Flows

When the oncoming flow is sheared, it is known to reduce the amplitude of vibration [36.59, 60]. An important effect of shear, however, in long tethers and risers is to cause the appearance of multifrequency response. Indeed, various parts of the structures are subject to excitation at different frequencies, since the oncoming velocity varies. Energy is input in a certain location – at or near the local Strouhal frequency – travels along the structure, and then is dissipated at another location. This necessitates the introduction of complex modes in the numerical prediction of VIV [36.27, 58]. In a classical standing wave analysis, energy cannot travel beyond nodes, and hence a different methodology is needed, one that allows standing waves, or traveling waves, or a combination of the two. This is achieved by assuming the amplitude to be complex, which introduces an amplitude and a constant phase at each location of the structure. An outstanding problem today is the partitioning of energy among the multiple frequencies existing in a long structure in shear flow.

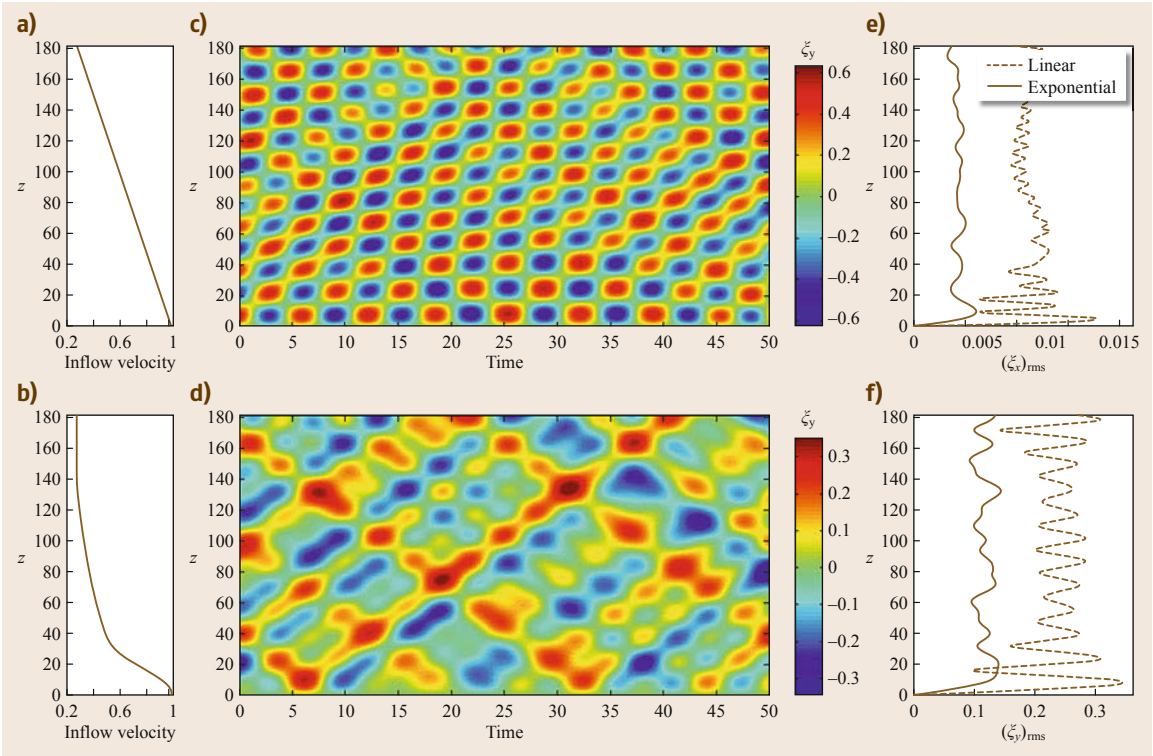


Fig. 36.11 **(a)** Linear and **(b)** exponential inflow velocity profiles, **(c,d)** selected time series of the cross-flow displacement along the span in the **(c)** linear and **(d)** exponential shear flow cases, **(e,f)** RMS value of the **(e)** in-line and **(f)** cross-flow displacements, along the span (after [36.61])

Gopalkrishnan et al. [36.62] treated the problem of a rigid cylinder section undergoing multifrequency vibration (two or three simultaneous frequencies). The total energy transfer from fluid to structure, under these conditions, is divided among the participating frequencies; also, the lift and the drag coefficients are affected. There is no general theory, however, available to model the partition of energy among the (several) frequencies that a very long cable or riser may be subjected to in a sheared flow.

Ocean engineering structures, such as marine cables or risers, are often exposed to currents with nonuniform velocity profiles. The VIV of slender flexible cylinders in nonuniform currents generally involve high structural modes and consist of a mixture of standing and traveling wave patterns [36.42, 63–67].

To illustrate the typical vibrations observed in this context, the structural responses obtained from coupling the system (36.18) with direct numerical simulation of the 3-D Navier–Stokes equations, are presented in Fig. 36.11. In these simulations, the cylindrical tensioned beam has a length-to-diameter aspect ratio equal to 200; it is immersed in sheared cross-flow with two distinct velocity profiles, either a linear or an exponential profile, as shown in Fig. 36.11a,b. In both cases, the Reynolds number based on the maximum inflow velocity is equal to 330. Selected time series of the cross-flow displacement are plotted along the span for the linear and exponential flow velocity profiles in Fig. 36.11c,d. In the case of linear shear, the structural waves appear mainly oriented from the high to the low inflow velocity regions. On the contrary, no preferential orientation of the waves can be identified in the exponential shear case. This aspect will be explained by the spanwise distribution of the excitation/damping regions in Sect. 36.2.2. Root mean square (RMS) values of the vibration amplitudes along the span are presented in Fig. 36.11e,f. The absence of vibration nodes along the span confirms the strong traveling-wave nature of the structural responses. These plots also illustrate the impact of the velocity profile shape on the response amplitudes.

Spatio-temporal spectral analysis based on a two-dimensional fast Fourier transform of the structural responses allows us to clarify the link between excited frequencies and structural wavenumbers/wavelengths, what we refer to as *modes*. In Fig. 36.12, the cross-flow displacement power spectral densities (PSD) associated

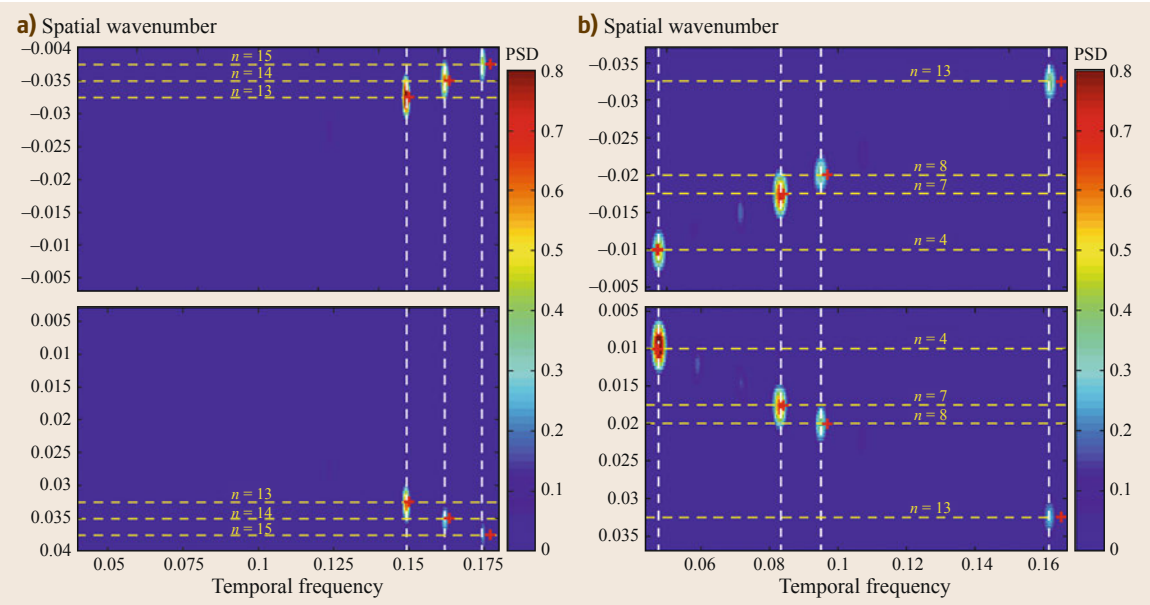


Fig. 36.12a,b Spatio-temporal spectral analysis of the cross-flow displacement in the **(a)** linear and **(b)** exponential shear flow cases. Selected vibration frequencies are identified by *white vertical dashed lines*. The wavenumbers of selected modes are indicated by *yellow horizontal dashed lines*. *Red crosses* denote the natural frequencies associated with these wavenumbers (after [36.61])

with the above vibrations are plotted as functions of the temporal frequency and spatial wavenumber. Both inflow velocity profiles lead to vibrations at multiple frequencies, but the structural responses differ in the width of the excited frequency band. Three main frequencies are excited along the span within a narrow band in the linear profile case, while a broadband response is observed in the exponential profile case. As a consequence, both *narrowband* and *broadband* VIV may be excited, depending on the shape of the oncoming flow velocity profile. It should be mentioned that, even under sheared current conditions, responses at a single frequency are also possible, in particular when the effect of the flow shear is counterbalanced by the curvature of the body [36.68]. Each excited frequency is generally associated with a single structural wavenumber.

In order to estimate the natural frequencies of the excited structural modes when the structure is submerged in the fluid, the dispersion relation in vacuum (36.19) may be modified as follows,

$$f'_{\mathrm{n}} = f_{\mathrm{n}} \sqrt{\frac{m}{m + c_{\mathrm{m}} \rho \pi \frac{D^2}{4}}} \,, \tag{36.20}$$

where c_{m} is the added mass coefficient, which is assumed to be constant along the span to derive this new relation. The natural frequencies predicted by considering an added mass coefficient $c_{\mathrm{m}} = 1$ (the theoretical value in potential flow) are indicated by red crosses in Fig. 36.12. Significant departures from the actual peaks may be observed; this is related to the strong variability

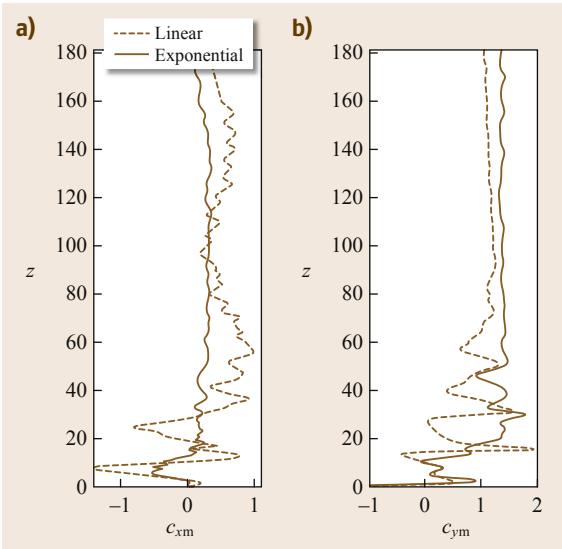


Fig. 36.13a,b Effective **(a)** in-line and **(b)** cross-flow added mass coefficients along the span. The potential flow value is equal to 1 (after [36.69])

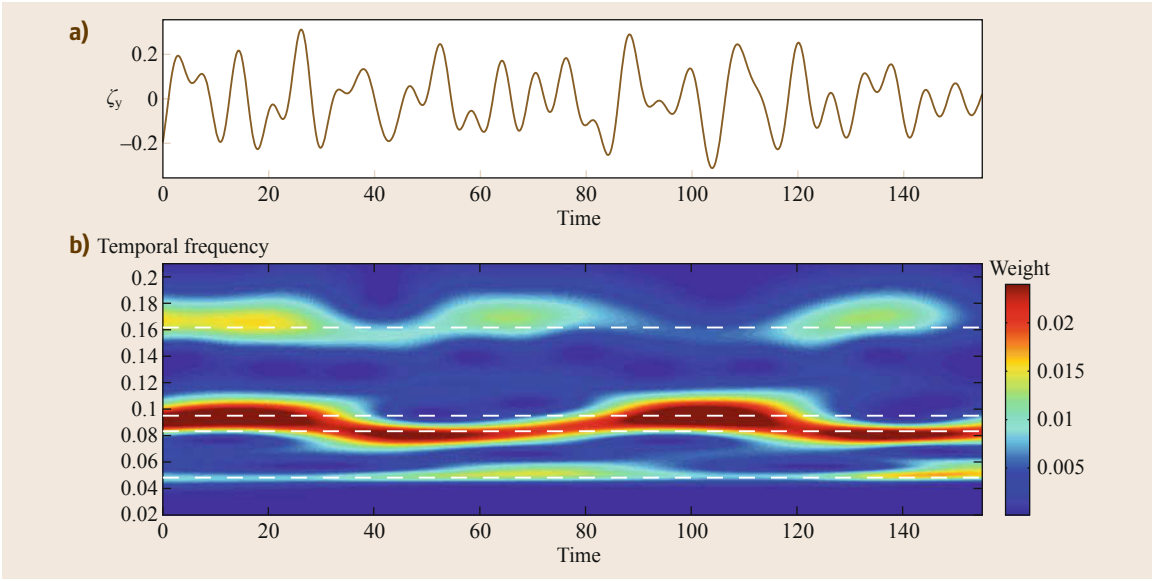


Fig. 36.14 **(a)** Selected time series of the cross-flow displacement and **(b)** corresponding frequency content as a function of time (scalogram). *Dashed lines* denote the predominant frequencies identified in Fig. 36.12

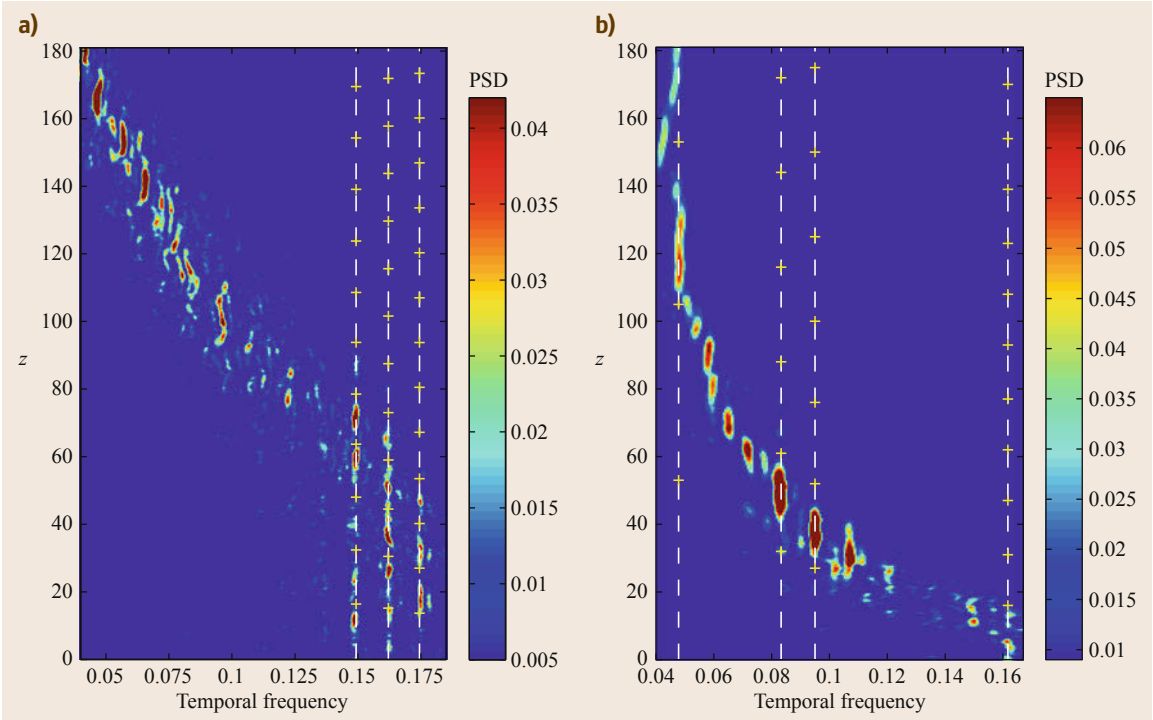


Fig. 36.15a,b Vortex shedding frequency identified by the PSD of the cross-flow component of flow velocity along a spanwise line in the wake, in the **(a)** linear and **(b)** exponential shear flow cases. *Dashed lines* indicate the frequencies of the cross-flow vibrations

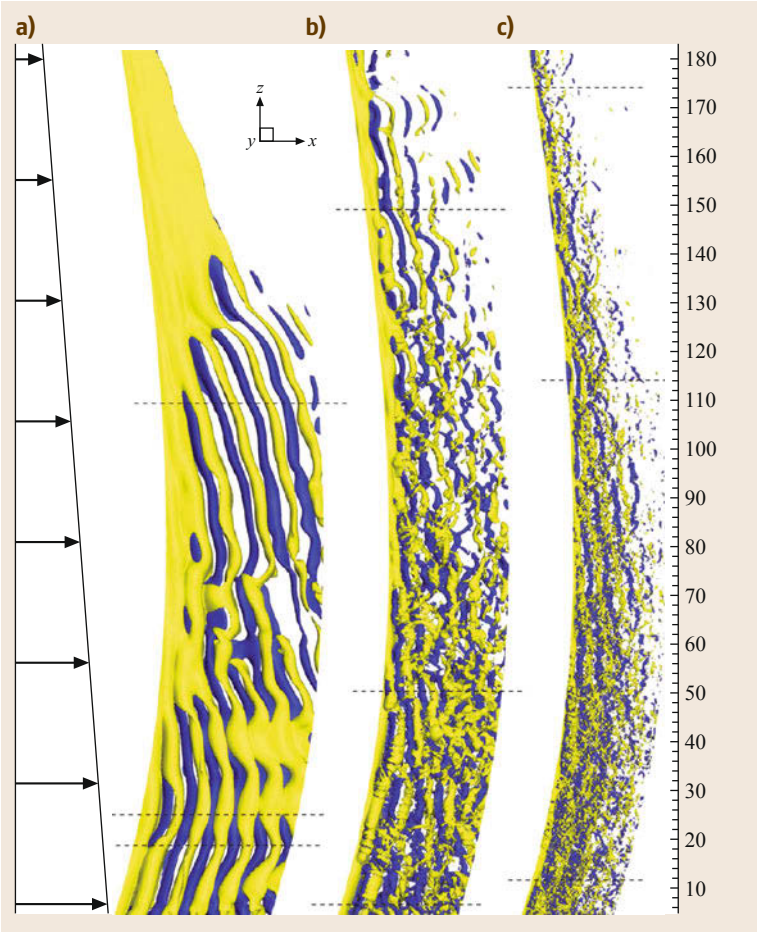


Fig. 36.16a–c Instantaneous isosurfaces of spanwise vorticity dowstream of a tensioned beam at **(a)** $Re = 110$, $\omega_z = \pm 0.13$, **(b)** $Re = 330$, $\omega_z = \pm 0.3$, **(c)** $Re = 1100$, $\omega_z = \pm 0.8$. *Arrows* represent the linear shear oncoming flow (after [36.67])

of the effective added mass as also noted for rigid-cylinder VIV. As can be observed in Fig. 36.13, the effective in-line and cross-flow added mass coefficients present large spanwise modulations and differ substantially from the potential flow value of 1.

As also noted in the rigid cylinder case, each excited frequency in the cross-flow direction can be related to an excited frequency in the in-line direction through a ratio of 2. Contrary to the pairs of in-line/cross-flow response frequencies, the corresponding spatial wavenumbers (structural modes) generally exhibit a ratio different from 2; this behavior is expected because of the nonlinear dispersion relation of tensioned beams, relating frequency and wavenumber.

The spectral analysis shows the existence of multi-frequency responses, where different structural modes are excited. At a given spanwise location, several structural wavenumbers can thus contribute to the total vibration. The question that arises is whether the response is instantaneously mono-frequency, with a shifting frequency in time, or it is multifrequency at all times. Typical time series of the cross-flow displacement in a spanwise region where several vibration frequencies exhibit significant contributions is plotted in Fig. 36.14a; the corresponding scalogram, which represent the instantaneous frequency content as a function of time using wavelets, is plotted in Fig. 36.14b. A temporal shift can be noted between frequencies associated with adjacent structural modes. However, if all vibration frequencies are considered, it clearly appears that multiple frequencies, and thus structural modes, can respond simultaneously. As a consequence, VIV of a flexible body can be *instantaneously multifrequency*.

36.2.2 Lock-In and Fluid-Structure Energy Transfer Within Shear Flow

For a freely vibrating rigid cylinder, the lock-in condition, under which energy is transferred from the flow to the structure, is defined as the VIV associated with synchronization between the vortex shedding and the body vibration. This definition can be extended to the case

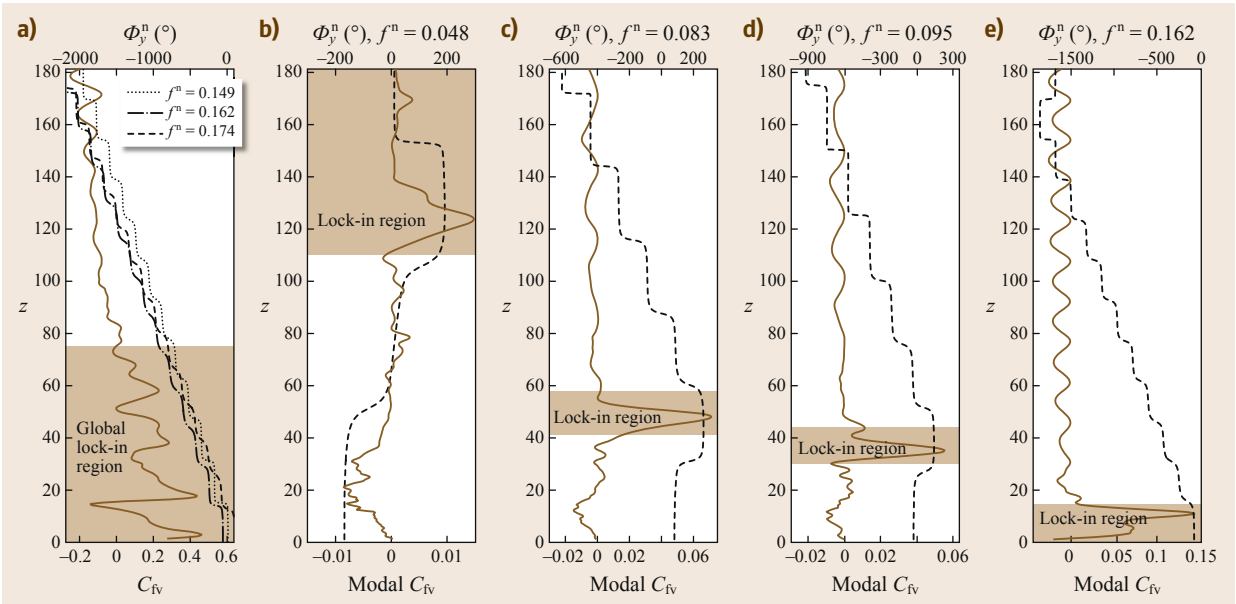


Fig. 36.17a–e Force coefficient in phase with the structure velocity (lower axis, *plain brown line*) and spatial phase of the cross-flow vibration (upper axis, *discontinuous black line*), along the span: **(a)** total c_{lv} and spatial phases of the three predominant vibration components in the narrowband response case, **(b–e)** modal c_{fv} and spatial phase of the vibration components in the broadband response case. The global lock-in region (in **(a)**) and the lock-in region associated with each vibration frequency (in **(b–e)**) are colored in *brown* (after [36.61])

of slender flexible bodies subjected to multifrequency VIV. The lock-in condition is defined at each spanwise location and each cross-flow vibration frequency as the local synchronization between the frequency of body oscillation and the local frequency of vortex formation; otherwise we refer to a nonlock-in condition. The vortex shedding frequency is plotted along the span in the previously considered linear and exponential shear flow cases in Figs. 36.15a,b respectively. The predominant cross-flow vibration frequencies are indicated by vertical dashed lines in these figures.

As also noted for stationary cylinders with varying spanwise conditions, that is, with either varying body diameter or nonuniform velocity profile, the wake exhibits a discontinuous pattern composed of cells of constant vortex shedding frequency [36.70–72]. The decrease of the vortex shedding frequency between two adjacent cells is globally driven by the decrease of the inflow velocity through the Strouhal relation. In the narrow-band response cases (Fig. 36.15a), the lock-in condition occurs over a limited length, on the high-inflow velocity side, while vortex shedding and body oscillation are not synchronized on the rest of the body. In contrast with this *localized* occurrence of the wake-body synchronization, the broadband vibration case (Fig. 36.15b) is characterized by a lock-in condition which is *distributed* along the span. Within a sheared current, the lock-in condition may thus appear simultaneously along most of the structure length, leading to broadband vibrations. As previously observed for rigid cylinder VIV, large deviations from the Strouhal shedding frequency are noted under lock-in. Under the lock-in condition, the wake is generally synchronized with a single structural frequency at each point of the span, even if several vibration frequencies coexist. As a consequence, the lock-in phenomenon remains a locally mono-frequency event, even in the case of broadband structural responses [36.61].

For a better visualization of the wake patterns, instantaneous iso-surfaces of spanwise vorticity downstream of a flexible cylinder subjected to VIV in linear shear flow, at three Reynolds numbers ($Re \in \{110, 330, 1100\}$), are presented in Fig. 36.16. The partitioning of the wake into spanwise cells of constant vortex shedding frequency leads to an oblique orientation of the vortex rows as they form, due to the variation of the vortex convection velocity within each cell. To ensure the continuity of the spanwise vortex filaments while the vortex shedding frequency is discontinuous, vortex splitting events similar to the *vortex dislocations* [36.69, 71, 73, 74] occur between adjacent cells.

As for the rigid cylinder, the transfer of energy between the flow and the oscillating structure is quantified by means of the fluid force coefficient in phase

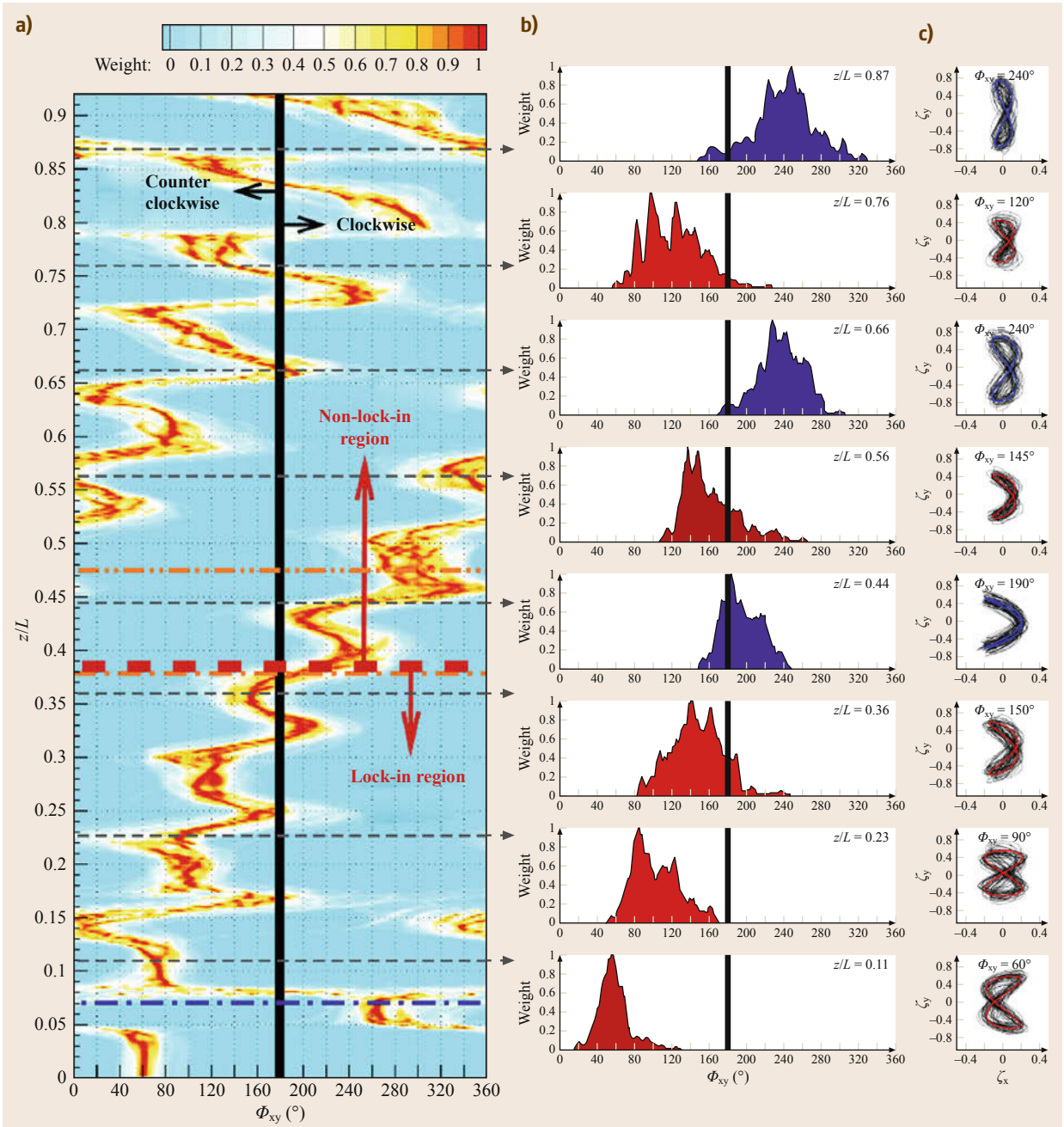


Fig. 36.18a–c Spanwise evolution of the histogram of phase difference between the in-line and cross-flow displacements (Φ_{xy}): **(a)** numerical simulation, **(b)** experiment. In **(a)**, the limit of lock-in region is indicated by a *red dashed line* (simulation); for the experiment, the estimated upper limits are indicated by *orange dashed-dotted* and *dashed-double-dotted lines*, lower limit by a *blue dashed-dotted line*. Locations of measurement points in **(b)** are indicated by gray *dashed arrows* in **(a)**. **(c)** Cross-flow versus in-line displacements. In **(b,c)** histograms and characteristic orbits are colored according to the preferential orientation: counter-clockwise/clockwise in *red/blue* (after [36.68])

with the body velocity. As shown in Fig. 36.17, where the spanwise evolution of this quantity is plotted for cases of narrowband and broadband vibrations, zones of positive energy transfer, that is, where the flow excites the structural vibrations, are located within the lock-in regions. Outside of the lock-in regions, the flow damps the body oscillations. Unlike the narrowband case, for broadband response it can be noted that the excitation zones are distributed along the length of the structure; high-wavenumber vibrations are excited from the high flow velocity side and low-wavenumber responses from the low inflow velocity side. For each excited frequency and regardless of the narrowband or broadband nature of the vibrations, the structural waves travel from the excitation region toward zones of damping, as illustrated in Fig. 36.17 by the orientation of the spatial phases of the waves. As a consequence, a broadband structural response may be composed of waves traveling in both directions, that is, toward decreasing or increasing inflow velocity directions. This behavior contrasts with the narrowband response case where the traveling waves are all oriented from the high to the low inflow velocity regions.

36.2.3 Synchronization of In-Line and Cross-Flow VIV

As in the case of rigid cylinder VIV, the in-line and cross-flow responses of a flexible body are coupled nonlinearly through the fluid forces and their synchronization is closely related to the occurrence of the lock-in condition and therefore to the energy transfer between the structure and the fluid. The in-line and cross-flow vibration components generally exhibit a relative frequency ratio of 2. In the case of mono-frequency responses, this leads to figure-eight trajectories in the plane perpendicular to the body axis, similarly to the rigid cylinder case. The shape and orientation of the figure-eight trajectories are controlled by the phase difference between the in-line and cross-flow displacements. Compared to the rigid body case, the flexibility of the structure allows this phase difference to vary along the span. The difference between the instantaneous phases of the in-line and cross-flow displacements (ϕ_x and ϕ_y respectively) may be evaluated as follows [36.75],

$$\Phi_{xy} = [\phi_x - 2\phi_y, \quad \mod 360^\circ] \,. \tag{36.21}$$

Values of Φ_{xy} in the range 0–180° (180–360°, respectively) correspond to figure-eight orbits where the beam moves upstream (downstream respectively) when reaching the cross-flow oscillation maxima. These two types of trajectories are referred to as *counter-clockwise* and *clockwise*, respectively [36.47].

Examples of spanwise evolutions of the in-line/cross-flow response phase difference issued from numerical simulation and experiment are presented in Fig. 36.18. In both cases, the studied system is composed of a long tensioned beam placed in linear shear flow; the structure exhibits mono-frequency VIV in each direction. A striking phenomenon is that the spanwise region where the lock-in condition is established is dominated by counter-clockwise orbits. Hence in-line and cross-flow vibrations are locked to a specific phase difference range within the lock-in region. The clockwise orbits primarily occur outside the lock-in region and are thus principally associated with damping fluid forces. The mechanism of phase-locking of the vibration components under wake-body synchronization is found to extend to the case of multifrequency VIV, including the case of broadband responses, where both high and low structural wavenumbers are excited: a Fourier expansion of the in-line and cross-flow responses and the association of in-line to cross-flow components with frequencies in the ratio of 2 to 1, allows the identification of a relative phase difference for each such pair; the phase difference angle is contained within the same range, associated with counter-clockwise orbits, as for mono-frequency responses [36.32].

36.3 Experimental Studies and Fatigue Analysis

Several experiments have been conducted to study VIV of flexible cylinders, in a controlled lab environment or in the field. In these tests, structural response at various points along the length of the cylinder is measured using strain gauges or accelerometers. Some of these tests are available online at the MIT VIV data repository [36.76]

The Norwegian Deepwater Programme (NDP) Riser High Mode VIV tests [36.77] are one of the most comprehensive laboratory experiments conducted to study VIV of flexible cylinders. The tests comprise uniform flow cases, where the flow velocity is constant along the riser and varies from $U = 0.3$ to $2.2\,\mathrm{m/s}$ with $0.1\,\mathrm{m/s}$ steps in various test cases; and linearly sheared flow cases, where the flow velocity is zero at one end and increases linearly along the length, with maximum flow velocity varying from $U = 0.3$ to $2.2\,\mathrm{m/s}$ with $0.1\,\mathrm{m/s}$ steps. Physical properties of the riser used in the experi-

Table 36.2 Physical properties of the NDP riser

Outer diameter (D)	0.027 m
Length (L)	38 m
Displaced water/length (m_a)	0.576 kg/m
Mean tension (T)	4000–6000 N
Modulus of elasticity (E)	2.25 GPa

Table 36.3 Physical properties of the Gulf Stream riser

Inner diameter	0.0249 m
Outer diameter	0.0363 m
EI	613 Nm^2
EA	3.322 MN
Weight in seawater	0.1972 kg/m
Weight in air	0.760 kg/m
Density of pipe material	1383 $\mathrm{kg/m}^3$
Effective mean tension	3225 N
Material	Glass fiber
Length	152.524 m

ments are given in Table 36.2. The model riser was taut horizontally and towed at various speeds, producing either uniform flow profiles if the two ends were towed at the same speed, or linearly sheared flow profiles if one end was fixed and the other end was towed. The in-line strain was measured at 40 points along the length, the cross-flow strain at 24 points, and the in-line and cross-flow accelerations at 8 points. There are 8 points along the length with measurements of both strain and acceleration, and in both cross-flow and in-line directions. The experiments were conducted on bare risers as well as risers with varying strake coverage.

Another series of experiments are the Gulf Stream tests [36.30, 78], which were conducted using a composite circular cylinder with properties listed in Table 36.3 [36.78] to observe high mode numbers VIV (modes greater than approximately 10th mode). The experiments were conducted by towing the cylinder from a boat. A railroad wheel was attached at the bottom end of the riser to provide tension. The riser was a simply supported tension dominated one, where universal joints were used at its two ends to create the pinned connections. The oscillations were measured using fiber optic strain gauges. An acoustic Doppler current profiler (ADCP) was used to measure the magnitude and direction of the incident flow profile.

36.3.1 Riser Orbital Motions and Excitation Region

The riser response data are not temporally statistically stationary. However, there are data segments consisting of dozens of cycles which are stationary. For such a stationary segment, the orbital motion of the structure – obtained by plotting the cross-flow oscillations versus the inline oscillations – shows either a figure-eight or a crescent like motion. If the signal is cut into subintervals, each corresponding to one cross-flow and two in-line periods, the direction of motion of the riser orbit for each subsignal can be found by calculating the phase difference between the cross-flow and the in-line displacements. The short period considered in each subsignal makes it possible to assume that the signal is sinusoidal. Therefore, one can assume

$$y = Y \sin(\omega t) \,, \tag{36.22}$$

$$x = X \sin(2\omega t + \psi) \,, \tag{36.23}$$

where ψ can be obtained by measuring the time difference between the cross-flow displacement zero up-crossing (the first point in each subsignal) and the following in-line displacement zero up-crossing (δ). Once this is measured, the phase difference can be calculated as

$$\psi = 2(2\pi/\delta)/(\Delta t) \,. \tag{36.24}$$

Based on the orbit and the calculated phase differences, the motion is characterized as clockwise or counterclockwise; counterclockwise is defined when the riser is at the top of its trajectory and moving upstream with the cross-flow from left to right. For the NDP data cases, the orbital motion of the riser at every point, where acceleration measurements were obtained, was calculated. The counterclockwise figure eight motions are observed mainly in the region with high local flow velocity.

36.3.2 Higher Harmonic Strain and Acceleration Components in Flexible Structures

In many experimental VIV data, if the fundamental frequency of oscillation for the cross-flow strain and acceleration signals is denoted by f_{CF}, a peak at three times this frequency, that is, around $f_{3\mathrm{CF}} = 3f_{\mathrm{CF}}$ is also observed. This is called a higher harmonic component. Higher harmonics are observed in various experimental data [36.79, 80].

As an example, Fig. 36.19 shows a sample PSD plot for an acceleration signal of one of the uniform flow cases with $U = 1.3\,\mathrm{m/s}$ from the NDP data. The fundamental frequency is at $f_{\mathrm{CF}} \simeq 7\,\mathrm{Hz}$, which is in the neighborhood of the expected Strouhal shedding frequency, $f = U\mathrm{St}/D \simeq 8\,\mathrm{Hz}$. A large third-harmonic peak exists at $f_{3\mathrm{CF}} \simeq 21\,\mathrm{Hz} = 3f_{\mathrm{CF}}$ and a rather smaller one at $f_{5\mathrm{CF}} \simeq 35\,\mathrm{Hz} = 5_{\mathrm{fCF}}$. The main frequency in the in-line direction is around $f_{\mathrm{IL}} = 2f_{\mathrm{CF}}$, as expected. The

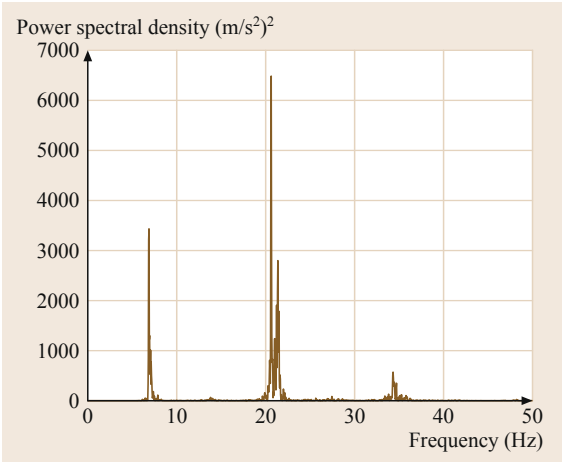


Fig. 36.19 A typical PSD plot of a measurement point with a large peak at the third harmonic (around 21 Hz) and a small peak at the fifth harmonic (around 35 Hz) (after [36.80])

contribution of higher harmonic components in the transverse direction is determined by evaluating the ratio of the area under the higher harmonic peak in the PSD plot over the total area under the PSD. For example, in Fig. 36.19, if the areas under the third-harmonic and the fifth-harmonic peaks are denoted by A_3 and A_5, respectively, then the ratios A_3/A_{Total} and A_5/A_{Total} are the measures of the relative importance of these higher harmonic components, where A_{Total} is the total area under the PSD.

For the NDP uniform current cases, a third-harmonic component of around 0.5 (50%) is widely observed. For linearly sheared cases, the third-harmonic contribution is mainly around 30–40% with some exceptional points where a relatively large contribution of up to 80% is observed. The influence of a third harmonic of around 0.3 (30%) on fatigue life is enough to be nonnegligible. For the fifth-harmonic components a contribution of mainly 15–20% is observed both for the uniform and the linearly sheared cases, with some cases of the fifth harmonic up to 30%.

36.3.3 Periodic (Type-I) and Chaotic (Type II) Signals

The riser response data are not always statistically stationary; in fact, the VIV of long flexible risers are characterized by time intervals of chaotic response, followed or preceded by periods of statistically stationary response. Regions of chaotic response contain distinctly different response features and have significant implications on riser fatigue analysis. While regions of statistically stationary response are characterized by nearly mono-frequency traveling waves, with small standing wave contributions, near the ends of the riser, and possibly accompanied by sharply peaked third- and fifth-force harmonics, the chaotic response is characterized by a wide-band spectrum with several individual peaks and a mix of traveling and standing waves.

Figure 36.20a shows the time history of a rare case which was specifically selected to exhibit the statistically stationary response that is typically assumed to occur. Figure 36.20b shows a more commonly occurring response, exhibiting features of a chaotic signal.

Using experimental data of VIV of long flexible cylinders, three types of VIV behavior are observed:

i) Cases where the entire signal is mainly quasi-periodic
ii) Cases where the signal exhibits quasi-periodic oscillations interrupted by bursts of chaos
iii) Cases where the entire signal is chaotic and even the smallest subsignals confirm chaotic features.

Focusing on the Strouhal region of the response spectrum gives adequate results for the statistically stationary response, if the higher harmonics are also accounted for, but it is inadequate for the chaotic response, whose fatigue properties are influenced by the entire broad-band spectrum. It is remarkable that both sheared and uniform current profiles cause both quasi-periodic and chaotic responses. Chaotic oscillations are

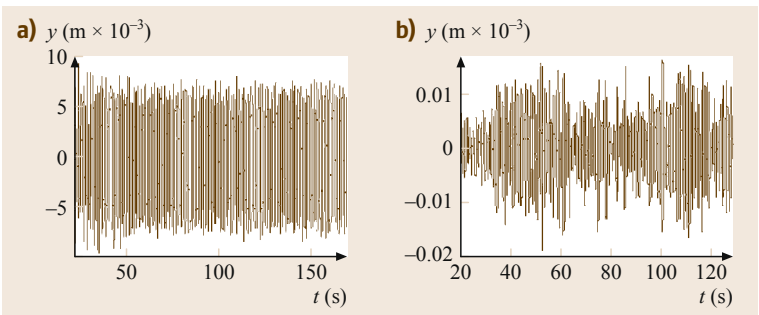


Fig. 36.20a,b The entire signal at **(a)** point 1 ($z \simeq 4$ m), case 2310 of the NDP data, with a linearly sheared flow profile, $U_{\max} = 0.3$ m/s, showing steady-state VIV behavior, which is rare in practice; **(b)** point 1 ($z \simeq 4$ m), case 2320 of the NDP data, a linearly sheared flow case, showing chaotic response, a commonly found VIV behavior in field tests (after [36.81])

observed not only at various intervals of time, but also at various locations of the riser, and both for the uniform and linearly sheared flow cases.

36.3.4 Response Reconstruction Using Experimental Data

Response measurements are conducted at some points along the length of a flexible structure, typically using strain gauges and accelerometers. A continuous VIV response can be reconstructed using data from limited number of sensors placed along its length. When the number of sensors (N_s) is sufficient (namely, $N_s \geq N_m + 1$, where N_m is the number of spatial harmonics), the problem of reconstructing the riser VIV response can be posed as a spatial Fourier decomposition. The displaced shape of the riser at any instant of time is written as a spatial Fourier series. The reconstruction problem is posed as a system of linear equations, where the unknown spatial Fourier coefficients at every instant of time are evaluated. Since the number of sensors is larger than the critical number, this is an over-determined system of linear equations. This system is solved using a pseudo-inverse method, where the least-squares error is minimized. This method allows estimation of the continuous motion of the riser (at every point along its length) and therefore fatigue life estimation is not limited to only the measurement points [36.82]. There are three main sources of uncertainties in this reconstruction method:

1. Uncertainty due to the presence of noise in experimental data
2. Uncertainty due to the use of limited number of sine and cosine terms, and
3. Uncertainty due to the presence of both acceleration and strain measurements.

Due to these uncertainties, the amplitude of reconstructed response at every point along the length will have an upper bound and a lower bound.

36.3.5 Fatigue Calculations

In order to calculate fatigue damage of the riser an S-N curve of the form $N = A(1/S)^B$ is used, where N is the number of cycles, S is stress in MPa, and A and B are constants, depending on the material. For the riser used in the NDP tests, these parameters are $A = 4.8641 \times 10^{11}$ and $B = 3.00$ [36.83]. Fatigue damage calculations are conducted based on rainflow fatigue prediction method [36.84].

The Influence of Higher Harmonics on Fatigue Damage

In the NDP experimental cases, higher harmonic components of strain and acceleration are significant. Fatigue damage calculations, by taking into account the higher harmonic components, give fatigue damage values of up to one order of magnitude larger than those calculated based on only first-harmonic signals. Two examples of uniform and linearly sheared NDP cases are discussed in what follows.

Fatigue calculations can be conducted using the measured signals directly. Using the unfiltered signals, the most realistic fatigue damage can be calculated. These are shown by triangles in Fig. 36.21 for two sample cases of NDP data. Using the signals filtered around their fundamental frequency, a smaller fatigue damage is calculated (circles shown in Fig. 36.21), since the influence of higher harmonic components is not considered. These points can be compared with the continuous line which is fatigue damage calculated based on the reconstructed (and therefore purely first harmonic) signals. The limits due to uncertainties arising in reconstruction procedure are shown by thin lines. Once the fatigue damage is calculated using the purely first-harmonic signal, the following procedure can be used in order to estimate a more realistic fatigue damage. Assuming a worst case scenario in which the magnitude of the total stress is equal to the summation of the magnitudes of the stress due to each of the harmonics, the total stress is written as

$$S_T = S_1 + S_3 + S_5 \tag{36.25}$$

or

$$S_T = S_1(1 + H_3 + H_5)\,, \tag{36.26}$$

where, S_1, S_3, and S_5 are magnitudes of the first, third- and fifth-harmonic stress, respectively and are not functions of time, and H_3 and H_5 are the higher harmonic components calculated using the unfiltered signals. Therefore the life can be predicted as $N = A(1/S_T)^B$. The asterisks in Fig. 36.21 show the estimated fatigue damage following this procedure for the measured signals. These values can be compared relatively well with the damage predicted using the unfiltered signals. In this method, it is assumed that the amplitude of the third-harmonic stress is equal to H_3 times the amplitude of the first-harmonic stress, and H_3 is a time-averaged value over the entire time. The use of these time-averaged higher harmonic coefficients leads to the differences between estimated (asterisks) and actual (triangles) fatigue life values shown in Fig. 36.21.

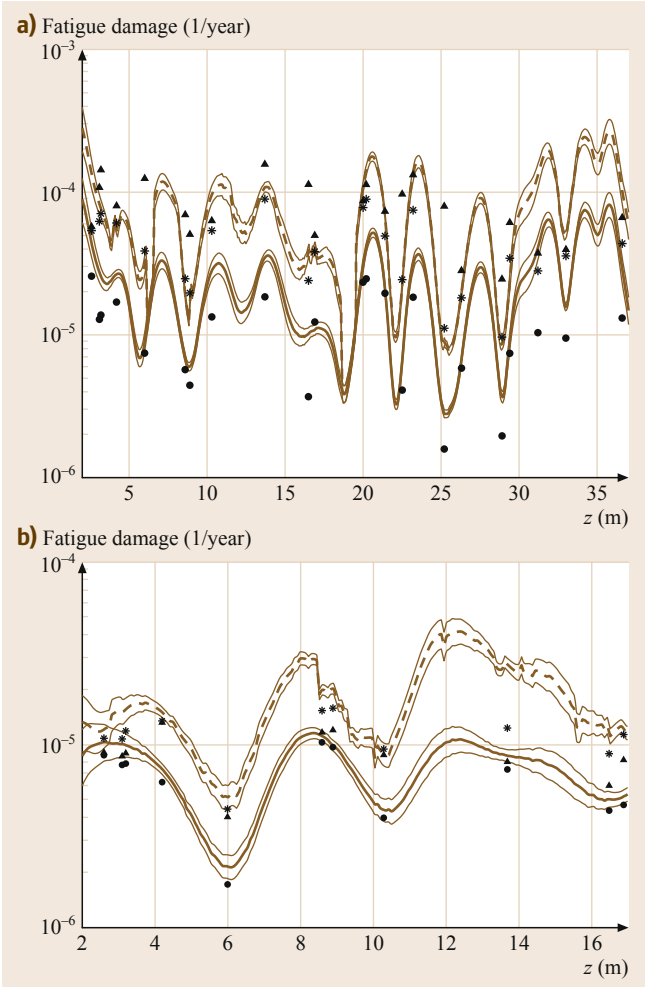


Fig. 36.21a,b Fatigue damage for two sample NDP cases of **(a)** uniform and **(b)** linearly sheared flow. Values are calculated based on purely first-harmonic reconstructed cross-flow (CF) motion (*continuous line*), and based on the first, third and fifth harmonics using higher harmonic predictions of the forced database (*dashed line*), while the *thin lines* correspond to the limits due to uncertainties arising in reconstruction procedure; direct fatigue calculations are conducted using the measured CF strain signals based on: only the first harmonic (•); the first, third, and fifth harmonics (∗); and the unfiltered signal (△)

This method can also be applied to the reconstructed response; in this case H_3 and H_5 are estimated using the rigid cylinder database. The dashed lines in Fig. 36.7 are the resulting fatigue damage calculations, while the thin lines around them show the limits due to uncertainties in the first-harmonic response reconstruction. It can be observed that the maximum fatigue damage is around 3.3×10^{-4} years^{-1} at $z\simeq 35$ m for the uniform case of Fig. 36.21a, while if one is limited to using the filtered signals and only at the measurement points, the maximum damage is calculated as 0.2×10^{-4} years^{-1} at $z\simeq 20$ m. The interesting fact about the predicted critical fatigue damage is that it occurs at a point where no measurement point is available ($z\simeq 35$ m). Similar observation can be made for the linearly sheared case of Fig. 36.21b. The critical fatigue damage is predicted as 3.3×10^{-5} years^{-1} at $z\simeq 12$ m, where the filtered measurement points would give 1×10^{-5}*years*$^{-1}$ at $z\simeq 9$ m.

Influence of Chaotic Signals on Fatigue Damage

If a chaotic signal is approximated through an *equivalent* periodic signal containing primarily the Strouhal frequency, as is often the practice in semi-empirical methods, fatigue life calculations seriously underpredict the fatigue damage.

Chaotic signals have relatively small third- and fifth-harmonic components. As an example, the third-harmonic component of the strain for case 2480 of the NDP data, a sheared flow case with $U_{\max}=2$ m/s, is 0.05 of the total strain. One could then argue that because this is a negligible higher harmonic component, one can consider only frequencies near the Strouhal frequency. If this is done, however, the resulting fatigue life will be over-predicted by 70%, as the filtered strain provides a fatigue life of 13.7×10^4 years, while the actual fatigue life is 4.1×10^4 years. Hence, in general, it is not safe to use the filtered part of the strain (around the Strouhal frequency) in order to calculate fatigue life, even when higher strain harmonics are small, because the signal might be chaotic. The difference between fatigue life calculated using a filtered VIV signal and the actual fatigue life is large when the signal is chaotic, mainly because the spectrum is broad-band.

36.3.6 VIV Prediction Tools

The most widely used methods for predicting the riser VIV are semi-empirical prediction programs, such as VIVA [36.29], SHEAR7 [36.85], and VIVANA [36.86]. Such programs typically consist of two parts:

1. A fluid–structure interaction model, and
2. Laboratory-derived or empirical databases, often adjusted to account for scale effects, as a result of comparison with full scale or pilot scale studies.

The fluid–structure interaction model consists of a fluid force model and the equations of structural dynamics.

The hydrodynamic databases primarily contain hydrodynamic information in the form of the lift force

coefficients (lift coefficient in phase with velocity and added mass coefficient) and drag coefficients. These force coefficients are often obtained from extensive laboratory experiments using rigid cylinders. It is a common practice [36.29] to use a strip theory approach to subdivide the flexible structure into small segments, which act similar to a series of interconnected, elastically mounted rigid cylinders, and estimate the vortex-induced forces on each of these segments: the excitation force in phase with velocity and the excitation force in phase with acceleration. As shown by *Triantafyllou* and *Grosenbaugh* [36.58], estimating these forces requires databases of added mass coefficient and lift coefficient in phase with velocity. From a structural dynamics viewpoint, the structure is adequately modeled as a tensioned beam with appropriate boundary conditions, acted upon by the external hydrodynamic forces induced by the forming vortical structures.

The available extensive databases used in VIVA were obtained in the MIT Towing Tank [36.87] and in high Reynolds number (supercritical regime) experiments at Memorial University in Newfoundland [36.35]. The lift coefficients and added mass coefficients are functions of reduced velocity and amplitude of response. The databases were obtained (i) at a Reynolds number around 10 000, and (ii) for cylinders undergoing harmonic motion at a single frequency and restricted to the cross-flow direction only.

To overcome these limitations, *Mukundan* et al. [36.82] developed a systematic method to extract information from field and experimental data from risers and update the existing databases. A formal procedure to optimize this process was discussed by *Mukundan* et al. [36.82] and a heuristic method, whereby a small number of parameters were allowed to vary, was used by *Chasparis* et al. [36.88].

36.4 Effectiveness of Vortex Canceling Devices

As we have seen in the previous sections, the effects of VIV on long slender structures may include significant fatigue damage to the structure due to regular oscillations as well as increased static loading on the structure due to amplification of the mean drag coefficient. These problems can lead to structural failure or unacceptable deflections of risers for offshore operations; hence in many cases, it is desirable to cancel VIV altogether. Vortex cancellation devices are designed with the purpose of addressing three main issues associated with VIV: (1) oscillatory motions, (2) increased drag coefficient, and (3) possible omni-directionality of the flow. In designing vortex cancellation devices, it is therefore desirable for devices to reduce the amplitude of vibratory motions, have a low drag coefficient, and be capable of handling changing directionality of the flow.

Vortex cancellation devices are used to inhibit VIV by either reducing the magnitude of oscillatory forces on the body, or by reducing the correlation of forces along the length of the body, or, finally, by avoiding separation on the body altogether. Vortex cancellation devices may be divided into two subgroups: passive devices and active devices. Passive devices do not require any energy input and are either stationary alterations to the shape of the body or passive structures that align to the oncoming flow, as reviewed in [36.3, 6]. Examples of passive cancellation devices include strakes and fairings.

In contrast, active suppression devices require energy input to control vortex formation. The devices may involve momentum injection into the boundary layer or active manipulation of the body surface. Since most active control devices are expensive to implement in practice, they are not typically found in use on large scale ocean structures. We will therefore focus primarily on passive cancellation devices, which have widespread use in industry.

36.4.1 Helical Strakes

A strake is a triangular, rectangular, or trapezoidal protrusion attached to the surface of a cylinder and wrapped along the length of the cylinder in a helix. Strakes are often implemented in a triple helix form, where three strakes are wrapped along the length of the cylinder. Effective VIV suppression occurs for triple strakes with pitch to diameter ratios in a range between 15 and 17; single strakes are effective with pitch ratio close to 5. It is interesting that accounting for a triple strake configuration, the effective pitch to diameter ratio is in a range between 5 and 6. The height of the strake must be sufficiently high to extend past the boundary layer on the cylinder. A height to diameter ratio of 0.15 has been observed to be effective at suppressing VIV.

Strakes do not suppress vortex shedding, but rather define a sharp separation point for which vortex shedding will occur. By implementing the strake in a helical pattern along the length of the cylinder, the strake provides this clear separation point at different angular positions along the length of the cylinder. At any

given cross-section, the phasing of vortex shedding will be slightly different than the phasing at different span locations on the cylinder. This alteration of the vortex shedding phasing leads to uncorrelated oscillatory forces along the length of the cylinder, reducing the amplitude of the oscillatory forces. The resulting response of the body may be significantly reduced with appropriate orientation of the strakes.

Although strakes may be extremely effective at reducing the vibrational response cylindrical structures, this reduction in vibration amplitude comes at a cost of increased drag. As mentioned previously, a vibrating cylinder will experience an increased mean drag coefficient due to the large wake deficit behind the oscillating structure. Since strakes significantly reduce the oscillation of the structure, the wake deficit is less than the deficit observed when the structure is oscillating. However, due to the presence of the protruding strakes on the cylinder, the wake deficit is still larger than if the cylinder were stationary in the flow with no strakes. For example, a flexible cylinder at subcritical flow speeds oscillating due to VIV with a maximum oscillation amplitude around one diameter may experience a mean drag coefficient between 1.8 and 3, depending on the shear in the flow. The same cylinder held stationary in the flow would experience a drag coefficient of 1.2, while the equivalent cylinder, outfitted with strakes with height equal to 25% of the diameter, oscillates little but incurs a drag coefficient around 1.7. This increased drag coefficient may be undesirable in certain operating conditions, hence strakes are not always suitable for VIV suppression.

Strakes are characterized by reduced excitation lift forces in phase with velocity [36.31], which can be modeled in semiempirical codes to provide predictions of straked or partially straked risers [36.89].

36.4.2 Fairings

A fairing is a streamlined, usually foil-shaped fitting that surrounds the cylinder. It works by alleviating the adverse pressure gradient on the downstream side of the body, reducing the possibility of separation and eliminating the formation of large scale correlated vortex structures in the wake of the body; substantial reduction in drag coefficient is also obtained.

Since a fairing is effective only if aligned with the flow velocity, it must be free to rotate, otherwise it can cause severe problems such as kiting or galloping. For example, a fairing on a towed cable may become stuck if the cable is curved and hence become ineffective. When the fairing is not aligned with the flow, steady lift develops, causing the cable to deviate in the cross-flow direction, in what is called *kiting*. At the same time, the drag force can increase substantially on the fairing, while a torque moment develops that can twist the cable.

The unsteady lift and torque caused by a misaligned fairing can induce galloping under certain conditions. The asymmetry in the streamlines around the misaligned fairing can cause increased force and moment as the cable vibrates; the positive feedback can result in large amplitude vibrations above a critical speed that depends on the structural damping, or galloping, that can lead to catastrophic failure of the structure. Galloping, which is not self-limiting, occurs at much lower frequencies than in self-limiting VIV.

36.4.3 Other VIV Suppression Devices

In addition to strakes and fairings, *Zdravkovich* [36.3] compiles data on several passive VIV suppression devices. One group of devices includes helical grooves made along the length of the body to achieve a similar result to applying extruding strakes. Helically wound cables are also presented as a variation on the strake as the cable itself acts as a helical protrusion on the body.

Zdravkovich [36.3] presents several other suppression devices including devices where small cylinders are attached to a the larger main cylinder. Appropriately placing a small control cylinder near the main cylinder has been shown to alter the fluid momentum in the wake of the large cylinder, leading to unseparated flow under the right placement conditions of the control cylinder [36.90]. Similar to the placement of control cylinders, *Grimminger* [36.91] tested various arrangements of vanes on a cylinder to redirect flow in the wake of the cylinder. These vanes were found to eliminate VIV and reduce drag, however they still required a directional orientation to the flow direction. More recently, *Galvao* et al. [36.92] tested various combinations of attached foil sections near a cylinder for redirecting flow in the wake of the cylinder.

Zdravkovich [36.3] also introduces shrouds as possible VIV suppression devices. A shroud is an outer covering that may include a pattern of holes to allow fluid to pass through the shroud and around the cylinder. Various percentages of hole distribution lead to varying effectiveness of the shroud in suppressing VIV. Additional variations on VIV suppression devices include splitter plates, ribbon fairings, haired fairings, and open channels that allow fluid to pass through the cylinder. These devices work with a variety of success dependent on the orientation and flow conditions [36.93].

Other active and passive means to decorrelate vortex shedding by introducing periodic variations along the span of a cylinder, such as in [36.94–99], have

proved effective in reducing the mean drag force by about 20% and the unsteady lift force significantly, for certain parametric combinations. *Lee* and *Nguyen* [36.100] note that when the wavelength to mean diameter ratio, λ/D_m is close to 2, their effectiveness is maximized. *Lam* et al. [36.96] noted that there is a second optimal value of $\lambda/D_m = 6$, which provides even better performance. The effectiveness of the spanwise undulations is attributed to the formation of an undulating wake, which decorrelates the induced forces, as well as the formation of streamwise vortices. The latter value, $\lambda/D_m = 6$, agrees with the reported optimal wavelength to diameter ratio for strakes, which is around 5 for single strakes, and between 15 and 17 for triple start strakes, that is, effectively, between 5 and 5.6.

36.5 Multiple Interfering Bluff Bodies

Cylinders with parallel axes placed in a cross-flow are subject to wake interference phenomena. This problem arises in many applications, such as arrays of offshore risers, multiple mooring lines, and power transmission lines. For nonoscillating cylinders, distinct flow regimes exist, which depend on the axial and transverse separation distance to diameter ratios [36.101–103]. For small axial separation to diameter ratios, s, below typically 2.5 [36.104], the reattachment of the leading cylinder's shear layer to the trailing cylinder can lead to bistable gap flows. Quasi-steady recirculation cells, with coupled vortex formation, occur for larger separation ratios, up to 4, while vortex roll-up from the leading cylinder occurs for distance ratios higher than 4, where such coupling is diminished.

For forced vibrations of two tandem cylinders, a wake lock-in exists for the extreme motion phase angles of zero and 180°, and lock-in of the wake to the motion occurs over a dramatically expanded region of amplitude and frequency, for small values of s [36.105]. In tests where both cylinders are compliantly mounted, large-amplitude vibrations of both cylinders can occur when the separation ratio is about 5–7, and for lateral offset ratio up to 1.5 [36.102]. The vibrations are limited to a specific range of reduced velocities, typically beginning at lower values than for a single cylinder. *Bokaian* and *Geoola* [36.106] considered the case of a fixed leading cylinder, and a compliantly mounted trailing cylinder and find initially a vortex-induced region, occurring over a specific range of reduced velocities, and then galloping instabilities, where motions persist for high reduced velocities. These two types of responses can occur independently or can overlap, depending on the separation distance.

The wake frequency behind the leading cylinder is typically unaffected by the proximity to the trailing cylinder, but the wake of the trailing cylinder exhibits a lower Strouhal frequency, consistent with reduced mean flow due to shielding effects. Hence, frequency lock-in occurs at a higher value than for a single cylinder [36.107]. *Hover* and *Triantafyllou* [36.108] considered compliantly mounted cylinders in Reynolds number 3×10^4, finding vibratory response of the trailing cylinder up to separation ratios of 17.

Assi et al. [36.109] showed that the galloping-like response is caused by a physical mechanism that they call *wake stiffness*, a spring-like restoring force caused by the impinging vortical structures from the upstream to the downstream cylinder. The mechanism explains well the observed behavior of the downstream cylinder.

36.6 Effect of Reynolds Number

The Reynolds number, Re, has a significant effect on the Strouhal number and drag coefficient of a fixed smooth cylinder in a cross-flow only over two regions: First, at low Re, below typically 1000, when the first instability occurs at $Re = 56$ leading to the formation of the Karman street, and then the secondary instabilities appear starting at around $Re \approx 250$ leading to turbulence in the wake. Second, in the *critical* Reynolds number regime, which starts at around $Re \approx 250\,000$, where the transition of the boundary layer at the flow separation points transitions from laminar to turbulent flow, affecting substantially the location of the separation points and the vortex formation process in the wake. The drag coefficient undergoes a dramatic drop from around $c_d = 1.2$ to $c_d = 0.25$, in what is termed the *drag crisis*. The drag recovers from these low values, as Re increases, reaching a value of around $c_d = 0.7$ for $Re \approx 10^6$, where the critical regime ends. Outside these ranges, the effect of Re is gradual and at most moderate.

Similarly, when a cylinder is allowed to oscillate transverse to the flow direction or with combined cross-flow and in-line motion, the Reynolds number plays an important role in the development of the boundary layer and separation points on the dynamically moving cylinder. Recent studies have shown that at subcritical Reynolds numbers, Re affects the fluid forces exerted

on the body by altering wake state transitions. Small variations in the shedding point on the surface of the cylinder can significantly alter the phasing of vortex shedding in the wake with respect to the cylinder motions [36.110].

Since the phasing of vortex shedding may be altered slightly for subcritical Reynolds numbers, this alteration leads to a change in the forces exerted on a cylinder, leading to changes in the resulting cylinder motions. For subcritical Reynolds numbers, it has been observed that unsteady lift force magnitudes increase with increasing Reynolds number, resulting in increased cylinder motions [36.111]. *Govardhan* and *Williamson* [36.111] estimate limits of the response amplitude for an oscillating cylinder at various subcritical Reynolds numbers based on the mass-damping characteristics of the system. It has also been shown that for cross-flow free vibrations, the maximum amplitude of oscillation is purely a function of system damping and the Reynolds number [36.112].

Few experimental data exist for VIV at critical and supercritical Reynolds numbers. For a freely oscillating cylinder in cross-flow, it has been observed that a cross-flow response in excess of two diameters may be achieved at the beginning of the critical Reynolds number regime [36.34]. These large amplitude motions at the onset of the critical Reynolds number regime have been observed even in cases with large mass-damping [36.113]. The use of roughness on the cylinder surface to influence the boundary layer and hence cause effects similar to those observed in the critical regime, can spark large amplitude motions at different Reynolds numbers as well [36.114]. Well within the critical Reynolds number regime, a smooth cylinder exhibits very small VIV; however, introducing even small amounts of roughness or external stream turbulence alters this condition, leading to substantial VIV. Hence, the exact VIV response shape in the critical regime is highly dependent on the surface roughness of the body [36.34].

At supercritical Reynolds numbers, cross-flow VIV is always present with an amplitude of response smaller than in the subcritical region, not exceeding 0.9 diameters. When combined in-line and cross-flow oscillations of a circular cylinder occur, similar trends are observed, that is, the amplitude of body oscillations is observed to be smaller than for a comparable system at subcritical Reynolds numbers [36.35]. Although motions are slightly smaller due to Reynolds number effects, it must be noted that the resonant phenomenon of lock-in still occurs, such that the same dynamic relations between vortex shedding and body motion occur at supercritical Reynolds numbers as those observed at subcritical Reynolds numbers.

References

36.1 T. Sarpkaya: Vortex-induced oscillations, J. Appl. Mech. **46**, 241–257 (1979)

36.2 R.D. Blevins: *Flow-induced vibrations* (Nostrand Reinhold Co., New York, 1977)

36.3 M.M. Zdravkovich: Review and classification of various aerodynamic and hydrodynamic means for suppressing vortex shedding, J. Wind Eng. Ind. Aerodyn. **7**, 145–189 (1980)

36.4 P.W. Bearman: Vortex shedding from oscillating bluff bodies, Annu. Rev. Fluid Mech. **16**, 195–222 (1984)

36.5 C.H.K. Williamson: Vortex shedding in the cylinder wake, Annu. Rev. Fluid Mech. **28**, 477–539 (1996)

36.6 M.M. Zdravkovich: *Flow Around Circular Cylinders* (Oxford Univ. Press, Oxford 1997)

36.7 C.H.K. Williamson, R. Govardhan: Vortex-induced vibrations, Annu. Rev. Fluid Mech. **36**, 413–455 (2004)

36.8 T. Sarpkaya: A critical review of the intrinsic nature of vortex-induced vibrations, J. Fluids Struct. **19**, 389–447 (2004)

36.9 T.L. Morse, C.H.K. Williamson: Prediction of vortex-induced vibration response by employing controlled motion, J. Fluid Mech. **634**, 5–39 (2009)

36.10 P.W. Bearman: Circular cylinder wakes and vortex-induced vibrations, J. Fluids Struct. **27**, 648–658 (2011)

36.11 C.M. Alexander: The complex vibrations and implied drag of a long oceanographic wire in cross-flow, Ocean Eng. **8**, 379–406 (1981)

36.12 D.A. Chapman: Towed cable behaviour during ship turning maneuvers, Ocean Eng. **11**, 327–361 (1984)

36.13 P.L. Bourget, D. Marichal: Remarks about variations in the drag coefficient of circular cylinders moving through water, Ocean Eng. **17**, 569–585 (1990)

36.14 D.R. Yoerger, M.A. Grosenbaugh, M.S. Triantafyllou, J.J. Burgess: Drag forces and flow-induced vibrations of a long vertical tow cable – Part I: Steady-state towing conditions, J. Offshore Mech. Arct. Eng. **113**, 117–127 (1991)

36.15 M.A. Grosenbaugh, D.R. Yoerger, M.S. Triantafyllou, F.S. Hover: Drag forces and flow-induced vibrations of a long vertical tow cable – Part II: Unsteady towing conditions, J. Offshore Mech. Arct. Eng. **113**(3), 199–204 (1991)

36.16 S. Welch, M.P. Tulin: An experimental investigation of the mean and dynamic tensions in towed

strumming cables, J. Offshore Polar Eng. **3**, 213–218 (1993)
36.17 E. Huse, K. Matsumoto: Mooring line damping due to first and second order vessel motion, Offshore Technol. Conf. (1989), OTC 5676
36.18 W. Koterayama: Viscous damping forces for slow drift oscillation of the floating structure acting on the hull and the mooring lines, Proc. 18th Int. Conf. Offshore Mech. Arct. Eng. (1989)
36.19 M.S. Triantafyllou, D.K.P. Yue, D.Y.S. Tein: Damping of moored floating structures, Offshore Technol. Conf. (1994), OTC 7489
36.20 M.S. Triantafyllou: Cable dynamics for offshore applications. In: *Developments in Offshore Engineering: Wave Phenomena and Offshore Topics*, ed. by J.B. Herbich (Gulf Publ., Houston 1999)
36.21 I.H. Brooks: A pragmatic approach to vortex-induced vibrations of a drilling riser, Offshore Technol. Conf. (1987), OTC 5522
36.22 D.W. Allen: Vortex-induced vibration analysis of the Auger TLP production and steel catenary export risers, Offshore Technol. Conf. (1999), OTC 7821
36.23 D.W. Allen, D.L. Henning: Prototype VIV tests for production risers, Offshore Technol. Conf. (2001), OTC 13114
36.24 T. Sarpkaya: *Vortex Shedding and Resistance in Harmonic Flow About Smooth and Rough Circular Cylinders at High Reynolds Numbers*, *Tech. Rep. NPS-59SL76021* (NSF, Washington 1976)
36.25 C.H.K. Williamson, A. Roshko: Vortex formation in the wake of an oscillating cylinder, J. Fluids Struct. **2**, 355–381 (2004)
36.26 A. Ongoren, D. Rockwell: Flow structure from an oscillating cylinder part 1. Mechanisms of phase shift and recovery in the near wake, J. Fluid Mech. **191**, 197–223 (1988)
36.27 C. Evangelinos, G.E. Karniadakis: Dynamics and flow structures in the turbulent wake of rigid and flexible cylinders subject to vortex-induced vibrations, J. Fluid Mech. **400**, 91–124 (1999)
36.28 R. Govardhan, C.H.K. Williamson: Modes of vortex formation and frequency response of a freely vibrating cylinder, J. Fluid Mech. **420**, 85–130 (2000)
36.29 M.S. Triantafyllou, G.S. Triantafyllou, D.Y.S. Tein, B.D. Ambrose: Pragmatic riser VIV analysis, Offshore Technol. Conf. (1999), OTC 10931
36.30 J.K. Vandiver, S. Swithenbank, V. Jaiswal, V. Jhingran: Fatigue damage from high mode number vortex-induced vibration, 25th Int. Conf. Offshore Mech. Arct. Eng. (2006)
36.31 C.M. Larsen, R. Yttervik, K. Vikestad, E. Passano: Lift and damping characteristics of bare and straked cylinders at riser scale Reynolds numbers, 20th Int. Conf. Offshore Mech. Arct. Eng. (2001)
36.32 R. Bourguet, G.E. Karniadakis, M.S. Triantafyllou: Phasing mechanisms between the in-line and cross-flow vortex-induced vibrations of a long tensioned beam in shear flow, Comput. Struct. **122**(0), 155–163 (2013)
36.33 W.C.L. Shih, C. Wang, D. Coles, A. Roshko: Experiments on flow past rough circular cylinders at large Reynolds Numbers, J. Wind Eng. Ind. Aero. **49**, 351–368 (1993)
36.34 Z.J. Ding, S. Balasubramanian, R.T. Lokken, T.-W. Yung: Circular cylinder wakes and vortex-induced vibrations, Offshore Technol. Conf. (2004), OTC 16341
36.35 J.M. Dahl, F.S. Hover, M.S. Triantafyllou, O.H. Oakley: Dual resonance in vortex-induced vibrations at subcritical and supercritical Reynolds numbers, J. Fluid Mech. **643**, 395–424 (2010)
36.36 W. Koch: Local instability characteristics and frequency determination of self-excited wake flows, J. Sound Vib. **99**, 53 (1985)
36.37 G.S. Triantafyllou, M.S. Triantafyllou, C. Chryssostomidis: On the formation of vortex streets behind stationary cylinders, J. Fluid Mech. **170**, 461–477 (1986)
36.38 P. Huerre, P.A. Monkewitz: Local and global instabilities in spatially developing flows, Annu. Rev. Fluid Mech. **22**, 473–537 (1990)
36.39 A.E. Perry, M.S. Chong, T.T. Lim: The vortex-shedding process behind two-dimensional bluff bodies, J. Fluid Mech. **116**, 77–90 (1982)
36.40 F.S. Hover, J.T. Davis, M.S. Triantafyllou: Is mode transition three-dimensional?, Conf. Bluff Body Wakes Vortex-Induced Vib. (2006)
36.41 F.S. Hover, A.H. Techet, M.S. Triantafyllou: Forces on oscillating uniform and tapered cylinders in crossflow, J. Fluid Mech. **363**, 97–114 (1998)
36.42 J.N. Newman: *Marine Hydrodynamics* (MIT Press, Cambridge 1977)
36.43 T.L. Morse, C.H.K. Williamson: Fluid forcing, wake modes, and transitions for a cylinder undergoing controlled oscillations, J. Fluids Struct. **25**, 697–712 (2009)
36.44 J. Carberry, J. Sheridan, D. Rockwell: Controlled oscillations of a cylinder: forces and wake modes, J. Fluid Mech. **538**, 31–69 (2005)
36.45 M.S. Triantafyllou, R. Gopalkrishnan, M.A. Grosenbaugh: Vortex-induced vibrations in a sheared flow: A new predictive method. In: *Hydroelasticity in Marine Technology*, ed. by O. Faltinsen, C.M. Larsen, T. Moan, A.A. Balkema (Taylor Francis, Rotterdam 1994)
36.46 N. Jauvtis, C.H.K. Williamson: The effect of two degrees of freedom on vortex-induced vibration at low mass and damping, J. Fluid Mech. **509**, 23–62 (2004)
36.47 J.M. Dahl, F.S. Hover, M.S. Triantafyllou, S. Dong, G.E. Karniadakis: Resonant vibrations of bluff bodies cause multivortex shedding and high frequency forces, Phys. Rev. Lett. **99**, 144503 (2007)
36.48 A.H. Techet, F.S. Hover, M.S. Triantafyllou: Vortical patterns behind tapered cylinders oscillating transversely to a uniform flow, J. Fluid Mech. **363**, 79–96 (1988)
36.49 D. Lucor, L. Imas, G.E. Karniadakis: Vortex dislocations and force distribution of long flexible cylinders subjected to sheared flows, J. Fluids Struct. **15**, 641–650 (2001)
36.50 A. Fage, J.H. Warsap: *The Effects of Turbulence and Surface Roughness on the Drag of a Circular Cylin-*

der, Reports and Memoranda of the GBARC, No. 1283 (HMSO, London 1929)
36.51 D.F. James, Q.S. Truong: Wind load on cylinder with spanwise protrusion, Proc. ASCE J. Eng. Mech. Div. **98**, 1573–1589 (1972)
36.52 T. Igarashi: Effect of tripping wires on the flow around a circular cylinder normal to an airstream, Bull. Jpn. Soc. Mech. Eng. **29**, 2917–2924 (1986)
36.53 M.S. Triantafyllou, M.A. Grosenbaugh: Prediction of flow-induced vibrations in sheared flows, Proc. 6th Int. Conf. Flow-Induc. Vib. (1994)
36.54 M.S. Triantafyllon, F.S. Hover, A.H. Techet, D.K.P. Yue: Vortex-induced vibrations of slender structures in shear flow, Proc. IUTAM Symp. Integr. Modeling Fully Coupled Fluid-Struct. Interact. Using Anal. Comput. Exp. (2003) pp. 313–327
36.55 T. Sarpkaya: Hydrodynamic damping, flow induced oscillations and biharmonic response, J. Offshore Mech. Arct. Eng. **177**, 232–238 (1995)
36.56 J.M. Dahl, F.S. Hover, M.S. Triantafyllou: Forces and wake modes of an oscillating cylinder, J. Fluids Struct. **22**, 807–818 (2001)
36.57 M.M. Bernitsas, K. Raghavan, G. Duchene: On hydrodynamic coefficients for combined cross-flow and in-line vortex induced vibrations, Proc. ASME 27th Int. Conf. Offshore Mechanics Arctic Eng. (OMAE) (2008)
36.58 G.S. Triantafyllou, M.A. Grosenbaugh: Vortex induced vibrations of flexible structures, Summer Meet. Am. Soc. Mech. Eng. (1998)
36.59 P.K. Stansby: The locking-on of vortex shedding due to the cross-stream vibration of circular cylinders in uniform and shear flows, J. Fluid Mech. **74**, 641–665 (1976)
36.60 J.A. Humphreys, D.H. Walker: Vortex excited response of large scale cylinders in sheared flow, J. Offshore Mech. Arct. Eng. **110**(3), 272–277 (1988)
36.61 R. Bourguet, G.E. Karniadakis, M.S. Triantafyllou: Multi-frequency vortex-induced vibrations of a long tensioned beam in linear and exponential shear flows, J. Fluids Struct. **41**(0), 33–42 (2013)
36.62 R. Gopalkrishnan, M.A. Grosenbaugh, M.S. Triantafyllou: Influence of amplitude modulation on the fluid forces acting on a vibrating cylinder in cross-flow, Proc. 3rd Int. Symp. Flow Induced Vib. Noise (ASME) (1992)
36.63 J.R. Chaplin, P.W. Bearman, Y. Cheng, E. Fontaine, J.M.R. Graham, K. Herfjord, F.J. Huera Huarte, M. Isherwood, K. Lambrakos, C.M. Larsen, J.R. Meneghini, G. Moe, R.J. Pattenden, M.S. Triantafyllou, R.H.J. Willde: Blind predictions of laboratory measurements of vortex-induced vibrations of a tension riser, J. Fluids Struct. **21**, 25–40 (2005)
36.64 H. Lie, K.E. Kaasen: Modal analysis of measurements from a large-scale {VIV} model test of a riser in linearly sheared flow, J. Fluids Struct. **22**(4), 557–575 (2006)
36.65 D. Lucor, H. Mukundan, M.S. Triantafyllou: Riser modal identification in CFD and full-scale experiments, J. Fluids Struct. **22**, 905–917 (2006)
36.66 J. Kim Vandiver, V. Jaiswal, V. Jhingran: Insights on vortex-induced, traveling waves on long risers, J. Fluids Struct. **25**(4), 641–653 (2009)
36.67 R. Bourquet, G. Karniadakis, M.S. Triantafyllou: Vortex-induced vibrations of a long flexible cylinder in shear flow, J. Fluid Mech. **677**, 342–382 (2011)
36.68 R. Bourguet, Y. Modarres-Sadeghi, G.E. Karniadakis, M.S. Triantafyllou: Wake-body resonance of long flexible structures is dominated by counterclockwise orbits, Phys. Rev. Lett. **107**, 134502 (2011)
36.69 R. Bourguet, G.E. Karniadakis, M.S. Triantafyllou: Distributed lock-in drives broadband vortex-induced vibrations of a long flexible cylinder in shear flow, J. Fluid Mech. **717**, 361–375 (2013)
36.70 M. Gaster: Vortex shedding from circular cylinders at low Reynolds numbers, J. Fluid Mech. **46**, 749–756 (1971)
36.71 P.S. Piccirillo, C.W. Van Atta: An experimental study of vortex shedding behind linearly tapered cylinders at low Reynolds number, J. Fluid Mech. **246**, 163–195 (1993)
36.72 A. Mukhopadhyay, P. Venugopal, S.P. Vanka: Numerical study of vortex shedding from a circular cylinder in linear shear flow, J. Fluids Eng. **121**, 460–468 (1992)
36.73 C.H.K. Williamson: The natural and forced formation of spot-like in the transition of a wake, J. Fluid Mech. **243**, 393–441 (1992)
36.74 H.Q. Zhang, U. Fey, B.R. Noack, M. König, H. Eckelmann: On the transition of the cylinder wake, Phys. Fluids **7**(4), 779–794 (1995)
36.75 F.J. Huera-Huarte, P.W. Bearman: Wake structures and vortex-induced vibrations of a long flexible cylinder – Part 1: Dynamic response, J. Fluids Struct. **25**, 969–990 (2009)
36.76 MIT: VIV data repository, http://oe.mit.edu/VIV/ (2007)
36.77 H. Braaten, H. Lie: *NDP Rise High Mode VIV Tests Main Report*, Report No. 512394.00.01 (Norwegian Marine Technology Research Institute, Marintek 2004)
36.78 J.K. Vandiver, S. Swithenbank, V. Jaiswal, V. Jhingran: *Final Report, DeepStar 8402: Gulf Stream Experiment* (Deepstar, Miami 2006)
36.79 V. Jhingran, J.K. Vandiver: Incorporating the higher harmonics in VIV fatigue predictions, Proc. 26th Int. Conf. Offshore Mech. Arct. Eng. (2007) pp. 891–899
36.80 Y. Modarres-Sadeghi, H. Mukundan, J.M. Dahl, F.S. Hover, M.S. Triantafyllou: The effect of higher harmonic forces on fatigue life of marine risers, J. Sound Vib. **329**, 43–55 (2010)
36.81 Y. Modarres-Sadeghi, F. Chasparis, M.S. Triantafyllou, M. Tognarelli, P. Beynet: Chaotic response is a generic feature of vortex-induced vibrations of flexible risers, J. Sound Vib. **330**, 2562–2579 (2011)
36.82 H. Mukundan, Y. Modarres-Sadeghi, F.S. Hover, M.S. Triantafyllou: Monitoring fatigue damage on marine risers, J. Fluids Struct. **25**, 617–628 (2008)

36.83 A.D. Trim, H. Braaten, H. Lie, M.A. Tognarelli: Experimental investigation of vortex-induced vibration of long marine risers, J. Fluids Struct. **21**(3), 335–361 (2005)

36.84 L. Christian: *Mechanical Vibration and Shock* (Taylor Francis, New York 2002)

36.85 J.K. Vandiver, H.M. Marcollo: High mode number VIV experiments, Proc. IUTAM Symp. Integr. Modeling Fully Coupled Fluid-Struct. Interact. Using Anal. Comput. Exp. (2003)

36.86 C.M. Larsen, K. Vikestad, R. Yttervik, E. Passano, G.S. Baarholm: *VIVANA Theory Manual* (MARINTEK, Trondheim 2005)

36.87 R. Gopalkrishnan: *Vortex Induced Forces on Oscillating Bluff Cylinders* (MIT, Massachusetts 1993)

36.88 F. Chasparis, Y. Modarres-Sadeghi, F. Hover, M.S. Triantafyllou, Y. Contantinides, H. Mukundan: Hydrodynamic data extraction from field data, Proc. 28th Int. Conf. Offshore Mech. Arct. Eng. (2009) pp. 891–899

36.89 H. Zheng, R. Price, Y. Modarres-Sadeghi, G. Triantafyllou, M. Triantafyllou: Vortex-induced vibration analysis (VIVA) based on hydrodynamic databases, Proc. 30th Int. Conf. Offshore Mech. Arct. Eng. (2011)

36.90 P.J. Strykowski, K.R. Sreenivasan: On the formations and suppression of vortex *shedding* at low Reynolds numbers, J. Fluid Mech. **218**, 71–107 (1990)

36.91 G. Grimminger: *The effect of rigid guide vanes on the vibration and drag of a towed circular cylinder*, Tech. Rep. 504 (David Taylor Model Basin, Washington 1945)

36.92 R. Galvao, E. Lee, D. Farrel, F. Hover, M.S. Triantafyllou, N. Kitnev, P. Beynet: Flow control in flow-structure interaction, J. Fluids Struct. **24**, 1216–1226 (2008)

36.93 V. Jacobsen: Vibration suppression devices for long, slender tubulars, Offshore Technol. Conf. (1996), OTC 8156

36.94 A. Ahmed, B. Byram: Transverse flow over a wavy cylinder, Phys. Fluids A: Fluid Dyn. **4**(9), 1959–1967 (1992)

36.95 K. Lam, F.H. Wang, R.M.C. So: Three-dimensional nature of vortices in the near wake of a wavy cylinder, J. Fluids Struct. **19**, 815–833 (2004)

36.96 K. Lam, Y.F. Lin, L. Zhou, Y. Liu: The effect of wavy surface on vortex shedding from an inclined cylinder in turbulent flow, Int. J. Offshore Polar Eng. **3**(1), 1343–1350 (2009)

36.97 E.J. Lee: Airfoil Vortex Induced Vibration Suppression Devices, Ph.D. Thesis (Massachusetts Institute of Technology, Massachusetts 2007)

36.98 J.C. Owen, A.A. Szewczyk, P.W. Bearman: Suppression of Karman vortex shedding, Phys. Fluids **12**(9), 1–13 (2000)

36.99 M.M. Zhang, L. Cheng, Y. Zhou: Closed-loop manipulated wake of a stationary square cylinder, Exp. Fluids **39**, 75–85 (2005)

36.100 S.J. Lee, A.T. Nguyen: Experimental investigation on wake behind a wavy cylinder having sinusoidal cross-sectional area variation, Fluid Dyn. Res. **39**, 292–304 (2007)

36.101 M.M. Zdravkovich, D.L. Pridden: Interference between two circular cylinders; series of unexpected discontinuities, J. Ind. Aerodyn. **2**, 255–270 (1977)

36.102 M.M. Zdravkovich: Flow induced oscillations of two interfering circular cylinders, J. Sound Vib. **101**, 511–521 (1985)

36.103 T. Igarashi: Characteristics of the flow around two circular cylinders arranged in tandem (First report), Bull. Jpn. Soc. Mech. Eng. **24**, 323–331 (1981)

36.104 G.V. Papaioannou, D.K.P. Yue, M.S. Triantafyllou, G.E. Karniadakis: Evidence of holes in the Arnold Tongues of flow past two oscillating cylinders, Phys. Rev. Lett. **96**, 014501 (2006)

36.105 N. Mahir, D. Rockwell: Vortex formation from a forced system of two cylinders. Part 1: tandem arrangement, J. Fluids Struct. **10**, 473–489 (2008)

36.106 A. Bokaian, F. Geoola: Wake-induced galloping of two interfering circular cylinders, J. Fluid Mech. **146**, 383–415 (1984)

36.107 D. Brika, A. Laneville: The flow interaction between a stationary cylinder and a downstream flexible cylinder, J. Fluids Struct. **13**, 579–606 (1999)

36.108 F.S. Hover, M.S. Triantafyllou: Galloping response of a cylinder with upstream wake interference, J. Fluids Struct. **15**, 503–512 (2001)

36.109 G.R.S. Assi, P.W. Bearman, B.S. Carmo, J.R. Meneghini, S.J. Sherwin, R.H.J. Willden: The role of wake stiffness on the wake-induced vibration of the downstream cylinder of a tandem pair, J. Fluid Mech. **718**, 210–245 (2013)

36.110 J. Carberry, J. Sheridan, D.O. Rockwell: Forces and wake modes of an oscillating cylinder, J. Fluids Struct. **15**, 523–532 (2001)

36.111 R. Govardhan, C.H.K. Williamson: Defining the modified Griffin plot in vortex-induced vibration: revealing the effect of Reynolds number using controlled damping, J. Fluid Mech. **561**, 147–180 (2006)

36.112 J.T. Klamo, A. Leonard, A. Roshko: The effects of damping on the amplitude and frequency response of a freely vibrating cylinder in cross-flow, J. Fluids Struct. **22**, 845–856 (2006)

36.113 K. Raghavan, M.M. Bernitsas: Experimental investigation of Reynolds number effect on vortex induced vibration of rigid circular cylinder on elastic supports, Ocean Eng. **38**, 719–731 (2011)

36.114 M.M. Bernitsas, K. Raghavan, G. Duchene: Induced separation and vorticity using roughness in VIV of circular cylinders at 8×10^3 to Re 1.5×10^5, Proc. Int. Conf. Offshore Mech. Arctic Eng. (2008) pp. 15–20

37. Structural Dynamics

H. Ronald Riggs, Solomon Yim

This chapter presents the basics of structural dynamics theory that is especially relevant to dynamics of flexible ocean structures. The primary focus herein is on linear structural dynamics, but the extension to nonlinear systems is also discussed. Section 37.1 introduces the single-degree-of-freedom, spring-mass-damper system. The equations of motion in the time and frequency domains are given, viscous and hysteretic damping models are discussed, and general solutions are provided. Topics of particular interest to ocean structures are discussed, followed by a discussion of numerical schemes for time-domain response. The generalization of these concepts to multi-degree-of-freedom systems is covered in Sect. 37.2. Hydroelasticity involves coupling the multi-degree-of-freedom, flexible structural model with the fluid model, such as used in Chap. 35, to obtain the coupled fluid–structure response. This is discussed in Sect. 37.3, in which the concepts of added mass and hydrodynamic damping are extended to the flexible structural model. Section 37.3.6 discusses the linear response to regular waves, and Sect. 37.4 extends this to random seas. Nonlinear hydroelasticity, viz. fluid–structure interaction, and corresponding solution strategies are discussed briefly in Sect. 37.5.

37.1 **Single Degree-of-Freedom System** 852
37.1.1 Equation of Motion 852
37.1.2 Response in the Time Domain – Free Vibration 854
37.1.3 Response in the Time Domain – Harmonic Loading 854
37.1.4 Response in the Frequency Domain 856
37.1.5 Structural Damping 856
37.1.6 Time Domain Response from Frequency Domain Response 858
37.1.7 Fluid Contribution 858
37.1.8 Nonlinear Systems 860
37.2 **Multi-Degree of Freedom Systems** 861
37.2.1 Equations of Motion 861
37.2.2 Modal Superposition 862
37.2.3 Numerical Time-Domain Solution 863
37.3 **Linear Hydroelasticity for Inviscid Fluid Flow** 864
37.3.1 Finite Element Structural Modeling 864
37.3.2 Hydrostatic Stiffness 864
37.3.3 Added Mass, Damping, and Exciting Forces 865
37.3.4 Reduced Basis Solution 865
37.3.5 Mapping from Structural Mesh to Fluid Mesh 866
37.3.6 Linear Response to Regular Waves – RAOs 866
37.3.7 Time-Domain Response 867
37.4 **Linear Response to Random Seas** 867
37.4.1 Specification of Long-Crested Random Seas 867
37.4.2 Specification of Short-Crested, Directional Random Seas 868
37.4.3 Response in Random Seas 869
37.4.4 Estimation of Extreme and Fatigue Response in Random Seas 869
37.5 **Nonlinear Hydroelasticity (Nonlinear FSI)** 870
37.5.1 Nonlinear Frequency Domain Analysis 871
37.5.2 Nonlinear Time-Domain Analysis 872
References 873

Although the basics of structural dynamics are covered, it is a rich topic and many books delve into it far more deeply than can be done here. A brief survey of some references is provided, as they may be useful for more detailed study (as well as providing a historical sense of the development of structural dynamics). One of the first books on the general topic of structural dynamics was [37.1]. It covered basic

linear elastic problems including linear vibrations of single-degree-of-freedom (SDOF) and multi-degrees-of-freedom (MDOF) discrete systems, as well as continuous systems. A much later edition added coverage on vibrations of rings, membranes, and plates [37.2]. In addition to the foundations of the structural dynamics of discrete and continuous systems, *Thomson* [37.3] covered nonlinear and random vibrations. An introduction to the application of the finite element method (FEM) to structural dynamics was added in later editions [37.4, 5]. *Hurty* and *Rubinstein* [37.6] covered the energy differential equation and integral equation methods. They also covered in detail nonproportional damping and response to random excitations. An even more detailed treatment of the dynamics of SDOF and MDOF discrete systems and continuous systems can be found in [37.7].

Structural response to earthquakes (seismic response) is an important application area for structural engineering. Starting in the 1970s, structural dynamics books began to cover this application. Specifically, *Clough* and *Penzien* [37.8, 9] provide an excellent introduction to structural dynamics, including random vibrations and seismic response. The work of *Chopra* [37.10], which first appeared in 1995, is a more specialized introduction to structural dynamics as it relates to seismic response of structures. *Tedesco* et al. [37.11] cover earthquake and blast loads and also some content related to the dynamics of ocean structures.

For a more offshore structure focus, *Brebbia* and *Walker* [37.12] present the fundamentals of random seas, forces on slender members, and diffraction problems for SDOF and MDOF discrete structural systems. *Hooft* [37.13] describes fluid motion and properties and hydrodynamic excitation forces, and covers motions of linear and nonlinear SDOF and MDOF systems. *Wilson* [37.14] provides detailed descriptions of deterministic and stochastic ocean waves, applications to SDOF and MDOF structural systems including pile foundation supports, fracture mechanics, and fatigue behavior of offshore structures. Books that focus on hydrodynamic loading include [37.15, 16]. *Chakrabarti* [37.17–20] contributed significantly to the topic of dynamics of offshore structures via a series of books. [37.17] covers hydrodynamics of offshore structures focusing on fluid loads and effects. [37.19] provides an in-depth description of physical modeling of offshore structures in large-scale wave laboratory experiments. He followed up with another book [37.20] that describes the theory and practice of hydrodynamics and vibration, which provides an excellent synopsis of the two topics. Nonlinear behavior of ocean structures is a challenging topic. Because ocean waves are inherently stochastic, random vibration behavior has to be included. [37.18] describes the short and long-term stochastic characteristics of wind, wave and current, and the corresponding motion responses of nonlinear offshore structures. For a detailed summary on the behavior of stochastically excited nonlinear ocean structures, see [37.21].

37.1 Single Degree-of-Freedom System

37.1.1 Equation of Motion

The spring-mass-damper system in Fig. 37.1a is a schematic of the classic single-degree-of-freedom (SDOF) system used to introduce structural dynamics. The SDOF classification means that there is a single variable $u(t)$ that uniquely and completely describes the motion of the mass, which as indicated is restricted to move in one direction only. The velocity $\dot{u}(t)$ and acceleration $\ddot{u}(t)$ follow. The spring represents the structural restoring forces, the structural stiffness; these forces f_r are principally a function of displacement. However, if f_r is a multi-valued function of u, then it will also depend on the history of the displacement (path dependence) and potentially on $\dot{u}(t)$. Energy dissipation is represented by the damper, and the force that the damper applies to the mass is denoted f_d. In the most common formulations in the time domain, the damping force is assumed to depend on the velocity, although it is possible that it may depend on the displacement as well. The excitation depicted in the figure is an applied external force $p(t)$. Figure 37.1b is a free-body diagram of the system. Unless otherwise noted, the discussion below will refer to this system.

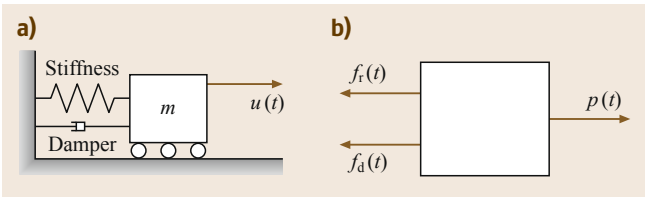


Fig. 37.1 (a) Spring-mass-damper system and (b) free-body diagram

The equation of motion in the time domain follows directly from Fig. 37.1b and Newton's second law; specifically, the sum of the external forces is equal to the change in momentum. Generally, in problems in

structural dynamics the mass does not vary with time. Therefore, the equation of motion can be written fairly generally as

$$p(t) - f_d - f_r = f_I \,, \tag{37.1}$$

in which f_I represents the product of mass and acceleration. In structural dynamics, the equation is more often written as

$$f_I + f_d + f_r = p(t) \,, \tag{37.2}$$

and f_I is referred to as the inertia force à la d'Alembert's principle. The equation can then be interpreted as the sum of the *internal* resistive forces is equal to the sum of the external forces. This gives rise to the concept of *dynamic equilibrium*, which is frequently used in structural dynamics.

Equation (37.2) covers both linear and nonlinear systems. A linear system requires a linear spring, such that $f_r(t) = ku(t)$, in which the constant stiffness k has units of force per unit displacement. It also requires a linear dashpot, such that $f_d(t) = c\dot{u}(t)$, in which the damping constant c has units of force per unit velocity. For a system with constant mass m, $f_I(t) = m\ddot{u}(t)$, where m has units of mass, or force per unit acceleration. The equation of motion becomes

$$m\ddot{u}(t) + c\dot{u}(t) + ku(t) = p(t) \,. \tag{37.3}$$

Note that this discussion assumes linear viscous damping; as discussed subsequently, an alternative damping model may be more convenient for ocean structures.

The above form assumes that the dynamic excitation of the structure is a result of an externally applied dynamic force. It is also possible that the excitation is a result of support motion instead, such as from earthquake ground motion. In Fig. 37.1a, let $p(t) = 0$ and assume that the support undergoes a *ground* displacement $u_g(t)$ and associated acceleration $\ddot{u}_g(t)$. The total displacement of the mass can be written as the sum of a rigid body motion $u_g(t)$ and the motion relative to the ground, $u(t)$. The inertia force depends on the total acceleration, but the structure's resistive forces f_r and f_d depend only on the relative motion. The linear equation of motion (37.3) can then be written as

$$m\ddot{u}(t) + c\dot{u}(t) + ku(t) = -m\ddot{u}_g(t) \,. \tag{37.4}$$

The term $-m\ddot{u}_g(t)$ can be viewed as an effective load. Note that the equation is written in terms of relative motion, and if commercial software is being used, the user should understand whether the relative displacements or the total displacements are being reported.

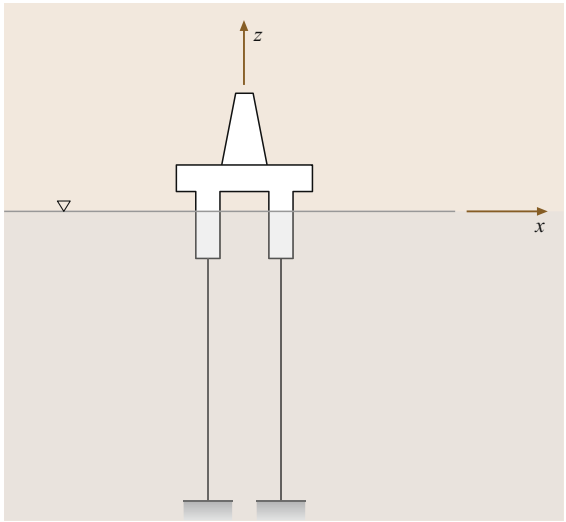


Fig. 37.2 Schematic of a TLP

Although the system in Fig. 37.1a may not look like any practical structure, it provides an excellent basis to understand structural dynamics. Furthermore, some systems can be reasonably well represented by a single degree-of-freedom, at least for initial studies. For example, Fig. 37.1a may be a reasonable model to study the linear vertical response of a tension leg platform (TLP), if that motion is assumed uncoupled from motion in other directions and the mass of the tendons is either ignored or partially lumped with the TLP mass; see the schematic in Fig. 37.2. If the fluid is ignored temporarily, m would be the mass of the platform, which would be assumed rigid; k would be the sum of the vertical stiffness of the tendons (again, ignoring buoyancy at the moment); and c a constant that would have to be specified. Dynamic energy dissipation in structures results from a variety of physical mechanisms that, for the most part, are not well understood quantitatively and are not modeled explicitly. Rather, a damping model, such as the viscous damper shown in Fig. 37.1a, is given equivalent properties to provide approximately the same level of energy dissipation that is found in field tests.

A slight modification to the free body in Fig. 37.1b makes it applicable to the uncoupled pitch motion of the TLP. In that case, u represents the rotational displacement (in radians) of the TLP, the forces are moments, m the rotational inertia of the TLP, and k the resistance to pitch provided by the tendons (moment per radian). Again, c would have to be based on experimental data. Hence, all the equations in this section apply to both uncoupled heave motion and pitch motion. This TLP application will be used to give more physical interpretation to the subsequent discussions. It also illustrates that practical structures will have more

than one degree-of-freedom; multi-degree-of-freedom structures are considered subsequently.

37.1.2 Response in the Time Domain – Free Vibration

The primary focus herein is on linear dynamic response because 1) analytically tractable solutions are possible, and 2) an understanding of linear structural behavior is critical prior to understanding nonlinear behavior.

The term free vibration in structural dynamics refers to the response of the system to initial displacement $u(0)$ and initial velocity $\dot{u}(0)$ (initial conditions), but with no loading, i. e., $p(t) = 0$. It is convenient to divide (37.3) by m to obtain

$$\ddot{u}(t) + 2\xi\omega_n\dot{u}(t) + \omega_n^2 u(t) = 0\,, \tag{37.5}$$

in which $\omega_n = \sqrt{k/m}$ is the undamped natural frequency (radians per second) and $\xi = c/(2m\omega_n)$ is the (dimensionless) ratio of the damping coefficient c to the critical damping of the system. The critical damping is the minimum level of damping that prevents oscillation during free vibration. For most practical structures, damping is relatively small and is often referred to as a percent of critical damping, such as *the structure has 2% damping*. If the damping is less than critical, the solution to (37.5) is

$$u(t) = \mathrm{e}^{-\xi\omega t}\left(\frac{\dot{u}(0) + u(0)\xi\omega_n}{\omega_D}\sin\omega_D t + u(0)\cos\omega_D t\right)\,, \tag{37.6}$$

in which $\omega_D = \omega_n\sqrt{1-\xi^2}$ is the damped natural frequency. The undamped and damped natural periods are $T_n = 2\pi/\omega_n$ and $T_D = T_n/\sqrt{1-\xi^2}$, respectively. It is also convenient to write (37.6) in a different form

$$u(t) = \rho\mathrm{e}^{-\xi\omega_n t}\cos(\omega_D t - \theta)\,, \tag{37.7a}$$

with

$$\rho = \sqrt{\left(\frac{\dot{u}(0) + u(0)\xi\omega_n}{\omega_D}\right)^2 + [u(0)]^2}\,, \tag{37.7b}$$

$$\theta = \tan^{-1}\frac{\dot{u}(0) + u(0)\xi\omega_n}{\omega_D u(0)}\,. \tag{37.7c}$$

Several aspects of the response can be seen from (37.6) and (37.7). First, for nonzero levels of damping, the response decays exponentially with time, whereas when there is no damping the response continues harmonically with no decay in amplitude. Second, with no damping, the structure vibrates with frequency ω_n and period T_n. In the case of damping, the structure vibrates with frequency ω_D and period T_D. The small levels of structural damping in structures of interest, often less than 5%, mean that the damped and undamped natural frequencies are essentially the same for many practical purposes. The undamped, 2 and 5% damped structural responses are illustrated in Fig. 37.3 for about five cycles of motion. Whereas viscous damping lengthens the period a little bit, the differences in periods between the damped and undamped systems for these levels of damping are very small. The amplitude decay as a result of damping is also evidenced in the figure. It is straightforward to show that the natural log of the ratio of successive peak amplitudes, which is defined as the logarithmic decrement δ, is given by

$$\delta \triangleq \ln\frac{u_j}{u_{j+1}} = \frac{2\pi\xi}{\sqrt{1-\xi^2}} = 2\pi\xi\frac{\omega_n}{\omega_D}\,, \tag{37.8}$$

in which u_j is the peak displacement at the j-th peak. At 5% damping, the amplitude is reduced by about 50% in two cycles of vibration, whereas for 2% damping it takes over five cycles to reduce by 50%.

In the TLP example discussed above, we saw that our SDOF model can be applied separately to vertical motion and pitch motion as long as they are uncoupled, i. e., if heave does not cause pitch and pitch does not cause heave. It is, therefore, clear that there will be a heave natural frequency and a pitch natural frequency. Generally, as we shall see, structures have as many natural frequencies as there are dynamic degrees-of-freedom.

37.1.3 Response in the Time Domain – Harmonic Loading

The general equation of motion, including loading, is

$$\ddot{u}(t) + 2\xi\omega_n\dot{u}(t) + \omega_n^2 u(t) = \frac{p(t)}{m}\,. \tag{37.9}$$

Analytical solutions are usually expressed as a sum of the complementary solution and a particular solution. The complementary solution is obtained when $p(t) = 0$ and the particular solution is the solution for the specific $p(t)$ for a given problem. The complementary solution has two undetermined constants, which are evaluated so that the combined solution satisfies the initial conditions at $t = 0$.

Wave loading on offshore structures is often random and nonlinear. However, for linear structural response analysis the wave loading can be considered a sum of harmonic (sinusoidal) forces with random phases, as in

$$p(t) = \sum_q p_q\cos(\omega_q t + \varepsilon_q)\,, \tag{37.10}$$

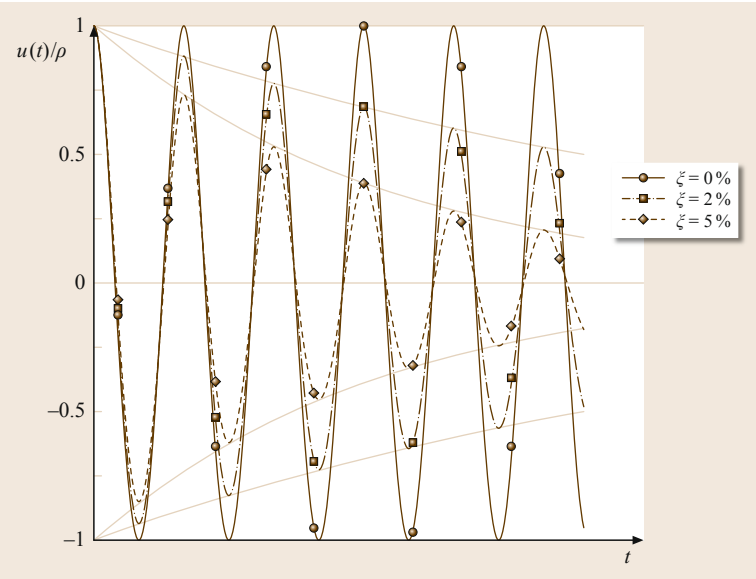


Fig. 37.3 Free vibration response for three values of the damping ratio. The *dashed bounding lines* indicate the decay in amplitude with damping

in which ε_q is the random phase angle for the q-th component. The structural response is then the sum of the response to each harmonic wave component (Sect. 37.5). For this reason, it is advantageous to first study the response of a structure to periodic loading. Periodic loadings can be decomposed into a Fourier series as a summation of sine and cosine terms. Similarly, because of linearity the response can be represented as the summation of the response to each harmonic component. It is, therefore, sufficient at this point to focus the discussion on the response to a single harmonic component of the load. If we assume, therefore, that the loading is $p(t) = p_q \cos(\omega_q t)$, in which p_q is the magnitude for the excitation frequency component ω_q, the response is

$$u(t) = \mathrm{e}^{-\xi\omega t}\left(A \sin \omega_\mathrm{D} t + B \cos \omega_\mathrm{D} t\right) + \frac{p_q}{k} \frac{\left(1-\beta^2\right)\cos\left(\omega_q t\right) + 2\xi\beta \sin(\omega_q t)}{\left(1-\beta^2\right)^2 + (2\xi\beta)^2}\,, \tag{37.11}$$

in which $\beta = \omega_q/\omega_\mathrm{n} = T_\mathrm{n}/T_q$. The solution for a sine variation in force can be obtained from (37.11) by replacing $\sin(\omega_q t)$ with $-\cos(\omega_q t)$ and $\cos(\omega_q t)$ with $\sin(\omega_q t)$. The first term, the complementary solution, involves the coefficients A and B; it is straightforward to determine them such that the complete solution satisfies the initial conditions. However, damping causes the complementary solution to decay relatively quickly with time, because it is multiplied by $\mathrm{e}^{-\xi\omega_\mathrm{n} t}$, and it is, therefore, referred to as the transient response. The second term, the particular solution, is referred to as the steady-state (harmonic) response.

For ocean structures, the excitations from waves occur over a long period of time, and the transient response is assumed to have decayed such that it can be ignored. Note that this may not be the case for other types of loading, such as slamming, collision, blast or ground (seismic) motion. However, we are concerned primarily with wave-induced response, and, therefore, we will focus on the steady-state solution and assume that the transient response can be ignored. We do note that for nonlinear wave response, one has to include the initial conditions and the transient response; the simulation must be run sufficiently long for the transients to decay and for the solution to converge to one of the (possibly multiple) stable solutions as a result of the path-dependence of the structure.

The peak displacement of the steady-state response, obtained from (37.11), is

$$\frac{u_\mathrm{peak}}{(p_q/k)} = \frac{1}{\sqrt{\left(1-\beta^2\right)^2 + (2\xi\beta)^2}} = R_\mathrm{d}\,, \tag{37.12}$$

in which the displacement has been nondimensionalized by the displacement that would result if the structure responded statically, that is, p_q/k. R_d is called the deformation response factor, the displacement response factor, or sometimes just the dynamic amplification factor. R_d is shown in Fig. 37.4 for various damping levels. As $\beta \to 0$, i. e., when the load varies very slowly relative to the structure's natural frequency, the response approaches the static displacement ($R_\mathrm{d} \to 1$). This provides the justification for static or quasi-static analysis for such slowly-varying loads. The maximum displacement response is obtained when $\omega_q = \omega_\mathrm{n}\sqrt{1-2\xi^2}$. For

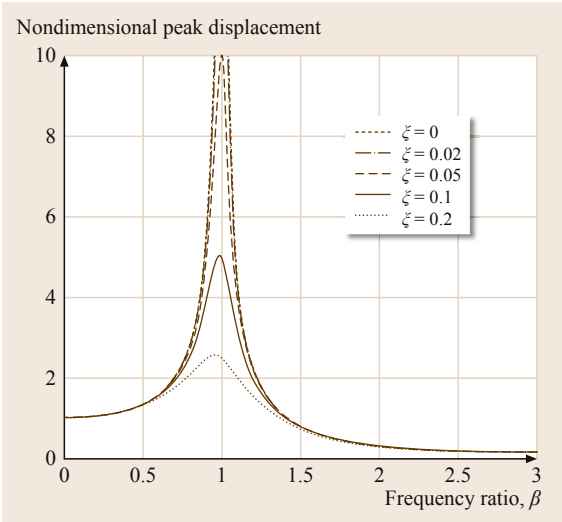


Fig. 37.4 Displacement response factor

practical values of structural damping, the excitation frequency corresponding to the peak displacement is very close to the undamped natural frequency. If $\beta = 1$ is substituted into (37.12), the peak response is $1/(2\xi)$ times the static response. Because damping is small, this means that the structure's response at resonance can be very large relative to the static response to the same load magnitude. Hence, especially for periodic excitation such as wave excitation, it is important to ensure that the structure's natural frequency is not in the frequency range of significant excitation. As the excitation frequency increases to values much larger than the natural frequency (large β), the displacement response goes to zero. A convenient physical explanation for this phenomenon is to imagine a large mass on a soft spring; as the load varies at high frequency, the mass simply does not have time to respond as the load changes direction.

Considering Fig. 37.4, design of ocean structures often involves ensuring that the important natural periods of the structure are either above or below the periods of significant wave energy. For example, for fixed jacket platforms, the significant natural frequencies are usually higher than the dominant wave energy; that is, they are to the left of $\beta = 1$ in the figure. For moored floating structures, such as semisubmersibles and tension-leg platforms, the significant natural frequencies are smaller than the dominant wave frequencies and they are to the right of $\beta = 1$.

37.1.4 Response in the Frequency Domain

As mentioned above, if the loading is periodic then it can be represented as a summation of harmonic components and the response can be represented as the summation of the response to each harmonic component. The situation is the same as considered above, that is, for damped systems the initial transients die out over a sufficient number of loading periods and the long-term (or steady-state) response is harmonic with the same frequency as the load. We again focus the discussion on the response to a single harmonic component of the load, $p_q e^{i\omega_q t}$, in which p_q is the magnitude for the excitation frequency component ω_q. Therefore, the displacement corresponding to this excitation component will be of the form $u_q e^{i\omega_q t}$. Substitution of these expressions into (37.3) results in the equation of motion for each component in the frequency domain

$$\left(-\omega_q^2 m + i\omega_q c + k\right) u_q = p_q \,. \tag{37.13}$$

Equation (37.13) is an algebraic equation that is much easier to solve than the differential (37.3). It is straightforward to obtain the solution as

$$u_q = \frac{p_q}{k} \frac{1}{1-\beta^2 + i(2\xi\beta)} \,, \tag{37.14}$$

in which $\beta = \omega_q/\omega_n$. We should note that (37.14) must be solved for each frequency component ω_q and that (37.14) and u_q are complex. The magnitude of the peak displacement from (37.14) is the same as that from (37.12). We therefore obtain the same definition for R_d.

If p_q corresponds to a unit value of a load parameter (unit wave amplitude), then evaluation of (37.14) for a range of wave (excitation) frequencies results in the complex, frequency-domain transfer function for the displacement, also frequently referred to as the complex frequency response function in structural dynamics. The magnitude of the transfer function is called the response amplitude operator (RAO) in ocean engineering. The symbol H is often used for the transfer function. If it is determined for a unit value of p_q, then

$$u_q = H_u\left(\omega_q\right) p_q \,. \tag{37.15}$$

An important advantage of (37.15) as compared to (37.3) is that it is much easier to treat the situation in which the properties m, c and/or k are frequency dependent. As we shall see, this is usually the case for m and c with ocean structures.

37.1.5 Structural Damping

Energy dissipation in structures excluding hydrodynamic effects is a combination of a number of mechanisms, such as internal energy dissipation in the material, friction in structural components and connections, and possibly concrete cracking. The amount of dissipation under linear behavior is typically quite small,

amounting to damping that is just a few percent of critical damping. There do not exist the same physics-based models for this energy dissipation as for determining the stiffness, and sophisticated modeling would be quite difficult. Fortunately, because it is small the energy dissipation can be modeled fairly crudely. In the time domain, a linear viscous damping model is adopted because it is so convenient. The equations remain linear and analytical solutions are tractable. However, viscous damping is not necessarily the best model for the physics involved in energy dissipation. The actual energy dissipation is more nearly proportional to displacement than velocity [37.9], and in the frequency domain an alternative model, sometimes called linear hysteretic damping, is easily accommodated. In linear hysteretic damping, the dissipation force f_d is still out of phase with the displacement, and can be modeled by $f_d = i2\xi k u_q$. Replacing the linear viscous damping in (37.13) with this expression results in

$$\left[-\omega_q^2 m + (1 + i2\xi)k\right] u_q = p_q \,, \tag{37.16}$$

the solution of which is

$$u_q = \frac{p_q}{k} \frac{1}{1 - \beta^2 + i(2\xi)} \,. \tag{37.17}$$

For practical levels of damping, the peak magnitude of (37.17) is very nearly the same as (37.14), although the peak is obviously now at the undamped natural frequency. The two models result in the same energy dissipation per cycle of vibration when the excitation frequency ω_q is equal to the natural frequency ω_n; otherwise, they are different. Actually, hysteretic damping is typically superior because it results in rate-independent energy dissipation. That is, the energy dissipated in a cycle of motion is independent of the frequency, which often reflects reality better than linear viscous damping, in which the energy dissipation varies with frequency. When the frequency domain approach is applicable, it is often preferable to use the hysteretic damping model. However, when the time-domain approach is used, linear viscous damping is used for convenience.

Both linear viscous damping and hysteretic damping require specification of the damping ratio ξ. Typical values usually come from field tests on similar structures. One common method to determine the level of damping in a structure is to measure the decay in the amplitude of motion during free vibration. For example, based on (37.8), it is straightforward to show that

$$\ln\left(\frac{u_j}{u_{j+m}}\right) = 2\pi m \xi \frac{\omega_n}{\omega_D} \,. \tag{37.18}$$

With this equation one can measure the peak amplitude of the motion over multiple cycles of vibration and obtain an estimate of the damping ratio. Of course, it must be practical to initiate the free vibration for this to be an option.

An alternative approach is based on the response to harmonic excitation. From Fig. 37.4 it is clear that damping not only reduces the peak amplitude, but it also broadens the shape of R_d as a function of frequency. The half-power bandwidth method utilizes this characteristic to estimate the damping. The structure is excited harmonically, the transients are allowed to die out, and the peak steady-state response is measured. This must be done over a range of frequencies to obtain an experimental plot of peak displacements vs. frequency. For small levels of damping, the frequency at peak response can be taken as ω_n. Then the frequency ω_L, to the left of the resonant frequency, that corresponds to a displacement magnitude equal to $1/\sqrt{2}$ of the peak amplitude is determined. Similarly, the frequency ω_R, to the right of the resonant frequency, that corresponds to a displacement magnitude equal to $1/\sqrt{2}$ of the peak amplitude is determined. For small levels of damping, the damping can be estimated as

$$\xi = \frac{\omega_R - \omega_L}{2\omega_n} \,. \tag{37.19}$$

Of course, one must be able to excite the structure harmonically to apply this procedure. Vibration generators, which consist of counter-rotating eccentric weights, have been developed to do this [37.10]. There are other approaches to determine damping from field data, such as from response from random ambient vibrations; see, for example, [37.22].

The issue of specifying the appropriate level of damping for marine structures can be challenging. The discussion by *Bishop* and *Price* [37.23] on the difficulties and unknowns is largely still valid. In that work, they provide a summary of experimentally determined damping ratios obtained by different researchers for different classes of ships. These hull-damping ratios range from about 0.5 to about 2%. Ambient and forced vibration measurements on a steel offshore platform in the North Sea indicated damping ratios between 1 and 3% [37.24]. Some of the difficulties of measuring damping of offshore structures is discussed by *Ruhl* [37.25], who measured damping of four jacket platforms. He also found a range of 1 to 5%, depending on what approach was used to obtain the damping coefficient. According to him, the most credible values seemed to be 2 to 3%, which agree well with *Gundy* et al. [37.24]. Although these are values for structures in water, and therefore reflect some the fluid, it is likely that most of the damping measured in these cases

is from structural damping. In any event, these small damping ratios indicate how small structural damping can be.

37.1.6 Time Domain Response from Frequency Domain Response

The transfer function $H_u(\omega_q)$ is the response in the frequency domain for a unit harmonic load, assuming the load has acted *forever*, so there are no remaining transients. The response in the frequency domain can be used to obtain the response in the time domain for an arbitrary periodic load. Let the periodic load $p(t)$ have a period T_p. It is assumed that the load has acted *forever*, so that the response in any given period is equal to the response in all other periods.

Because the load is periodic, it can be represented as a Fourier series. To match the period T_p, the smallest cyclic frequency in the series is $f_0 = 1/T_p$, and all other frequency components must be integer multiples of f_0, otherwise the load would not have period T_p. Therefore,

$$p(t) = \sum_{n=-\infty}^{\infty} P_n e^{i(2\pi n f_0)t} , \tag{37.20}$$

in which

$$P_n = \frac{1}{T_p} \int_0^{T_p} p(t) e^{-i(2\pi n f_0)t} dt . \tag{37.21}$$

Of course, practical analysis requires 1) truncation of the series in (37.20), and 2) usually a discretization of the load. One approach to obtain the discrete Fourier series representation is to use the trapezoidal method for the integral in (37.21). The period is divided into N segments, such that $\Delta t = T_p/N$, the inverse of which gives the largest cyclic frequency considered. The value of N must be large enough such that it 1) captures the frequency content of the loading, and 2) captures the largest natural frequency of the structure that needs to be included in the response. The load is sampled at discrete times $t_s = s\Delta t$, with $s = 0, 1, \ldots N-1$, because of the periodicity, $p(t_0) = p(t_N)$. Numerical evaluation of (37.21) results in

$$P_n = \frac{1}{N} \sum_{s=0}^{N-1} p(t_s) e^{-i\left(\frac{2\pi ns}{N}\right)}, \quad n = 0, \ldots \frac{N}{2} . \tag{37.22}$$

For real $p(t)$, P_{-n} is the complex conjugate of P_n.

Once the frequency domain representation of the load has been obtained, the frequency domain response is simply

$$u_n = H_u(2\pi n f_0) P_n . \tag{37.23}$$

To obtain the response in the time domain, one only needs to take the inverse Fourier series

$$u(t_s) = u_0 + 2 \sum_{n=1}^{N/2} \mathrm{Re}\left(u_n e^{i\left(\frac{2\pi ns}{N}\right)}\right) . \tag{37.24}$$

The approach requires the transfer functions to be evaluated at $N/2$ frequencies, which can be a large number. If the corresponding computations are excessive, one option is to use some interpolation scheme to reduce the frequencies for which the transfer functions must be computed. Practical applications of this approach also invariably use the fast Fourier transform (FFT). The classic FFT algorithm was developed by *Cooley* and *Tukey* [37.26], although others have been developed subsequently, for example [37.27].

The discrete Fourier transform can also be used to obtain the solution for finite duration loading that is not periodic. The basic strategy is to extend the *period* of the loading beyond the time at which the actual load ends. The load in this extended section is zero. The extension must be sufficiently long that the response from the previous *cycle* has damped to nearly zero. *Veletsos* and *Ventura* [37.28] discuss the procedure and some of the pitfalls associated with its use. *Chopra* [37.10] also provides a concise summary of errors associated with the discrete Fourier transform for nonperiodic loads. Reference [37.29] discusses the application of this approach to the seismic response of a structure in the context of soil–structure and fluid–structure interaction.

37.1.7 Fluid Contribution

Static Effect – Hydrostatic Stiffness

The resistance to heave and pitch displacement of the TLP comes both from structural elements, the tendons, but also from the fluid as the hydrostatic pressure acting on the structure changes with vertical displacement. Integration of these pressure changes over the wetted surface results in the hydrostatic restoring force coefficients, or hydrostatic stiffness. Whether for heave or pitch of the TLP, we denote the corresponding stiffness by k_f. The total system stiffness is the sum of the structure stiffness from the tendons, now denoted k_s, and the hydrostatic stiffness, such that k in the above equations is now given by $k = k_s + k_f$. Of course, (37.16) would have to be modified slightly because the damping term should only multiply k_s; the water provides no displacement-related energy dissipation.

Dynamic Effect – Added Mass, Damping, and Exciting Forces

To determine the dynamic effects of the fluid, a specific fluid model needs to be adopted. This is the same

situation as in rigid body hydrodynamics in Chap. 34. For the linear analysis of large ocean structures, linear potential theory is usually adopted, and this will be the basis of the discussion here. For the TLP model, the heave or pitch added mass, m_f, the hydrodynamic damping, c_f, as well as the wave exciting (incident and diffraction) forces, F_f, are the same as would be calculated in Chap. 34 if the tendons were ignored.

Because the added mass, hydrodynamic damping, and wave exciting forces are functions of the wave frequency ω_q, an obvious difficulty arises with including them in the equation of motion in the time domain, Here, the advantage of the frequency domain approach becomes apparent, because there is no difficulty to include the frequency dependence. Equation (37.13) now becomes

$$\left[-\omega_q^2\left(m_s+m_f\right)+\mathrm{i}\left(\omega_q c_f+2\xi k_s\right)+\left(k_s+k_f\right)\right]u_q = F_f\,. \tag{37.25}$$

Equation (37.25) is easily solved to obtain the response in the frequency domain and to develop the RAO of the response. Of course, hidden within the equation is the necessity to determine the fluid coefficients, which can entail a large computational effort, as discussed in Chap. 34.

Equation (37.25) is often adequate for the wave response of linear structures, but when it is desired to introduce nonlinearities, for example, in the structure stiffness or the forces involved in active control systems, then a time-domain solution is required. The frequency dependence of the fluid effects can be transferred to the time domain through an inverse Fourier transform. In structural dynamics, it is not uncommon to transform the linear response in the frequency domain to the response in the time domain, and this can be done conveniently using the fast Fourier transform. The situation here is a little different, however, because it is not the response that is transformed. The cosine transform proposed by *Cummins* [37.30] and discussed further by *Ogilvie* [37.31] is generally used; see, for example [37.32], while *Taghipour* et al. [37.33] and *Sheng* and *Lewis* [37.34] provide good summaries. For zero mean speed, the basic equation of motion becomes

$$\left(m_s+m_{f\infty}\right)\ddot{u}(t)+\int_0^t K\left(t-\tau\right)\dot{u}(t)\mathrm{d}\tau + \left(k_s+k_f\right)u(t)=F_f(t)\,, \tag{37.26}$$

in which $m_{f\infty}$ is the infinite frequency added mass and the retardation function $K(t)$ is given by

$$K(t)=\frac{2}{\pi}\int_0^\infty c_f(\omega)\cos(\omega t)\mathrm{d}\omega\,. \tag{37.27}$$

The convolution integral in (37.26), which is often referred to as the fluid memory effect, must be evaluated at each time step. This can be computationally significant.

Recent work on this approach seems to have focused on wave-energy conversion devices, including their control. Optimization and control of wave energy extraction may be performed most efficiently with a state-space formulation of the equation of motion [37.35]. Examples can be found in [37.33, 36–39].

Dry and Wet Natural Frequencies

The so-called dry or *in-air* (undamped) natural frequency of the system is obtained when the fluid is ignored

$$\omega_n=\sqrt{\frac{k_s}{m_s}}\,. \tag{37.28}$$

The wet natural frequency when the structure is in water is

$$\omega_n=\sqrt{\frac{k_s+k_f}{m_s+m_f(\omega_n)}}\,, \tag{37.29}$$

in which the functional dependence of the added mass on the frequency of motion is made explicit. (37.29) is clearly nonlinear, and an iterative process is required to find the wet natural frequency. This process requires repeated solution of the radiation problem to obtain m_f. The iteration would cease once the frequency for which m_f was calculated is sufficiently close to ω_n given by (37.29). Because of the need to solve multiple radiation problems, finding the wet natural frequency can be computationally significant. One option to approximate the wet natural frequency is to ignore the added mass, such that

$$\omega_n\approx\sqrt{\frac{k_s+k_f}{m_s}}\,. \tag{37.30}$$

There is probably little reason to do this for the SDOF model. However, it could be useful in multiple degrees-of-freedom models, as discussed subsequently.

Natural frequencies are determined primarily for two reasons. First, they can help to understand the dynamic response. Knowing the natural frequency helps understand which excitation frequencies will cause large response and which will cause small response, viz. Fig. 37.4. For multiple degrees-of-freedom (MDOF) systems, one also gains an understanding of how the structure will deform. While the RAOs, computed for a range of wave frequencies, will reveal the frequency-dependence of the response, the shape of the deformations is admittedly more difficult to obtain from the RAOs. The second reason to solve for the natural frequencies is to develop solution procedures in the time domain that are computationally more efficient than would otherwise be possible. This is not as relevant to the SDOF system, but can be very useful for structural dynamics of linear MDOF systems. However, because of the frequency dependence of added mass and damping, the computational efficiency that would otherwise be provided by the wet natural frequencies is typically lost. As a result, the wet natural frequencies are usually not calculated explicitly, and resonance is identified through peaks in the RAOs. One can also approximate the wet natural frequencies with (37.29) by using an added mass for a frequency relatively close to the calculated natural frequency.

Direct Coupling for Fluid–Structure Interaction

The above discussion has assumed the most common approach to consider the coupled fluid–structure problem. For example, a typical diffraction (panel) code is used to obtain the hydrodynamic coefficients, which are then included in the structural model. A more tightly coupled approach could be used, in which the two systems are solved simultaneously. See [37.40] for an example of such an approach. The direct coupling obviates the iteration discussed above to find the wet natural frequencies.

The discussion herein will focus on the former strategy, which is still the most common approach for linear hydroelasticity of ocean structures. The reason for that may be as much historical as anything. Linear hydroelasticity evolved out of linear rigid body dynamics based on diffraction codes [37.23, 41, 42]. These codes were well developed to calculate the added mass, damping, and wave exciting forces for rigid ocean structures. As will be seen, the equations involved in these codes need only slight modifications to be applied to linear flexible structures. A completely separate, existing structural analysis code can be used for the structural model. With this architecture, the loose coupling only requires some modest interchange of data. Furthermore, the solution of (37.25) to obtain the wave-induced response does not require iteration. Therefore, there does not appear to be a disadvantage to the loose coupling if one does not need the wet natural frequencies.

37.1.8 Nonlinear Systems

The discussion so far has focused on linear systems. Analysis of practical structures involving nonlinearities almost always involves a numerical solution in the time domain. An incremental equation of motion can be written such that the solution is progressed from displacement $u(t)$ at time t to displacement $u(t+\Delta t) = u(t) + \Delta u$ at time $t+\Delta t$. Equation (37.2) is written at time $t+\Delta t$, and then linearized about time t. The resulting incremental equation of motion is

$$m\Delta\ddot{u} + c_T\Delta\dot{u} + k_T\Delta u = p(t+\Delta t) - m\ddot{u}(t) - f_d(t) - f_r(t)\,. \tag{37.31}$$

k_T is the tangent stiffness at time t; that is, it is the slope of the $f_r - u$ function at $u(t)$. Most structural materials such as steel and concrete are nonlinear hysteretic. Therefore, the tangent will depend also on the previous path, or history of the displacement, as well as on whether the material is loading or unloading. (Consider, for example, the nonlinear uniaxial stress-strain curve for steel, as it yields, unloads, and then is loaded past yielding again.) Similarly, c_T is the tangent damping coefficient. When only typical energy dissipation in the structure is considered, this is very often assumed to be a constant c. If equilibrium is satisfied exactly at time t, then the right-hand-side of (37.31) is just Δp. However, equilibrium is rarely satisfied exactly and using the form in (37.31) can help to control errors. Nevertheless, iteration is often required.

An important source of nonlinearity for offshore structures is the nonlinearity from the fluid loading, even if the structure itself remains linear elastic. The Morison equation formulation is a practical formulation of nonlinear fluid forces on *small structures* [37.43]. For wave flow around an oscillating slender structure, such as the members of a jacket platform, the presence of the structure induces a wake immediately in the downstream of the instantaneous wave flow but does not significantly affect the flow field away from the vicinity of the structure [37.15, 17]. Under this *small structure* assumption, the fluid loads on the structure can be represented by the Morison inertia force f_{fI} and drag force f_{fD}

$$f(t) = f_{fI} + f_{fD}\,, \tag{37.32a}$$

$$f_{fI}(t) = C_M A_I \dot{U}(t) - C_A A_i \ddot{u}(t)\,, \tag{37.32b}$$

$$f_{fD}(t) = C_D A_D\,[U(t) - \dot{u}(t)]\,|U(t) - \dot{u}(t)|\,, \tag{37.32c}$$

in which C_M is the inertia force coefficient, and C_A and C_D are the inertia and drag coefficients, A_I and A_D are the projected areas of the structure in the direction of the fluid flow, and $U(t)$ and $\dot{U}(t)$ are the (absolute) fluid flow velocity and acceleration, respectively. The above equation is called the relative velocity Morison equation. In the literature, the density of water sometimes is written explicitly in the latter two equations, and a factor of $1/2$ is included in the third expression. However, often these factors are absorbed into the inertia and drag coefficients [37.17, 44], which is the case here. The Morison equation works well when the flow velocity is oscillatory, such as for a small structure in periodic waves [37.17, 44]. If the fluid flow consists of a combination of periodic waves and a steady current, the Morison drag force can be written as

$$f_{fD}(t) = C_D A_D \left(U(t) + U_C(t) - \dot{u}(t)\right) \left|U(t) + U_C(t) - \dot{u}(t)\right| , \qquad (37.33)$$

where $U_C(t)$ is the steady current. It is clear from these equations that the system is nonlinear because of the drag term.

There are several options available to include these Morison forces acting on the structure in the incremental equation of motion. For simplicity, we will assume that the only forces are the Morison forces, and hence $p(t+\Delta t)$ is equal to $f(t+\Delta t)$. We will assume further that equilibrium has been achieved at time t such that the right-hand side is just $\Delta p(t+\Delta t) = f(t+\Delta t) - f(t)$. Substitution into (37.31) results in

$$(m + C_A A_I)\,\Delta\ddot{u} + c_T \Delta\dot{u} + k_T \Delta u = C_M A_I \Delta\dot{U} + f_{fD}(t+\Delta t) - f_{fD}(t) . \qquad (37.34)$$

Although the fluid inertia term is handled easily, the nonlinear drag term is not. The right-hand side depends on the unknown $\Delta\dot{u}$. To resolve this problem, the simplest approach is to calculate $f_{fD}(t+\Delta t)$ using $\dot{u}(t)$ instead of $\dot{u}(t+\Delta t)$. Errors resulting from this approximation can likely be controlled by using a sufficiently small time step. Another option is to use iteration within a time step such that at iteration i, $f_{fD}(t+\Delta t)$is calculated using the value of $\dot{u}(t+\Delta t)$ obtained in iteration $i-1$.

A more involved approach to deal with nonlinearity is to linearize the drag term about time t, and then move the term involving $\Delta\dot{u}$ to the left-hand side. The difficulty here is that the coefficient on $\Delta\dot{u}$ will involve $U(t+\Delta t)$ and $\dot{u}(t)$. Because the coefficient will be a function of time, it will need to be updated for each time increment. *Sarpkaya* and *Isaacson* [37.15] and *Veletsos* et al. [37.45] discuss options for such formulations in more detail.

It must be emphasized that the Morison force equations work for cases that correspond to the conditions for which the empirical inertia and drag coefficients have been determined experimentally. For significantly different flow conditions, a different set of *Morison coefficients* C_M and C_D will have to be determined experimentally. There is no analytical means of extending the range of applicability of these empirical coefficients.

37.2 Multi-Degree of Freedom Systems

37.2.1 Equations of Motion

The response of very few structures can be represented adequately by a single displacement. As a result, models that involve multiple degrees-of-freedom are required, just as a floating rigid body requires six degrees of freedom to completely define its motion. As such, the equation of motion in the time domain for a flexible structure becomes

$$\mathbf{M}_s\ddot{\boldsymbol{u}}(t) + \mathbf{C}_s\dot{\boldsymbol{u}}(t) + \mathbf{K}_s\boldsymbol{u}(t) = \boldsymbol{P}(t) , \qquad (37.35)$$

in which $\boldsymbol{u}(t)$ is a vector of N nodal displacements that are required to approximate sufficiently the motion of the structure. $\mathbf{M}_s$, $\mathbf{C}_s$, and $\mathbf{K}_s$ are the $N \times N$ structural mass, damping, and stiffness matrices, and $\boldsymbol{P}(t)$ is the vector of N forces corresponding to the N displacements. The term displacement is generalized to include both translational displacements and rotational displacements, and hence, the term force is generalized to include both actual forces and moments. The structure is most often modeled by the finite element method, the details of which can be found in a number of texts; for example, see [37.46]. The method provides a methodology to obtain the mass and stiffness matrices and the load vector. However, no such formulation is generally available to obtain the damping matrix based on a physical model of the damping mechanisms. One approach to construct $\mathbf{C}_s$ is to use Rayleigh damping, wherein the matrix is formed from a linear combination of $\mathbf{M}_s$ and $\mathbf{K}_s$ to obtain a target proportional damping ξ. This issue will be discussed in more detail subsequently.

As before, the equation of motion can be written in the frequency domain as

$$\left[-\omega_q^2\mathbf{M}_s + (1 + \mathrm{i}2\xi)\mathbf{K}_s\right]\boldsymbol{u}_q = \boldsymbol{P}_q , \qquad (37.36)$$

This formulation again has the advantage of a simpler representation of damping. It should be noted that (37.36) indicates that a single value of ξ is used to multiply the stiffness matrix, but in fact, it is quite easy to use different values for different parts of the structure, in case they exhibit different energy dissipation characteristics. In that case, the stiffness matrices from the different parts of the structure are simply multiplied by different values of ξ to assemble the imaginary component of the now complex stiffness.

37.2.2 Modal Superposition

When the homogenous, undamped form of (37.36) is considered (no damping and $\boldsymbol{P}(t) = \boldsymbol{0}$), then the solution is harmonic and an eigenvalue equation results

$$\left[\mathbf{K}_\mathrm{s} - \omega_j^2 \mathbf{M}_\mathrm{s}\right] \boldsymbol{\psi}_j = \boldsymbol{O} \,, \tag{37.37}$$

in which there are up to N eigenvalues ω_j^2 and eigenvectors, or normal mode shapes, $\boldsymbol{\psi}_j$, $j = 1, \ldots N$. The undamped natural frequencies are usually ordered such that $\omega_1 \le \omega_2 \le \ldots \le \omega_N$. These natural frequencies have the same interpretation as the undamped natural frequency of the SDOF system, except now the vibration shape at frequency ω_j is described by $\boldsymbol{\psi}_j$.

The natural frequencies and (normalized) mode shapes satisfy two very important characteristics

$$\boldsymbol{\psi}_j^\mathrm{T} \mathbf{K}_\mathrm{s} \boldsymbol{\psi}_q = \begin{cases} \omega_q^2, j = q \\ 0, j \neq q \end{cases} \,, \tag{37.38a}$$

$$\boldsymbol{\psi}_j^\mathrm{T} \mathbf{M}_\mathrm{s} \boldsymbol{\psi}_q = \begin{cases} 1, j = q \\ 0, j \neq q \end{cases} \,. \tag{37.38b}$$

That is, the mode shapes are orthogonal with respect to both the mass and stiffness matrices. If Raleigh damping is assumed, they are also orthogonal with respect to the damping matrix. The *magnitude* of an eigenvector is not specified by (37.37). Equation (37.38b) has assumed that the eigenvectors have been mass normalized, such that $\boldsymbol{\psi}_j^\mathrm{T} \mathbf{M}_\mathrm{s} \boldsymbol{\psi}_j = 1$, a convention that is adopted by virtually all structural dynamic computer programs.

It is important to note that the N mode shapes form a basis for the same vector space in which $\boldsymbol{u}(t)$ exists, and therefore we can represent the displacements as a linear combination of these basis vectors with no loss in accuracy,

$$\boldsymbol{u}(t) = \sum_{j=1}^{N} \boldsymbol{\psi}_j Y_j(t) \,, \tag{37.39}$$

in which $Y_j(t)$ is the nondimensional generalized coordinate of mode shape j. With (37.38), the equations of motion (37.35) can be decoupled as

$$\ddot{Y}_j(t) + 2\xi_j \omega_j \dot{Y}_j(t) + \omega_j^2 Y_j(t) = \boldsymbol{\psi}_j^\mathrm{T} \boldsymbol{P}(t) = P_j^*(t) \,, \quad j = 1, \cdots N \,, \tag{37.40}$$

in which ξ_j is the modal damping ratio for mode j. These can be specified independently for each mode, but it is not uncommon to assume that they are the same for all modes. Note that with this approach, the damping matrix $\mathbf{C}$ need never be defined, much less formed. The modal force P_j^* is a measure of how much the spatial distribution of the load excites mode j.

Although it is a significant amount of work to solve (37.37) for the ω_j and $\boldsymbol{\psi}_j$, it can be well worth the effort given this result. Specifically, we have converted a large system of coupled equations into a set of uncoupled equations, which we can solve using the solution strategies and even the analytical solutions outlined for SDOF systems. That notwithstanding, the primary reason for doing this is that for relatively slowly-varying loading, such as with wave loading, the response of the structure is dominated by the response of the lower modes, and typically we let $j = 1, \ldots m$, $m \ll N$. That is, we represent the displacement vector with a reduced set of basis vectors. In this way, the effort to determine the natural frequencies and mode shapes can be justified. The process described here is referred to as modal superposition in structural dynamics. It should be noted that the high frequency modes, i. e., mode shapes greater than $j = m$, while they may not respond dynamically (Fig. 37.4 for small β), may respond statically. Hence, procedures to include this so-called static correction may be used to add the higher mode static response to the dynamic response.

The mode shapes from (37.37) are not the only alternative basis in which to represent the response. In earthquake structural dynamics in particular, so-called Ritz vectors are also used very successfully, and may be even more efficient than the normal modes because they are based on the spatial variation of the loading. However, the application of Ritz vectors to wave loading appears limited because the spatial variation of the loading changes with the wave frequency. Hence, this approach will not be discussed further.

Equation (37.40) illustrates an important difference between the SDOF system and the MDOF system. Dynamic response of the MDOF system depends not just on the ratio of excitation frequency to natural frequency, β, as shown in Fig. 37.4, but also on how the spatial distribution of the load excites a given mode shape, as quantified by the modal load P_j^*. That is, while R_d provides a measure of how much the loading frequency excites a given natural frequency of the structure, P_j^*

provides a measure of how much the spatial variation of the load excites the associated normal mode of the structure. For example, consider a structure with a natural frequency and a mode shape in one plane. A force in the orthogonal plane, even at the natural frequency of the structure, will not excite that mode because $P_j^* = 0$.

37.2.3 Numerical Time-Domain Solution

In Sect. 37.1.2, it was possible to obtain an analytical solution for the linear response in the time domain of a SDOF system subjected to a simple harmonic loading when the mass, damping, and stiffness are constant. For large MDOF systems, this approach is not practical, and in any event the loading is rarely a simple analytical function and often there is some source of nonlinearity, both of which preclude a simple analytical solution, even for SDOF systems. Therefore, we must numerically integrate the equations of motion to obtain the solution at discrete points in time, specifically at times Δt, $2\Delta t$, $3\Delta t$, etc., where Δt is the time step (assumed here to be constant for simplicity).

There are many different numerical schemes to obtain the time history of the response. Integration schemes can be categorized broadly as either explicit or implicit. Only implicit schemes can be unconditionally stable, their numerical stability does not require an upper limit on the size of the time step. Explicit schemes require for stability the time step to be based on the smallest natural period of the structural model. For very short duration loading, such as blast and impact that involve wave propagation in the structure, a small time step may be required for accuracy, and explicit time integration methods are frequently adopted. For such applications in structures, the central difference method is a common choice.

As mentioned, wave loading (and earthquake ground motion) has frequency content that is relatively low compared to the largest natural frequencies (smallest periods) contained in MDOF structural models. As such, the dynamic response of the higher modes in the structural model is not significant, and explicit methods are not efficient. As a result, for such low-frequency loading, unconditionally stable implicit schemes are almost always used for structural dynamics. One of the most widely-used schemes is Newmark's method, which we will consider here. Accuracy always depends on the time step, however, and of course the integration time step should be no larger than the time step required to resolve adequately the time variation of loading. In addition, the structural natural frequencies that are important for the response also limit the time step; a common rule of thumb is that the time step should not be larger than about one-tenth of the structure's smallest natural period that is significant to the response, although we prefer an even smaller time step.

Newmark's method is a single-step method, which means that we only use the information at time $i\Delta t$ to step to time $(i+1)\Delta t$. Actually, Newmark's method refers to a family of methods, but the parameters most often chosen result in the average (thus constant) acceleration member of the family. The acceleration in the time step is assumed to be the average of the accelerations at each end of the time step, that is, $(\ddot{\boldsymbol{u}}_i + \ddot{\boldsymbol{u}}_{i+1})/2$. It is then straightforward to obtain

$$\Delta\dot{\boldsymbol{u}}_i = \Delta t\ddot{\boldsymbol{u}}_i + \frac{\Delta t}{2}\Delta\ddot{\boldsymbol{u}}_i \,, \tag{37.41a}$$

$$\Delta\boldsymbol{u}_i = \Delta t\dot{\boldsymbol{u}}_i + \frac{1}{2}(\Delta t)^2\ddot{\boldsymbol{u}}_i + \frac{1}{4}(\Delta t)^2\Delta\ddot{\boldsymbol{u}}_i \,, \tag{37.41b}$$

in which, for example, $\Delta\boldsymbol{u}_i = \boldsymbol{u}_{i+1} - \boldsymbol{u}_i$.

The equation of motion (37.35) written at time step $i+1$ in terms of the increments in displacement, velocity, and acceleration is

$$\begin{aligned}&\mathbf{M}_\mathrm{s}\Delta\ddot{\boldsymbol{u}}_i + \mathbf{C}_\mathrm{s}\Delta\dot{\boldsymbol{u}}_i + \mathbf{K}_\mathrm{T}\Delta\boldsymbol{u}_i \\ &\quad = \boldsymbol{P}_{i+1} - \mathbf{M}_\mathrm{s}\ddot{\boldsymbol{u}}_i - \mathbf{C}_\mathrm{s}\dot{\boldsymbol{u}}_i - (\boldsymbol{F}_\mathrm{r})_i \,,\end{aligned} \tag{37.42}$$

in which it has been assumed for simplicity that the mass and damping are constant while the force–displacement relationship of the structure is represented by $\boldsymbol{F}_\mathrm{r}(\boldsymbol{u})$ and may be nonlinear. The first-order expansion is represented here, such that $(\boldsymbol{F}_\mathrm{r})_{i+1} \approx (\boldsymbol{F}_\mathrm{r})_i + \mathbf{K}_\mathrm{T}\Delta\boldsymbol{u}_i$, in which $\mathbf{K}_\mathrm{T}$ is the tangent stiffness matrix. For a linear system, $\mathbf{K}_\mathrm{T} = \mathbf{K}_\mathrm{s}$ and $(\boldsymbol{F}_\mathrm{r})_i = \mathbf{K}_\mathrm{s}\boldsymbol{u}_i$. However, for a linear system it is not necessary to write the equation in incremental form.

The displacement, velocity, and acceleration at time i are known, and the only unknowns are the increments. Equations (37.41) can be used to write $\Delta\dot{\boldsymbol{u}}_i$ and $\Delta\ddot{\boldsymbol{u}}_i$ in terms of $\Delta\boldsymbol{u}_i$. These relations can then be used in (37.42) to obtain an equation in $\Delta\boldsymbol{u}_i$ only. This is the typical strategy used in the time integration of the equations of motion. Of course, if (37.42) is nonlinear, iteration within a time step may be necessary to control errors.

Newmark's method as outlined above can be modified to handle the convolution integral as in (37.26). See, for example, [37.47].

It should be noted that the form of Newmark's method described above is equivalent to the trapezoidal method for numerical integration of an ordinary differential equation. It, therefore, has the accuracy and convergence characteristics of that method, as well as the same lack of numerical damping.

It is clear that to use (37.42), an explicit damping matrix $\mathbf{C}$ is required. As mentioned earlier, Rayleigh

damping is often assumed

$$\mathbf{C}_s = a_0\mathbf{M}_s + a_1\mathbf{K}_s\,, \tag{37.43}$$

in which a_0 and a_1 are coefficients. Note that this form of damping ensures that the normal modes of vibration are also orthogonal with respect to the damping matrix. With Rayleigh damping, the damping in mode n with natural frequency ω_n is given by

$$\xi_n = \frac{a_0}{2}\frac{1}{\omega_n} + \frac{a_1}{2}\omega_n\,. \tag{37.44}$$

The values of the coefficients are determined by requiring (37.44) to provide target values of modal damping for two modes important to the response.

It is obvious from (37.44) that for low natural frequencies the mass-proportional term dominates, whereas for larger natural frequencies the stiffness-proportional term dominates (subject to the values of the coefficients, of course). It should also be noted that the mass-proportional term provides damping to rigid body modes, which may not be desired for floating structures. The stiffness-proportional term does not damp rigid body modes, but the modal damping does increase as the natural frequency increases. This may not be altogether bad, however, because typically one wants the higher modes that are not important to the dynamic response to be damped out anyway so as not to affect the solution. Because of this *physical* damping, it is probably not necessary to use a numerical method that has numerical damping to accomplish the same thing [37.48].

Rayleigh damping, as well as some other forms of damping, are discussed in most structural dynamics books; see, for example, [37.10].

37.3 Linear Hydroelasticity for Inviscid Fluid Flow

37.3.1 Finite Element Structural Modeling

Structural analysis of offshore structures almost always involves the finite element method (FEM). In the displacement-based FEM, the displacements everywhere in the structure are related to the N nodal displacements through interpolation functions. Specifically

$$\boldsymbol{u}\,(x_1, x_2, x_3, t) = \mathbf{N}\boldsymbol{u}(t)\,, \tag{37.45}$$

in which the notation $x_1 - x_2 - x_3$ is used for the coordinate axes for convenience in the following. The 3×1 vector $\boldsymbol{u}\,(x_1, x_2, x_3, t)$ contains the displacements u_1, u_2, u_3 in the coordinate directions. The $3\times N$ matrix of interpolation functions $\mathbf{N}$ includes the spatial variation of the displacements, while the nodal displacements $\boldsymbol{u}(t)$ have the temporal variation. Equation (37.45) also applies to rigid body motion as well. In that case, the $\boldsymbol{u}(t)$ would be the six degrees-of-freedom surge, sway, heave, roll, pitch, and yaw, and $\mathbf{N}$ would contain the kinematic constraints that relate the six rigid body displacements to the displacement at any point in the body. In the case of flexible structures, however, the interpolation functions include deformation of the body.

Equation (37.45) provides the necessary mechanism to incorporate the effect of the fluid in the equations of motion, specifically, $\mathbf{K}_f$, $\mathbf{M}_f$, $\mathbf{C}_f$, and $\boldsymbol{F}_f$, the matrix equivalents of the hydrostatic stiffness, added mass, hydrodynamic damping, and wave exciting force in (37.35).

37.3.2 Hydrostatic Stiffness

The hydrostatic stiffness in rigid body hydrodynamics has two components. One component results from the change in hydrostatic pressure acting on the wetted surface as well as the change in orientation of the wetted surface as a result of rotation. The second component comes from a shifting of the gravitational loads in the structure as a result of rotation. The same is true for flexible structures, although the formulation is more complex.

Several formulations for the hydrostatic stiffness have been proposed, but at this point it appears that the most accurate formulation for the i, j element of the hydrostatic stiffness matrix is [37.49]

$$K_{fij} = -\rho g \int_{S_o} N_q^i \left(N_3^j + x_3 N_{l,l}^j\right) n_q \mathrm{d}S + \rho g \int_{S_o} x_3 N_l^i N_{q,l}^j n_q \mathrm{d}S + \int_{\Omega_s} \sigma_{lm} N_{q,l}^i N_{q,m}^j \mathrm{d}\Omega\,, \tag{37.46}$$

in which ρ is the mass density of water, g is gravitational acceleration, S_o is the wetted surface, N_q^i is the interpolation function for the displacement in the q-th direction for nodal degree-of-freedom i, n_q is the q-th component of the outward directed unit normal to the wetted surface, $()_{,q} \stackrel{\text{def}}{=} \partial()/\partial x_q$; Ω_s is the structural volume, σ_{lm} is the structural stress under gravitational

loads in calm seas, repeated indices imply summation, q, l, m range from 1 to 3, i, j range from 1 to N, and it is assumed that the origin of the coordinate system is at the still water level and x_3 is directed vertically up. The last integral in (37.46) is the well-known geometric stiffness matrix [37.46]. This term is an integral component of the hydrostatic stiffness matrix for flexible structures, as it contains the effect of the shifting of the gravitational loads as a result of rotation. Although including the geometric stiffness in $\mathbf{K}_\mathrm{f}$ instead of $\mathbf{K}_\mathrm{s}$ may seem odd to a structural engineer, it is done so that (37.46) is consistent with the traditional hydrostatic stiffness of rigid bodies. (37.46) is somewhat complex, and there is some evidence that practical results can be obtained with simplified implementations [37.50].

37.3.3 Added Mass, Damping, and Exciting Forces

To obtain added mass, hydrodynamic damping, and exciting forces, it is necessary to adopt a specific fluid model. Linear hydroelasticity of offshore structures is based predominantly on an inviscid fluid model, and specifically the potential flow formulation, such as discussed in Chap. 34. The previous comments on direct coupling of fluid and structure are also valid here.

The approach to determine the added mass matrix $\mathbf{M}_\mathrm{f}$, the hydrodynamic damping matrix $\mathbf{C}_\mathrm{f}$, and the wave exciting force vector $\boldsymbol{F}_\mathrm{f}$ is the same as is used in Chap. 34 to find the corresponding quantities for a rigid body. For each nodal displacement j that causes a displacement of the wetted surface $\phi_{\mathrm{R}j}$, the radiation velocity potential for a unit amplitude motion of displacement j, must be determined for wave frequency ω_q. The only difference is that the body-boundary displacements used to determine the radiation potentials are based on the modes of motion that include deformation, not just rigid body hydrodynamics. The i, j terms of $\mathbf{M}_\mathrm{f}$ and $\mathbf{C}_\mathrm{f}$ are, respectively

$$\mu_{ij} = \rho \mathrm{Re} \int_{S_o} \phi_{\mathrm{R}j} n_i^* \, \mathrm{d}S \,, \tag{37.47}$$

$$\lambda_{ij} = -\rho \omega_q Im \int_{S_o} \phi_{\mathrm{R}j} n_i^* \, \mathrm{d}S \,, \tag{37.48}$$

in which n_i^* is the *generalized normal*, the normal displacement of the wetted surface as a result of a unit nodal displacement of degree-of-freedom i: $n_i^* = \boldsymbol{n} \cdot \boldsymbol{u}_i(x_1, x_2, x_3)$. As mentioned above, $\boldsymbol{n}$ is the outward-directed unit normal of the wetted surface. $\boldsymbol{u}_i(x_1, x_2, x_3)$ is obtained from (37.45) when displacement i is 1 and all other nodal displacements are 0. Similarly, the j component of $\boldsymbol{F}_\mathrm{f}$ is

$$F_{\mathrm{f}j} = \mathrm{i}\rho\omega_q \int_{S_o} (\phi_\mathrm{I} + \phi_\mathrm{D}) n_j^* \, \mathrm{d}S \,, \tag{37.49}$$

in which ϕ_I and ϕ_D are the same incident and diffraction potentials as for the rigid body and $\mathrm{i} = \sqrt{-1}$. Application of these equations can be found in many references, e.g., [37.51].

The equation of motion in the frequency domain for wave frequency ω_q then becomes

$$\left[-\omega_q^2 \left(\mathbf{M}_\mathrm{s} + \mathbf{M}_\mathrm{f}\right) + \mathrm{i}\left(\omega_q \mathbf{C}_\mathrm{f} + \mathbf{C}_\mathrm{s}\right) + \left(\mathbf{K}_\mathrm{s} + \mathbf{K}_\mathrm{f}\right)\right] \boldsymbol{u}_q = \boldsymbol{F}_\mathrm{f} \,, \tag{37.50}$$

which is the MDOF counterpart to (37.25). $\mathbf{C}_\mathrm{s}$ is the structural damping matrix. As mentioned, in a frequency domain formulation, this will usually be hysteretic damping, but (37.50) admits viscous damping as well (with ω_q folded into the matrix).

37.3.4 Reduced Basis Solution

Equation (37.50) in practice does not present an efficient, and in many cases a tractable, solution scheme. The structural model often involves tens of thousands, and even hundreds of thousands, of degrees-of-freedom. These size problems are tractable, even for common computing, because the structural matrices are sparse. There are two problems with applying the above formulation for hydroelastic analysis. First, the radiation potential for each displacement degree-of-freedom that causes displacement of the wetted surface is required. Second, the matrices $\mathbf{M}_\mathrm{f}$ and $\mathbf{C}_\mathrm{f}$ are full. As a result, the computational requirements could be enormous. The typical solution to this problem is to use the reduced basis strategy of modal superposition. The wave response of a structure will be dominated by the lower natural frequencies, both because of the frequency content of ocean waves and the spatial variation of the loading. Hence, we can choose a reduced basis of $m \ll N$ dry modes and form the $N \times m$ modal matrix $\boldsymbol{\Psi}$, in which column j is $\boldsymbol{\psi}_j$. The displacement is then approximated as

$$\boldsymbol{u}_q \approx \boldsymbol{\Psi} \mathbf{Y}_q \,, \tag{37.51}$$

in which $\mathbf{Y}_q$ is the vector of modal displacements Y_j for wave frequency ω_q. Substitution of (37.51) into (37.50) and premultiplying by $\boldsymbol{\Psi}^\mathrm{T}$ results in

$$\left[-\omega_q^2 \left(\mathbf{M}_\mathrm{s}^* + \mathbf{M}_\mathrm{f}^*\right) + \mathrm{i}\left(\omega_q \mathbf{C}_\mathrm{f}^* + \mathbf{C}_\mathrm{s}^*\right) + \left(\mathbf{K}_\mathrm{s}^* + \mathbf{K}_\mathrm{f}^*\right)\right] \boldsymbol{Y}_q = \boldsymbol{F}_\mathrm{f}^* \,, \tag{37.52}$$

in which $\mathbf{M}_\mathrm{s}^* = \boldsymbol{\Psi}^\mathrm{T} \mathbf{M}_\mathrm{s} \boldsymbol{\Psi}$ is the identity matrix and $\mathbf{K}_\mathrm{s}^* = \boldsymbol{\Psi}^\mathrm{T} \mathbf{K}_\mathrm{s} \boldsymbol{\Psi}$ is a diagonal matrix with the square of

the natural frequencies on the diagonal. $\mathbf{C}_s^* = \boldsymbol{\Psi}^T \mathbf{C}_s \boldsymbol{\Psi}$, but it may not be necessary to form $\mathbf{C}_s$ directly. If hysteretic damping is used, and ξ is constant such that $\mathbf{C}_s = 2\xi \mathbf{K}_s$, then $\mathbf{C}_s^*$ is a diagonal matrix with $2\xi\omega_j^2$ in diagonal j. A similar result can be obtained for viscous damping. $\mathbf{K}_f^* = \boldsymbol{\Psi}^T \mathbf{K}_f \boldsymbol{\Psi}$, which is not diagonal but is a much smaller matrix than $\mathbf{K}_f$.

The modal matrices $\mathbf{M}_f^*$, $\mathbf{C}_f^*$, $\boldsymbol{F}_f^*$ are formed directly by obtaining the radiation potential ϕ_{Rj} for mode shape $\boldsymbol{\psi}_j$, and using the generalized normal for mode shape $\boldsymbol{\psi}_i$ in (37.47)–(37.49). In this way, only $m \ll N$ radiation potentials need be solved.

The above assumes that $\boldsymbol{\psi}_j$ is obtained from structural mass and stiffness properties only using (37.37). If the sum of the structural and fluid stiffness matrices $(\mathbf{K}_s + \mathbf{K}_f)$ is used instead of just the structural stiffness matrix $\mathbf{K}_s$ to obtain the natural frequencies and mode shapes, then the sum $(\mathbf{K}_s^* + \mathbf{K}_f^*)$ is a diagonal matrix with the square of the natural frequencies. There is no computational advantage to doing this, however.

It is not practical to use $\mathbf{M}_f$ in (37.37) to obtain the wet natural frequencies and mode shapes. However, they can be approximated by solving the homogeneous, undamped form of (37.52). As discussed previously for SDOF systems, this involves an iterative procedure because of the frequency dependence of the added mass. Some researchers have then used the wet modes for the reduced basis. However, in this approach the wet modes are simply linear combinations of the dry modes. Therefore, if the same number of dry and wet modes are used, the calculated response should be the same, everything else being equal.

37.3.5 Mapping from Structural Mesh to Fluid Mesh

To solve the coupled fluid–structural dynamics problem, there will be a finite element mesh of the structure and a panel mesh of the fluid–structure interface. To evaluate the radiation potentials and the integrals in (37.47)–(37.49), the displacement on the wetted surface from the structural model must be mapped to the fluid panels, typically either the panel nodes or the panel centers. The fluid mesh requires a much higher fidelity to the geometry of the wetted surface than does the structural model. For example, depending on the application a reasonable model of a TLP might involve beam elements for the columns and pontoons, whereas the panel mesh would follow the wetted surface. Alternatively, a very fine structural mesh may be required to resolve stress concentrations. Unfortunately, there exists neither an easy nor unique mapping strategy to adopt. Significant research on this problem has been carried out not just in hydroelasticity, but also in aeroelasticity. Nonlinear hydroelasticity also requires the pressures from the fluid panels to be mapped back to the structural mesh.

The simplest solution is to use a structural mesh that conforms to the fluid panel mesh, such that the mapping between structural elements and fluid panels is one-to-one. This will usually require shell finite elements for the structure. While such a model may be more detailed and complex than would normally be adopted for the structure under other circumstances, it may be worth the effort to simplify the mapping. It is not always possible for practical applications, however.

37.3.6 Linear Response to Regular Waves – RAOs

As mentioned previously, the linear response for periodic waves is almost always found in the frequency domain. That is, (37.52) is solved for N_w wave frequencies to obtain the complex $m \times N_w$ matrix $\mathbf{Y}$, the q-th column of which is $\boldsymbol{Y}_q$. Note that because $\boldsymbol{F}_f$ and $\boldsymbol{F}_f^*$ depend on the wave direction, this process must be done for each wave direction of interest. This is typically done for a unit wave amplitude, and $\mathbf{Y}$ is then the matrix of transfer functions for the modal displacements.

Because (37.52) is a relatively small set of equations, the computational effort to solve it is not significant. The computational demand of the solution is dominated first by the solution of the velocity potentials. The m radiation velocity potentials must be determined for each wave frequency, but the diffraction potential must be solved for each wave frequency and wave direction. The solution of the structural eigenvalue problem can also be significant, but it must be done only once.

The transfer function for each nodal displacement is given by

$$\mathbf{H}_u = \boldsymbol{\Psi}\mathbf{Y}\,, \tag{37.53}$$

in which $\mathbf{H}_u$ is an $N \times N_w$ complex matrix, the q-th row of which is the transfer function for nodal displacement q. The magnitude (absolute value) of the elements of $\mathbf{H}_u$ result in the RAOs. For every response quantity of interest, such as displacement, force, or stress at a specific point, the RAO must be determined explicitly, similarly to (37.53). For example, let $\mathbf{H}_\tau$ be a $p \times N_w$ matrix of transfer functions for p response quantities, which might be a mixture of displacements, forces and stresses. Then

$$\mathbf{H}_\tau = \boldsymbol{\Psi}_\tau \mathbf{Y}\,, \tag{37.54}$$

in which $\boldsymbol{\Psi}_\tau$ is a $p \times m$ real matrix, component i,j of which is the value of response quantity i in mode shape $\boldsymbol{\psi}_j$.

For an application involving different formulations of hydroelasticity, see [37.51].

37.3.7 Time-Domain Response

Time-domain calculation of the response in linear hydroelasticity is not very common. The discussion in Sect. 37.1.7 regarding the transform of the frequency-dependent coefficients to the time domain for solution therein is equally valid for MDOF structures as SDOF structures. Now, everything has to be done on a matrix basis, rather than the scalars indicated earlier. In fact, this is the most common case, because even for rigid bodies multiple degrees-of-freedom are almost always considered. For larger, flexible structures, the computations become more extensive. For wave-induced response, the frequency domain approach is more efficient. However, for some class of loading, such as impulsive and short duration transient loading, the frequency domain approach has its own limitations. Although the authors are not aware of a lot of work done in this area for linear hydroelasticity, there are certainly some applications. For example, *Kashiwagi* [37.52] considered the transient response of a floating runway during aircraft take-off and landing.

37.4 Linear Response to Random Seas

The linear response of offshore structures to random seas is presented in this section. First, the characteristics of long-crested random seas are described in Sect. 37.4.1 and extended to include shorted-crested random seas in Sect. 37.4.2. Computation of the response of MDOF systems to random seas is then described in Sect. 37.4.3, and finally estimation of extreme response and mean time to fatigue failure are presented in Sect. 37.4.4.

37.4.1 Specification of Long-Crested Random Seas

The sea surface often shows a random behavior composed of waves with a large variety of wave lengths moving independently in arbitrary directions. A first step to capture this randomness is to represent waves of different lengths, but all moving in the same direction. To characterize such random behavior we introduce the concept of the *significant wave*, which is represented by the significant wave height, H_s or $H_{1/3}$, and the significant wave period, T_s or $T_{1/3}$. These are defined as the average wave height and average wave period, respectively, of the highest one-third of all waves that are observed over an appropriate time interval. In fact, the significant wave is found to be close to the wave reported by sight observation. The significant wave is obtained from wave-by-wave analysis, which divides the random wave into individual waves by means of a zero-up crossing or zero-down crossing method. The significant wave is convenient to characterize the scale of the wave field because it expresses the wave field with a single wave parameter, but it does not provide information on the wave-height scatter or the corresponding wave energy distribution.

To represent the frequency distribution of wave energy in random seas, a frequency spectrum is used. The frequency spectra can be obtained from observational wave data by means of a Fourier transform. Many design spectra have been developed to describe ocean waves [37.17, 44]. They are generally based on one or more parameters, e.g., significant wave height, wave period, and peak frequency. Of those spectra, the Pierson–Moskowitz, Bretschneider, and JONSWAP spectra are perhaps the most commonly used at present. These spectra are defined for deep water.

The Pierson–Moskowitz (P–M) spectrum is defined as

$$S(\omega) = 0.0081 g^2 \omega^{-5} \exp\left[-1.25\left(\frac{\omega}{\omega_0}\right)^{-4}\right], \tag{37.55}$$

where ω denotes the spectral peak.

The Bretschneider spectrum is given in terms of the significant wave height H_s and the significant wave frequency ω_s

$$S(\omega) = 0.1687 H_s^2 \omega_s^4 \omega^{-5} \exp\left[-0.675\left(\frac{\omega}{\omega_s}\right)^{-4}\right]. \tag{37.56}$$

It can be applied to both fully developed and partially developed wave fields.

The JONSWAP spectrum (derived from the Joint North Sea Wave Project) is given as

$$S(\omega) = 0.0081 g^2 \omega^{-5} \exp\left[-1.25\left(\frac{\omega_p}{\omega}\right)^4\right] \times \gamma^{\exp\left[\frac{(\omega-\omega_p)^2}{2\tau^2\omega_p^2}\right]}, \tag{37.57}$$

where γ represents the peakedness parameter or the ratio of the maximum spectral density to that of the corresponding P–M spectrum. The values of ω_p denote the peak frequency at which $S(\omega)$ is a maximum, and τ denotes the shape parameter. The quantities of the parameters are given by

$$\tau = \begin{cases} 0.07 & \text{for } \omega \le \omega_p \,, \\ 0.09 & \text{for } \omega > \omega_p \,. \end{cases} \tag{37.58}$$

The value of γ generally varies from 1 to 7; a frequently used value is 3.3.

To obtain the stochastic response in the frequency domain, the wave spectrum is all that is required to define the wave field. However, to simulate the structural response in the time domain, time series of the wave fields of interest may be reconstructed from a wave spectrum, such as the one shown in Fig. 37.5. We assume that the random waves observed at a spatial point are composed of a large number, N_w, of sinusoidal waves

$$\eta(t) = \sum_{q=1}^{N_w} A_q \cos\left(\omega_q t + \varepsilon_q\right) \,, \tag{37.59}$$

in which A_q is the amplitude $(H_q/2)$, ε_q is the (random) phase, and ω_q is the frequency of the q-th wave component. The wave spectrum is divided into N_w frequency bins. One such bin is the shaded area in Fig. 37.5. The $\Delta\omega$ may be chosen either uniformly or such that the energy contained in each bin (the shaded area) is constant. For the q-th bin, ω_q is the average frequency in the bin. The corresponding (single-sided) wave amplitude is given by

$$A_q = \sqrt{2S\left(\omega_q\right)\Delta\omega_q} \,. \tag{37.60}$$

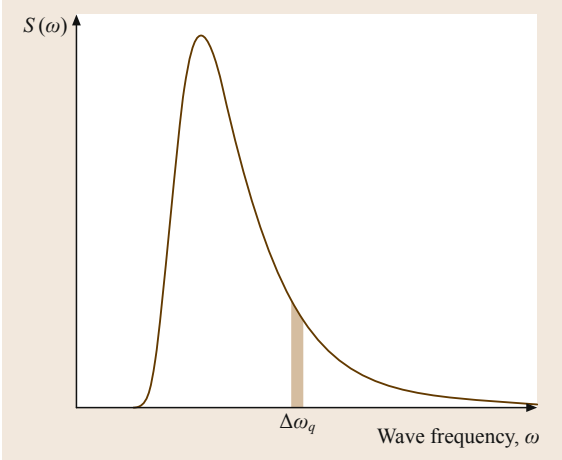


Fig. 37.5 Sample wave spectrum

A phase angle ε_q that is uniformly distributed in the range of $(0, 2\pi)$ is selected randomly. The wave number k_q is calculated from the frequency, and the wave profile is computed from

$$\eta(x,t) = \sum_{q=1}^{N} A_q \cos\left(k_q x - \omega_q t + \varepsilon_q\right) \tag{37.61}$$

to define a long-crested wave field.

37.4.2 Specification of Short-Crested, Directional Random Seas

In actual seas, the wave profile varies in both spatial directions, such that the actual wave crests are *short*. A means to characterize these fluctuations is through a stochastic description of the wave profile as

$$\eta(x,y,t) = \sum_{q=1}^{\infty} A_q \cos\left[\left(k_q \cos\theta_q\right) x + \left(k_q \sin\theta_q\right) y - \omega_q t + \varepsilon_q\right] \,, \tag{37.62}$$

where θ_n represents the wave incident angle. Then a two-dimensional (2-D) directional spectrum is defined as

$$S(\omega,\theta)\,\Delta\omega\Delta\theta = \sum_{\omega}^{\omega+\Delta\omega} \sum_{\theta}^{\theta+\Delta\theta} \frac{1}{2} A_q^2 \,. \tag{37.63}$$

The integral of a two-dimensional directional spectrum from 0 to 2π consistent with a frequency spectrum $S(\omega)$ is defined as

$$S(\omega) = \int_0^{2\pi} S(\omega,\theta)\mathrm{d}\theta \,. \tag{37.64}$$

It is convenient to separate a frequency spectrum for engineering use, thus

$$S(\omega,\theta) = S(\omega)G(\omega,\theta) \,. \tag{37.65}$$

$G(\omega)$ is called a directional spreading function with

$$\int_0^{2\pi} G(\omega,\theta)\mathrm{d}\theta = 1 \,. \tag{37.66}$$

Various expressions for $G(\omega,\theta)$ have been proposed [37.15]. $G(\omega,\theta)$ is, in general, a function of

ω, but the simplest expression considers G to be independent of ω. One possibility is the cosine-squared function

$$G(\theta) = \begin{cases} \frac{2}{\pi}\cos^2\theta & \text{for } |\theta| < \frac{\pi}{2}\,, \\ 0 & \text{otherwise}\,, \end{cases} \tag{37.67}$$

where θ is referred to as the base direction.

The wave profile is computed from

$$\eta(x,y,t) = \sum_{q=1}^{N_w}\sum_{m=1}^{M} A_{q,m}\cos\left[(k_q\cos\theta_m)x + (k_q\sin\theta_m)y - \omega_q t + \varepsilon_{q,m}\right]\,. \tag{37.68}$$

The amplitude of a wave at a particular frequency and direction is given by

$$A_{q,m} = \sqrt{2S\left(\omega_q,\theta_m\right)\Delta\omega\Delta\theta}\,. \tag{37.69}$$

Phase angles $\varepsilon_{q,m}$ are typically chosen to be uniformly distributed in the range of $(0, 2\pi)$, but sometimes they are assumed to be constant for a given frequency, ε_q.

37.4.3 Response in Random Seas

The procedure to compute the response of an offshore structure to random seas is the same, whether it is for long-crested (unidirectional) seas or short-crested (multidirectional) seas. In the case of linear behavior, the response to wave excitation composed of the sum of a number of frequencies is the sum of the response to the individual frequencies. Thus the response to each wave frequency can be solved independently. That is, for each excitation frequency ω_q, we first determine the response from (37.52)

$$\left[-\omega_q^2\left(\mathbf{M}_s^* + \mathbf{M}_f^*\right) + \mathrm{i}\left(\omega_q\mathbf{C}_f^* + \mathbf{C}_s^*\right) + \left(\mathbf{K}_s^* + \mathbf{K}_f^*\right)\right]\boldsymbol{Y}_q = \boldsymbol{F}_f^*\,. \tag{37.70}$$

Note that $\mathbf{M}_f^*$, $\mathbf{C}_f^*$ and $\boldsymbol{F}_f^*$ are functions of the wave frequency, and $\boldsymbol{F}_f^*$ is also a function of wave direction. If there is no directional spreading, then $M = 1$ in (37.68), and the equation corresponding to this value of ω_q needs to be solved only once. Otherwise, the same equation is solved for each spreading angle θ_m for $m = 1,\ldots M$, and the responses for frequency ω_q for each spreading angle are then summed. This process needs to be repeated for each excitation frequency q for $q = 1,\ldots,N_w$, and the total response to random seas is thus the sum of all the responses to each excitation frequency including all spreading angles θ_m for $m = 1,\ldots M$.

In some situations, it is not necessary to know the full time history of a particular response quantity. Rather, a measure of the stochastic response suffices. For response quantity τ, for any combination of wave frequency q and wave angle θ_j, the corresponding RAO for τ_{qj} can be obtained from (37.54), and the spectral value of the response can be computed by

$$S_{\tau_{qj}}\left(\omega_q,\theta_j\right) = |H_{\tau_{qj}}|^2 S\left(\omega_q,\theta_j\right)\,, \tag{37.71}$$

where $S\left(\omega_q,\theta_j\right)$ is the directional wave spectral density at ω_q and θ_j. The corresponding variance of the response τ_j for the wave angle θ_j, $\sigma_{\tau_j}^2$, is the area under the spectrum, $S_{\tau_{qj}}\left(\omega_q,\theta_j\right)$, of the response quantity τ_j. Assuming that the response of quantities τ_j to each wave angle θ_j are pairwise mutually independent, then the variance σ_τ^2 of response quantity τ is given by

$$\sigma_\tau^2 = \sum_1^{N_A}\sigma_{\tau_j}^2\,, \tag{37.72}$$

where N_A is the number of wave angles. This is a conservative assumption since usually there is a positive but weak correlation between adjacent wave angles.

37.4.4 Estimation of Extreme and Fatigue Response in Random Seas

The stochastic characteristics of the response of a MDOF structure to directional random seas including extreme and fatigue responses can be estimated when the corresponding characteristics for each individual response quantity τ are known using some stochastic combinational analyses such as (37.70) above. For this reason, we will focus on estimation of stochastic properties of SDOF systems in this section.

For random waves describable by linear wave theory, it is often assumed that the wave excitation is Gaussian with the characteristics described in Sects. 37.4.1–37.4.2. It is assumed that the transfer function for the response quantity τ_k has been obtained as outlined above. Then the statistics for any particular response quantity, e.g., the mean m_{τ_k}, $m_{\dot\tau_k}$ and standard deviation σ_{τ_k}, $\sigma_{\dot\tau_k}$, of the k-th response quantity τ, i. e., τ_k, respectively, and the cross-correlation coefficient between the displacement and velocity, $\rho_{\tau_k\dot\tau_k}$, can be computed. If the response quantities τ_k are defined as deviations from the mean position of the structure in random seas, then the mean displacement and velocity are zero [37.53], i. e., $m_{\tau_k} = 0$, and $m_{\dot\tau_k} = 0$. The standard deviation of the response quantity, nodal displacement, τ_k, and ve-

locity, $\dot{\tau}_k$ can be computed as

$$\sigma_{\tau_k} = \sqrt{\int_0^\infty S_{\tau_k}(\omega)\mathrm{d}\omega}\,, \tag{37.73}$$

where $S_{\tau_k}(\omega)$ is the single-sided spectral density of response quantities τ_k. Note that $S_{\tau_k}(\omega)$ may be obtained from (37.70). Similarly, the standard deviation of the nodal velocity $\dot{\tau}_k$ is given by

$$\sigma_{\dot{\tau}_k} = \sqrt{\int_0^\infty \omega^2 S_{\dot{\tau}_k}(\omega)\mathrm{d}\omega}\,. \tag{37.74}$$

Because displacement and velocity of a Gaussian dynamical quantity are statistically independent, the cross-correlation coefficient, $\rho_{\tau_k\dot{\tau}_k}$, is zero. The terms displacement and velocity were used above to refer to the response quantity τ_k, and its time rate of change, $\dot{\tau}_k$. However, τ_k may not be displacement but, e.g., stress. Regardless, the time rate of change is needed to compute positive crossing rates for extreme and fatigue failure computations, as shown below.

The probability density function $p(\tau_k)$ of the response quantity τ_k and the joint probability density function $p(\tau_k, \dot{\tau}_k)$ of the response quantity τ_k and velocity $\dot{\tau}_k$, respectively, are given by

$$p(\tau_k) = \frac{1}{\sqrt{2\pi}\sigma_{\tau_k}} \mathrm{e}^{-\frac{\tau_k^2}{2\sigma_{\tau_k}^2}}\,, \tag{37.75}$$

$$p(\tau_k, \dot{\tau}_k) = \frac{1}{2\pi\sigma_{\tau_k}\sigma_{\dot{\tau}_k}} \mathrm{e}^{-\frac{1}{2}\left(\frac{\tau_k^2}{\sigma_{\tau_k}^2} + \frac{\dot{\tau}_k^2}{\sigma_{\dot{\tau}_k}^2}\right)}\,. \tag{37.76}$$

The positive crossing rate (or crossing frequency), ν_{a}^+, at a level a_k of the random variable τ_k can be computed from the joint probability density function, and for a Gaussian process $\tau_k(t)$, the positive crossing rate is given by [37.53]

$$\nu_{\mathrm{a}k}^+ = \frac{\sigma_{\dot{\tau}_k}}{2\pi\sigma_{\tau_k}} \mathrm{e}^{-\frac{1}{2}\frac{a_k^2}{\sigma_{\tau_k}^2}}\,. \tag{37.77}$$

The probability distribution of the peaks of the response quantity τ_k is Rayleigh,

$$p_{\mathrm{p}}(a_k) = \frac{a_k}{\sigma_{\tau_k}^2} \mathrm{e}^{-\frac{a_k^2}{2\sigma_{\tau_k}^2}}\,, \quad 0 \le a \le \infty\,. \tag{37.78}$$

The expected time to failure (or mean life time to failure), T_a, or for a Gaussian random process $\tau_k(t)$ to exceed an extreme value a_k, is given by

$$T_a = \frac{1}{\nu_{\mathrm{a}k}^+ \int_0^\infty \left(\frac{1}{N(S)}\right) p_{\mathrm{p}}(S)\mathrm{d}S}\,, \tag{37.79}$$

where $N(S)$ is the S-N curve for the material under consideration [37.53]. Correspondingly, the probability of failure P_{f} due to the crossing of an extreme value a_k, for a Gaussian process τ_k discussed above is given by [37.53]

$$P_{\mathrm{f}}(T_{a_k}) = 1 - \mathrm{e}^{-\nu_{a_k}^+ T_{a_k}}\,. \tag{37.80}$$

As mentioned earlier, once the probabilistic characteristics of the response quantities τ_k are determined, the global response probabilistic characteristics of a MDOF system can be determined using a stochastic combination procedure.

37.5 Nonlinear Hydroelasticity (Nonlinear FSI)

The discussions in the previous sections involve only small amplitude wind-wave or earthquake excitations and small amplitude responses, and linear superposition applies. For large wave excitations, the waves will need to be described by a higher-order nonlinear theory. The response to each excitation frequency ω_q may also become nonlinear, or the amplitude of the total response may become too large, and the response among different frequencies ω_l and ω_k, $l \neq k$, may be coupled. In this case, the principle of superposition may cease to apply and some form of nonlinear analysis is required. For small nonlinearity, the perturbation method that is used to derive the linear equations of motion can again be used by including second-order terms in the expansion series of both the waves and the structure responses, thus including sum and difference frequency terms in the wave loads and the structural responses. However, for vary large wave excitations, hurricanes, tsunamis, and storm surges, the nonlinearity may become so large that the perturbation method, if it remains applicable, will involve too many terms in the expansion to be practical, and a direct coupled fluid-dynamics/structural mechanics computational approach will have to be used.

37.5.1 Nonlinear Frequency Domain Analysis

For large amplitude or shallow water waves, the linear theory of sinusoidal waves may no longer be adequate to describe the wave field, and a nonlinear wave theory that takes into account the steeper and narrower peaks, and wider and shallower troughs of the physical waves will have to be employed [37.18]. The nonlinear wave phenomenon (and the associated incident wave potential) can be described by higher-order perturbation wave theory. Correspondingly, theories of equal order perturbation have to be developed for the associated diffraction and radiation potentials for consistent and accurate treatment of the structural response computation. A second-order theory is briefly described in this section to demonstrate the gist of nonlinear perturbation expansion formulation and solution of the wave–structure interaction problem.

Higher Order Wave Description

General theories of nonlinear waves, including Stokes higher-order theory, Cnoidal theory, and stream functions, may be found in [37.17, 54–56]. For gravity waves described by potential flow, the incident wave potential can be expressed in a perturbation form as

$$\phi_{\mathrm{I}}(x,y,x;t) = \varepsilon^1\phi_{\mathrm{I}}^1(x,y,x;t) + \varepsilon^2\phi_{\mathrm{I}}^2 + \varepsilon^3\phi_{\mathrm{I}}^3 + \cdots , \tag{37.81}$$

where ε is the perturbation expansion parameter, and ϕ_{I}^j is the j-th order incident wave potential.

For Stokes second-order wave theory, the horizontal water particle velocity can be expressed as

$$u(x,t) = \frac{\pi H}{T}\frac{\cosh ky}{\sinh kd}\cos(kx-\omega t) + \frac{3}{4c}\left(\frac{\pi H}{T}\right)^2\frac{\cosh 2ky}{\sin h^4 kd}\cos 2(kx-\omega t) , \tag{37.82}$$

where H, T, k, ω, and $c = L/T$ are the first-order wave height, wave period, wave number, wave frequency, and celerity, respectively, with L being the first-order wave length, and d is the water depth. Note that the first term, which corresponds to that of linear wave theory, is proportional to $\frac{\pi H}{T}$, while the second term, which is proportional to $\left(\frac{\pi H}{T}\right)^2$, is a second-order correction term, and $\frac{\pi H}{T}$ can be considered the perturbation parameter. Similarly, the vertical water particle velocity is expressed as

$$v(x,t) = \frac{\pi H}{T}\frac{\sinh ky}{\sinh kd}\sin(kx-\omega t) + \frac{3}{4c}\left(\frac{\pi H}{T}\right)^2\frac{\sinh 2ky}{\sinh^4 kd}\sin 2(kx-\omega t) . \tag{37.83}$$

To determine the structural response to Stokes second-order wave excitations, these expressions would be used for wave load computations.

Higher Order Wave Load and Large-Body Structural Response Computation

Corresponding to incident wave perturbation expansion, the diffracted wave and radiation wave potentials can also be described as

$$\phi_{\mathrm{D}}(x,y,x;t) = \varepsilon^1\phi_{\mathrm{D}}^1(x,y,x;t) + \varepsilon^2\phi_{\mathrm{D}}^2 + \varepsilon^3\phi_{\mathrm{D}}^3 + \cdots , \tag{37.84}$$

and

$$\phi_{\mathrm{R}}(x,y,x;t) = \varepsilon^1\phi_{\mathrm{R}}^1(x,y,x;t) + \varepsilon^2\phi_{\mathrm{R}}^2 + \varepsilon^3\phi_{\mathrm{R}}^3 + \cdots . \tag{37.85}$$

The formulation for the second-order problem is similar to that of the first-order (linear) problem presented in the previous sections in this chapter. However, there are several major differences in the derivation of the second-order transfer functions and computation of the second-order responses that the readers need to be aware of.

First, the transfer functions (see $\mathbf{H}_\tau$ in (37.54)) of the second-order responses are functions of frequency pairs (ω_i, ω_j), instead of just a single frequency. The derivation of these second-order functions are beyond the scope of this chapter. Interested readers may want to consult the literature [37.18].

Second, it is important that in formulating the second-order load and response computations, second-order effects originating from first-order quantities have to be included. It is well known that (second-order) drift forces are caused by first-order (linear) waves, and play an important role in the low frequency response of compliant structures.

Third, other nonlinear effects that may cause second-order responses include large displacement and rotation (geometric effects) and need to be included in the perturbation expansion of the equations of structural motion.

As one can observe, computation of second-order response in the frequency domain can be quite complex. The study of stochastic response and extreme load

and motion, and fatigue failure predictions are active research topics.

Stochastic Higher Order Wave Load and Small-Body Structural Response Computation

An investigation of a second-order stochastic method for representing the second-order Morison hydrodynamic drag forces on a SDOF marine structure subjected to random waves showed that the method may be used to obtain good statistical approximation of the nonlinear effects via analysis in the frequency domain [37.57]. A distinct advantage of the approach is that a linear analysis of the structure in the frequency domain can be used to acquire detailed statistical information on the response to the nonlinear drag force. The resulting response statistics can then be used to obtain long-term fatigue life estimation and extreme value estimates for the response, based on the procedures described in Sect. 37.4.4.

37.5.2 Nonlinear Time-Domain Analysis

An alternative to obtain nonlinear fluid–structure interaction responses is to solve the coupled fluid–structure interaction problem as an initial boundary value problem numerically in the time domain. This method can capture the nonlinear transient and steady-state effects under arbitrary high sea states (at least theoretically) without any limit on the form of nonlinearity. However, the computational effort required can be very high (see discussion below).

Potential-Flow Field Approximation

For moderately high sea states where there is no breaking, the wave field can be assumed inviscid and irrotational, thus a potential flow model can be used to describe the hydrodynamics. The fully nonlinear hydrodynamic flow field can be solved, usually with the boundary-element method (BEM) without making any assumption on the degree of nonlinearity. In theory, the free surface can even be multi-valued, except the fluid domain has to be simply connected, thus no breaking and reentry/reconnection are allowed. The computational effort for fully nonlinear potential flow coupling with the structural dynamics of the large body (floating platform) and the mooring system is rather expensive. In general, some approximations on the flow field and the fluid–structure interaction problem are assumed to make the problem computationally efficient for practical industrial use.

An efficient approximation of the coupled three-dimensional (3-D) fluid–structure interaction (hybrid) time-domain solution under regular wave excitation is to keep the solution of the fluid linear (obtained in the frequency domain as shown in previous sections), but model the wave forces on the (rigid) structure in the time domain as follows

$$\mathbf{M}_s\ddot{x}(t) = \boldsymbol{F}_r(\omega,t) + \boldsymbol{F}_d(\omega,t) + \boldsymbol{F}_h(t) + \boldsymbol{F}_f(t) + \boldsymbol{F}_{vis}(t) + \boldsymbol{F}_g(t) + \boldsymbol{F}_m(t)\,, \tag{37.86}$$

where $\boldsymbol{x}(t)$ is the structural displacement vector, t is time, ω is the incident wave frequency, $\mathbf{M}_s$ is the structural mass matrix, $\boldsymbol{F}_r(\omega,t)$ is the frequency and time dependent radiation force vector, $\boldsymbol{F}_d(\omega,t)$ is the frequency and time dependent diffraction force vector, $\boldsymbol{F}_h(t)$ is the instantaneous hydrostatic force vector, $\boldsymbol{F}_f(t)$ is the instantaneous Froude–Krylov force vector, $\boldsymbol{F}_{vis}(t)$ is the viscous force vector, $\boldsymbol{F}_g(t)$ is the instantaneous gravity force vector, and $\boldsymbol{F}_m(t)$ is the instantaneous mooring force vector.

Under this formulation, the radiation and diffraction forces are calculated using linear wave theory and BEM, while the nonlinear hydrostatic, Froude–Krylov, gravity and mooring forces are calculated based on their instantaneous position. The nonlinear viscous force, which represents the effect of viscous dissipation, is often modeled in a Morison form with the drag coefficient as a parameter obtained from experimental results.

Mooring forces are usually quite nonlinear. They can be obtained using a cable dynamic model and solved numerically using the finite-element method. This approach can also be used for flexible structure models.

Navier–Stokes Flow Field Approximation

For highly nonlinear transient effects, an increasingly popular approach is to solve numerically the fluid and structural domain responses of the coupled fluid–structure interaction problem directly using high-performance computing.

One approach to include nonlinearity in the waves is to solve the potential problem directly in the time domain using a step-by-step solution procedure. The nonlinear potential flow problem can be solved using either the boundary-element method (BEM) see [37.58] or the finite-element method (FEM) see [37.59]. For computational fluid-structure interaction (FSI), both the BEM and the FEM approaches coupling with FEM structural modeling have seen success [37.60]. However, potential flow can only approximate flows that are inviscid and irrotational (mostly for diffraction problems of *large* structures). For highly nonlinear violent flow around a structure, such as wave slamming, these assumptions will no longer be valid, and viscous flow governed by the Navier–Stokes equation (NSE) will

need to be employed. When compressibility is important (wave–structure impact), the governing equations for conservation of mass and momentum of the NSE (as shown below) have to be used

$$\frac{\partial \rho}{\partial t} + \rho \nabla \cdot \boldsymbol{u} = 0\,,$$
$$\frac{\partial \boldsymbol{u}}{\partial t} + \boldsymbol{u} \cdot \nabla \boldsymbol{u} = -\frac{1}{\rho}(\nabla P + \frac{2}{3}\mu \nabla \cdot \boldsymbol{u}) + \frac{\mu}{\rho}\nabla^2 \boldsymbol{u} + \boldsymbol{g}\,, \qquad (37.87)$$

where ρ is the fluid density; $\boldsymbol{u}$ is the fluid particle velocity vector, P is the fluid pressure, μ is the bulk viscosity, and $\boldsymbol{g}$ is the body force vector due to gravity. However, under many applications the water can be assumed incompressible and the conservation of mass equation can be simplified to $\nabla \cdot \boldsymbol{u} = 0$.

The fully nonlinear wave forces on a rigid structure can be calculated based on the fluid pressure which is obtained by solving the NSE equation above.

An advantage of using a NSE fluid solver is its capability to couple with a computational structural mechanics solver, thus allowing the direct computation of fully coupled fluid–flexible structure interaction. Currently, there are a number of commercially available coupled fluid–structure interaction software applications using the FEM solution for computational structural mechanics (CSM) and a computational fluid dynamic (CFD) solver using either FEM or the finite-volume method (FVM). Some of these CSM-CFD coupling software applications are maturing and ready for standard practice. References for commercial software can be found readily on the internet. Readers who are interested in learning the details of these solvers should consult the CFD and CSM literature in general, and coupled fluid–structure interaction (FSI) literature in particular.

References

37.1 S. Timoshenko: *Vibration Problems in Engineering*, 1st edn. (Van Nostrand Reinhold, New York 1928)
37.2 S. Timoshenko, D.H. Young, W. Weaver Jr.: *Vibration Problems in Engineering*, 4th edn. (Wiley, New York 1974)
37.3 W.T. Thomson: *Mechanical Vibrations* (Prentice Hall, New York 1948)
37.4 W.T. Thomson: *Theory of Vibration with Applications*, 4th edn. (Prentice Hall, Englewood Cliffs 1993)
37.5 W.T. Thomson, M.D. Dahleh: *Theory of Vibration with Applications*, 5th edn. (Prentice Hall, Upper Saddle River 1998)
37.6 W.C. Hurty, M.F. Rubinstein: *Dynamics of Structures* (Prentice Hall, Englewood Cliffs 1964)
37.7 J.L. Humar: *Dynamics of Structures* (Prentice Hall, Englewood Cliffs 1990)
37.8 R.W. Clough, J. Penzien: *Dynamics Of Structures* (McGraw-Hill, New York 1975)
37.9 R.W. Clough, J. Penzien: *Dynamics Of Structures*, 2nd edn. (McGraw-Hill, New York 1993)
37.10 K.A. Chopra: *Dynamics of Structures, Theory and Applications to Earthquake Engineering*, 4th edn. (Prentice Hall, Upper Saddle River 2012)
37.11 J.W. Tedesco, W.G. McDougal, C.A. Ross: *Structural Dynamics: Theory and Applications* (Addison Wesley, Menlo Park 1999)
37.12 C.A. Brebbia, S. Walker: *Dynamics of Offshore Structures* (Newnes-Butterworths, London 1979)
37.13 J.P. Hooft: *Advanced Dynamics of Marine Structures* (Wiley, New York 1982)
37.14 J.F. Wilson: *Dynamics of Offshore Structures* (Wiley-Interscience, New York 1984)
37.15 T. Sarpkaya, M. Isaacson: *Mechanics of Wave Forces on Offshore Structures* (Van Nostrand Reinhold, New York 1981)
37.16 T. Sarpkaya: *Wave Forces on Offshore Structures* (Cambridge Press, Cambridge 2012)
37.17 S.K. Chakrabarti: *Hydrodynamics of Offshore Structures* (WIT Press, Southampton 1987)
37.18 S.K. Chakrabarti: *Nonlinear Methods in Offshore Engineering* (Elsevier, Amsterdam 1990)
37.19 S.K. Chakrabarti: *Offshore Structures Modeling* (World Scientific, Singapore 1994)
37.20 S.K. Chakrabarti: *Theory and Practice of Hydrodynamics and Vibration* (World Scientific, Singapore 2002)
37.21 M.F. Schlesinger, T.F. Swean: Stochastically Excited Nonlinear Ocean Structures (World Scientific, Singapore 1998)
37.22 M. Olagnon, M. Prevosto: The variations of damping ratios with sea conditions for offshore structure under natural excitation, Offshore Technol. Conf. (1984) pp. 57–66
37.23 R.E.D. Bishop, W.G. Price: *Hydroelasticity of Ships* (Cambridge Univ. Press, Cambridge 1979)
37.24 W.E. Gundy, T.D. Scharton, R.L. Thomas: Damping measurements on an offshore platform, Offshore Technol. Conf. (1980), OTC 3863-MS
37.25 J.A. Ruhl: Offshore platforms: Observed behavior and comparison with theory, Offshore Technol. Conf. (1976) pp. 333–352, OTC 2553
37.26 J.W. Cooley, J.W. Tukey: An algorithm for the machine calculation of complex Fourier series, Math. Comput. **19**, 297–301 (1965)
37.27 J.F. Hall: An FFT Algorithm for Structural Dynamics, Earthq. Eng. Struct. Dyn. **10**, 797–811 (1982)
37.28 A.S. Veletsos, C.E. Ventura: Dynamic analysis of structures by the DFT method, J. Struct. Eng. **111**, 2625–2642 (1985)
37.29 US Army Corps of Engineers: *Time-History Dynamic Analysis of Concrete Hydraulic Structures*, EM 1110-

2-6051 (US Army Corps of Engineers, Washington 2003)
37.30 W.E. Cummins: The impulse response function and ship motions, Schiffstechnik **47**(9), 101–109 (1962)
37.31 T.F. Ogilvie: Recent progress toward the understanding and prediction of ship motions, 5th Symp. Nav. Hydrodyn. (1964) pp. 3–80
37.32 J.O. de Kat, J.R. Paulling: The simulation of ship motions and capsizing in severe seas, Trans. Soc. Naval Arch. Mar. Eng. **97**, 139–168 (1989)
37.33 R. Taghipour, T. Perez, T. Moan: Hybrid frequency–time domain models for dynamic response analysis of marine structures, Ocean Eng. **35**(7), 685–705 (2008)
37.34 W. Sheng, A. Lewis: Assessment of wave energy extraction from seas: Numerical validation, J. Energy Res. Technol. **134**(4), 041701 (2012)
37.35 J. Falnes: *Ocean Waves and Oscillating systems – Linear Interaction Including Wave-Energy Extraction* (Cambridge Univ. Press, Cambridge 2002)
37.36 Z. Yu, J. Falnes: State-space modelling of a vertical cylinder in heave, Appl. Ocean Res. **17**, 265–275 (1995)
37.37 E. Kristiansen, Å. Hjulstad, O. Egeland: State-space representation of radiation forces in time-domain vessel models, Ocean Eng. **32**(17/18), 2195–2216 (2005)
37.38 T. Duarte, M. Alves, J. Jonkman, A. Sarmento: State-space realization of the wave-radiation force within FAST, ASME 32nd Int. Conf. Ocean Offshore Arct. Eng. (2013)
37.39 J.D. Nolte, R.C. Ertekin: Wave power calculations for a wave energy conversion device connected to a drogue, J. Renew. Sustain. Energy **6**(1), 013117 (2014)
37.40 K.-T. Kim, P.-S. Lee, K.C. Park: A direct coupling method for 3D hydroelastic analysis of floating structures, Int. J. Num. Methods Eng. **93**(13), 842–866 (2013)
37.41 Y. Wu: Hydroelasticity of Floating Bodies, Ph.D. Thesis (Brunel University, Uxbridge 1984)
37.42 W.G. Price, Y. Wu: Hydroelasticity of marine structures. In: *Theoretical and Applied Mechanics*, ed. by F.I. Niordson, N. Olhoff (Elsevier, Amsterdam 1985)
37.43 J.R. Morison, M.P. O'Brien, J.W. Johnson, S.A. Schaaf: The force exerted by surface waves on piles, AIME Pet. Trans. **189**, 149–157 (1950)
37.44 R.T. Hudspeth: *Waves and Wave Forces on Coastal and Ocean Structures* (World Scientific, Singapore 2006)
37.45 A.S. Veletsos, A.M. Prasad, G. Hahn: Fluid-structure interaction effects for offshore structures, Earthq. Eng. Struct. Dyn. **16**, 631–652 (1988)
37.46 R.D. Cook, D.S. Malkus, M.E. Plesha, R.J. Witt: *Concepts and Applications of Finite Element Analysis*, 4th edn. (Wiley, New York 2002)
37.47 A.M. Puthanpurayil, A.J. Carr, R.P. Dhakal: A generic time domain implementation scheme for non-classical convolution damping models, Eng. Struct. **71**, 88–98 (2014)
37.48 E.L. Wilson: *Three-Dimensional Static and Dynamic Analysis of Structures*, 3rd edn. (Computers and Structures Inc., Berkeley 2002)
37.49 H.R. Riggs: Comparison of formulations for the hydrostatic stiffness of flexible structures, J. Offshore Mech. Arct. Eng. **131**, 024501 (2009)
37.50 I. Senjanović, N. Hadžić, F. Bigot: Finite element formulation of different restoring stiffness issues in the ship hydroelastic analysis and their influence on response, Ocean Eng. **59**, 198–213 (2013)
37.51 H.R. Riggs, H. Suzuki, R.C. Ertekin, J.W. Kim, K. Iijima: Comparison of hydroelastic computer codes based on the ISSC VLFS benchmark, Ocean Eng. **35**, 589–597 (2008)
37.52 M. Kashiwagi: Transient responses of a VLFS during landing and take-off of an airplane, J. Mar. Sci. Technol. **9**(1), 14–23 (2004)
37.53 D.E. Newland: *An Introduction to Random Vibrations, Spectral and Wavelet Analysis*, 3rd edn. (Longman, Singapore 1993)
37.54 J.D. Fenton: A firth-order Stokes theory for steady waves, J. Waterw. Port Coast. Ocean Eng. **3**(2), 216–234 (1985)
37.55 R.G. Dean: Stream function representation of nonlinear ocean waves, J. Geophys. Res. **70**, 4561–4572 (1965)
37.56 R.G. Dean: Relative validities of water wave theories, J. Waterw. Harb. Coast. Eng. **96**, 105–119 (1970)
37.57 A. Naess, S.C. Yim: Stochastic response analysis of dynamically sensitive offshore structures excited by drag forces, J. Eng. Mech. **122**(5), 442–448 (1996)
37.58 S.B. Nimmala, S.C. Yim, S.T. Grilli: An efficient parallelized 3-D-FNPF numerical wave tank for large-scale wave basin experiment simulation, J. Offshore Mech. Arct. Eng. **135**(2), 021104 (2013)
37.59 G.X. Wu, Z. Hu: Numerical simulation of viscous flow around unrestrained cylinders, J. Fluids Struct. **22**, 371–390 (2006)
37.60 X.J. Chen, T. Moan, S.X. Fu, W.C. Cui: A second-order hydroelastic analysis of a floating plate in multi-directional irregular waves, Int. J. Nonlinear Mech. **41**, 1206–1218 (2006)

38. Cable Dynamics for Marine Applications

Ioannis K. Chatjigeorgiou, Spyros A. Mavrakos

It is the purpose of this chapter to provide most of the information associated with the dynamic behavior of slender cables (mooring lines), with particular emphasis on marine applications. The basics of its content are based on the personal fundamental research of the authors as regards the applications of cables in the marine environment, giving special consideration to the details of their time varying response.

The dynamics of cables have been investigated for decades, but admittedly, the specific subject attracted profound attention by the global research community with respect to the use of cables in offshore deep water applications. The interested reader can find an enormous amount of very important and thorough studies in the literature, which address relevant problems, from statics to very complicated nonlinear dynamics, using various mathematical formulations, and from simple to sophisticated numerical solution schemes.

The scope of the present effort is to associate the aspect of dynamics, linear and nonlinear, with the most typical practical applications of cables in the marine environment and especially in the offshore sector. The methodology that was chosen to perform this task relies on a global formulation for the governing dynamical systems. A generic analysis that is based on the assumption of line dynamics exploiting the fact that relevant representations may be effectively considered as two-point boundary value problems.

38.1 **Mathematical Formulation** 876
38.1.1 Derivatives in Time and Space 877
38.1.2 Dynamic Equilibrium 878
38.1.3 Final System 878
38.1.4 Reduced Model Omitting Bending Stiffness 879
38.1.5 Equivalent Linearization 880

38.2 **The Eigenvalue Problem of a Catenary Mooring Cable** 882

38.3 **High Tension Cables; Snap-Slack Conditions** 885

38.4 **Dynamics of Catenary Moorings** 888
38.4.1 A Single Mooring Cable 888
38.4.2 Effect of Attached Intermediate Buoys 890
38.4.3 Mooring Line-Induced Damping . 892

38.5 **Second-Order Nonlinear Dynamics of Cables Formulated as Euler–Bernoulli Beams** 894
38.5.1 Background 894
38.5.2 The Governing 2-D Nonlinear System 895
38.5.3 Perturbation Series Expansion 895
38.5.4 The Frequency Domain Approach 896
38.5.5 Numerical Solution 898
38.5.6 The Bottom–Structure Interaction Approach 899
38.5.7 Results of Second-Order Double-Frequency Effects 900

References 903

Cables and elastic lines are used in a plethora of technologically advanced applications, on land, in the air, and at sea. The present effort aims at investigating the dynamic behavior of cables as regards their marine applications. Examples of the use of cables in naval architecture, marine, and offshore engineering include, but are not limited to, towing applications, elastic hoses, guides of remote operating vehicles, deployment of fiber ropes, cables for deployment of instrumentation packages on the seafloor, mooring systems, etc. Among the preceding list, the latter is the most frequent and at the same time the most demanding application.

A cable is virtually a slender structure that, depending on the application, can be (considered) inelastic (metallic) or very elastic (synthetic). However, contrary to analogous applications of slender structures where the bending stiffness is the prevailing component (risers and pipelines), the main concern for cables is the axial force, which will be referred herein as the tension or effective tension. Due to their small cross-sectional area the bending stiffness of cables admits very small values, meaning that cables are not capable of undertaking bending effects and, accordingly, the tension becomes the most crucial component. Ten-

sion is important for cables for two major reasons: firstly, tension is the term that stabilizes the lengthy structure and secondly it affects its integrity. If a cable is going to fail, that will occur due to extreme axial loading when it exceeds the breaking tension. In specific applications, such as mooring cables, the tension is the component that provides the necessary restoring characteristics to the moored vessel and restrains its displacements. Hence, the analysis of the dynamic equilibrium of cables intends primarily that task, namely the anticipation of the most possible accuracy of the component that adds to the static tension (when applicable). The variability of the dynamic components is tightly correlated with fatigue and, hence, relevant effects should be properly and accurately addressed.

Admittedly, the ability of omitting the bending stiffness from the group of physical and mechanical properties of cables leads to significant simplifications of the theoretical formulation. It is also true that, as far as the cable dynamics is concerned, the inclusion or not of the actual bending stiffness will not change the axial loading computations. In fact, this is exactly the approach that is taken by the majority of the research endeavors in the field. Nevertheless, the inclusion of the bending effects, regardless of their insignificant contribution on the dynamics of cables, resolves many problems in the numerical simulation, giving special attention to singular applications such as high or low tension cables. In addition, it should be admitted that the incorporation of all relevant effects in a system is conceptually the most accurate and advisable approach.

Marine applications of cable structures involve several peculiarities and crucial aspects. These arise from the fact that similar applications require the employment of lengthy cables, whilst in addition the cables operate in an extremely hostile and variable environment. Cables in the sea are always under motion and variable configuration, making their dynamic behavior the most important and, at the same time, the most interesting part of their design and analysis processes. Areas of interest and concern as far as the dynamics of marine cables are concerned are resonance regimes (both linear and nonlinear), low and high tension impacts, zero tension singularities, cable bottom interaction effects, mooring-induced damping effects, vortex-induced vibrations (VIV), induced fatigue, etc. It should be admitted that from the previous list specific parts are adapted to specific applications.

For the purposes of the present contribution, the authors attempt to address most of the preceding features, excluding VIV effects. The text that follows should not be regarded as material that focuses on a specific field. Is rather a survey that tries to cover the details of marine cables' dynamic behavior in the majority of their typical operations, and it is based on the long-term knowledge, experience, and efforts of the authors in the field.

The chapter on cable dynamics of the handbook at hand is structured as follows. Section 38.1 outlines the mathematical formulation of the problem and derives the complete nonlinear governing set that is accordingly reduced to provide the simplified linearized formulas. Section 38.2 treats the eigenvalue problems of both in-plane and out-of-plane vibrations and this is extended to highlight the differences that occur when a cable line has a part lying on the seafloor. Section 38.3 is dedicated to high and low tension regimes of marine cables; in other words, the snap-slack conditions as they are encountered in highly demanding applications such as in tendons of tension leg platforms. Section 38.4 analyzes the dynamics of mooring cables and the effect of intermediate buoys attached along their suspended length, whilst it exploits the employed solution to quantify and discuss the so-called mooring-induced damping phenomenon that contemporarily attracts major attention in mooring systems of wave energy converters. Section 38.5 presents a novel frequency domain-based approach for assessing the nonlinear dynamics of cables combined with a sophisticated methodology that allows accounting for cable bottom interaction effects.

The authors have avoided providing extensive numerical computations, and thus the results shown in each section are indicative of the application considered. The goal of the carefully selected plots is to provide evidence on the details of the dynamic response of cables and, obviously, to support the relevant discussion.

38.1 Mathematical Formulation

The cable is modeled as a Euler–Bernoulli beam with the following physical and mechanical properties: mass per unit unstretched length m, weight per unit unstretched length w_0, diameter d_O, cross-sectional area A, polar moment I_p, second moment I, material density ρ_c, modulus of elasticity E, and shear modulus G. It is noted that the quantities d_O, A, I_p, and I refer to the unstretched condition. The balance of the loading components (moments and forces) is shown schematically in Fig. 38.1. The stretched differential length of the element is $(1+e)\,\mathrm{d}s$, where $\mathrm{d}s$ is the unstretched differential length, e is the axial strain, and s denotes the Lagrangian coordinate that takes values along the unstretched length of the cylindrical element. All vari-

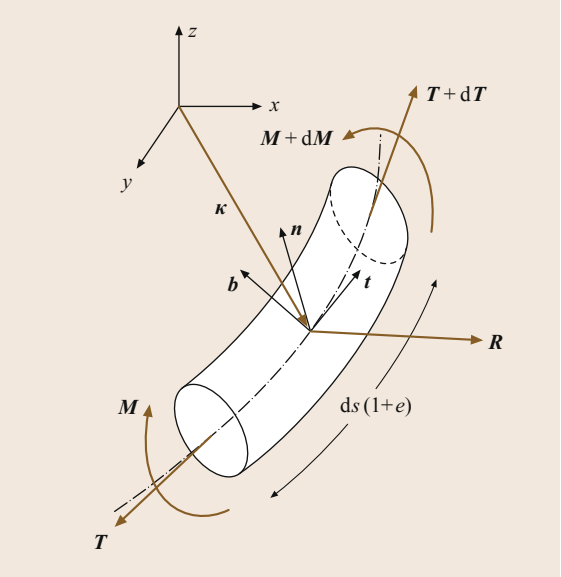


Fig. 38.1 Dynamic equilibrium of the stretched element

ables are defined with respect to the local Lagrangian coordinate system, which is determined by the unit vectors $\boldsymbol{t}$, $\boldsymbol{n}$, and $\boldsymbol{b}$. The unit vector $\boldsymbol{t}$ is tangent to the axis, $\boldsymbol{n}$ is perpendicular to $\boldsymbol{t}$, and the unit vector $\boldsymbol{b}$ is defined such that the system of vectors is orthogonal and right handed. In the following, the motions and the velocities parallel to $\boldsymbol{n}$ and $\boldsymbol{b}$ will be referred to as *normal* and *bi-normal*, respectively. By analogy, the loading components parallel to $\boldsymbol{n}$ and $\boldsymbol{b}$ will be referred to as *in plane* and *out of plane*.

In Fig. 38.1 $\boldsymbol{T} = T\boldsymbol{t} + S_\mathrm{n}\boldsymbol{n} + S_\mathrm{b}\boldsymbol{b}$, $\boldsymbol{M} = M_1\boldsymbol{t} + M_2\boldsymbol{n} + M_3\boldsymbol{b}$ denote respectively the vectors of forces and moments. Furthermore, T, S_n, S_b, M_1, M_2, M_3 denote, respectively, the tension, the in-plane and the out-of-plane shear forces, the moment due to torsion, the out-of-plane bending moment, and the in-plane bending moment. Finally, let $\boldsymbol{R}$ denote the vector summation of the distributed forces applied on the cable.

38.1.1 Derivatives in Time and Space

The material derivatives in time and space are complicated to some extent due to the selection of the Lagrangian coordinate system. This can be traced back to the fact that aside from the variation of the dynamic components in time and space, the coordinate system itself also varies with time t and the local coordinate s. The associated derivatives will first be expressed for an arbitrary vector $\boldsymbol{H}$, which in terms of the Lagrangian coordinate system of Fig. 38.1 is given by

$$\boldsymbol{H} = H_1\boldsymbol{t} + H_2\boldsymbol{n} + H_3\boldsymbol{b} \,. \tag{38.1}$$

The material derivative of $\boldsymbol{H}$ with respect to s is written as

$$\frac{\mathrm{D}\boldsymbol{H}}{\mathrm{D}s} = \frac{\partial \boldsymbol{H}}{\partial s} + H_1\frac{\partial \boldsymbol{t}}{\partial s} + H_2\frac{\partial \boldsymbol{n}}{\partial s} + H_3\frac{\partial \boldsymbol{b}}{\partial s} \,. \tag{38.2}$$

If $\boldsymbol{\Omega}$ denotes the Darboux vector of rotation for a space curve [38.1]

$$\boldsymbol{\Omega} = \Omega_1\boldsymbol{t} + \Omega_2\boldsymbol{n} + \Omega_3\boldsymbol{b} \,, \tag{38.3}$$

then for calculating the spatial derivatives of the unit vectors in (38.2), the symmetrical properties of the Darboux vector will be employed. These are [38.1]

$$\frac{\partial \boldsymbol{k}}{\partial s} = \boldsymbol{\Omega} \times \boldsymbol{k} \,, \tag{38.4}$$

for $\boldsymbol{k} = \boldsymbol{t}, \boldsymbol{n}, \boldsymbol{b}$.

After introducing (38.4) into (38.2), the later becomes

$$\frac{\mathrm{D}\boldsymbol{H}}{\mathrm{D}s} = \frac{\partial \boldsymbol{H}}{\partial s} + \boldsymbol{\Omega} \times \boldsymbol{H} \,. \tag{38.5}$$

With regard to the material derivative with respect to t, $\mathrm{D}/\mathrm{D}t$, of the arbitrary vector $\boldsymbol{H}$, the equivalent of the Darboux vector is the rotation vector of the classical dynamics

$$\boldsymbol{\omega} = \omega_1\boldsymbol{t} + \omega_2\boldsymbol{n} + \omega_3\boldsymbol{b} \,, \tag{38.6}$$

where ω_1, ω_2 and ω_3 are the associated angular velocities. Thus, the material derivative of $\boldsymbol{H}$ with respect to the time is written as [38.2]

$$\frac{\mathrm{D}\boldsymbol{H}}{\mathrm{D}t} = \frac{\partial \boldsymbol{H}}{\partial t} + \boldsymbol{\omega} \times \boldsymbol{H} \,. \tag{38.7}$$

The components of vectors $\boldsymbol{\omega}$ and $\boldsymbol{\Omega}$ are expressed in terms of the Euler angles ϕ, θ, and ψ as

$$\omega_1 = \frac{\partial \psi}{\partial t} - \frac{\partial \phi}{\partial t}\sin\theta \,, \tag{38.8}$$

$$\omega_2 = \frac{\partial \theta}{\partial t}\cos\psi + \frac{\partial \phi}{\partial t}\cos\theta\sin\psi \,, \tag{38.9}$$

$$\omega_3 = \frac{\partial \phi}{\partial t}\cos\theta\cos\psi - \frac{\partial \theta}{\partial t}\sin\psi \,, \tag{38.10}$$

$$\Omega_1 = \frac{\partial \psi}{\partial s} - \frac{\partial \phi}{\partial s}\sin\theta \,, \tag{38.11}$$

$$\Omega_2 = \frac{\partial \theta}{\partial s}\cos\psi + \frac{\partial \phi}{\partial s}\cos\theta\sin\psi \,, \tag{38.12}$$

$$\Omega_3 = \frac{\partial \phi}{\partial s}\cos\theta\cos\psi - \frac{\partial \theta}{\partial s}\sin\psi \,, \tag{38.13}$$

where ψ is the angle of torsion, ϕ is the angle which is formed between the tangent of the cable and the horizontal in the plane of reference, and θ is the out-of-plane angle. It is also noted that the elements of the

vector of moments are associated with the curvatures according to

$$\begin{aligned}\boldsymbol{M} &= M_1\boldsymbol{t} + M_2\boldsymbol{n} + M_3\boldsymbol{b} \\ &= GI_\mathrm{p}\Omega_1\boldsymbol{t} + EI\Omega_2\boldsymbol{n} + EI\Omega_3\boldsymbol{b}\,. \end{aligned} \tag{38.14}$$

38.1.2 Dynamic Equilibrium

Applying Newton's law to the stretched cable element along with the conservation of mass yields

$$m\frac{\mathrm{D}\boldsymbol{V}}{\mathrm{D}t} = \frac{\mathrm{D}\boldsymbol{T}}{\mathrm{D}s} + \boldsymbol{R}(1+e)\,. \tag{38.15}$$

Next, we expand the material derivatives in the local coordinate system to yield

$$m\left(\frac{\partial \boldsymbol{V}}{\partial t} + \boldsymbol{\omega}\times\boldsymbol{V}\right) = \frac{\partial \boldsymbol{T}}{\partial s} + \boldsymbol{\Omega}\times\boldsymbol{T} + \boldsymbol{R}(1+e)\,. \tag{38.16}$$

In (38.15) and (38.16), $\boldsymbol{V}$ denotes the velocity vector, which is expressed as

$$\boldsymbol{V} = u\boldsymbol{t} + v\boldsymbol{n} + w\boldsymbol{b}\,, \tag{38.17}$$

where u, v and w denote the velocities of the cable element parallel to the unit vectors $\boldsymbol{t}$, $\boldsymbol{n}$, and $\boldsymbol{b}$ of the local Lagrangian coordinate system. Also, after having defined the Euler angles and recalling that the distributed forces are composed by the gravitational components, the added mass terms and the nonlinear drag forces $\boldsymbol{R}(1+e)$ will be given by the following expression

$$\begin{aligned}\boldsymbol{R}(1+e) = &-(w_0\sin\phi\cos\theta - R_{\mathrm{d}t})\boldsymbol{t} \\ &-\left(w_0\cos\phi + m_\mathrm{a}\frac{\partial v_{2\mathrm{r}}}{\partial t} - R_{\mathrm{d}n}\right)\boldsymbol{n} \\ &-\left(w_0\sin\phi\sin\theta + m_\mathrm{a}\frac{\partial v_{3\mathrm{r}}}{\partial t} - R_{\mathrm{d}b}\right)\boldsymbol{b}\,, \end{aligned} \tag{38.18}$$

where

$$R_{\mathrm{d}t} = -\frac{1}{2}\pi\rho d_\mathrm{O} C_{\mathrm{d}t} v_{1\mathrm{r}}\,|v_{1\mathrm{r}}|\,\sqrt{1+e}\,, \tag{38.19}$$

$$R_{\mathrm{d}n} = -\frac{1}{2}\rho d_\mathrm{O} C_{\mathrm{d}n} v_{2\mathrm{r}}\sqrt{v_{2\mathrm{r}}^2 + v_{3\mathrm{r}}^2}\sqrt{1+e}\,, \tag{38.20}$$

$$R_{\mathrm{d}b} = -\frac{1}{2}\rho d_\mathrm{O} C_{\mathrm{d}b} v_{3\mathrm{r}}\sqrt{v_{2\mathrm{r}}^2 + v_{3\mathrm{r}}^2}\sqrt{1+e} \tag{38.21}$$

are the drag forces in axial, normal, and bi-normal directions depending on the associated drag coefficients $C_{\mathrm{d}t}$, $C_{\mathrm{d}n}$, and $C_{\mathrm{d}b}$, while $v_{1\mathrm{r}}$, $v_{2\mathrm{r}}$, and $v_{3\mathrm{r}}$ are the relative velocities in the case of the present of current. Clearly, ρ denotes the water density.

Further, the balance of moments equation is obtained by taking moments about the left-hand side of the differential element and expressing the results in terms of unstretched quantities. This yields [38.3]

$$\begin{aligned}\frac{1}{1+e}\frac{\mathrm{D}}{\mathrm{D}t}[\rho_\mathrm{c}\mathbf{I}\boldsymbol{\omega}] = &\frac{1}{(1+e)^2}\frac{\mathrm{D}\boldsymbol{M}}{\mathrm{D}s} + \mathrm{d}\boldsymbol{\kappa} \\ &\times\boldsymbol{R}(1+e) + \frac{\mathrm{d}\boldsymbol{\kappa}}{\mathrm{d}s}\times\boldsymbol{T}\,, \end{aligned} \tag{38.22}$$

where $\boldsymbol{\kappa}$ denotes the vector from the origin to the beam element (Fig. 38.1). In the limit $\mathrm{d}\boldsymbol{\kappa}\to 0$ it holds that

$$\mathrm{d}\boldsymbol{\kappa}\times\boldsymbol{R}\to 0\,, \tag{38.23}$$

$$\frac{1}{1+e}\frac{\mathrm{d}\boldsymbol{\kappa}}{\mathrm{d}s}\to\boldsymbol{t}\,. \tag{38.24}$$

Thus, (38.22) renders the following simplified form that provides the balance of moments equations

$$\frac{1}{1+e}\frac{\mathrm{D}}{\mathrm{D}t}[\rho_\mathrm{c}\mathbf{I}\boldsymbol{\omega}] = \frac{1}{(1+e)^2}\frac{\mathrm{D}\boldsymbol{M}}{\mathrm{D}s} + \boldsymbol{t}\times\boldsymbol{T}(1+e)\,, \tag{38.25}$$

where $\rho_\mathrm{c}\mathbf{I}$ is a diagonal 3×3 matrix with $\mathrm{diag}[\rho_\mathrm{c}\mathbf{I}] = (\rho_\mathrm{c}I_\mathrm{p}, \rho_\mathrm{c}I, \rho_\mathrm{c}I)$. After expanding the material derivatives, (38.25) is transformed into

$$\begin{aligned}\frac{\rho_\mathrm{c}\mathbf{I}}{1+e}\left(\frac{\partial\boldsymbol{\omega}}{\partial t} + \boldsymbol{\omega}\times\boldsymbol{\omega}\right) = &\frac{1}{(1+e)^2}\left(\frac{\partial\boldsymbol{M}}{\partial s} + \boldsymbol{\Omega}\times\boldsymbol{M}\right) \\ &+\boldsymbol{t}\times\boldsymbol{T}(1+e)\,. \end{aligned} \tag{38.26}$$

Finally, the compatibility relation will be derived, which requires that the element is free of discontinuities. Therefore, the vector $\boldsymbol{\kappa}(s,t)$ must be a continuous function of both independent variables. In other words,

$$\frac{\mathrm{D}}{\mathrm{D}t}\left[\frac{\mathrm{D}\boldsymbol{\kappa}}{\mathrm{D}s}\right] = \frac{\mathrm{D}}{\mathrm{D}s}\left[\frac{\mathrm{D}\boldsymbol{\kappa}}{\mathrm{D}t}\right]. \tag{38.27}$$

Taking into account that (Fig. 38.1) $\boldsymbol{V} = \mathrm{D}\boldsymbol{\kappa}/\mathrm{D}t$, $(1+e)\boldsymbol{t} = \mathrm{D}\boldsymbol{\kappa}/\mathrm{D}s$ and assuming a linear stress-strain relation $e = T/EA$, (38.27) is cast into

$$\frac{1}{EA}\frac{\partial T}{\partial t}\boldsymbol{t} + (1+e)\boldsymbol{\omega}\times\boldsymbol{t} = \frac{\partial\boldsymbol{V}}{\partial s} + \boldsymbol{\Omega}\times\boldsymbol{V}\,. \tag{38.28}$$

38.1.3 Final System

The governing dynamic equilibrium system is derived by expanding the cross products in (38.16), (38.26), and (38.28). As a result, a system of partial differential equations is obtained, which is given in detail in the following,

$$\begin{aligned}&\frac{\partial T}{\partial s} + S_\mathrm{b}\Omega_2 - S_\mathrm{n}\Omega_3 - w_0\sin\phi\cos\theta - m\frac{\partial u}{\partial t} \\ &\quad - mw\left(\frac{\partial\theta}{\partial t}\cos\psi + \frac{\partial\phi}{\partial t}\cos\theta\sin\psi\right) \\ &\quad + mv\left(\frac{\partial\phi}{\partial t}\cos\theta\cos\psi - \frac{\partial\theta}{\partial t}\sin\psi\right) + R_{\mathrm{d}t} = 0\,, \end{aligned} \tag{38.29}$$

$$\frac{\partial S_\mathrm{n}}{\partial s} + \Omega_3 T - \Omega_1 S_\mathrm{b} - w_0 \cos\phi - m\frac{\partial v}{\partial t} - mu\left(\frac{\partial \phi}{\partial t}\cos\theta\cos\psi - \frac{\partial \theta}{\partial t}\sin\psi\right) + mw\left(\frac{\partial \psi}{\partial t} - \frac{\partial \phi}{\partial t}\sin\theta\right) + R_{dn} = 0\,, \tag{38.30}$$

$$\frac{\partial S_\mathrm{b}}{\partial s} + \Omega_1 S_\mathrm{n} - \Omega_2 T - w_0 \sin\phi\sin\theta - m\frac{\partial w}{\partial t} - mv\left(\frac{\partial \psi}{\partial t} - \frac{\partial \phi}{\partial t}\sin\theta\right) + mu\left(\frac{\partial \theta}{\partial t}\cos\psi + \frac{\partial \phi}{\partial t}\cos\theta\sin\psi\right) + R_{db} = 0\,, \tag{38.31}$$

$$\frac{\partial u}{\partial s} + \Omega_2 w - \Omega_3 v - \frac{1}{EA}\frac{\partial T}{\partial t} = 0\,, \tag{38.32}$$

$$\frac{\partial v}{\partial s} + \Omega_3 u - \Omega_1 w - \frac{\partial \phi}{\partial t}\cos\theta\cos\psi + \frac{\partial \theta}{\partial t}\sin\psi = 0\,, \tag{38.33}$$

$$\frac{\partial w}{\partial s} + \Omega_1 v - \Omega_2 u + \frac{\partial \theta}{\partial t}\cos\psi + \frac{\partial \phi}{\partial t}\cos\theta\sin\psi = 0\,, \tag{38.34}$$

$$GI_\mathrm{p}\frac{\partial \Omega_1}{\partial s} - \rho_\mathrm{c} I_\mathrm{p}\left(\frac{\partial^2 \psi}{\partial t^2} - \frac{\partial^2 \phi}{\partial t^2}\sin\theta - \frac{\partial \phi}{\partial t}\frac{\partial \theta}{\partial t}\cos\theta\right) = 0\,, \tag{38.35}$$

$$EI\frac{\partial \Omega_2}{\partial s} + (GI_\mathrm{p} - EI)\Omega_1\Omega_3 - S_\mathrm{b} - \rho_\mathrm{c} I\left(\frac{\partial^2 \theta}{\partial t^2}\cos\psi - \frac{\partial \theta}{\partial t}\frac{\partial \psi}{\partial t}\sin\psi + \frac{\partial^2 \phi}{\partial t^2}\cos\theta\sin\psi - \frac{\partial \phi}{\partial t}\frac{\partial \theta}{\partial t}\sin\theta\sin\psi + \frac{\partial \phi}{\partial t}\frac{\partial \psi}{\partial t}\cos\theta\cos\psi\right) = 0\,, \tag{38.36}$$

$$EI\frac{\partial \Omega_3}{\partial s} + (EI - GI_\mathrm{p})\Omega_1\Omega_2 + S_\mathrm{n} - \rho_\mathrm{c} I\left(\frac{\partial^2 \phi}{\partial t^2}\cos\theta\cos\psi - \frac{\partial \phi}{\partial t}\frac{\partial \theta}{\partial t}\sin\theta\cos\psi - \frac{\partial \phi}{\partial t}\frac{\partial \psi}{\partial t}\cos\theta\sin\psi - \frac{\partial^2 \theta}{\partial t^2}\sin\psi - \frac{\partial \theta}{\partial t}\frac{\partial \psi}{\partial t}\cos\psi\right) = 0\,. \tag{38.37}$$

A similar, simplified system omitting torsion effects was derived by *Howell* [38.4].

Clearly, EI and GI_p in the above relations were assumed to be independent of s. Further, (38.29)–(38.37) are combined with the curvatures, (38.11)–(38.13), to form a system of 12 partial differential equations with the same number of unknowns. For slender structures with very small cross-sectional areas, such as cables, torsional effects can be considered negligible. The flexural rigidity is retained, however, for practical reasons to eliminate possible zero tension singularities when solving the dynamics of the governing equilibrium system. Thus, after eliminating the terms that govern torsion, the system is reduced to

$$\frac{\partial T}{\partial s} + S_\mathrm{b}\Omega_2 - S_\mathrm{n}\Omega_3 - w_0 \sin\phi\cos\theta + R_{dt} - m\frac{\partial u}{\partial t} - m\left(w\frac{\partial \theta}{\partial t} - v\frac{\partial \phi}{\partial t}\cos\theta\right) = 0\,, \tag{38.38}$$

$$\frac{\partial S_\mathrm{n}}{\partial s} + \Omega_3(T + S_\mathrm{b}\tan\theta) - w_0\cos\phi + R_{dn} - m\frac{\partial v}{\partial t} - m\frac{\partial \phi}{\partial t}(u\cos\theta + w\sin\theta) - m_\mathrm{a}\frac{\partial v}{\partial t} = 0\,, \tag{38.39}$$

$$\frac{\partial S_\mathrm{b}}{\partial s} - \Omega_2 T - \Omega_3 S_\mathrm{n}\tan\theta - w_0\sin\phi\sin\theta + R_{db} - m\frac{\partial w}{\partial t} + m\left(v\frac{\partial \phi}{\partial t}\sin\theta + u\frac{\partial \theta}{\partial t}\right) - m_\mathrm{a}\frac{\partial w}{\partial t} = 0\,, \tag{38.40}$$

$$\frac{\partial u}{\partial s} + \Omega_2 w - \Omega_3 v - \frac{1}{EA}\frac{\partial T}{\partial t} = 0\,, \tag{38.41}$$

$$\frac{\partial v}{\partial s} + \Omega_3(u + w\tan\theta) - (1+e)\frac{\partial \phi}{\partial t}\cos\theta = 0\,, \tag{38.42}$$

$$\frac{\partial w}{\partial s} - \Omega_3 v\tan\theta - \Omega_2 u + (1+e)\frac{\partial \theta}{\partial t} = 0\,, \tag{38.43}$$

$$EI\frac{\partial \Omega_2}{\partial s} + EI\Omega_3^2\tan\theta - S_\mathrm{b}(1+e)^3 = 0\,, \tag{38.44}$$

$$EI\frac{\partial \Omega_3}{\partial s} - EI\Omega_2\Omega_3\tan\theta + S_\mathrm{n}(1+e)^3 = 0\,, \tag{38.45}$$

$$\frac{\partial \theta}{\partial s} - \Omega_2 = 0\,, \tag{38.46}$$

$$\frac{\partial \phi}{\partial s}\cos\theta - \Omega_3 = 0\,. \tag{38.47}$$

38.1.4 Reduced Model Omitting Bending Stiffness

The basic mathematical model is composed of (38.38)–(38.47) and can describe effectively the global three-dimensional (3-D) nonlinear dynamics of any nonspinning slender structure, regardless of whether or not flexural rigidity is considered significant. It can also account for variable physical and mechanical properties

along its length. The latter can be realized by letting w_0, m, m_a, EA, and EI to be functions of the coordinate s.

Typically, when we are referring on cables, or chains, it is usual to assume negligible bending stiffness. This, in turn, will lead to a significant simplification of the governing set, allowing faster calculations. In particular, by letting $EI \to 0$, the generic nonlinear system of (38.38)–(38.47) is reduced to only six partial differential equations with an equal number of unknowns, the elements T, ϕ, θ, u, v, and w. The reduced system can be compacted to the following matrix form

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & -v\cos\theta & w & 1 & 0 & 0 \\ 0 & \cos\theta(u+\tan\theta) & 0 & 0 & 1 & 0 \\ 0 & -v\sin\theta & -u & 0 & 0 & 1 \end{bmatrix} \frac{\mathrm{d}}{\mathrm{d}s} \begin{bmatrix} T \\ \phi \\ \theta \\ u \\ v \\ w \end{bmatrix}$$

$$= \begin{bmatrix} w_0\cos\theta\sin\phi - R_{\mathrm{d}t} \\ \dfrac{w_0}{T}\dfrac{\cos\phi}{\cos\theta} - \dfrac{R_{\mathrm{d}n}}{T}\dfrac{1}{\cos\theta} \\ -\dfrac{w_0}{T}\sin\theta\sin\phi + \dfrac{R_{\mathrm{d}b}}{T} \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$+ \begin{bmatrix} 0 & -mv\cos\theta & mw & m & 0 & 0 \\ 0 & \dfrac{m}{T}(u+w\tan\theta) & 0 & 0 & \dfrac{m+m_a}{T\cos\theta} & 0 \\ 0 & \dfrac{m}{T}v\sin\theta & \dfrac{m}{T}u & 0 & 0 & -\dfrac{m+m_a}{T} \\ \dfrac{1}{EA} & 0 & 0 & 0 & 0 & 0 \\ 0 & (1+e)\cos\theta & 0 & 0 & 0 & 0 \\ 0 & 0 & -(1+e) & 0 & 0 & 0 \end{bmatrix}$$

$$\times \frac{\mathrm{d}}{\mathrm{d}t} \begin{bmatrix} T \\ \phi \\ \theta \\ u \\ v \\ w \end{bmatrix} . \qquad (38.48)$$

It is immediately apparent that as far as the cable dynamics are concerned, the assumption that the model cannot undertake any bending effects leads univocally to the probability of the occurrence of a singularity, usually referred as the *zero tension singularity*. Virtually, this singularity characterizes only the theoretical formulation of the cable dynamical problem, meaning that when the total tension at any point along the cable becomes zero, then the theoretical model and, accordingly, the numerical solution scheme collapses. There is a practical, mostly empirical, way to avoid the associated singularity during numerical calculations by stopping the computations during the associated time interval and artificially putting the tension equal to zero. This is the approach taken by *Papazoglou* et al. [38.5] and, accordingly, by *Johanning* et al. [38.6], although in the latter study the authors do not arrive at a condition that fully cancels the effective tension. It should be admitted, however, that it is conceptually more accurate to simulate a cable as a slender beam given the fact that it undertakes, admittedly to a marginal extent, some bending effects. With this approach, the zero tension singularity, which may occur in specific applications such as in low tension cables, is virtually eliminated.

38.1.5 Equivalent Linearization

The basic dynamical model that is considered herein is the complete 3-D nonlinear continuous system given by (38.38)–(38.47). It is apparent that the solution of that system can be achieved only by applying appropriate time domain solution techniques. Typically, the solution in the time domain requires post-processing of the time histories of the dynamic components, which inevitably results in additional computational effort. Most of the time, however, the actual dynamic behavior can be accurately approximated by only the linear components, whereas the nonlinear terms are significant only in specific conditions. Hence, it is advisable for any nonlinear dynamic system to start from the study of its linearized equivalent, which is obtained by ignoring or by linearizing the nonlinear terms. For the nonlinear system at hand, the linearization procedure begins by considering that the time varying terms are given by the summation of static and dynamic counterparts. Thus, the vector of the unknowns (38.38)–(38.47)

$$\boldsymbol{Y}(s,t) = \left[T \;\; S_{\mathrm{n}} \;\; S_{\mathrm{b}} \;\; u \;\; v \;\; w \;\; \Omega_2 \;\; \Omega_3 \;\; \theta \;\; \phi \right]^{\mathrm{T}} , \qquad (38.49)$$

is written as

$$\boldsymbol{Y}(s,t) = \boldsymbol{Y}_0(s) + \boldsymbol{Y}_1(s,t) , \qquad (38.50)$$

where

$$\boldsymbol{Y}_0(s) = \left[T_0 \;\; S_{\mathrm{n}0} \;\; S_{\mathrm{b}0} \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; \Omega_{30} \;\; 0 \;\; \phi_0 \right]^{\mathrm{T}} , \qquad (38.51)$$

and

$$\boldsymbol{Y}_1(s,t) = \left[T_1 \;\; S_{\mathrm{n}1} \;\; S_{\mathrm{b}1} \;\; u \;\; v \;\; w \;\; \Omega_{21} \;\; \Omega_{31} \;\; \theta_1 \;\; \phi_1 \right]^{\mathrm{T}} . \qquad (38.52)$$

The static quantities for the out-of-plane angle θ_0 and the out-of-plane curvature Ω_{20} were set equal to zero as in the static equilibrium position the cable was considered to have a perfect two-dimensional (2-D) configuration. It is noted that in the context of the linear problem,

the velocities u, v, and w can be immediately replaced by the corresponding motions $u = \partial p/\partial t$, $v = \partial q/\partial t$, and $w = \partial r/\partial t$. After introducing (38.50) into the nonlinear governing equation (38.38)–(38.47) and omitting higher order terms, the original set is recast to

$$m\frac{\partial^2 p}{\partial t^2} = \frac{\partial T_1}{\partial s} - S_{n1}\Omega_{30} - S_{n0}\Omega_{31} - w_0\cos\phi_0\phi_1 \,, \tag{38.53}$$

$$(m+m_a)\frac{\partial^2 q}{\partial t^2} = \frac{\partial S_{n1}}{\partial s} + T_1\Omega_{30} + \Omega_{31}T_0 + w_0\sin\phi_0\phi_1 - c\omega\,|q|\,\frac{\partial q}{\partial t} \,, \tag{38.54}$$

$$(m+m_a)\frac{\partial^2 r}{\partial t^2} = \frac{\partial S_{b1}}{\partial s} - S_{n0}\Omega_{30}\theta_1 - T_0\Omega_{21} - w_0\sin\phi_0\theta_1 - c\omega\,|r|\,\frac{\partial r}{\partial t} \,, \tag{38.55}$$

$$T_1 = EA\left(\frac{\partial p}{\partial s} - \Omega_{30}q\right) , \tag{38.56}$$

$$\phi_1 = \frac{\partial q}{\partial s} + \Omega_{30}p \,, \tag{38.57}$$

$$\theta_1 = -\frac{\partial r}{\partial s} \,, \tag{38.58}$$

$$EI\frac{\partial \Omega_{21}}{\partial s} - EI\Omega_{30}\theta_1 = S_{b1} \,, \tag{38.59}$$

$$EI\frac{\partial \Omega_{31}}{\partial s} = -S_{n1} \,, \tag{38.60}$$

$$\frac{\partial \theta_1}{\partial s} = \Omega_{21} \,, \tag{38.61}$$

$$\frac{\partial \phi_1}{\partial s} = \Omega_{31} \,, \tag{38.62}$$

where it was assumed that $1 + e \approx 1$ and the drag force in axial direction was omitted as it depends on a very small drag coefficient. In the realm of linear theory, the nonlinear drag forces must be linearized as well, and that was exactly the approach that was taken herein as it is manifested by the last terms of (38.54)–(38.55). The linearization requires that the drag forces must be approximated by a Fourier series expansion from which only the linear components are retained [38.7]. Here, c denotes the linearized damping coefficient $c = (4/3\pi)\rho C_d d_O$, where d_O is the outer diameter. Clearly, for a cylindrical structure the drag coefficients in normal and bi-normal directions are equal, i. e., $C_d = C_{dn} = C_{db}$.

A careful inspection of the system of equations (38.53)–(38.62) reveals that the out-of-plane and the in-plane motions are virtually decoupled. In fact, the assumption that enabled decoupling was the omission of torsion. For more information the interested reader is referenced to [38.8]. Hence, (38.53)–(38.62) can be split into two independent sets that can be treated separately. Accordingly, the in-plane and the out-of-plane linear dynamics of the cable are governed by the sets of equations (38.53), (38.54), (38.56), (38.57), (38.60), (38.62), and (38.55), (38.58), (38.59), and (38.61), respectively.

The vectors of the associated unknown dynamic components for the in-plane and the out-of-plane linear systems are, respectively, given by

$$\boldsymbol{X}(s,t) = \begin{bmatrix} T_1 & S_{n1} & p & q & \Omega_{31} & \phi \end{bmatrix}^T \,, \tag{38.63}$$

$$\boldsymbol{Z}(s,t) = \begin{bmatrix} S_{b1} & r & \Omega_{21} & \theta_1 \end{bmatrix}^T \,. \tag{38.64}$$

The response is considered harmonic and the problem is treated in the complex space $\mathbb{C}$. Thus, the vectors of (38.63)–(38.64) are expressed as

$$\boldsymbol{X}(s,t) = \mathrm{Re}\{\boldsymbol{x}(s)e^{i\omega t}\} \,, \tag{38.65}$$

$$\boldsymbol{Z}(s;t) = \mathrm{Re}\{\boldsymbol{z}(s)e^{i\omega t}\} \,, \tag{38.66}$$

where ω is the frequency of excitation and

$$\boldsymbol{x}(s) = \begin{bmatrix} \bar{T}_1(s) & \bar{S}_{n1}(s) & \bar{p}(s) & \bar{q}(s) & \bar{\Omega}_{31}(s) & \bar{\phi}(s) \end{bmatrix}^T \,, \tag{38.67}$$

$$\boldsymbol{z}(s) = \begin{bmatrix} \bar{S}_{b1}(s) & \bar{r}(s) & \bar{\Omega}_{21}(s) & \bar{\theta}_1(s) \end{bmatrix}^T \,. \tag{38.68}$$

The spatial vectors $\boldsymbol{x}(s)$ and $\boldsymbol{z}(s)$ are complex.

Next, (38.67) and (38.68) are introduced into the in-plane and the out-of-plane linear systems. After separating real and imaginary parts, the following matrix equations are obtained

$$\frac{\mathrm{d}}{\mathrm{d}s}\begin{bmatrix} \bar{T}_1 \\ \bar{S}_{n1} \\ \bar{p} \\ \bar{q} \\ \bar{\Omega}_{31} \\ \bar{\phi}_1 \end{bmatrix} = \begin{bmatrix} 0 & \Omega_{30} & -\omega^2 m & 0 & S_{n0} & w_0\cos\phi_0 \\ -\Omega_{30} & 0 & 0 & -(m+m_a)\omega^2 + ic\omega^2|\bar{q}| & -T_{30} & -w_0\sin\phi_0 \\ \frac{1}{EA} & 0 & 0 & \Omega_{30} & 0 & 0 \\ 0 & 0 & -\Omega_{30} & 0 & 0 & 1 \\ 0 & \frac{-1}{EI} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \times \begin{bmatrix} \bar{T}_1 \\ \bar{S}_{n1} \\ \bar{p} \\ \bar{q} \\ \bar{\Omega}_{31} \\ \bar{\phi}_1 \end{bmatrix} \,, \tag{38.69}$$

$$\frac{\mathrm{d}}{\mathrm{d}s}\begin{bmatrix}\bar{S}_{\mathrm{b}1}\\ \bar{r}\\ \bar{\Omega}_{21}\\ \bar{\theta}_1\end{bmatrix} = \begin{bmatrix} 0 & -(m+m_{\mathrm{a}})\omega^2 + \mathrm{i}c\omega^2|\bar{r}| & T_0 & S_{\mathrm{n}0}\Omega_{30} + w_0\sin\phi_0 \\ 0 & 0 & 0 & -1 \\ \dfrac{1}{EI} & 0 & \Omega_{30} & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \times \begin{bmatrix}\bar{S}_{\mathrm{b}1}\\ \bar{r}\\ \bar{\Omega}_{21}\\ \bar{\theta}_1\end{bmatrix}. \tag{38.70}$$

The reduced systems that represent the eigenvalue problem the solution of which provides the natural frequencies and the mode shapes of the cable, are derived through (38.69)–(38.70) after omitting the imaginary terms that express the contribution of the nonlinear drag forces. The solution of the eigenvalue problems requires the employment of valid assumptions regarding the boundary conditions. For a cable structure it is sufficient to assume that the motions as well as the curvatures at both ends are zero. A rational conclusion that is reasonably drawn from (38.69) and (38.70) is that the mode shapes, as well as the natural frequencies, strongly depend on the static components, namely the initial configuration of the cable. This feature is specifically highlighted to note that the popular sinusoidal mode shapes which are frequently adopted in the literature may lead to erroneous results, as they omit the all-important weight of the cable that provides the pattern of the static configuration. The weight effect on the eigenvalue problem of a string simulating a cable was thoroughly discussed by *Triantafyllou* and *Triantafyllou* [38.9], who also introduced the bending system to avoid zero tension singularity on the lower boundary layer of the assumed hanging string. An analogous procedure that relied on singular perturbations aiming at treating the eigenvalue problem of a vertical slender structure was applied recently by *Chatjigeorgiou* [38.10].

38.2 The Eigenvalue Problem of a Catenary Mooring Cable

A thorough investigation of the dynamic equilibrium system should obviously include the solution of the associated eigenvalue problem. This will provide the mode shapes and, most importantly, the natural frequencies that detect the locations of potential resonances. As far as the dynamics of cables are concerned the eigenvalue problem was systematically investigated in series of studies led by *Triantafyllou* et al. [38.11–14] (see also [38.9]).

For a continuous system such as a cable structure the number of eigenfrequencies and, accordingly, the number of eigenmodes are infinite. The calculation of the shapes of eigenmodes is crucial from many viewpoints. Firstly, they provide generic information on how the structure will vibrate. Secondly, they allow the construction of series expansions for the solution of the associated dynamic problem, which accordingly can be simplified by applying the orthogonality relations of the eigenmodes resulting in reduced systems of ordinary differential equations. The latter procedure could provide insight on the details of the nonlinear oscillations of the system investigated. Analytic solutions for the eigenvalue problem are, indeed, feasible provided that several simplistic assumptions are made [38.9] and [38.15]. Nevertheless, they have specific limitations in their range of application.

The details of the eigenvalue problem strongly depend on the initial configuration of the structure investigated, namely, the parameters that govern its static equilibrium. Cables in marine applications typically obtain a catenary shape either as mooring lines or as towing cables of underwater structures, or during deployment. Here, a catenary mooring cable is considered with the following physical and the mechanical properties: length 1000 m, mass 94.5 kg/m, added mass 16 kg/m, weight 770 N/m density 6052 kg/m^3, Young's modulus 207 GPa, diameter 0.141 m, static tension at the top 780 kN, and installation depth 375 m. According to the above, approximately 212 m of the cable will lie on the assumed horizontal bottom. The specific configuration was purposely selected to investigate potential differences on the natural frequencies and the mode shapes due to the part lying o the bottom. The out-of-plane eigenvalue problem was also considered. Nevertheless, the latter concerned only the suspended part.

Both the in-plane and the out-of-plane eigenvalue problems incorporated the bending stiffness which, however, was set to a very small (but not zero) value. Apart from the *principally accurate* approach, the inclusion of flexural rigidity allows considering the out-of-plane problem as well. The sought solution was achieved by treating the mathematical models ((38.69) and (38.70), letting $c = 0$) as a two-point boundary value problem. Zero motions and curvatures at the end points were employed, whilst the required calculations were performed using the dedicated *bvp4c* routine of

Table 38.1 In-plane eigenfrequencies [rad/s]

Mode no	1	2	3	4	5
Resting on bottom	0.4270	0.6619	0.8637	1.1116	1.3362
Suspended part	0.5479	0.8231	1.1470	1.4251	1.7357

Table 38.2 Out-of-plane eigenfrequencies [rad/s]

Mode no	1	2	3	4	5
Suspended part	0.2294	0.5847	0.8762	1.1682	1.4607

Matlab which integrates the system of equations. An important aspect as regards the in-plane problem is the formulation of the part lying on the bottom. To overcome this difficulty, the solution of the static equilibrium problem considered in that region is neutrally buoyant, resulting in a straight line. The same assumption was made for the dynamic eigenvalue problem as well, which additionally required that the effective weight must be variable along the complete length of the cable.

Results from the eigenvalue problem are given in Tables 38.1 and 38.2 and in Figs. 38.2–38.7 (first five modes). Tables 38.1 and 38.2, respectively, list the natural frequencies in the in-plane and the out-of-plane directions. In Table 38.1 it is immediately apparent that the inclusion of the bottom resting part profoundly changes the values of the eigenfrequencies. In fact, these shift towards lower values. The changes observed are indicative of the different configurations of the dynamic systems, which obey different particulars. Accordingly, it can be safely said that approximating a mooring cable with only its suspended part could be a deficient approach. Table 38.2 provides the natural frequencies for the out-of-plane dynamics (suspended part only), which are sufficiently lower than the in-plane ones.

Evidence of the dynamic response of the mooring cable concerned are provided by the mode shapes, which are depicted in Figs. 38.2–38.5 for the in-plane vibrations and in Figs. 38.6–38.8 for the out-of-plane

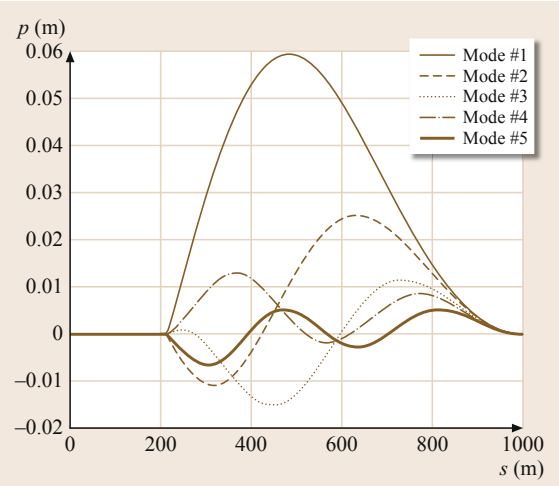


Fig. 38.2 In-plane axial motion eigenmodes of a cable resting on the bottom of the sea

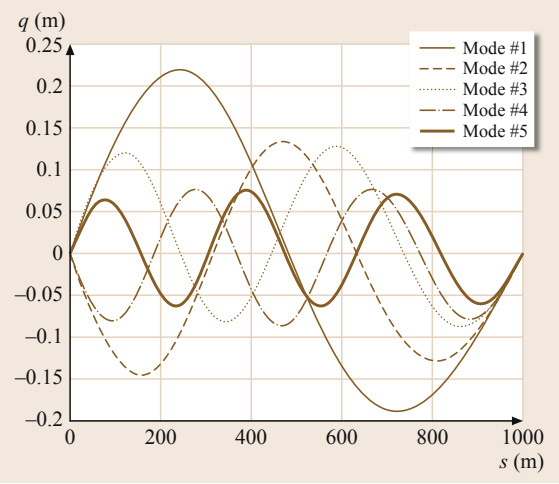


Fig. 38.3 In-plane normal motion eigenmodes of a cable resting on the bottom of the sea

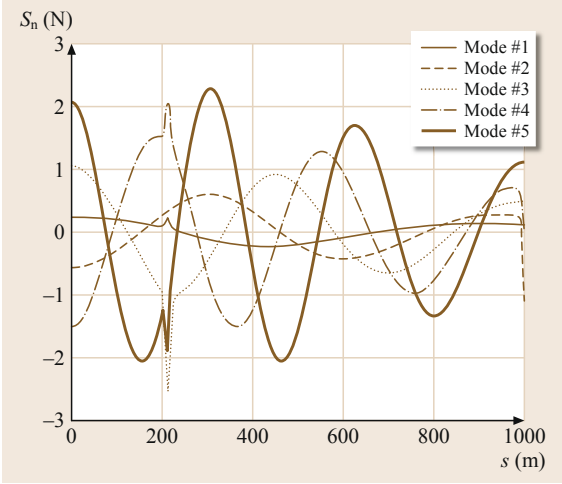


Fig. 38.4 In-plane shear force eigenmodes of a cable seated at the bottom

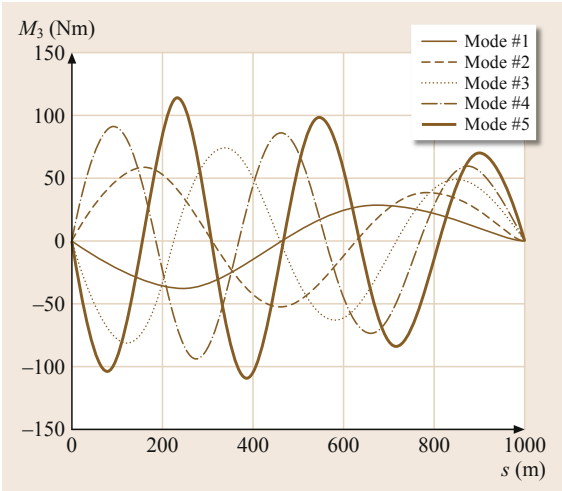


Fig. 38.5 In-plane bending moment eigenmodes of a cable seated at the bottom

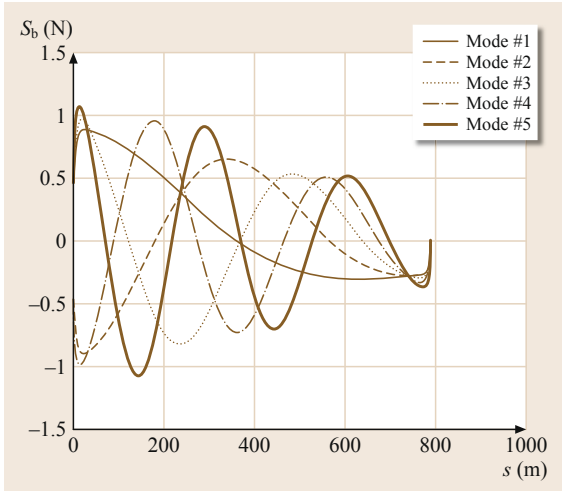


Fig. 38.7 Out-of-plane shear force eigenmodes of the suspended part

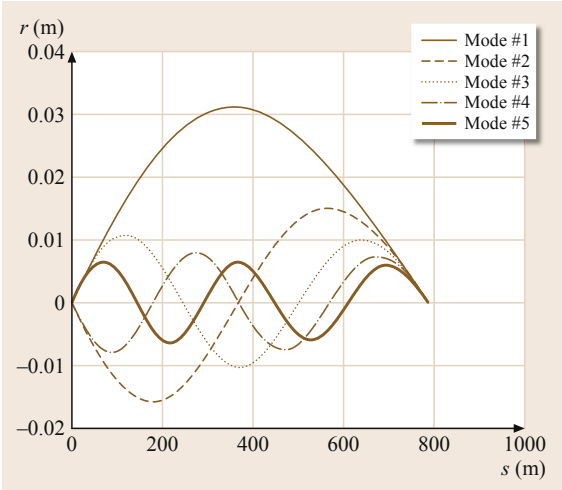


Fig. 38.6 Bi-normal motion eigenmodes of the suspended part

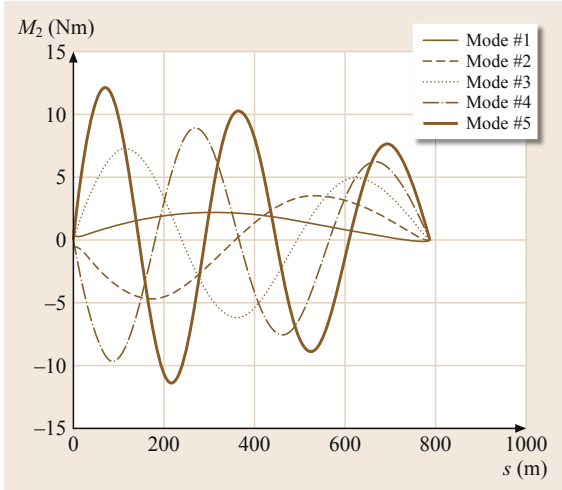


Fig. 38.8 Out-of-plane bending moment eigenmodes of the suspended part

vibrations. Although insignificant due to the selection of a negligible bending stiffness, the bending moments and the shear forces are provided as well (Figs. 38.4 and 38.5). The very small values of the (dimensionalized) shear force and bending moment are confirmative of the insignificant impact of bending as far as the dynamics of cables are concerned.

Here, only the bottom resting case results are shown, as they contain the most striking characteristics. The axial motions (the stretching) of the cable are examined with the aid of Fig. 38.2. This is the only depiction that explicitly shows the lying part and demonstrates that the axial dynamic stretching along that portion is zero. It should be said that the axial modes provide the pattern for the variation of the dynamic tension amplification as well. Along the suspended, part the axial mode shapes obey the rule of $n-1$ crossings with the horizontal, with n being the successive mode number. An important feature that is common with the normal motions as well (Fig. 38.3) is related to the progressive decay of the mode magnitude for increasing mode number.

The normal modes, however, do not cross the horizontal axis (the s coordinate) $n-1$ times. On the contrary, there are n crossings for the n-th mode. This immediately distinguishes the actual mode shapes from the simplified sinusoidal modes, implying that the latter are indeed a deficient approach that involves a high

rate of uncertainty in terms of the final accuracy as it omits the all-important weight of the cable. It should be also observed that the complete length of the cable, including the part lying on the bottom, is set into motion in the normal direction. This is the outcome of the assumption of a perfectly horizontal bottom. Evidence of the fact that a part of the cable resting on the bottom is not shown in Fig. 38.3. The exact location of the touch-down point, however, is immediately detectable a fortiori in the mode shapes of the shear force and secondarily in the modes of the bending moment. The touch-down point can be realized as a point of discontinuity, and that is clearly demonstrated in Fig. 38.4.

The out-of-plane vibrations in the bi-normal direction are examined through the depictions of Figs. 38.6–38.8, which provide, respectively, the motions, the shear force, and the out-of-plane bending moment. The last two parameters are again very small and indicative of the fact that bending effects are negligible for cables. For line structures such as cables, however, the flexural rigidity can be effectively used as a tool for avoiding numerical difficulties in regimes that are characterized by occurrences of potential singularities. The magnitude of the out-of-plane modes (Fig. 38.6) is again reduced for increasing mode number. That remark assures beyond any doubt that the employment of the actual mode shapes in a series expansion of potential solutions will result in a convergent formulation. Further, it should be mentioned that the out-of-plane mode shapes are characterized by asymmetry, contrary to the frequently employed harmonic shapes which are symmetrical with respect to the midpoint. The maxima, which are more easily detectable in the first mode, shift towards the bottom point, which, accordingly, implies that the maximum motions should be expected to occur at that region. The last remark is also valid for the bending moments and the shear forces as well (Figs. 38.7 and 38.8).

38.3 High Tension Cables; Snap-Slack Conditions

The present section investigates an indicative *practical* case that experiences the impact of high dynamic tension amplification. The cable is purposely guided to an excitation regime that eventually cancels the total tension applied on the cable. The term *high* herein is correlated with dynamic tension and it is distinguished from static tension. In fact, when the latter is sufficiently large, it is increasingly difficult for the total tension to be canceled, thus discouraging possible dynamic instabilities.

In several practical applications high tension and low tension realms may occur simultaneously, say during a period of the response. Other practical applications involve only low tension cables, e.g., deployed cables. In this section, emphasis is on the former, which, among others, allow investigation of the so-called snap-slack phenomenon. In particular, snapping occurs during maximum tension amplification and can be interpreted as a shock loading that incurs the complete length of the cable, and unwanted slacking occurs whenever the tension is canceled. Relevant examples are the highly tensioned tendons of tension leg platforms or cables that are used for the deployment of instrumentation packages on the seafloor or cables that tow underwater vehicles.

To avoid tension cancelation and the inevitable collapse of the numerical scheme that treats the dynamic equilibrium problem, the most advisable method is to include the bending effects in the mathematical formulation. This allows the cable to undertake the so-called compression loading that is governed by the induced negative tension [38.16]. In other words, the cable stops functioning as a chain and becomes a beam that experiences the action of bending. Nevertheless, in reality the very small flexural rigidity is incapable of allowing large compression that would lead to buckling. Practically, the tension never becomes zero.

There have been studies, mostly experimental, that explicitly show the occurrence of the concerned phenomenon. For more details, the interested reader is referred to the works due to *Papazoglou* et al. [38.5] and *Mavrakos* et al. [38.17]. The most difficult part, however, as far as the computations are concerned is the simulation of the time interval during which the total tension approaches zero at specific location(s) along the cable. Relevant studies avoid the numerical difficulties involved by artificially letting the tension be zero [38.5, 18, 19]. This approximation, however, has specific limitations and does not provide answers for the small negative tension that is applied on cables.

It is apparent that in order to capture the compression loading regime, a robust and reliable methodology should be developed and employed. The complete system assumes the involvement of the flexural rigidity and, in fact, this is the proper way to proceed. To obtain successful computations an inevitable price has to be paid. This is associated with the employment of an implicit numerical scheme and the use of dense discretizations in both time and space. For the purposes of the present contribution this is achieved by employing an unconditionably stable implicit methodology.

Without providing exhaustive details, it is admitted that the most popular solution techniques rely on finite element methods. The latter have provided the tools to de facto standard software programs in the maritime and offshore industry for analyses of submerged cables, e.g., RIFLEX [38.20]. The general finite element method has been widely covered in the literature during the last decades and, thus, only two indicative examples are cited herein. These are the studies due to *Wang* et al. [38.21] and *Aamo* and *Fosen* [38.22].

Use of the method of finite differences on cable dynamics problems was first proposed by *Ablow* and *Schechter* [38.23], and their method was accordingly modified by *Milinazzo* et al. [38.24]. Improved schemes of the finite difference approximation were subsequently developed and employed in a series of studies on a variety of cable dynamics problems, such as low tension cables [38.25, 26], cable deployment [38.27], free hanging lines [38.28], highly extensible cables [38.29], and mooring line dynamics [38.30]. The finite differences method is explicitly mentioned here as it is the method that was applied for the derivation of the present numerical results. In particular, both the nonlinear and linearized systems were treated by the centered differences scheme [38.31]. For the nonlinear system, the centered differences were applied for the discretization of both temporal and the spatial nodes, resulting in a second-order accurate $O(\Delta t^2) + O(\Delta s^2)$ algorithm. For the linear system the centered differenced approximation is required only for the spatial discretization, again resulting in a second-order accurate scheme. It should be mentioned that the method developed focuses on two-point boundary value problems and accounts for only first derivatives in time and space. If additionally second-order derivatives have to be considered, say if torsion is included, then a more complicated scheme has to be constructed. Here the final system was elaborated numerically by the relaxation method [38.22].

The numerical results presented in this section concern a TLP tendon with the following physical and mechanical properties: length 100 m; installation depth 100 m; mass 20.8 kg/m, added mass 5.15 kg/m; Young's modulus 0.8×10^9 N/m^2; diameter 0.08 m; submerged weight 153.5 N/m; normal drag coefficient 1.0; and tangential drag coefficient 0. The pretension at the top was equal to the total submerged weight of the cable, whereas the cable was assumed to be excited by a combined monochromatic excitation in surge and heave in its plane of reference with equal frequency 1.5 rad/s and zero phase amplitudes 1 m and 2.5 m, respectively.

The solution to the particular problem was sought by applying the complete 3-D nonlinear system of (38.38)–(38.47); albeit the physical problem is defined only in 2-D. Simply, no binormal motion was employed. For the solution of the system the governing equations were supplemented by the boundary conditions, which assumed pinned ends and zero motions at the bottom, whilst the external motions at the top explicitly defined the top boundary conditions. Only indicative results are shown herein, which emphasize the most interesting regime, namely the slacking condition. If the effective tension including its dynamic amplification is going to be canceled, that will occur primarily at the bottom point where the static tension admits the lowest value. Hence, particular attention is paid to the lower connection of the cable; the exerted loading is depicted in Figs. 38.9 and 38.10, where the total effective tension and the shear force are given, respectively. The cable was discretized by 600 nodes and the computations were run with a time step equal to 0.01 s. It should be admitted that the preceding figures correspond to a relatively coarse grid and sufficiently large time step. However, convergence at each time step is achieved with only five to six iterations. These remarks imply a very efficient, stable, and robust numerical scheme, contrary to the criticism that the box (centered-centered) method received by *Gobat* and *Grosenbaugh* [38.32].

Coming back to the physical model and the results on its dynamic behavior, the first (and the most important) feature that attracts our attention is associated with the zero effective tension (here shown for the bottom end point). The zero tension interval when the cable slacks, is followed by a rapid increase in tension, which can be characterized as the snapping interval. It is also

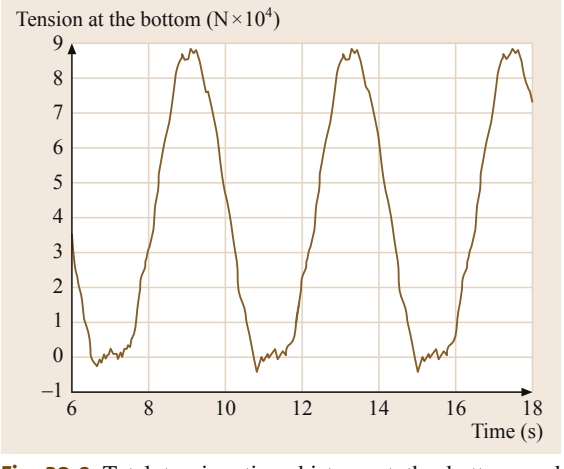


Fig. 38.9 Total tension time history at the bottom end point of a vertical tendon under snap-slack loading (amplitude of excitation = 1 m surge, 2.5 m heave; freq. = 1.5 rad/s)

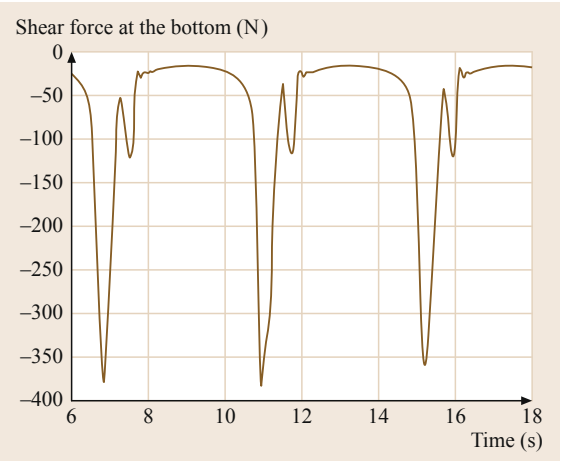


Fig. 38.10 Total shear force time history at the bottom end point of a vertical tendon under snap-slack loading (amplitude of excitation = 1 m surge, 2.5 m heave; freq. = 1.5 rad/s)

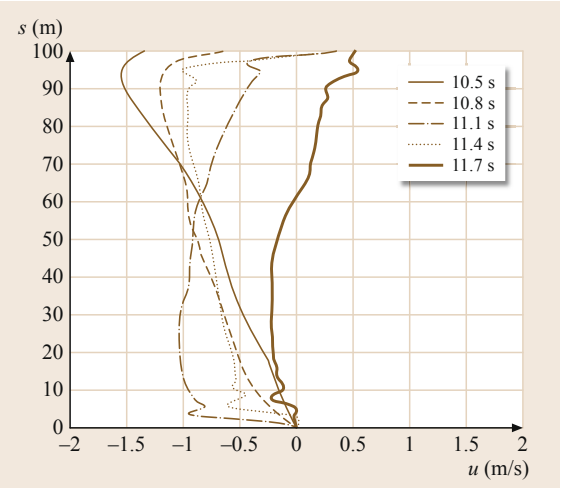


Fig. 38.11 Snapshots of the axial structural velocity along the cable during the slack condition interval. The snapshots correspond to 10.5, 10.8, 11.1, 11.4, and 11.7 s (see also Fig. 38.9)

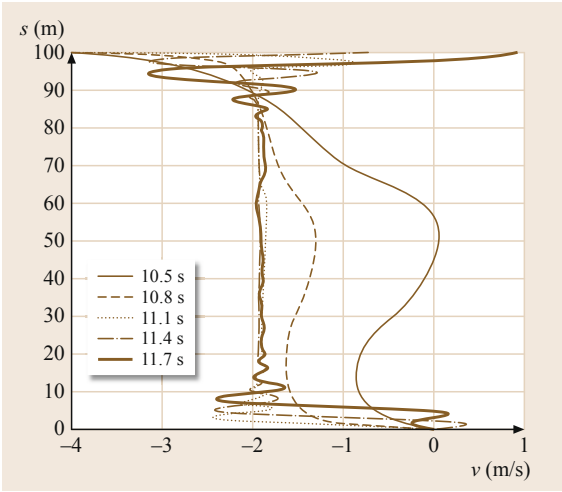


Fig. 38.12 Snapshots of the normal structural velocity along the cable during the slack condition interval. The snapshots correspond to 10.5, 10.8, 11.1, 11.4, and 11.7 s (see also Fig. 38.9)

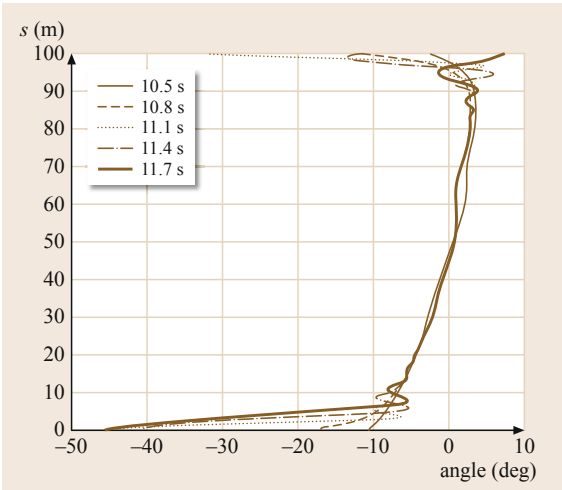


Fig. 38.13 Snapshots of the angle along the cable during the slack condition interval. The snapshots correspond to 10.5, 10.8, 11.1, 11.4, and 11.7 s (see also Fig. 38.9)

important to note that the cable indeed admits a small negative tension during which a slight compression occurs. During the extreme stretching of the cable the shear force approaches zero (Fig. 38.10), whereas the mathematical formulation during slacking is supported by the bending terms (see the sharp increase in the shear force of Fig. 38.10).

Details on the actual configuration of the tendon during the zero tension interval are given in Figs. 38.11–38.13, where the axial structural velocity, the normal structural velocity and the angle with respect to the horizontal are shown. The configuration of the cable during slacking can be assessed through Fig. 38.12, where the normal (along the x axis) velocity is provided. According to that depiction, the cable can be assumed vertical as along the major part of its length the velocity is nearly constant, but the lower region slacks profoundly. Analogous characteristics can be seen in the snapshots of the angle which reaches approximately 45° with respect to the perpendicular. During the interval concerned the loss of tension causes the destabilization of the cable.

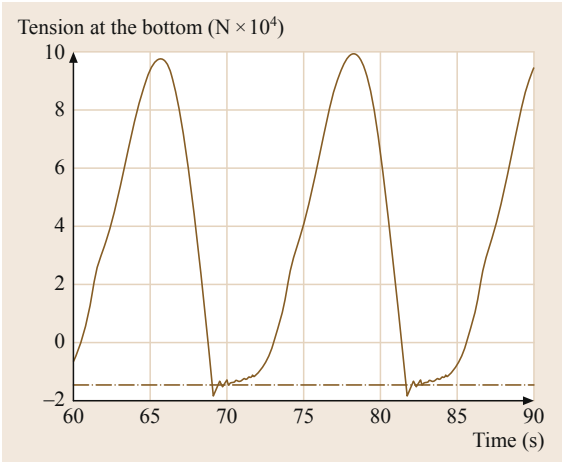


Fig. 38.14 Dynamic tension time history at the bottom end point of a vertical tendon under snap-slack loading (amplitude of excitation = 3 m surge, 3.5 m heave; freq. = 0.5 rad/s)

Further, a more visible slack cable condition is investigated and discussed. Here the frequency of excitation, always applied at the top, was reduced to 0.5 rad/s to obtain an extended interval during which the dynamic tension becomes negative and cancels the static tension. Due to the slower vibrations of the cable, the amplitudes of the excitation were considered to be sufficiently large, i. e., 3.0 and 3.5 m along the x and the z directions, respectively. Focus is on the dynamic tension amplification time history, and the associated signal is provided in Fig. 38.14. The chained line was artificially introduced to show the level of the static tension. Again, it can be deduced that virtually the cable is not capable of undertaking any bending effects and thus responds with nearly zero effective tension. This is due to the very small bending stiffness that originates from the small cross-sectional area. It should be mentioned that we would not be able to obtain the depicted results if the bending stiffness were omitted. The omission of EI would result in an ill-posed problem (according to the term used by *Triantafyllou* and *Howell* [38.16]), which becomes well posed after the inclusion of the bending stiffness. In Fig. 38.14 it is easy to see that the total effective tension instantly becomes negative and accordingly reverts to positive (although small) values. In fact, the zero tension interval is not practically *zero*. The total tension advances with very small positive increments until extreme stretching is attained. It should be noted that the depicted behavior has been verified experimentally by *Papazoglou* et al. [38.5].

38.4 Dynamics of Catenary Moorings

The technological advances of offshore industry to deeper water fields required the use of moored facilities that are restrained in waves, not necessarily with tendons or taut inclined cables. In fact, the most popular mooring method is the use of catenary mooring systems that provide the necessary restoring characteristics to the moored platform through the submerged weight of the mooring lines. Admittedly there is a huge amount of relevant studies in the literature which deal with both statics and dynamics of catenary moorings, the majority of which have been presented at specialized conferences. Nevertheless, the classical book of *Irvine* [38.33] should be mentioned.

Typically, the dynamics of catenary moorings are investigated in terms of the slowly varying and the wave frequency motions of the moored floater. The former can be investigated quasi-statically, meaning that the loading exerted on the catenary can be approximated by the static equilibrium solution, except in the case where slowly-varying drift loads are taken into account. The wave frequency motions, however, cannot admit that approach and, hence, the solution of the dynamic equilibrium system is required. This can be performed either by using time or frequency domain techniques. Here, both solution schemes have been developed but the numerical results presented in the sequel were obtained only by the frequency domain approach in (38.69) and (38.70), which provides a fast estimate of the actual dynamics of a mooring cable. This section is divided into three parts: the first part is dedicated to the dynamics of catenaries alone, the second part considers intermediate buoys attached on the mooring line, and the last part addresses the so-called mooring-induced damping phenomenon. Only indicative results are presented in an attempt to cover the most of the important characteristics as regards the dynamic behavior of mooring cables in a wet environment.

38.4.1 A Single Mooring Cable

The results and the subsequent discussion concern a single mooring line with uniform physical and mechanical properties along its complete length, taken from the work of *Mavrakos* and *Chatjigeorgiou* [38.34]. In particular, the cable is 3000 m long installed in a 375 m deep water field. The horizontal pretension at the top is 7.8×10^6 N. Also, the cable has the following properties: mass 94.5 kg/m, added mass 16 kg/m, submerged

weight 770 N/m, diameter 0.141 m, and elastic stiffness 10^9 N. The normal and the tangential drag coefficients were set equal to 1.0 and 0.0, respectively. Finally, a single excitation amplitude of 2 m was applied separately in the horizontal and bi-normal directions for the in-plane and the out-of-plane problems, respectively. It should be mentioned that the static configuration of the model concerned has a small part (approximately equal to 200 m) lying on the assumed horizontal seafloor. This could be approximated by a spring and a damper to simulate the elastic stiffness and the Coulomb friction. However, here we rely on its relatively insignificant length compared to the suspended part and we omit the associated effects including the friction with the bottom. The cable was considered fully suspended and approximation was allowed by the relatively small part resting at the bottom. Given the fact that that part was not particularly formulated, it is expected that its occurrence will not be reflected on the variation graphs of the dynamic components along the cable.

The most important dynamic parameter as far as the dynamics of mooring cables are concerned is the tension applied at the top. This is because the dynamic tension amplification supplements the static counterpart, which obtains its maximum at the top terminal point. Relevant results for the dynamic tension amplification at three points along the cable are given in Fig. 38.15. The transfer function of the dynamic tension amplification of Fig. 38.15 must be seen in correlation with the corresponding curve of Fig. 38.20, as the solution methods employed are different. Figure 38.15 assumes the cable to be a slender beam with $EI = 1$ (38.69), whilst the numerical predictions were obtained by the centered differences approximation along with the relaxation method outlined briefly in Sect. 38.3. On the contrary, the corresponding results of Fig. 38.20 were delivered by the artificial centered differences method described in [38.35]. It can immediately be seen that the comparisons as far as the single cable case is concerned are extremely favorable, which in turn proves the efficacy and the accuracy of both methods.

There are several interesting characteristics in Fig. 38.15 that must be properly highlighted and discussed. First, all resonances predicted by the eigenvalue problem are canceled by the strong action of the drag forces. Further, the dynamic tension amplification as predicted by linear theory at the top end is quite low,

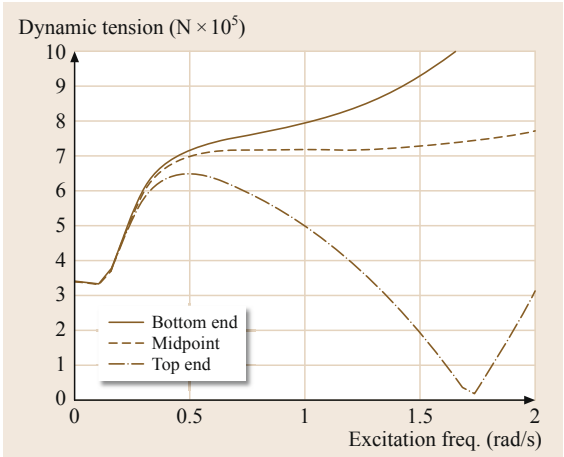


Fig. 38.15 Dynamic tension amplification for three points along a catenary cable (at the top, middle, and bottom end of the suspended part) under in-plane surge excitation at the top with amplitude $\alpha_x = 2$ m

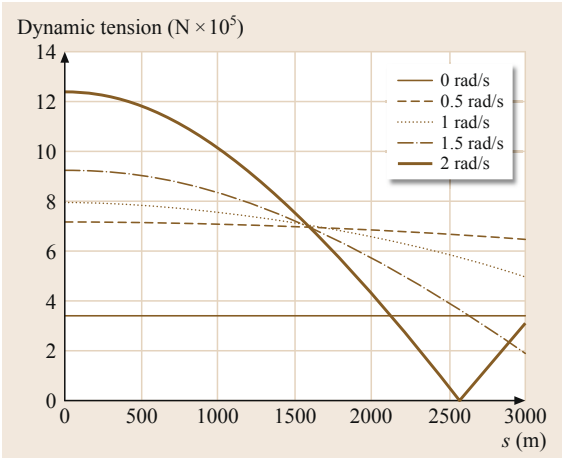


Fig. 38.16 Dynamic tension amplification along a catenary cable under in-plane surge excitation at the top with amplitude $\alpha_x = 2$ m and at four distinct frequencies

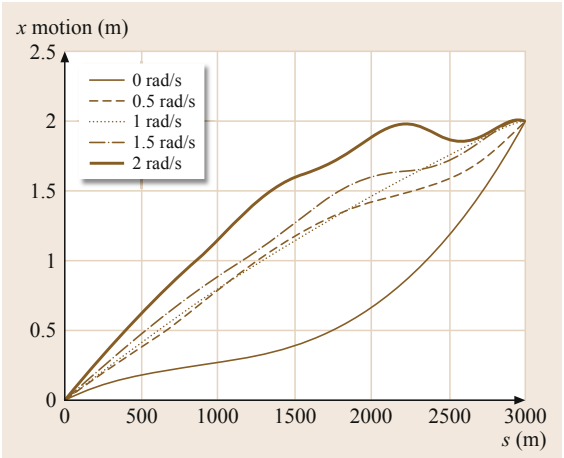


Fig. 38.17 Surge structural motions along a catenary cable under in-plane surge excitation at the top with amplitude $\alpha_x = 2$ m and at four distinct frequencies

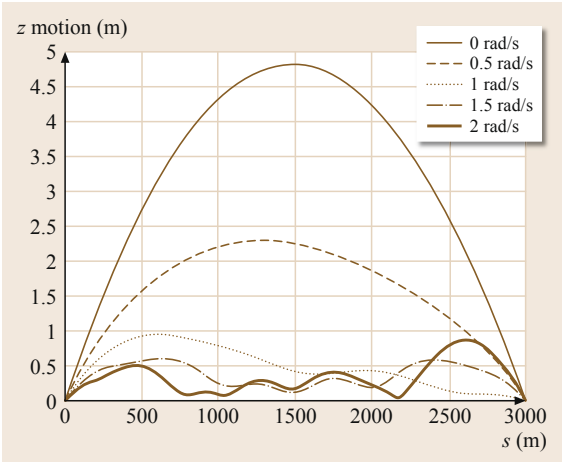


Fig. 38.18 Heave structural motions along a catenary cable under in-plane surge excitation at the top with amplitude $\alpha_x = 2\,\text{m}$ and at four distinct frequencies

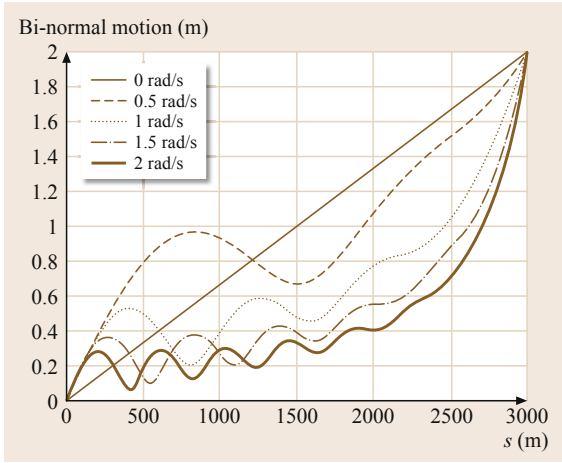


Fig. 38.19 Bi-normal structural motions along a catenary cable under out-of-plane excitation at the top with amplitude $\alpha_b = 2\,\text{m}$ and at four distinct frequencies

here approximately 10% of the static tension. In addition, it progressively decays at the high frequency range, reaches a lower cancelation point and, accordingly, exhibits a sharp increase at a frequency that coincides with the first elastic eigenfrequency.

Clearly, the dynamic tension is not equal along the cable, as implied by the three depicted curves. They admit comparable magnitudes only at the low frequency range and after that the dynamic tension obtains very large magnitudes close to the lower portion of the suspended part. In fact, the illustrated curves remedy the glitch of the assumption due to *Papazoglou* et al. [38.5], according to which the dynamic tension is constant along the suspended part, which in turn allowed the simplification of the formula for the dynamic tension calculation. In fact, this is true only for low frequencies, as is explicitly demonstrated in Figs. 38.15 and 38.16.

Typically for marine mooring cables, the variable of concern is the axial loading and especially the tension that is applied at the fairlead. However, it is undoubtedly interesting to realize how the group of governing parameters varies along the cable, and this task is performed with the aid of Figs. 38.16–38.19. Figures 38.16–38.18 concern the in-plane vibrations and provide the tension and the surge, and heave motions along the suspended part for five equally spaced frequencies. Again, it can be seen that the dynamic tension is nearly constant only for small frequencies. Strong variations occur for higher frequencies, whereas the tension at the bottom is increased for faster oscillations. The applied tension is obviously related with to induced motions that exhibit strong variations as the excitation frequency is increased (Figs. 38.17 and 38.18). The heave structural oscillations in particular tend to vanish for fast imposed motions as manifested by the 2.0 rad/s curve of Fig. 38.18.

It should be admitted that the vast majority of studies on mooring dynamics is performed in 2-D and the cable's plane of reference. This is due to the fact, as previously said, that the parameter of concern is the tension. To enrich the value of the present chapter, additional results concerning the bi-normal behavior are provided. These were obtained through the matrix system (38.70) and the bending stiffness was again set equal to unity. The linear system assumes decoupled in-plane and bi-normal motions leading to the consideration that the bi-normal behavior does not add to dynamic tension amplification.

This is indeed true and it can be demonstrated by the solution of the nonlinear problem in (38.38)–(38.47). In addition, the shear loading that is developed along the cable is inevitably small due to the negligible bending stiffness. The same is valid for the bending moment as well. Thus, only the out-of-plane motions are considered herein, and the results obtained for 2 m bi-normal top imposed excitation are shown in Fig. 38.19. The quasi-static excitation with $\omega \to 0$ delivers linear variation along the cable, as expected. Large bi-normal vibrations occur for small frequency excitations, which accordingly decay in magnitude and exhibit strong wavy trends for faster imposed motions.

38.4.2 Effect of Attached Intermediate Buoys

The use of intermediate buoyancy devices (buoys) attached along mooring cables is, in fact, a practical technique that aims at altering the static and dynamic

characteristics of the actual line through reducing its effective submerged weight. This technique results in immediate differentiations in the dynamics of the mooring line as compared to the same line without the buoys. Also, the formulation and the solution techniques should be updated in order to properly take into account the various segments that are formed between the buoys regardless of whether or not they have the same properties. Additionally, the complicated mooring line-buoys system requires considering the dynamics of the buoys as well, which depend, among others things, on their geometry, their physical properties, and the points of attachment.

The number of studies reported on the subject is significantly lower than for those dealing with simple moorings. Moorings with buoys have been studied by [38.34, 36–44], to mention a few.

Here, the outlined analysis is based on the theory and the solution method presented by the present authors in [38.34], who, among others, investigated potential impacts due to buoy rotation. The solution method adopted in that respect was based on a centered differences algorithm analogous to that developed by *Bliek* [38.35]. The numerical predictions discussed in the following refer to the linearized problem of the mooring line described in Sect. 38.4.1. For the purposes of the present section, the specific mooring line was assumed to have a buoy attached at 2450 m from the bottom. The buoy was assumed to be spherical with a drag coefficient equal to 0.8. The detailed properties of the buoy can be found in [38.34]. For a detailed discussion on the effect of multiple buoys attached at various locations along mooring cables with different segment properties along with experimental results, the interested reader can refer to the study due to *Mavrakos* et al. [38.36].

In the following, the results of three indicative cases are provided which are distinguished by the buoy's net buoyancy offered to the cable, i. e., 20, 35, and 50% of the cable's total submerged weight. The excitation at the top was assumed to be horizontal and of 2 m magnitude. For reference, the no-buoy case of the single mooring line was considered as well. The predictions shown in Figs. 38.20 and 38.21 depict the transfer functions of the dynamic tension amplification applied at the top (Fig. 38.20) and immediately below the buoy (Fig. 38.21).

It is immediately apparent that buoy's insertion at some point along the mooring line modifies the top applied tension (as well as the tension along the complete length of the line) profoundly and, accordingly, the line's restoring characteristics offered to the floating vehicle. The first specific remark concerns the differences in the tension at the top and immediately below the buoy. The latter are greater than the former, in accordance to the findings of the previous section. Moreover, the buoy, indeed, reduces the tension applied at the top, although not drastically.

The most important feature however, originates from the cancelation frequency and, accordingly, from the subsequent elastic frequency. The cancelation frequency shifts toward slower excitations and approaches the region of the typical wave frequencies encountered in the sea. The tension at the top appears to be influenced increasingly by the amount of the net buoyancy, whereas the tension immediately below the buoy is not

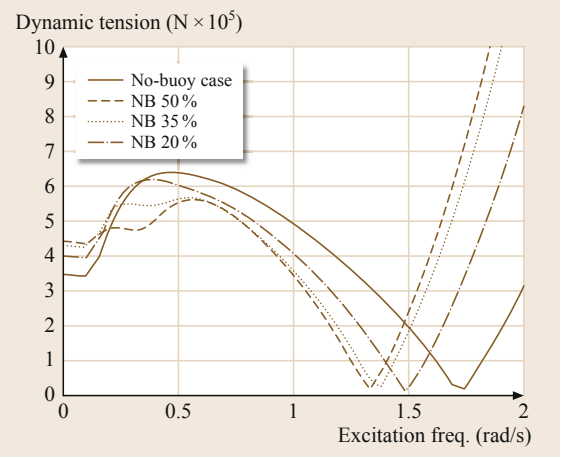


Fig. 38.20 Dynamic tension amplification at the top of the mooring cable under horizontal monochromatic excitation with amplitude $\alpha_x = 2\,\mathrm{m}$

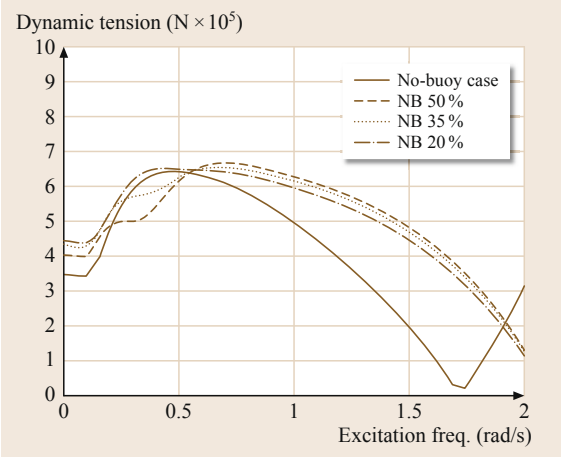


Fig. 38.21 Dynamic tension amplification immediately below the buoy. The cable was assumed excited at the top with horizontal monochromatic motion with amplitude $\alpha_x = 2\,\mathrm{m}$ (the nobuoy legend applies to the tension at the top)

affected so much by the parameter concerned. Further, contrary to what is happening at the top, the cancelation frequency for the tension immediately below the buoy admits higher values. It is apparent that this is due to the fact that the segment at this location can be regarded as a separate line with different properties and, accordingly, with different eigenvalue particulars.

Figures 38.20 and 38.21 indeed prove the efficacy of buoys at reducing the dynamic tension amplification. As the dynamic tension is the quantity that adds to the possibility of failure of a mooring cable, the reduction mechanism introduced by buoys is of paramount importance, especially in survival conditions. It should be mentioned, however, that the proposed technique requires a comprehensive survey in terms of the optimum size, number, and location of the attached buoys. It should be admitted that not much work has been done in this direction. A relevant example is the contribution mentioned in [38.45].

38.4.3 Mooring Line-Induced Damping

The concept of the so-called mooring line-induced damping phenomenon was introduced originally in the 1980s by *Huse* [38.46] and *Huse* and *Matsumoto* [38.47]. According to that, the oscillations of the mooring line into a wet environment, which are caused by the motions of the moored floater, induce energy dissipation that originates mainly from the damping mechanism of the drag forces. According to the inspirers of that idea, which was a breakthrough idea at the time, the loss of energy can be translated into a nonconservative force that absorbs energy and adds to the total damping on the moored structure. *Huse* [38.46] estimated that mooring line-induced damping can add up to 25% of the total damping rate and hence, the omission of that term overestimates the magnitudes of the floater motions.

There were several studies after *Huse* [38.46] that dealt with the mooring-induced damping phenomenon. *Liu* and *Bergdahl* [38.48] proposed an improvement to the *Huse* [38.46] model. *Brown* and *Lyons* [38.49] claimed that the mooring system may under certain circumstances provide up to 80% of the total damping available, and thus significantly reduce the peak line tensions. Further, *Brown* and *Mavrakos* [38.50] presented the results of a comparative study initiated by the International Ship and Offshore Structures Congress (ISSC 1997), Committee I2 (Loads), whereas *Bauduin* and *Naciri* [38.51] addressed the problem only quasi-statically. Contemporary, focus is made on the damping induced by moorings of wave energy converters (WEC) [38.52]. Recently, *Chatjigeorgiou* et al. [38.53] introduced the so-called risers-induced damping using the same formulation as for cables.

Mooring line-induced damping is quantified by an equivalent linearized damping coefficient that can be introduced into the dynamic equilibrium system of the moored floater, the latter being assumed to act as a single or multiple degree oscillator. A prerequisite is the solution of the dynamic equilibrium problem of the mooring line, and evidently in a multi-leg mooring system the total mooring line damping is composed by the superposition of the individual damping coefficients of all mooring lines. Hence the linearized coefficients can be obtained using either frequency domain or time domain techniques. The former require the solution of the linear dynamic equilibrium problem in complex space. Accordingly, the linearized damping coefficient is obtained through the imaginary part of the horizontal component of the dynamic tension applied at the top according to

$$c_x = \Im\left[\frac{F_x}{\alpha_x \omega}\right], \tag{38.71}$$

where α_x denotes the amplitude of the horizontal motion applied at the top and ω is the frequency of the excitation. Clearly, an analogous relation holds for the mooring-induced damping in heave direction. The same expression can be used for the calculation of linearized damping in the bi-normal direction as well, provided that F_x is replaced by the out-of-plane shear force $\bar{S}_{b1}$ applied at the top, and α_x by the associated imposed motion. It is evident, therefore, that out-of-plane linearized damping requires the solution of the homonymous linear problem and, accordingly, the inclusion of the bending stiffness into the governing equilibrium problem.

The derivation of the damping coefficient via time domain solution techniques requires the evaluation of the energy dissipated per cycle [38.50]. This is obtained by

$$E_k = \int_0^\tau F_k \frac{\mathrm{d}q_k}{\mathrm{d}t}\,\mathrm{d}t\,, \tag{38.72}$$

for $k = x, y, z$, and τ denotes the period of the response. Here, y denotes the out-of-plane direction. According to the notations used for formulating the nonlinear dynamic equilibrium problem, (38.72) admits the following expressions in a Cartesian system

$$E_x = \int_0^\tau (T\cos\phi)(u\cos\phi - v\sin\phi)\,\mathrm{d}t\,, \tag{38.73}$$

$$E_y = \int_0^\tau S_\mathrm{b} w \, \mathrm{d}t \,, \tag{38.74}$$

$$E_z = \int_0^\tau (T \sin\phi)(u \sin\phi + v \cos\phi) \, \mathrm{d}t \,. \tag{38.75}$$

Clearly, the contribution of the in-plane shear force was ignored in (38.73) and (38.75) due to the negligible in-plane shear force S_n compared to the extreme values of tension T. It should be noted that the integration can be performed either by using total quantities or their dynamic counterparts only. The results should be the same given the fact that the static component is constant. Accordingly, the linearized damping coefficients are given by

$$c_k = \frac{E_k \tau}{2\pi^2 a_k} \,, \tag{38.76}$$

for $k = x, y, z$, where a_k denotes the oscillation amplitude associated with the period τ.

Indicative results for the linearized mooring-induced damping coefficients and the energy dissipation loops are provided in Figs. 38.22–38.24. Again, the subject of investigation was the mooring cable described in Sect. 38.4.1. The in-plane dynamics are considered with and without buoys (Fig. 38.22), whereas the dynamic behavior in the bi-normal direction assumes no buoy connections (Fig. 38.23). Both figures depict calculations that have been obtained via the frequency domain approximation.

Observing the curves shown in Fig. 38.22, it is deduced that the variation pattern of the damping coefficients as a function of the excitation frequency is dictated by the trends of the dynamic tension (Fig. 38.20). The progressive reduction of the dynamic tension and, accordingly, of its horizontal component, before the cancelation point and the first elastic eigenfrequency, combined with the increased excitation frequency results in a relative sharp decrease of the damping coefficient. It is also interesting to note that the local maximum of the damping coefficient does not coincide with the maximum of the tension. In addi-

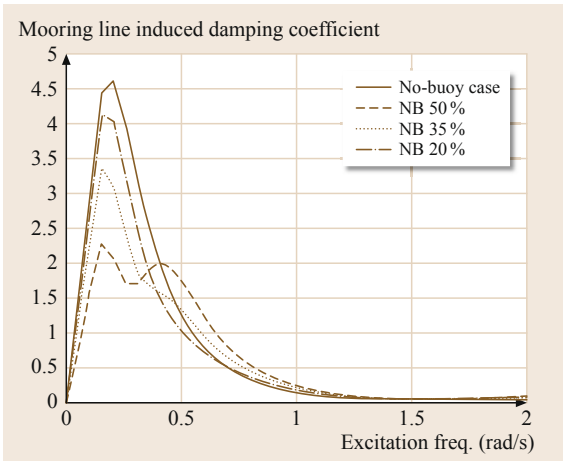


Fig. 38.22 Mooring line-induced damping coefficient in surge direction $c_x = \Im[F_x/\alpha_x\omega] * 10^{-5}$. The cable was assumed to be excited at the top with a horizontal monochromatic motion with amplitude $\alpha_x = 2\,\mathrm{m}$

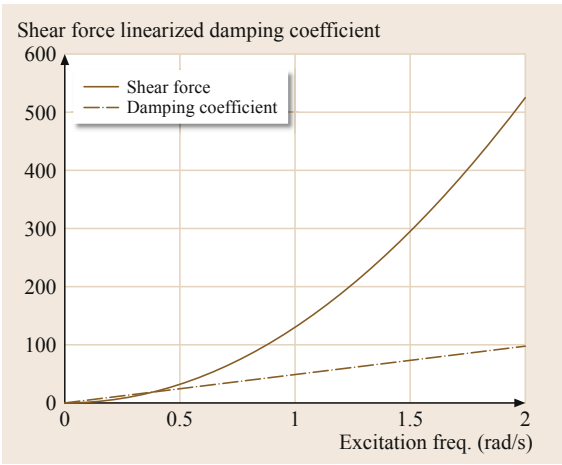


Fig. 38.23 Shear force $\bar{S}_{\mathrm{b}1}$ (N) and mooring line-induced damping coefficient in the bi-normal direction $c_\mathrm{b} = \Im[S_\mathrm{b}/\alpha_\mathrm{b}\omega]$ [N/m/s]. The cable was assumed excited at the top with a bi-normal monochromatic motion with amplitude $\alpha_\mathrm{b} = 2\,\mathrm{m}$

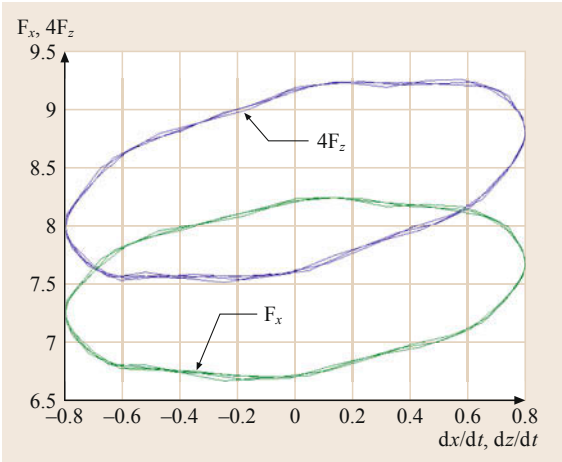


Fig. 38.24 Energy dissipation loops in x and z directions obtained via the 3-D time domain simulation of the dynamic response of a mooring cable subjected x, y, and z monochromatic motions at the top with equal amplitude 2 m and excitation frequency 0.4 rad/s

tion, the use of buoy(s) along the line, albeit it has a positive effect as far as the dynamic tension amplification is concerned, results in damping reduction. Figure 38.22 provides results for increasing net buoyancy up to 50% of the total submerged weight. Clearly, the largest net buoyancy case results in a bump around 0.3 rad/s. It should be noted that numerical tests performed for even larger net buoyancy indicate that the maximum damping coefficient (the local maximum) shifts toward higher frequencies. This is indeed the direct outcome of the transfer function of the dynamic tension amplification which follows the same trend. Hence the conclusion that is easily deduced is that the selection of buoys should be properly put into an optimization procedure that must take into account several dynamic effects. It is important to note the magnitude of the damping coefficient, which admittedly obtains relatively large values. It is apparent that in the linear equation of motion of an equivalent oscillator, this figure will be divided by the (typically) large mass of the floater. However, it should be mentioned that the total mooring line damping will be composed of the individual damping terms of the constituent mooring lines in a multi-leg arrangement, and hence, it is expected that a non-negligible term will be constructed.

Figure 38.23 shows the out-of-plane damping coefficient of the same mooring line, again as a function of the excitation frequency. The calculations depicted were obtained via the frequency domain approach of the out-of-plane equilibrium system in (38.70). For reference, the bi-normal shear force at the top of the cable is provided as well. Contrary to the findings that refer to the in-plane vibrations of the cable, here the damping coefficient demonstrates an increasing trend and in particular a linear variation. Nevertheless, it should be emphasized that the bi-normal damping offered to the vessel is extremely small compared to the surge motion mooring line damping.

The last part of the present discussion is dedicated to the energy dissipation due to cable's oscillation in the marine environment. As has been said, this is determined via time domain techniques, and since the solution is performed in the time domain it is advisable to use the original system for completeness. The loops depicted in Fig. 38.24 were derived through the solution of the fundamental nonlinear dynamic equilibrium system in (38.38)–(38.47). In fact, the energy dissipation is the measure of the mooring line damping and is represented by the area contained in the loops depicted for the surge and heave motions of the floater. In the case depicted, the energy dissipated per cycle in surge is approximately four times greater than the energy in heave. The loops that refer to the bi-normal (sway) direction are not provided due to the insignificant value of the damping in which they result.

38.5 Second-Order Nonlinear Dynamics of Cables Formulated as Euler–Bernoulli Beams

38.5.1 Background

When a study focuses on nonlinear impacts, the obvious approach is the employment of time domain solution techniques. This approach provides only time signal(s) of the output component(s). To provide insight into the details of the response, further analyses are required, typically using Fourier transformation techniques, which convert the time domain results into the frequency domain. In fact, even for the simplest possible excitation, e.g., harmonic fluctuations, the Fourier analyses of the output signals describing the response of nonlinear systems will most probably reveal the occurrence of several harmonics with, evidently, different magnitudes [38.54].

The previous discussion that refers to time domain numerical calculations, ostensibly disagree with reported theoretical formulations on the subject. For example, the complete three-dimensional dynamic system outlined in the present (38.38)–(38.47) that was derived using a Newtonian procedure involves nonlinear terms up to the third-order. Therefore, the question that arises concerns the mechanism that stimulates the, hypothetically, infinite harmonics that occur in the form of integer multiplications of a fundamental harmonic.

The present section is dedicated to the development of an advanced frequency domain methodology for studying the second-order double-frequency nonlinear responses of catenary cables formulated as Euler–Bernoulli beams. Emphasis is given to marine applications where the distributed drag forces are very influential. The process provides insight to the actual mechanism that stimulates the higher-order harmonics of the response, with the cable excited by a monochromatic motion. To this end, the governing nonlinear set is elaborated by employing a perturbation approach. In fact, the associated expansion stops at the second harmonic but it is noted that the employment of the same procedure can capture all subsequent harmonics. Here, only the in-plane vibrations are taken into considera-

tion, but it should be noted that the same procedure could be applied to capture the associated details of the out-of-plane dynamics as well.

The method results in two distinct systems, which are solved separately and successively. The second-order system requires knowing the first-order dynamic components and, accordingly, the solution of the n-th order system requires knowing the dynamic components of all previous orders up to $(n-1)$.

Also, the outlined theoretical formulation is combined with a novel mathematical approach aiming at assessing the impact of the bottom-cable interaction phenomenon. The theoretical model relies on the assumption that the bottom end point oscillates around an average location, which is assumed to coincide with the static equilibrium touchdown point (TDP). The proposed model is incorporated in both the linear and the second-order systems.

38.5.2 The Governing 2-D Nonlinear System

The complete 3-D nonlinear set of (38.38)–(38.47) is reduced to 2-D after omitting the out-of-plane components. The tangential drag coefficient was also omitted whilst it was assumed that $e \approx 0$. The reduced system reads

$$\frac{\partial T}{\partial s} - S_\mathrm{n}\Omega_3 - w_0 \sin\phi - m\frac{\partial u}{\partial t} + mv\frac{\partial \phi}{\partial t} = 0\,, \tag{38.77}$$

$$\frac{\partial S_\mathrm{n}}{\partial s} + \Omega_3 T - w_0 \cos\phi + R_{\mathrm{d}n} - (m+m_\mathrm{a})\frac{\partial v}{\partial t} - m\frac{\partial \phi}{\partial t}u = 0\,, \tag{38.78}$$

$$\frac{\partial u}{\partial s} - \Omega_3 v - \frac{1}{EA}\frac{\partial T}{\partial t} = 0\,, \tag{38.79}$$

$$\frac{\partial v}{\partial s} + \Omega_3 u - \frac{\partial \phi}{\partial t} = 0\,, \tag{38.80}$$

$$EI\frac{\partial \Omega_3}{\partial s} + S_\mathrm{n} = 0\,, \tag{38.81}$$

$$\frac{\partial \phi}{\partial s} - \Omega_3 = 0\,. \tag{38.82}$$

The unknown dynamic components of the above system are written in vector form according to

$$\boldsymbol{V} = \begin{bmatrix} T & S_\mathrm{n} & u & v & \Omega_3 & \phi \end{bmatrix}^\mathrm{T}\,. \tag{38.83}$$

Again, the elements of vector $\boldsymbol{V}$ denote total quantities, i. e., the summation of static and dynamic components. The associated vector that describes the static configuration of the cable, obtained by the solution of the 2-D static equilibrium problem, is written as

$$\boldsymbol{V}_0 = \begin{bmatrix} T_0 & S_{\mathrm{n}0} & 0 & 0 & \Omega_{30} & \phi_0 \end{bmatrix}^\mathrm{T}\,. \tag{38.84}$$

Accordingly, the dynamic amplification terms are the elements of the vector

$$\boldsymbol{V}_1 = \begin{bmatrix} T_1 & S_{\mathrm{n}1} & u & v & \Omega_{31} & \phi_1 \end{bmatrix}^\mathrm{T}\,, \tag{38.85}$$

and they are functions of both s and t. In (38.84) and (38.85), the indices 0 and 1 are used to denote static and dynamic terms, whereas it is evident that $\boldsymbol{V}$ is obtained by adding $\boldsymbol{V}_0$ and $\boldsymbol{V}_1$.

Next, after introducing (38.83)–(38.85) into the governing set of (38.77)–(38.82) and omitting the static equilibrium components, the following dynamic system is derived,which provides only the dynamic amplification terms and the velocities along the cable

$$\frac{\partial T_1}{\partial s} - \Omega_{30}S_{\mathrm{n}1} - S_{\mathrm{n}0}\Omega_{31} - S_{\mathrm{n}1}\Omega_{31} - w_0\cos\phi_0\phi_1 - m\frac{\partial u}{\partial t} + mv\frac{\partial \phi_1}{\partial t} = 0\,, \tag{38.86}$$

$$\frac{\partial S_{\mathrm{n}1}}{\partial s} + \Omega_{30}T_1 + T_0\Omega_{31} + \Omega_{31}T_1 + w_0\sin\phi_0\phi_1 + R_{\mathrm{d}n1} - (m+m_\mathrm{a})\frac{\partial v}{\partial t} - m\frac{\partial \phi_1}{\partial t}u = 0\,, \tag{38.87}$$

$$\frac{\partial u}{\partial s} - \Omega_{30}v - \Omega_{31}v - \frac{1}{EA}\frac{\partial T_1}{\partial t} = 0\,, \tag{38.88}$$

$$\frac{\partial v}{\partial s} + \Omega_{30}u + \Omega_{31}u - \frac{\partial \phi_1}{\partial t} = 0\,, \tag{38.89}$$

$$EI\frac{\partial \Omega_{31}}{\partial s} + S_{\mathrm{n}1} = 0\,, \tag{38.90}$$

$$\frac{\partial \phi_1}{\partial s} - \Omega_{31} = 0\,. \tag{38.91}$$

Equations (38.86)–(38.91) constitute the system that will be further elaborated in the frequency domain to investigate the contribution of the second-order nonlinear effects. Clearly, the treatment of the nonlinear system of (38.86)–(38.91) using a frequency domain technique requires a sophisticated approximation, as apart from the nonlinear drag force, which admittedly can be effectively linearized, the mathematical formulation contains products of time-dependent terms. To this end, the system under consideration will be further manipulated using an appropriate perturbation technique [38.55, 56].

38.5.3 Perturbation Series Expansion

Let ε be a measure of the amplitude of motion. Accordingly, it is assumed that all time-varying terms can be expanded in perturbation series according to

$$\boldsymbol{V}_1 = \varepsilon\boldsymbol{V}_1^{(1)} + \varepsilon^2\boldsymbol{V}_1^{(2)} + \cdots\,, \tag{38.92}$$

where for $k = 1, 2$

$$V_1^{(k)} = \left[T_1^{(k)} \quad S_{n1}^{(k)} \quad u^{(k)} \quad v^{(k)} \quad \Omega_{31}^{(k)} \quad \phi_1^{(k)} \right]^T . \tag{38.93}$$

The same holds for the drag force as well. Hence,

$$R_{dn1} = \varepsilon R_{dn1}^{(1)} + \varepsilon^2 R_{dn1}^{(2)} + \cdots . \tag{38.94}$$

The drag force is expressed in terms of Morison's formula and, consequently, the perturbation components will be given by

$$R_{dn1}^{(k)} = -b\,|v|\,v^{(k)} , \tag{38.95}$$

where $b = 1/2\rho C_d d_O$. Given the fact that the second-order components are expected to be significantly smaller than the corresponding linear terms, the magnitude of the normal velocity is approximated as $|v| = |\varepsilon v^{(1)}|$.

Next, after introducing (38.92)–(38.94) into the system of (38.86)–(38.91), omitting the $O(\varepsilon^3)$ terms, and equating the terms of like powers of ε, the following first and second-order systems are derived.

First-Order System $O(\varepsilon)$

$$\frac{\partial T_1^{(1)}}{\partial s} - \Omega_{30} S_{n1}^{(1)} - S_{n0}\Omega_{31}^{(1)} - w_0 \cos\phi_0 \phi_1^{(1)} - m\frac{\partial u^{(1)}}{\partial t} = 0 , \tag{38.96}$$

$$\frac{\partial S_{n1}^{(1)}}{\partial s} + \Omega_{30} T_1^{(1)} + T_0 \Omega_{31}^{(1)} + w_0 \sin\phi_0 \phi_1^{(1)} + R_{dn1}^{(1)} - (m + m_a)\frac{\partial v^{(1)}}{\partial t} = 0 , \tag{38.97}$$

$$\frac{\partial u^{(1)}}{\partial s} - \Omega_{30} v^{(1)} - \frac{1}{EA}\frac{\partial T_1^{(1)}}{\partial t} = 0 , \tag{38.98}$$

$$\frac{\partial v^{(1)}}{\partial s} + \Omega_{30} u^{(1)} - \frac{\partial \phi_1^{(1)}}{\partial t} = 0 , \tag{38.99}$$

$$EI\frac{\partial \Omega_{31}^{(1)}}{\partial s} + S_{n1}^{(1)} = 0 , \tag{38.100}$$

$$\frac{\partial \phi_1^{(1)}}{\partial s} - \Omega_{31}^{(1)} = 0 . \tag{38.101}$$

Second-Order System $O(\varepsilon^2)$

$$\frac{\partial T_1^{(2)}}{\partial s} - \Omega_{30} S_{n1}^{(2)} - S_{n0}\Omega_{31}^{(2)} - w_0 \cos\phi_0 \phi_1^{(2)} - m\frac{\partial u^{(2)}}{\partial t} = S_{n1}^{(1)}\Omega_{31}^{(1)} - m v^{(1)}\frac{\partial \phi_1^{(1)}}{\partial t} , \tag{38.102}$$

$$\frac{\partial S_{n1}^{(2)}}{\partial s} + \Omega_{30} T_1^{(2)} + T_0 \Omega_{31}^{(2)} + w_0 \sin\phi_0 \phi_1^{(2)} + R_{dn1}^{(2)} - (m + m_a)\frac{\partial v^{(2)}}{\partial t} = -\Omega_{31}^{(1)} T_1^{(1)} + m u^{(1)}\frac{\partial \phi_1^{(1)}}{\partial t} , \tag{38.103}$$

$$\frac{\partial u^{(2)}}{\partial s} - \Omega_{30} v^{(2)} - \frac{1}{EA}\frac{\partial T_1^{(2)}}{\partial t} = v^{(1)}\Omega_{31}^{(1)} , \tag{38.104}$$

$$\frac{\partial v^{(2)}}{\partial s} + \Omega_{30} u^{(2)} - \frac{\partial \phi_1^{(2)}}{\partial t} = -\Omega_{31}^{(1)} u^{(1)} , \tag{38.105}$$

$$EI\frac{\partial \Omega_{31}^{(2)}}{\partial s} + S_{n1}^{(2)} = 0 , \tag{38.106}$$

$$\frac{\partial \phi_1^{(2)}}{\partial s} - \Omega_{31}^{(2)} = 0 . \tag{38.107}$$

The sets of (38.96)–(38.101) (first-order) and (38.102)–(38.107) (second-order) compose two discrete systems that can be treated separately. Virtually, the first system is identical to the system of (38.53), (38.54), (38.56), (38.57), (38.60), and (38.62) with the note that it is expressed in terms of velocities instead of displacements. Clearly, the solution of the second-order system requires knowing the solution of the first-order system. Assuming that the cable is excited at one end by a monochromatic excitation (without reference yet to the other end), this motion constitutes a boundary condition for the first-order (linear) problem. Unquestionably, the solution of that problem will yield equal frequency variations for the motions, rotations, and internal loading components, whereas the frequency of the response will coincide with the excitation frequency. On the contrary, we cannot assume that the second-order system is subjected to an external excitation. Therefore, the question which arises concerns the nature of the excitation experienced by the second-order system. To provide an appropriate answer we must first inspect the right-hand side terms of (38.102)–(38.105). According to these terms, the mechanism that forces the cable to respond at second order and subsequently induces the double-frequency effects can be regarded as a distribution of generalized nonlinear springs along the complete length of the cable, which mainly originate from geometric nonlinearities. Evidently, the main contribution arises from the geometric nonlinearity $\Omega_{31}^{(1)} T_1^{(1)}$.

38.5.4 The Frequency Domain Approach

In the context of the linear problem the vector of the unknowns is written as

$$V_1^{(1)} = \mathrm{Re}\{\boldsymbol{\xi}_1^{(1)} e^{i\omega t}\} , \tag{38.108}$$

where

$$\boldsymbol{\xi}_1^{(1)} = \left[T_1^{(1)} \quad S_{\mathrm{n}1}^{(1)} \quad u^{(1)} \quad v^{(1)} \quad \Omega_{31}^{(1)} \quad \phi_1^{(1)} \right]^{\mathrm{T}}$$

is a complex vector depending on ω and the spatial coordinate s. Alternatively, (38.108) admits the following form

$$\boldsymbol{V}_1^{(1)} = \frac{1}{2}\left[\boldsymbol{\xi}_1^{(1)} \mathrm{e}^{\mathrm{i}\omega t} + \bar{\boldsymbol{\xi}}_1^{(1)} \mathrm{e}^{-\mathrm{i}\omega t} \right], \tag{38.109}$$

where the bar above ξ denotes the equivalent complex conjugate.

Next, (38.109) is introduced into the system of (38.96)–(38.101), which accordingly yields the following matrix form

$$\frac{\mathrm{d}\left(\boldsymbol{\xi}_1^{(1)} \mathrm{e}^{\mathrm{i}\omega t}\right)}{\mathrm{d}s} + cc = \mathbf{K}^{(1)} \left(\boldsymbol{\xi}_1^{(1)} \mathrm{e}^{\mathrm{i}\omega t}\right) + cc\,, \tag{38.110}$$

where cc stands for the equivalent complex conjugate. The 6×6 square matrix associated with the linear problem is given by

$$\mathbf{K}^{(1)} = \begin{bmatrix} 0 & \Omega_{30} & \mathrm{i}\omega m & 0 & S_{\mathrm{n}0} & w_0 \cos\phi_0 \\ -\Omega_{30} & 0 & 0 & \mathrm{i}\omega(m+m_{\mathrm{a}}) + b|v| & -T_0 & -w_0 \sin\phi_0 \\ \dfrac{\mathrm{i}\omega}{EA} & 0 & 0 & \Omega_{30} & 0 & 0 \\ 0 & 0 & -\Omega_{30} & 0 & 0 & \mathrm{i}\omega \\ 0 & \dfrac{-1}{EI} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}. \tag{38.111}$$

A brief inspection of the second-order system, (38.102)–(38.107), immediately confirms that the associated variables will respond with the double of the frequency of the excitation. It is also evident that the quadratic products at the right-hand side terms of the same equations will yield some time-independent terms that will amplify the magnitudes of the double-frequency (2ω) components by a constant amount. These components will originate by the products of terms depending on $\exp(-\mathrm{i}\omega t)$ and $\exp(\mathrm{i}\omega t)$. As far as the solution $\boldsymbol{V}_1^{(2)}$ of the second-order problem is concerned, the latter remark suggests letting

$$\boldsymbol{V}_1^{(2)} = \mathrm{Re}\left\{ \boldsymbol{\xi}_{\mathrm{c}}^{(2)} + \boldsymbol{\xi}_1^{(2)} \mathrm{e}^{2\mathrm{i}\omega t} \right\}, \tag{38.112}$$

where $\boldsymbol{\xi}_{\mathrm{c}}^{(2)}$ and $\boldsymbol{\xi}_1^{(2)}$ denote, respectively, the second-order constant and the double-frequency unknown complex vectors, which are expanded as

$$\boldsymbol{\xi}_{\mathrm{c}}^{(2)} = \left[T_{\mathrm{c}}^{(2)} \quad S_{\mathrm{nc}}^{(2)} \quad u_{\mathrm{c}}^{(2)} \quad v_{\mathrm{c}}^{(2)} \quad \Omega_{3\mathrm{c}}^{(2)} \quad \phi_{\mathrm{c}}^{(2)} \right]^{\mathrm{T}}, \tag{38.113}$$

$$\boldsymbol{\xi}_1^{(2)} = \left[T_1^{(2)} \quad S_{\mathrm{n}1}^{(2)} \quad u^{(2)} \quad v^{(2)} \quad \Omega_{31}^{(2)} \quad \phi_1^{(2)} \right]^{\mathrm{T}}. \tag{38.114}$$

The term *constant* implies that the elements of $\boldsymbol{\xi}_{\mathrm{c}}^{(2)}$ are independent of time. However, they are still functions of the Lagrangian coordinate s. The fact that some of the elements of $\boldsymbol{\xi}_{\mathrm{c}}^{(2)}$ might be zero is irrelevant to the general analysis that is conducted herein. In any event, their actual values depend on the application and the structural model under consideration and they are given by the solution of the second-order problem.

Again, (38.112) is alternatively written as

$$\boldsymbol{V}_1^{(2)} = \frac{1}{2}\left[\boldsymbol{\xi}_{\mathrm{c}}^{(2)} + \bar{\boldsymbol{\xi}}_{\mathrm{c}}^{(2)} \right] + \frac{1}{2}\left[\boldsymbol{\xi}_1^{(2)} \mathrm{e}^{2\mathrm{i}\omega t} + \bar{\boldsymbol{\xi}}_1^{(2)} \mathrm{e}^{-2\mathrm{i}\omega t} \right]. \tag{38.115}$$

The preceding equation is introduced into the system of (38.102)–(38.107), which after several mathematical manipulations provides the following matrix equation

$$\begin{aligned} &\frac{\mathrm{d}\boldsymbol{\xi}_{\mathrm{c}}^{(2)}}{\mathrm{d}s} + \frac{\mathrm{d}(\boldsymbol{\xi}_1^{(2)} \mathrm{e}^{2\mathrm{i}\omega t})}{\mathrm{d}s} + cc \\ &= \mathbf{K}^{(2)} \left(\boldsymbol{\xi}_1^{(2)} \mathrm{e}^{2\mathrm{i}\omega t}\right) + \frac{1}{2}\boldsymbol{N}^{(2)} \mathrm{e}^{2\mathrm{i}\omega t} + \frac{1}{2}\tilde{\boldsymbol{N}}^{(2)} + cc\,. \end{aligned} \tag{38.116}$$

The 6×6 square matrix and 6×1 vectors of the right-hand side of (38.116) are given by

$$\mathbf{K}^{(2)} = \begin{bmatrix} 0 & \Omega_{30} & 2\mathrm{i}\omega m & 0 & S_{\mathrm{n}0} & w_0 \cos\phi_0 \\ -\Omega_{30} & 0 & 0 & 2\mathrm{i}\omega(m+m_{\mathrm{a}}) + b|v| & -T_0 & -w_0 \sin\phi_0 \\ \dfrac{2\mathrm{i}\omega}{EA} & 0 & 0 & \Omega_{30} & 0 & 0 \\ 0 & 0 & -\Omega_{30} & 0 & 0 & 2\mathrm{i}\omega \\ 0 & \dfrac{-1}{EI} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \tag{38.117}$$

$$\boldsymbol{N}^{(2)} = \begin{bmatrix} S_{\mathrm{n}1}^{(1)} \Omega_{31}^{(1)} - \mathrm{i}\omega m v^{(1)} \phi_1^{(1)} \\ -T_1^{(1)} \Omega_{31}^{(1)} + \mathrm{i}\omega m u^{(1)} \phi_1^{(1)} \\ \Omega_{31}^{(1)} v^{(1)} \\ -\Omega_{31}^{(1)} u^{(1)} \\ 0 \\ 0 \end{bmatrix}, \tag{38.118}$$

$$\tilde{N}^{(2)} = \begin{bmatrix} S_{n1}^{(1)}\bar{\Omega}_{31}^{(1)} + i\omega m v^{(1)}\bar{\phi}_1^{(1)} \\ -\bar{T}_1^{(1)}\Omega_{31}^{(1)} + i\omega m\bar{u}^{(1)}\phi_1^{(1)} \\ \bar{\Omega}_{31}^{(1)} v^{(1)} \\ -\Omega_{31}^{(1)}\bar{u}^{(1)} \\ 0 \\ 0 \end{bmatrix} . \tag{38.119}$$

Further, we distinguish between the constant and the time-varying components. Thus, (38.116) is split into two systems of ordinary differential equations, both expressed in convenient matrix form. The first system provides the second-order constant displacement and internal loading terms, whilst the second system yields the respective time-varying double-frequency components. Here, emphasis is given to the dynamic behavior of the structural model under investigation and hence the solution of the constant system is omitted. It should be noted, however, that the occurrence of the elements of vector $\tilde{N}^{(2)}$ and, accordingly, the existence of a time-independent term in the global solution of the second-order problem imply that the double-frequency variables oscillate around a nonzero average. In particular, the system that provides the double-frequency components is

$$\frac{\mathrm{d}\left(\xi_1^{(2)} e^{2i\omega t}\right)}{\mathrm{d}s} + cc = \mathbf{K}^{(2)}\left(\xi_1^{(2)} e^{2i\omega t}\right) + \frac{1}{2}N^{(2)} e^{2i\omega t} + cc\,. \tag{38.120}$$

The forcing terms for that system are involved in $N^{(2)}$ and they virtually represent the excitation for the double-frequency system.

38.5.5 Numerical Solution

The linear and the second-order systems are treated separately as two discrete, two-point boundary value problems. The solution of the double-frequency system relies on the derivation of the first-order components. The systems are transformed in order to be expressed in terms of displacements instead of velocities. Thus,

$$\begin{aligned} u^{(1)} e^{i\omega t} &= i\omega p^{(1)} e^{i\omega t}\,, \\ u^{(2)} e^{2i\omega t} &= 2i\omega p^{(2)} e^{2i\omega t}\,, \\ v^{(1)} e^{i\omega t} &= i\omega q^{(1)} e^{i\omega t}\,, \\ v^{(2)} e^{2i\omega t} &= 2i\omega q^{(2)} e^{2i\omega t}\,, \end{aligned}$$

where again, $p^{(k)}$, $q^{(k)}$, and $k = 1, 2$ are, respectively, the first and second-order complex (ω, s) functions of the axial and normal displacements. In the context of the frequency domain approach, the drag forces in normal direction $R_{dn1}^{(k)}$, $k = 1, 2$ should be replaced by the linearized equivalent terms obtained using a Fourier series expansion technique. It can be shown that the associated coefficients, to first and second-order, are, respectively, equal to $8/(3\pi)$ and $112/(60\pi)$, and accordingly, the drag forces become

$$R_{dn1}^{(1)} = -ib^{(1)}\omega^2 |q|\, q^{(1)};\;\; R_{dn1}^{(2)} = -ib^{(2)}\omega^2 |q|\, q^{(2)}\,, \tag{38.121}$$

where

$$b^{(1)} = 4/(3\pi)\rho d_O C_d$$

and

$$b^{(2)} = 112/(60\pi)\rho d_O C_d\,.$$

Next, after separating into real and imaginary parts, a system of 12 ordinary differential equations with an equal number of unknowns is derived. In order to distinguish between real and imaginary parts, the latter will be denoted in the sequel as bold symbols. Thus, if

$$\boldsymbol{F}^{(k)} = \left[T_1^{(k)} \;\; \boldsymbol{T}_1^{(k)} \;\; S_{n1}^{(k)} \;\; \boldsymbol{S}_{n1}^{(k)} \;\; p^{(k)} \;\; \boldsymbol{p}^{(k)} \;\; q^{(k)} \;\; \boldsymbol{q}^{(k)} \;\; \Omega_{31}^{(k)} \;\; \boldsymbol{\Omega}_{31}^{(k)} \;\; \phi_1^{(k)} \;\; \boldsymbol{\phi}_1^{(k)}\right]^{\mathrm{T}}, \tag{38.122}$$

for $k = 1, 2$ (38.110) and (38.120) are rewritten as

$$\frac{\mathrm{d}\boldsymbol{F}^{(k)}}{\mathrm{d}s} = \mathbf{P}^{(k)}\boldsymbol{F}^{(k)} + \frac{1}{2}\boldsymbol{Q}^{(k)}\,, \tag{38.123}$$

where $\boldsymbol{Q}^{(1)} = 0$. The transformation matrices $\mathbf{P}^{(k)}$ and the excitation vectors $\boldsymbol{Q}^{(k)}$ originated directly from $\mathbf{K}^{(k)}$ and $\boldsymbol{N}^{(k)}$, respectively, after equating real and imaginary parts. The elements of the vector $\boldsymbol{Q}^{(2)}$ literally represent the excitation of the nonlinear system and can be regarded as nonlinear springs composed of products of the first-order components.

The systems denoted by (38.123) are solved using the centered differences method [38.31]. To this end, the length of the structure is discretized using n nodes $(l = 1, \ldots, n)$, and all terms of (38.31) are evaluated at $l - 1/2$. Thus,

$$F_l^{(k)} - F_{l-1}^{(k)} = \frac{\Delta s}{2}\left(P_l^{(k)}F_l^{(k)} + P_{l-1}^{(k)}F_{l-1}^{(k)}\right) + \frac{\Delta s}{4}\left(Q_l^{(k)} + Q_{l-1}^{(k)}\right)\,. \tag{38.124}$$

The index l defines the spatial grid points (the nodes). For n nodal points, (38.124) defines a system of $12 \times (n-1)$ equations to be solved for the $12 \times n$-dependent variables. The 12 equations needed to complete the problem are provided by boundary conditions. The conditions that govern the behavior of the upper end at $l = n$, $s = L$, i. e., where the excitation is applied, are defined explicitly by the following relations

$$p^{(1)} = p_a(t); \quad q^{(1)} = q_a(t); \quad \Omega_{31}^{(1)} = 0;$$
$$p^{(2)} = 0; \quad q^{(2)} = 0; \quad \Omega_{31}^{(2)} = 0\,. \qquad (38.125)$$

In other words it is assumed that the top terminal point is simply supported, which implies that the curvatures at both first and second orders are zero, whereas for a monochromatic excitation the linear motions are predefined functions of time (denoted by $p_a(t)$ and $q_a(t)$). In addition, the specified top end excitation requires that the motions at second order must be zero. This feature is in compliance with a previous remark according to which the excitation for the second-order system is provided by equivalent nonlinear springs formulated mathematically by the linear components.

The derivation of the remaining six boundary conditions at the lower end is combined with a novel approach that tries to formulate mathematically the bottom–structure interaction phenomenon. The relevant procedure is discussed in detail in the following section.

38.5.6 The Bottom–Structure Interaction Approach

Background

The mathematical formulation outlined in the preceding sections is extended to account for an additional impact that is known to occur in real applications. This is the interaction of the cable with the seafloor, which in the context of the present frequency domain approximation is treated with a novel mathematical approach.

There have been several studies reported in the literature that try to assess the impact of the cable–bottom interaction phenomenon from the mathematical perspective. In this context *Chatjigeorgiou* and *Mavrakos* [38.57] evaluated the time-varying TDP using quasi-static considerations and then they solved the dynamic equilibrium system along the time variable suspended part. *Gobat* and *Grosenbaugh* [38.58] investigated the interaction problem for catenary moorings both numerically and experimentally by relating the TDP speed with the transverse wave speed of the mooring line. *Demeio* and *Lenci* [38.59] transformed the free boundary value problem of the moving TDP into a fixed boundary value problem that was successively approached by a perturbation expansion method. In order to detect the main mechanical phenomenon, the authors considered terms up to the second order. In the work due to *Gatti-Bono* and *Perkins* [38.60] the seabed was simulated as an elastic foundation with a prescribed topology that assumes linear damping, whereas *Han* and *Grosenbaugh* [38.61] employed a simplified model based on the static equilibrium system and formulated the seabed by a coupled system consisting of mass, damper, and springs. Finally, although not explicitly connected with mooring lines, the work of *Pesce* et al. [38.62] should be mentioned, in which the authors proposed the employment of a boundary layer technique at TDP to approximate the dynamic curvature amplification at the TDZ.

Here, an alternative novel theoretical approach is taken, the details of which are outlined thoroughly in the following section.

A Mathematical Approach

The following analysis relies on the assumption of a perfectly horizontal bottom. It is apparent that under the external imposed excitation the catenary cable will continuously lift off and lie down. For a monochromatic excitation a lift-off/lie-down cycle will be equal to the displacement cycle. In fact, the length of the cable between the end nodes of maximum lift-off and maximum lie-down can be accurately determined by quasi-static considerations. Having as a base configuration the static equilibrium shape of the catenary cable, it should be noted that the lengths of the segments on either side of the TDP at $t = 0$ are not necessarily equal. That depends on the static configuration and the orientation of the excitation at the top. Nevertheless, it can be assumed that the time-varying TDP will oscillate horizontally around an average location that coincides with the TDP at $t = 0$ (the static configuration TDP). Also, it can be safely assumed that the TDP is instantly a pinned support.

The assumption of a horizontally moving pinned TDP around an average location, suggests the employment of a Taylor series expansion for the variables involved in the boundary conditions. First, using velocity terms and assuming that $f^{(k)}$, $k = 1, 2$ is a global notation for axial velocity, normal velocity, and curvature at k-th order, the Taylor series expansion of $f^{(k)}$ is written as

$$f^{(k)}|_0 + \left.\frac{\partial f^{(k)}}{\partial s}\right|_0 \Delta s + \cdots = 0\,. \qquad (38.126)$$

The index 0 in the above relation denotes that the term concerned is calculated at the static equilibrium TDP ($t = 0$). The $O(\Delta s^2)$ terms and above have been omitted.

Further, by using the perturbations technique, equating the terms of like powers of ε and employing (38.98)–(38.100) and (38.104)–(38.106) the following are derived

$$u^{(1)}|_0 + \left(\frac{1}{EA}\frac{\partial T_1^{(1)}}{\partial t} + \Omega_{30} v^{(1)}\right)\Bigg|_0 \Delta s = 0\,, \tag{38.127}$$

$$v^{(1)}|_0 + \left(\frac{\partial \phi_1^{(1)}}{\partial t} - \Omega_{30} u^{(1)}\right)\Bigg|_0 \Delta s = 0\,, \tag{38.128}$$

$$\Omega_{31}^{(1)}|_0 - \frac{S_{\mathrm{n1}}^{(1)}}{EI}\Bigg|_0 \Delta s = 0\,, \tag{38.129}$$

$$u^{(2)}|_0 + \left(\frac{1}{EA}\frac{\partial T_1^{(2)}}{\partial t} + \Omega_{30} v^{(2)} + \Omega_{31}^{(1)} v^{(1)}\right)\Bigg|_0 \Delta s = 0\,, \tag{38.130}$$

$$v^{(2)}|_0 + \left(\frac{\partial \phi_1^{(2)}}{\partial t} - \Omega_{30} u^{(2)} - \Omega_{31}^{(1)} u^{(1)}\right)\Bigg|_0 \Delta s = 0\,, \tag{38.131}$$

$$\Omega_{31}^{(2)}|_0 - \frac{S_{\mathrm{n1}}^{(2)}}{EI}\Bigg|_0 \Delta s = 0\,. \tag{38.132}$$

Again, using displacements instead of velocities and separating real and imaginary parts, (38.127)–(38.129) are transformed at $O(\varepsilon)$ into

$$p^{(1)}|_0 + \frac{T_1^{(1)}}{EA}\Bigg|_0 \Delta s = 0\,, \tag{38.133}$$

$$q^{(1)}|_0 + \phi_1^{(1)}|_0 \Delta s = 0\,, \tag{38.134}$$

$$\Omega_{31}^{(1)}|_0 - \frac{S_{\mathrm{n1}}^{(1)}}{EI}\Bigg|_0 \Delta s = 0\,. \tag{38.135}$$

It is noted that (38.133)–(38.135) apply identically for both real

$$\left[T_1^{(1)} \quad S_{\mathrm{n1}}^{(1)} \quad p^{(1)} \quad q^{(1)} \quad \Omega_{31}^{(1)} \quad \phi_1^{(1)}\right]$$

and imaginary parts

$$\left[\boldsymbol{T}_1^{(1)} \quad \boldsymbol{S}_{\mathrm{n1}}^{(1)} \quad \boldsymbol{p}^{(1)} \quad \boldsymbol{q}^{(1)} \quad \boldsymbol{\Omega}_{31}^{(1)} \quad \boldsymbol{\phi}_1^{(1)}\right].$$

Finally, for second-order $O(\varepsilon^2)$, the bottom end boundary conditions become

$$p^{(2)}|_0 + \left[\frac{T_1^{(2)}}{EA} + \frac{1}{4}\left(q^{(1)}\Omega_{31}^{(1)} - \boldsymbol{q}^{(1)}\boldsymbol{\Omega}_{31}^{(1)}\right)\right]_0 \Delta s = 0\,, \tag{38.136}$$

$$\boldsymbol{p}^{(2)}|_0 + \left[\frac{\boldsymbol{T}_1^{(2)}}{EA} + \frac{1}{4}\left(q^{(1)}\boldsymbol{\Omega}_{31}^{(1)} + \boldsymbol{q}^{(1)}\Omega_{31}^{(1)}\right)\right]_0 \Delta s = 0\,, \tag{38.137}$$

$$q^{(2)}|_0 + \left[\phi_1^{(2)} - \frac{1}{4}\left(p^{(1)}\Omega_{31}^{(1)} - \boldsymbol{p}^{(1)}\boldsymbol{\Omega}_{31}^{(1)}\right)\right]_0 \Delta s = 0\,, \tag{38.138}$$

$$\boldsymbol{q}^{(2)}|_0 + \left[\boldsymbol{\phi}_1^{(2)} - \frac{1}{4}\left(\boldsymbol{p}^{(1)}\Omega_{31}^{(1)} + p^{(1)}\boldsymbol{\Omega}_{31}^{(1)}\right)\right]_0 \Delta s = 0\,, \tag{38.139}$$

$$\Omega_{31}^{(2)}|_0 - \frac{S_{\mathrm{n1}}^{(2)}}{EI}\Bigg|_0 \Delta s = 0\,, \tag{38.140}$$

$$\boldsymbol{\Omega}_{31}^{(2)}|_0 - \frac{\boldsymbol{S}_{\mathrm{n1}}^{(2)}}{EI}\Bigg|_0 \Delta s = 0\,. \tag{38.141}$$

The first-order (38.133)–(38.135) imply that the bottom end boundary conditions are linear. Nevertheless, according to the proposed mathematical expansion the boundary conditions at $O(\varepsilon)$ become coupled as all dynamic components at the time-varying TDP at $l = 1$ are involved. Also, the present sophisticated formulation of the boundary conditions demonstrates that the displacements and the curvature at the original TDP are not zero. This theoretical finding is in accordance with the actual phenomenon as the original TDP (at $t = 0$) oscillates continuously.

Studies reported on the subject abound in verbal statements regarding the nonlinear nature of the interaction phenomenon. Nevertheless, relevant references are usually correlated with the nonlinear stiffness and the nonlinear stress–strain relations of soil. The mathematical analysis that is conducted herein demonstrates that the phenomenon is itself nonlinear and the nonlinearities involved are not explicitly driven by the distorting behavior of the soil. In fact, the nonlinear boundary conditions (38.136)–(38.139) show the occurrence of double-frequency components, which are stimulated by first-order quadratic terms.

38.5.7 Results of Second-Order Double-Frequency Effects

Succinctly, the analysis conducted in Sect. 38.5 implies that when a cable is subjected to monochromatic excitation it admits contributions from the integer multiplications of the associated frequency. Hypothetically, the excitation frequencies that contribute extend to infinity. The same would occur if, for instance, the cable were excited by two frequencies. In that case, the response would include contributions from all possible sum and minus combinations of these frequencies, in-

cluding combinations of the integer multiplications. Nevertheless, it can be safely said that the impact of the nonlinear effects decays for higher orders and the prevailing component remains the linear term. Nonlinear effects have visual and significant results only under specific conditions, say at nonlinear resonances.

To assess the effect of the nonlinearities on the dynamics of cables, the mooring line, already examined in Sect. 38.4.1, is used as the base structural model. Along with the double-frequency contributions, the discussion is extended to comment on bottom-interaction effects as they are determined with the novel approach outlined in detail in Sect. 38.5.6 herein. Here, the associated numerical predictions are labeled as BI. Two excitation cases are considered, namely in horizontal and vertical directions. In both cases, the amplitude was kept the same, i. e., 2 m.

First the horizontal excitation is considered. The effect of bottom interactions on the linear problem is shown in Fig. 38.25. In fact, the curves of the linear problem are the same as those shown in Fig. 38.15. The symbols denote the results obtained after incorporating the relations that govern the Taylor approach of the bottom-interaction phenomenon. It is immediately evident that for this particular case of excitation and the configuration of the mooring line, the interaction with the bottom has practically zero effect, at least as far as the tension amplification is concerned. It should be noted that to eliminate possible inaccuracies due to the numerical solution of the problem a sufficiently dense grid was employed and the cable was discretized by 1000 nodes along its complete length.

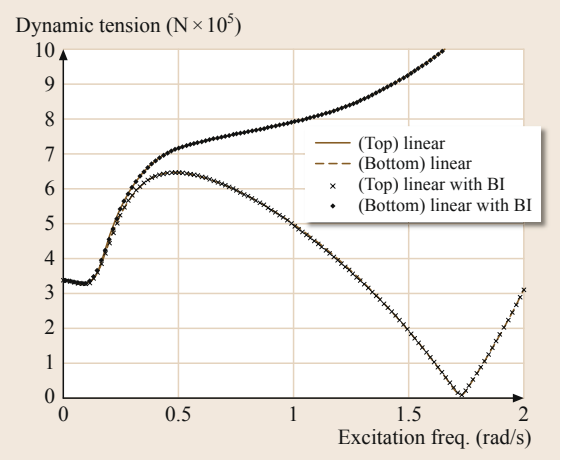


Fig. 38.25 Linear dynamic tension amplification at the top and the bottom of a mooring line, excited by a horizontal monochromatic excitation with amplitude $a_x = 2$ m (BI denotes that bottom interaction effects have been considered)

The next comment refers to the effect of the second-order components of the dynamic tension, which are applied at the top and bottom of the cable at double the frequency of the excitation. The associated quadratic transfer functions are depicted in Fig. 38.26. Here it is interesting to observe that the two curves nearly coincide along the whole frequency range. In fact, a depiction of the tension variation along the cable would reveal that, indeed, the tension exhibits a symmetrical variation about the middle of the structure and that occurs for all frequencies. As regards the details of the plotted quadratic transfer functions, the attention is immediately attracted by the two local maxima, which apparently coincide with specific eigenfrequencies of the dynamic system. The second peak is clearly a resonance that obtains a relatively large value, albeit approximately 10 times lower than the linear contribution. The important aspect is that the sharp amplification occurs at exactly the frequency where the linear dynamic tension is canceled.

Analogous behavior is observed in the linear and the quadratic transfer functions of the motions of the touch-down point (TDP). As has been said, the Taylor approximation allows the originally fixed TDP to move freely in the 2-D plane of reference. The kinematics of that point is described through the axial (stretching) and normal motions of the cable. The stretching is examined in Fig. 38.27 and the normal motions in the Fig. 38.28. The former reveals that the pattern of the dynamic tension amplification is provided directly by the stretching (axial motion). Also, the second-order component is practically zero as compared to the linear term, as it

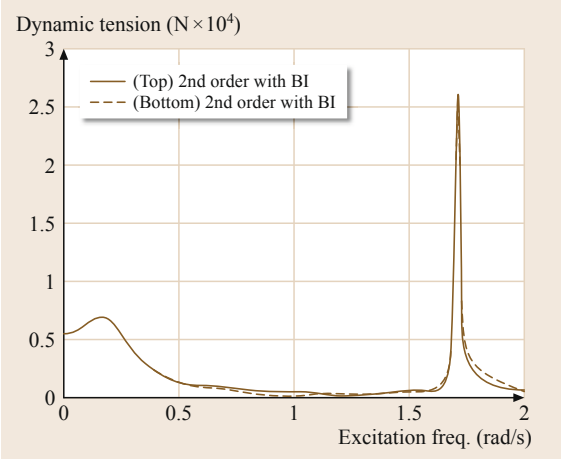


Fig. 38.26 Second-order double-frequency dynamic tension amplification at the top and the bottom of a mooring line, excited by a horizontal monochromatic excitation with amplitude $a_x = 2$ m (BI denotes that bottom interaction effects have been considered)

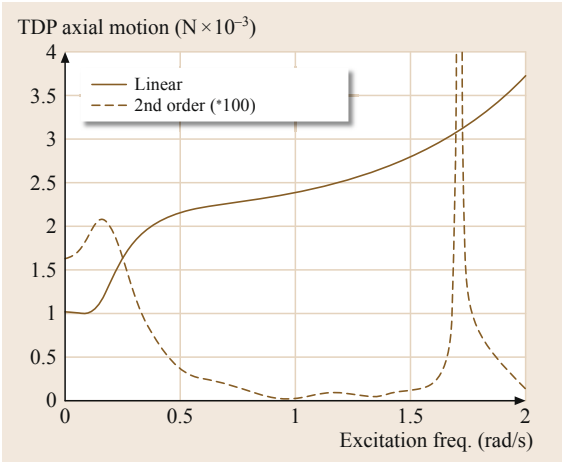


Fig. 38.27 Linear and second-order axial stretching at the bottom of a mooring line, excited by a horizontal monochromatic excitation with amplitude $a_x = 2\,\mathrm{m}$ (bottom interaction effects have been considered)

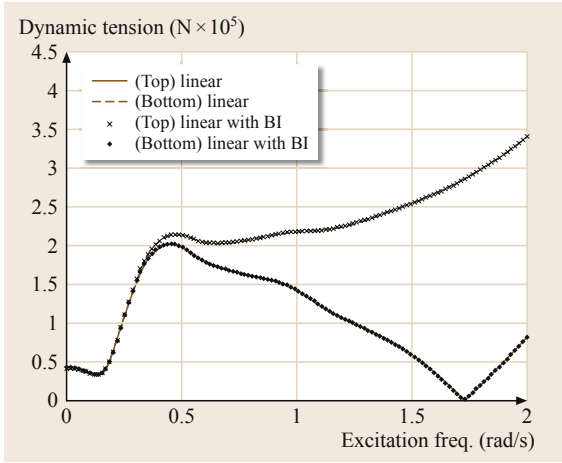


Fig. 38.29 Linear dynamic tension amplification at the top and the bottom of a mooring line, excited by a vertical monochromatic excitation with amplitude $a_z = 2\,\mathrm{m}$ (BI denotes that bottom interaction effects have been considered)

is approximately 100 times smaller, which in turn explains why the transfer functions shown in Fig. 38.25 coincide perfectly.

The last variable that is considered is the normal motion of the TDP, which as manifested in Fig. 38.28 is small but not negligible. According to the demonstrated trends the maximum uplift of the TDP from the bottom approximates 2 cm for the excitation amplitude and the structural model under consideration. It should be mentioned that the variation of the normal motion of the TDP as a function of the excitation frequency does not follow any predictable path complying to a certain rule.

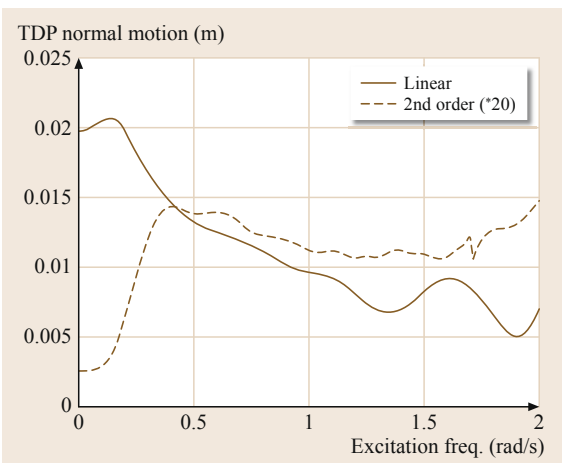


Fig. 38.28 Linear and second-order TDP normal motions of a mooring line, excited by a horizontal monochromatic excitation with amplitude $a_x = 2\,\mathrm{m}$ (bottom interaction effects have been considered)

Next, the heave excitation case is investigated. To this end, Figs. 38.29–38.32 are given, which depict analogous components. The transfer functions of the dynamic tension (linear problem) are shown in Fig. 38.29. The behavior is more or less the same as with the surge excitation case. Again, a cancelation point occurs for the dynamic tension at the top, whilst the dynamic tension is larger at the bottom. The interaction with the bottom has a negligible impact, although the associated curves do not coincide perfectly. However, the curves that coincide perfectly are the quadratic

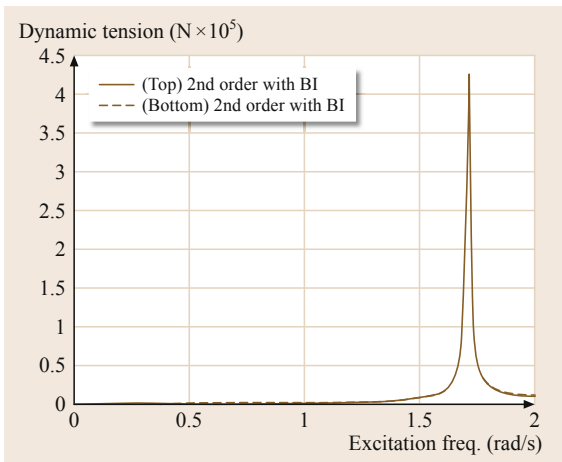


Fig. 38.30 Second-order double-frequency dynamic tension amplification at the top and the bottom of a mooring line, excited by a vertical monochromatic excitation with amplitude $a_z = 2\,\mathrm{m}$ (BI denotes that bottom interaction effects have been considered)

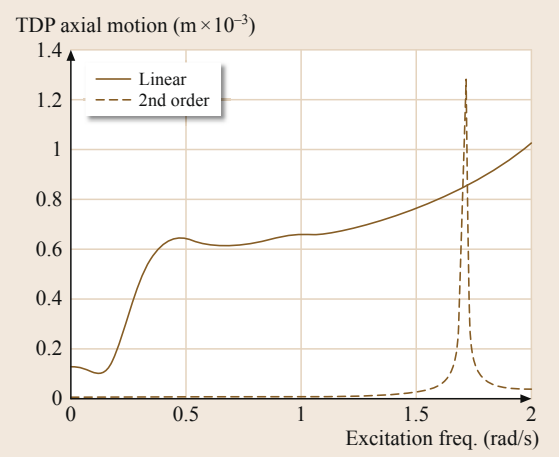


Fig. 38.31 Linear and second-order axial stretching at the bottom of a mooring line, excited by a vertical monochromatic excitation with amplitude $a_z = 2\,\mathrm{m}$ (bottom interaction effects have been considered)

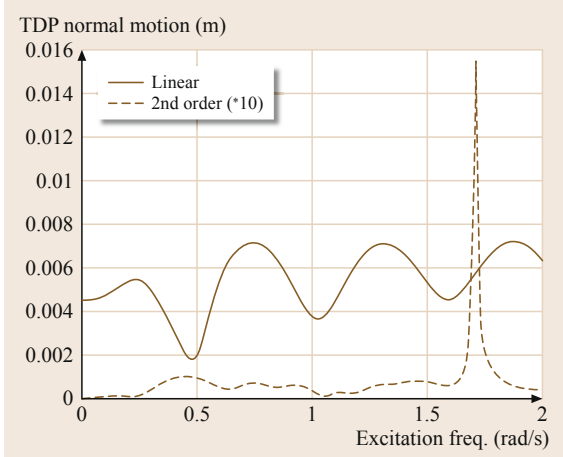


Fig. 38.32 Linear and second-order TDP normal motions of a mooring line, excited by a vertical monochromatic excitation with amplitude $a_z = 2\,\mathrm{m}$ (bottom interaction effects have been considered)

transfer functions of the dynamic tension applied at the top and the bottom (Fig. 38.30).

As has been said and still holds, the variation of double-frequency tension along the cable exhibits a symmetry with respect to the midpoint of the suspended part. The most important characteristic depicted in Fig. 38.30 is that the peak (the resonance) that occurs at the end of the frequency range obtains an unusually large value for a second-order term and reaches the values of the linear components (in fact, it exceeds the maximum linear component). This is indeed a very important finding, which could be correlated with the fatigue response of the cable, since it proves that apart from the linear loading, at a specified frequency the cable will be subjected to a very fast and very strong internal forcing.

With reference to Fig. 38.31 it can be seen that the pattern of the axial stretching at the TDP provides the variation pattern of the dynamic tension amplification in both linear and double-frequency orders. Stretching is one of the two degrees of freedom in the plane of reference. The other is the normal motion, which practically demonstrates the displacements performed by the TDP. The associated linear and quadratic transfer functions are given in Fig. 38.32. The linear displacements follow a wavy trend with respect to the frequency. The second-order normal motions are governed by the well-known peak at the high frequency range, which shows that the maximum expected normal motion of the TDP will not exceed 2 cm.

It is important to mention that the results shown in this section must not lead to simplistic conclusions such as, for instance, the interaction with the bottom or the nonlinear effects being negligible. It is emphasized that the numerical predictions are case sensitive, meaning that they are explicitly adapted to the model investigated and to the excitation conditions. The cable examined herein was sufficiently stretched without allowing the dynamic components to be fully developed. Hence, the computations showed a relatively minor effect arising from the dynamics of the cable. Also, the small motions of the TDP can be attributed to this particular configuration. The condition would definitely be different for structures with large bending stiffness, say for risers. In relevant cases, however, the parameter of concern would be the bending moment and not the tension. Even for the cable investigated, however, which possesses the properties of a typical mooring line, it is important that we have simulated the kinematics of the TDP and shown that is indeed released from the bottom.

References

38.1 M.S. Rahman: The Darboux vector characterizing Frenet's formulas for plane curves and straight lines, Int. J. Math. Educ. Sci. Tech. **25**, 764–767 (1994)

38.2 S.H. Crandall: *Engineering Analysis* (McGraw-Hill, New York 1956)

38.3 I.K. Chatjigeorgiou: On the effect of internal flow on vibrating catenary risers in three dimensions, Eng. Struct. **32**, 3313–3329 (2010)

38.4 C.T. Howell: Investigation of the Dynamics of Low Tension Cables, Ph.D. Thesis (Massachusetts Institute of Technology, Cambridge 1992)

38.5 V.J. Papazoglou, S.A. Mavrakos, M.S. Triantafyllou: Non-linear cable response and model testing in water, J. Sound Vibr. **140**, 103–115 (1991)

38.6 L. Johanning, G.H. Smith, J. Wolfram: Measurements of static and dynamic mooring line damping and their importance for floating WEC devises, Ocean Eng. **34**, 1918–1934 (2007)

38.7 L.P. Krolikowski, T.A. Gay: An improved linearization technique for frequency domain riser analysis, Proc. 12th OTC Conf., Houston (1980), OTC-3777-MS

38.8 I.K. Chatjigeorgiou: Linear out-of-plane dynamics of catenary risers, Proc. IMechE M J. Eng. Marit, Env. **224**, 13–27 (2010)

38.9 M.S. Triantafyllou, G.S. Triantafyllou: The paradox of the hanging string: An explanation using singular perturbations, J. Sound Vibr. **148**, 343–351 (1991)

38.10 I.K. Chatjigeorgiou: Solution of the boundary layer problems for calculating the natural modes of riser-type slender structures, J. Offshore Mech. Arct. Eng. **130**, 011003–11001 (2008)

38.11 M.S. Triantafyllou: The dynamics of taut inclined cables, Q. J. Mech. Appl. Math. **37**(3), 421–440 (1984)

38.12 M.S. Triantafyllou, A. Bliek, H. Shin: Dynamic analysis as a tool for open sea mooring system design, Trans. Soc. Nav. Archit. Mar. Eng. **93**, 303–324 (1985)

38.13 M.S. Triantafyllou, L. Grinfogel: Natural frequencies and modes of inclined cables, J. Struct. Eng. **112**, 139–148 (1986)

38.14 J.J. Burgess, M.S. Triantafyllou: The elastic frequencies of cables, J. Sound Vibr. **120**, 153–165 (1988)

38.15 I.K. Chatjigeorgiou: Application of the WKB method for catenary shaped slender structures, Math. Comput. Model. **48**, 249–257 (2008)

38.16 M.S. Triantafyllou, C.T. Howell: Dynamic response of cables under negative tension: An ill posed problem, J. Sound Vibr. **173**, 433–447 (1994)

38.17 S.A. Mavrakos, I.K. Chatjigeorgiou, G. Grigoropoulos, A. Marón: Scale experiments for the measurement of motions and wave run-up on a TLP model subjected to monochromatic waves, Proc. 14th ISOPE Conf., Vol. 1, Toulon (2004) pp. 382–389

38.18 J.M. Niedzwecki, S.K. Thampi: Snap-loading of marine cable systems, Appl. Ocean Res. **13**, 2–11 (1991)

38.19 I.K. Chatjigeorgiou, S.A. Mavrakos: Dynamic behavior of a marine cable under snap-loading conditions, Ship Tech. Res. Schiffstechnik **48**, 171–180 (2001)

38.20 I.J. Fylling, C.M. Larsen, N. Sødahl, H. Ormberg, A. Engseth, E. Passano, K. Holthe: *Riflex Theory Manual, Techn. Rep. STF70 F95219, MARINTEK, 22/9* (Marintek, Trondheim 1995)

38.21 P.H. Wang, R.F. Fung, M.J. Lee: Finite element analysis of a three dimensional underwater cable with time dependent length, J. Sound Vibr. **209**, 223–249 (1998)

38.22 O.M. Aamo, T.I. Fosen: Finite element modeling of moored vessels, Math. Comput. Model. Dyn. Syst. **7**, 47–75 (2001)

38.23 C.M. Ablow, S. Schechter: Numerical simulation of undersea cable dynamics, Ocean Eng. **10**, 443–457 (1983)

38.24 F. Miinazzo, M. Wilkie, S.A. Latchman: An efficient algorithm for simulating the dynamics of towed cable systems, Ocean Eng. **14**, 513–526 (1987)

38.25 C.T. Howell: Numerical analysis of 2-D nonlinear cable equations with applications to low-tension problems, Int. J. Offshore Mech. Arct. Eng. **2**, 110–113 (1992)

38.26 M.A. Grosenbaugh, C.T. Howell, C.T. Moxnes: Simulating the dynamics of underwater vehicles with low tension tethers, Int. J. Offshore Mech. Arct. Eng. **3**, 213–218 (1993)

38.27 J.J. Burgess: Bending stiffness in a simulation of undersea cable deployment, Int. J. Offshore Mech. Arct. Eng. **3**, 197–204 (1993)

38.28 J.I. Gobat: The Dynamics of Geometrically Compliant Mooring Systems, Ph.D. Thesis (Massachusetts Institute of Technology, Cambridge 2000)

38.29 A.A. Tjavaras, Q. Zhu, Y. Liu, M.S. Triantafyllou, D.K.P. Yue: The mechanics of highly extensible cable, J. Sound Vibr. **213**, 709–734 (1998)

38.30 I.K. Chatjigeorgiou, S.A. Mavrakos: Comparative evaluation of numerical schemes for 2D mooring dynamics, Int. J. Offshore Mech. Arct. Eng. **10**, 301–309 (2000)

38.31 J.D. Hoffman: *Numerical Methods for Engineers and Scientists* (McGraw-Hill, New York 1993)

38.32 J.I. Gobat, M.A. Grosenbaugh: Application of the generalized-α method to the time integration of the cable dynamic equations, Comput. Methods Appl. Mech. Eng. **190**, 4817–4829 (2001)

38.33 H.M. Irvine: *Cable Structures* (MIT Press, Cambridge 1981)

38.34 I.K. Chatjigeorgiou, S.A. Mavrakos: Dynamic behavior of deep water mooring lines with submerged buoys, Comput. Struct. **64**, 819–835 (1997)

38.35 A. Bliek: Dynamic Analysis of Single Span Cables, Ph.D. Thesis (Massachusetts Institute of Technology, Cambridge 1984)

38.36 S.A. Mavrakos, V.J. Papazoglou, M.S. Triantafyllou, I.K. Chatjigeorgiou: Deep water mooring dynamics, Mar. Struct. **9**, 181–209 (1996)

38.37 W. Raman-Nair, R.E. Baddour: Three-dimensional coupled dynamics of a buoy and multiple mooring lines: Formulation and algorithm, Q. J. Mech. Appl. Math. **55**, 179–207 (2002)

38.38 S. Surendran, M. Goutam: Reduction in the dynamic amplitudes of moored cable systems, Ships Offshore Struct. **4**, 145–163 (2009)

38.39 A. Tahar, M.H. Kim: Coupled-dynamic analysis of floating structures with polyester mooring lines, Ocean Eng. **35**, 1676–1685 (2008)

38.40 A. Montano, M. Restelli, R. Sacco: Numerical simulation of tethered buoy dynamics using mixed finite elements, Comput. Methods Appl. Mech. Eng. **196**, 4117–4129 (2007)

38.41 H. Joosten: Wave buoys and their elastic mooring, Int. Ocean Syst. **10**, 18–21 (2006)

38.42 A. Umar, T.K. Data: Nonlinear response of a moored buoy, Ocean Eng. **30**, 1625–1646 (2003)

38.43 R.J. Smith, C.J. MacFarlane: Statics of a three component mooring line, Ocean Eng. **28**, 899–914 (2001)

38.44 L.O. Garza-Rios, M.M. Bernitsas: Effect of size and position of supporting buoys on the dynamics of spread mooring systems, J. Offshore Mech. Arct. Eng. **123**, 49–56 (2001)

38.45 S.A. Mavrakos, I.K. Chatjigeorgiou, V.J. Papazogloy: Use of buoys for dynamic tension reduction in deep water mooring applications, Proc. 7th Conf. Beh. Offshore Struct., Boston (1994) pp. 417–426

38.46 E. Huse: Influence of mooring line damping upon rig motion, Proc. 18th Conf. Offshore Technol., Houston (1986), OTC-5204-MS

38.47 E. Huse, K. Matsumoto: Practical estimation of mooring line damping, Proc. 20th Conf. Offshore Technol., Houston (1988), OTC-5676-MS

38.48 Y. Liu, L. Bergdahl: Improvement of Huse's model for estimating mooring cable induced damping, Proc. 17th Conf. Offshore Mech. Arct. Eng., Lisbon (1998), OMAE-98-0353

38.49 D.T. Brown, G.L. Lyons, H.M. Lin: Advances in mooring line damping, J. Soc. Underw. Tech. **21**, 5–12 (1995)

38.50 D.T. Brown, S.A. Mavrakos: Comparative study on mooring line damping, Mar. Struct. **12**, 131–151 (1999)

38.51 C. Bauduin, M. Naciri: A contribution on quasi-static mooring line damping, J. Offshore Mech. Arct. Eng. **122**, 125–133 (2000)

38.52 L. Johanning, G.H. Smith, J. Wolfram: Measurements of static and dynamic line damping and their importance for floating WEC devises, Ocean Eng. **34**, 1918–1934 (2007)

38.53 I.K. Chatjigeorgiou, G. Damy, M. LeBoulluec: Numerical and experimental investigation of the dynamics of catenary risers and the riser-induced-damping phenomenon, Proc. 27th Conf. Offshore Mech. Arct. Eng., Lisbon (1998), OMAE 2008-57616

38.54 M.H. Ghayesh: Coupled longitudinal-transverse dynamics of an axially accelerating beam, J. Sound Vibr. **331**, 5107–5124 (2012)

38.55 J. Kevorkian, J.D. Cole: *Perturbation Methods in Applied Mathematics* (Springer, New York 1981)

38.56 A.H. Nayfeh: *Problems in Perturbation* (Wiley, New York 1985)

38.57 I.K. Chatjigeorgiou, S.A. Mavrakos: Assessment of bottom cable interaction effects on mooring line dynamics, Proc. 17th Conf. Offshore Mech. Arct. Eng., Lisbon (1998), OMAE-98-0355

38.58 J.I. Gobat, M.A. Grosenbaugh: Dynamics in the touchdown region of catenary moorings, Int. J. Offshore Polar Eng. **11**, 273–281 (2001)

38.59 L. Demeio, S. Lenci: Forced nonlinear oscillations of semi-infinite cables and beams resting on a unilateral substrate, Nonlinear Dyn. **49**, 203–215 (2007)

38.60 C. Gatti-Bono, N.C. Perkins: Numerical simulations of cable/seabed interaction, Int. J. Offshore Polar Eng. **14**, 118–124 (2004)

38.61 S. Han, M.A. Grosenbaugh: Modeling of seabed interaction of oceanographic moorings in the frequency domain, J. Waterw. Port Coast. Ocean Eng. **132**, 450–456 (2006)

38.62 C.P. Pesce, J.A.P. Aranha, C.A. Martins, O.G.S. Ricardo, S. Silva: Dynamic curvature in catenary risers at the touch down point region: An experimental study and the analytical boundary-layer solution, Int. J. Offshore Polar Eng. **8**, 303–310 (1998)

39. Offshore Geotechnics

Dong-Sheng Jeng

Offshore geotechnics is a multidisciplines area that cover conventional civil engineering disciplines such as fluid mechanics, coastal engineering, geotechnical engineering, structures engineering, etc. This area has attracted great attention among coastal and geotechnical engineers due to growing activities in marine environment worldwide. An appropriate design of foundations of marine infrastructures, such as breakwater, offshore pipeline, platforms, and offshore wind turbine systems plays an important role in the success of offshore engineering projects. The evaluation of the soil response due to hydrodynamic loading, such as waves and currents around foundation of marine structures and its resultant seabed instability is one of the key factors in the design of foundation.

This chapter is an attempt to give a comprehensive review of wave–seabed interaction around the marine structure. It also takes into consideration all state-of-the art knowledge. We start off with the basic models including a detailed review and summary of existing work. Then, we outline recent advances in the field and their engineering applications.

39.1 **Basic Models** 907
39.1.1 Simplified Models 908
39.1.2 Biot's Poroelastic Models 909
39.1.3 Poro-Elastoplastic Models 915
39.2 **Mechanisms of Seabed Dynamics** 916
39.2.1 Oscillatory Soil Response 917
39.2.2 Residual Soil Response 918
39.2.3 Integrated Model 919
39.2.4 Wave-Induced Residual Liquefaction 921
39.3 **Wave(Current)-Induced Soil Response in Marine Sediments** 922
39.3.1 Theoretical Formulations 922
39.3.2 Effects of Currents 923
39.3.3 Liquefaction of Seabed Under Combined Nonlinear Wave and Current Loading 924
39.4 **Seabed Stability Around Caisson Breakwaters** 925
39.4.1 Theoretical Models 925
39.4.2 Verifications 926
39.4.3 Dynamic Response of Seabed Around Caisson Breakwaters 928
39.4.4 Wave-Induced Liquefaction Around Breakwaters 929
39.5 **Remarks** 931
References 932

39.1 Basic Models

Numerous models for the wave-induced seabed response have been developed with various assumptions since the 1940s. The assumptions and the leading governing equations for each model will be summarized here:

- *Simplified model*: In the model, both pore fluid and soil are considered as incompressible medium, and the accelerations due to fluid and soil motion are ignored. This leads to an uncoupled model, in which Laplace's equation is the governing equation [39.1]. Another similar approach is to include the compressibility of pore fluid, which leads to diffusion equation [39.2].
- *Biot's poro-elastic models*: Basically, three poro-elastic models have been developed since the 1970s, including consolidation model (or quasi-static model), *u-p* approximation, and dynamic models.
  - *Consolidation model*: In the consolidation model, both pore fluid and soil skeleton are considered to be compressible, but the accelerations due to fluid and soil motion are ignored. Since the acceleration has been ignored, this model is also called as the quasi-static model [39.3].
  - *u-p approximation*: In this model, the acceleration due to pore fluid is ignored in the general formulation. This model was first proposed by

Zienkiewicz et al. [39.4] with a one-dimensional analysis, and then extended to two-dimensional analysis for the wave–seabed interaction [39.5, 6].
  - *Dynamic model*: The full set of governing equations established by Biot [39.7] is employed in the analysis, in which both accelerations due to pore fluid and soil motion are included. Since this model is rather complicated, only a few investigations are available in the literature [39.8, 9].
- *Poro-elastoplastic model*: Besides linear poro-elastic models, some advanced models such as the poro-elasto-plastic model for the wave-induced seabed response have been developed recently [39.10, 11]. The poro-elastoplastic models will provide better predictions of the potential of the wave-induced seabed instability, which is normally a large deformation.

In the aforementioned models, both analytical and numerical models have been employed to obtain the wave-induced soil response. In the following sections, most previous theoretical investigations for the wave-induced seabed response will be reviewed under the heading of each model.

39.1.1 Simplified Models

Uncoupled models have been used as the first approximation in the area of wave–seabed interaction. In the model, either pore fluid or soil skeleton has been considered as incompressible medium. The accelerations due to both pore fluid and soil motion are ignored in this approach. Since the governing equation is either Laplace's equation or the diffusion equation, for which the analytical solutions have been well developed, most previous investigations with these assumptions are analytical approximations.

Models with the Laplace Equation

Based on the assumptions of a rigid, permeable sandy seabed and incompressible pore fluid, Laplace's equation is the governing equation for the wave-induced pore pressure,

$$\nabla^2 p = \frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} + \frac{\partial^2 p}{\partial z^2} = 0\,, \tag{39.1}$$

where p is the pore pressure in the seabed.

Using a linear wave theory, *Putnam* presented a simple solution for an isotropic porous seabed of finite thickness and concluded that a significant loss of wave energy occurred in the presence of a porous sandy seabed due to viscous percolation of fluid [39.1]. The percolation was activated by the pressure variation at the interface between sea water and seabed. The solution indicated that the pressure distribution within the seabed depended only on the wave characteristics and geometry of the sand layer, and not on the properties of the seabed. However, a possible error of the wave height was observed in *Putnam*'s paper, which overestimated the dissipation function by a factor of 4 [39.12].

Based on the same assumptions as Putnam, the wave-induced pore pressure for a porous seabed of finite thickness with anisotropic permeability was examined by *Sleath* [39.13]. He also conducted laboratory experiments to verify the theory, resulting in the discovery of a phase lag (less than 10°) in the pore pressure. However, these experimental results were inconsistent with his theoretical results. This inconsistence is due to the assumptions of his theoretical approach.

Considering the viscous effect of the boundary layer and energy balance, *Liu* modeled the flow in a permeable bed and determined the damping rate for an infinite seabed [39.14]. The continuity of pressure and velocity is required as boundary conditions at the interface up to order $O(\sqrt{\nu})$, where ν is the viscosity of the pore fluid. His results indicated that there was no relationship between pore pressure and permeability, whilst fluid velocity depended on the porosity and permeability. However, *Liu*'s solution only considered the pressure condition whilst neglecting shear stress [39.14]. Thus, it may not be a complete analysis of viscous flow. With the same framework, *Liu* further developed a solution for the damping of the wave-induced pressure in a two-layered porous seabed [39.15]. Compared with the solution for an infinite seabed [39.14], the pore pressure was found to depend on both the permeability and thickness of the upper layer only to a small degree. However, he only considered the case of a two-layered seabed with uniform permeability in each layer. Later, based on the generalized Darcy's equation [39.16], the boundary layers between the seabed surface and the impervious stratum are included in the model [39.17]. It was concluded that the spatial damping rate depended strongly on the permeability and the water depth, when the physical wave number remained approximately the same over its rigid bottom.

From a different aspect, *Massel* took into account the nonlinear damping and the inertia terms in the momentum equation for a rigid porous seabed [39.18]. His results indicated that the effect of permeability on the pressure distribution in the seabed was negligible and that they were essentially the same as that from Laplace's equation.

In describing the wave-induced soil response, *Mallard* used a set of elastic displacement equations based

on the assumption of stress equilibrium within the soil, neglecting the effect of soil inertia on the response [39.19]. *Dawson* considered the soil inertia term in the model of Mallard and concluded that the effect of the soil inertia term could not generally be ignored without committing a serious error for the case of an incompressible soil [39.20].

Models with the Diffusion Equation

Another type of uncoupled model has been proposed by *Nakamura* et al. [39.2] and *Moshagen* and *Torum* [39.21], based on assumptions of compressible pore fluid and a nondeformable porous soil skeleton. This results in the heat conduction equation or diffusion equation for pore pressure,

$$\nabla^2 p - \frac{\gamma_w n \beta}{K_z} \frac{\partial p}{\partial t} = 0 \,, \tag{39.2}$$

where γ_w is the unit weight of water, n is the porosity, β is the compressibility of pore water that is defined in (39.4), and K_z is the permeability in the z-direction.

Among these, *Nakamura* et al. compared theoretical results of pore pressure with laboratory experiments in fine and coarse sandy beds [39.2]. The experimental results for the latter showed no phase lag and agreed reasonably well with the solution from Laplace's equation. The data for fine sand exhibited a large pressure attenuation and phase lag, which agreed reasonably well with the diffusion theory. However, an unexplainable pressure discontinuity exists near the seabed surface in their experimental data. As *Yamamoto* et al. pointed out, the waves generated in the experiments of *Nakamura* et al. were too steep [39.22]. Therefore, the state of stresses in sandy beds under wave crests and troughs might have reached the limit of equilibrium or the state of liquefaction, thus causing a large pressure drop. Furthermore, a critical error was found in their calculations. The compressibility of the water used in the calculation was 980 times that of the real water. *Yamamoto* et al. showed that the false agreement reported might be explained by the existence of a small amount of air in the sand used.

Moshagen et al. considered the wave-induced flow in a porous medium under the assumption of compressible pore fluid and an incompressible soil skeleton [39.21]. They found that the inclusion of pore fluid compressibility in the analysis of the wave-induced pore pressures in a porous soil significantly altered the vertical seepage forces acting on the soil. However, the assumption made by [39.21] regarding the relative compressibility of the pore fluid and the soil skeleton appeared somewhat unrealistic [39.23]. Thus, serious doubts arise about the validity of *Moshagen* and *Torum*'s [39.21] conclusion.

All the aforementioned theories assumed that the seabed is a rigid porous medium. Because these approaches do not permit the coupling of pore-fluid motion and soil skeleton motion, the governing equation for the pore pressure is Laplace's equation for incompressible fluid, or the diffusion equation for compressible pore fluid for hydraulically isotropic (isotropical permeability) seabeds. However, such solutions for pore pressure are limited to a particular case of soil and wave conditions, that is, Laplace's equation for very permeable beds such as coarse sand, or a diffusion equation for poorly permeable beds such as clay. Furthermore, these approaches provide no information for the effective stresses and soil displacements in the seabed.

39.1.2 Biot's Poroelastic Models

Coupled models generally treat both the soil and the pore fluid as compressible medium. They can more precisely describe the mechanical properties of porous medium and the soil–pore fluid interaction. Recently, the investigations of the consolidation or the dynamic response of seabed soil were conducted using this coupled models. Biot's theory is the coupled model most widely used for the soil–pore fluid interaction. Depending on the inclusion of acceleration of soil particles and pore fluid, and the relative displacement of pore fluid to soil particles, there are three types of expression of Biot's equation: quasi-static model, *u-p* approximation, and full dynamic mdoels.

Consolidation Models (Quasi-Static Model)

Biot first formulated a three-dimensional consolidation equation treating the soil as an isotropic poro-elastic porous medium with compressible pore water and deformable soil particles [39.3] with the following assumptions:

1. The soil is homogeneous and isotropic.
2. The stress–strain relation is reversible under final equilibrium (linear elastic).
3. Small deformation of solid and pore fluid are considered
4. The pore fluid in porous, and soil particles are compressible.
5. The water flows steadily within porous medium, Darcy's flow ($R_e \leq 1$ [39.24]), the quasi-static model was established [39.3].

Based on the conservation of mass, we have

$$\frac{K_x}{K_z}\frac{\partial^2 p}{\partial x^2} + \frac{K_y}{K_z}\frac{\partial^2 p}{\partial y^2} + \frac{\partial^2 p}{\partial z^2} - \frac{\gamma_w n \beta}{K_z}\frac{\partial p}{\partial t} = \frac{\gamma_w}{K_z}\frac{\partial \epsilon_s}{\partial t} \,, \tag{39.3}$$

for a hydraulically anisotropic seabed. In (39.3), n is the soil porosity, γ_w is the unit weight of the pore-water, K_x, K_y, and K_z are the soil permeabilities in the x-, y-, and z-directions, respectively; the compressibility of the pore fluid (β) and the volume strain (ϵ_s) are defined as

$$\beta = \frac{1}{K_w} + \frac{1-S}{P_{w0}}, \quad \text{and} \quad \epsilon_s = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z}, \tag{39.4}$$

where K_w is the true bulk modulus of elasticity of the pore water that is taken as $1.95\times10^9\,\mathrm{N/m^2}$ [39.22]; u, v, and w are the soil displacements in the x-, y-, and z-directions, respectively; S is the degree of saturation; P_{w0} is the absolute water pressure. If the soil skeleton is fully saturated, then, $\beta = 1/K_w$ since $S = 1$.

On the basis of force balance and Hooke's law, we have the governing equations for overall equilibrium in a pore-elastic medium as

$$G\nabla^2 u + \frac{G}{1-2\mu}\frac{\partial \epsilon_s}{\partial x} = \frac{\partial p}{\partial x}, \tag{39.5}$$

$$G\nabla^2 v + \frac{G}{1-2\mu}\frac{\partial \epsilon_s}{\partial x} = \frac{\partial p}{\partial y}, \tag{39.6}$$

$$G\nabla^2 w + \frac{G}{1-2\mu}\frac{\partial \epsilon_s}{\partial x} = \frac{\partial p}{\partial z}. \tag{39.7}$$

In Biot's poro-elastic theory, the flow of fluid in porous medium is generally treated as steady flow. Therefore, Darcy's law is used in formulating the equations. Furthermore, only the small deformation problem could be applicable. It is worth noting that *Biot*'s consolidation equations [39.3] ignored the inertia terms of the solid and fluid. This kind of simplification is acceptable for the consolidation process with small permeability or low-frequency loading problems as the acceleration of the solid or fluid is apparently small under this situation.

Most previous investigations with Biot's consolidation equations have been directly solved and obtain the wave-induced pore pressure, soil displacements, and effective stresses. This approach was first developed by *Yamamoto* et al. [39.22], *Yamamoto* [39.25] and *Madsen* [39.26], who considered compressible pore fluid in a compressible porous medium. A three-dimensional general consolidation equation [39.3] was adopted in these studies, in which only progressive waves were examined. Among these, *Madsen* [39.26] considered a hydraulically anisotropic and unsaturated porous bed, whilst *Yamamoto* et al. [39.22] studied an isotropic medium. Both considered only an infinite thickness. Moreover, *Yamamoto* [39.25] investigated soil response in a homogeneous soil of finite thickness under isotropic and partially saturated conditions. However, *Yamamoto*'s solution [39.25] was cast in a semianalytical manner that did not have a closed form.

Yamamoto et al. [39.22] concluded that when the stiffness of a porous medium is much smaller than that of the pore fluid (e.g., saturated soft soils), the soil response is independent of the permeability and has no phase lag. On the other hand, when the stiffness of a porous medium is much larger than that of the pore fluid (e.g., partially saturated dense sands), pore pressure attenuates rapidly. In the latter case, the phase lag increases linearly with the distance from the seabed surface.

Madsen [39.26] investigated a hydraulically anisotropic and partially saturated seabed. He found that this had an appreciable effect on the nature of the wave-induced effective stresses in coarse sand. For all soils, the effect of partial saturation on soil response may be significant.

Yamamoto [39.27] developed a semianalytical solution for a nonhomogeneous layered porous seabed, together with a comprehensive verification using data obtained from Mississippi Delta. *Yamamoto* [39.27] pointed out that a layer of concrete blocks had a significant effect on the wave-induced soil response. However, the assumption of treating the concrete blocks as soils seems unrealistic because the properties of concrete blocks are quite different from soil.

Okusa [39.28] used the compatibility equation under elastic conditions and reduced the governing equation of *Yamamoto* et al. [39.22] to a fourth-order differential linear equation. It is noted that *Okusa*'s experiment [39.28] was based on plane stress conditions, while *Yamamoto* et al. [39.22] was based on plane strain conditions. *Okusa* [39.28] found that the wave-induced pore pressure and effective stresses consisted of two parts. The first depended only on the wave characteristics and the second was related to both the sediment and the wave characteristics. He reported that the wave-induced soil response depended only on the wave conditions, not on the soil characteristics for a fully saturated and isotropic sandy seabed of infinite thickness. However, this conclusion is invalid for an isotropic seabed of finite thickness, even under a saturated condition [39.29, 30].

Rahman et al. summarized the previous work with the direct analytical framework in a semianalytical analysis [39.31]. In their model, a general layered seabed is considered, which is particularly important for the design of a cover layer for seabed protection.

The aforementioned investigations have been limited to an isotropic homogeneous seabed, which may be an idealized case. The major difficulty in analyzing the wave-induced soil response in a seabed with vari-

able permeability has been the governing equation that includes variable coefficients. By employing the VS (Varley-Seymour) function [39.32], *Seymour* et al. derived an analytical solution for such a condition [39.33]. In their study, only fine sand is considered. In their model, the first-order derivation of permeability with respect to vertical distance was excluded, which has been reported to play an important role in the evaluation of wave-induced soil response in coarser material [39.34]. Later, *Jeng* and *Seymour* further developed analytical solutions for general soils in seabed of both infinite and finite thicknesses [39.35, 36]. They concluded that the relative difference of the wave-induced pore pressure between variable and uniform permeability might go up to 23% of the amplitude of the wave pressure at the seabed surface.

Recently, another analytical solution for the wave-induced seabed response with variable permeability was proposed by *Kitano* and *Mase* [39.37]. However, their one-dimensional model is limited to the exponential distribution decay of the permeability, although it provides a simpler formulation than that of [39.35, 36].

The inclusion of cross-anisotropic soil behavior in the wave–seabed interaction can also be handled analytically. *Jeng* may have been the first to derive the analytical solution for the wave–seabed interaction in a cross-anisotropic seabed [39.38, 39]. His numerical results show that the conventional solution with the assumption of isotropic soil behavior may overestimate the pore pressure, but underestimate the effective stresses. Consideration of cross-anisotropic soil behavior is particularly important in determining the wave-induced soil displacements. An identical approximation is also proposed by *Yuhi* and *Ishida* [39.40].

A simplified analytical solution for the soil response in a cross-anisotropic seabed was proposed by *Yuhi* and *Ishida* [39.41]. In the model, two parameters related to boundary-layer thickness and the stiffness ratio were introduced. However, their model is only applicable in the region near the seabed surface (e.g., $|z|/L < 0.02$, L is the wavelength of ocean waves), outside this region, the relative difference between the simplified solution [39.41] and exact solution [39.38] is not negligible, as reported in [39.42].

Since the direct analytical solution for the wave-induced seabed response is involved with complicated mathematical presentations, especially for a seabed of finite thickness, it is even impossible to have a closed-form solution for a layered seabed. An alternative approximation, the boundary-layer approximation, was proposed by *Mei* and *Foda* [39.43]. The principle of the boundary-layer approximation is to divide the whole soil domain into two regions: inner and outer regions. In the inner region (near the seabed surface, defined by the boundary layer thickness), the full solution is required. On the other hand, a simplified solution is sufficient in the outer region. This solution agrees well with that of *Yamamoto* for fine sand. However, it may lose accuracy for all soils in unsaturated conditions and for coarse sand under saturated conditions [39.25]. This shortcoming may be attributed to the solution being only suitable for a seabed with low permeability, for which a scaling was carried out [39.44]. However, their solutions are more convenient for engineering application due to their much simpler form, compared to those of *Hsu* and *Jeng* [39.45].

Although the boundary-layer approximation proposed by *Mei* and *Foda* [39.43] is a simple yet fairly accurate analysis, it was restricted to low-frequency waves only. *Huang* and *Chwang* investigated Biot's equation for the acoustic problem and obtained three uncoupled Helmholtz equations to represent each of the three kinds of waves [39.46]. Their approach is applicable for the complete range of wave frequencies.

Later, *Kitano* and *Mase* derived another set of analytical solution to investigate the influence of cross-anisotropic soil behavior on the wave-induced soil response through the boundary-layer approximation [39.47]. Their model agreed well with the exact solution proposed by Jeng for fine sand.

Besides the direct analytical solution and boundary-layer approximation, [39.48] proposed a random-walk model for the wave-induced pore pressure in marine sediments. The principle of the model is based on the shear stress in the marine sediment to determine the build-up pore pressure. However, the determination of the shear stress relied on other analytical solutions such as that of *Hsu* and *Jeng* [39.45].

Madga developed a one-dimensional finite-difference model for the wave-induced pore pressure in a highly saturated sandy bed and applied to the cases of buried pipelines [39.49–51]. He concluded that the time phase in pressure generation is dominated by the degree of saturation, compressibility of the soil skeleton and soil permeability.

Zen and *Yamazak* simplified a two-dimensional boundary value problem to one dimension, based on the assumption of the seabed thickness being very small compared with the wavelength [39.52, 53]. A numerical model (finite difference method) was also established, which was only applicable to a single layer of porous seabed.

Gatmiri developed a simplified finite element model for the wave-induced effective stresses and pore pressure in an isotropic and saturated permeable seabed [39.29]. Two important conclusions were drawn from his paper. First, there exists a critical bed thickness

about 0.2 times the wavelength, in which the horizontal movement of the soil skeleton is maximum and where the unstable state occurs. Second, the soil response is affected by soil characteristics even in a hydraulically isotropic and saturated seabed of finite thickness. This result complemented the solution for a seabed of infinite thickness reported by Okusa. However, the general trend of pore pressure distribution versus the seabed thickness in [39.29] was found to be inconsistent [39.30].

Gatmiri further extended his numerical model to consider the soil response in a cross-anisotropic saturated seabed [39.54]. The numerical results showed that the effects of cross-anisotropic soil parameters are significant and the soil response was affected by the combined parameters in different ways. Compared with the cross-anisotropy, the effect of hydraulic anisotropic permeability on the variation of effective stresses may be insignificant.

A possible error in the results of Gatmiri's model may have stemmed from the boundary condition used. The lateral boundaries at $x/L = 0, 1$ in his paper, $v = 0$, $p = 0$ and u free, were used in his model. However, it has been proved that there is a phase lag in soil response in a fully saturated seabed of finite thickness [39.30]. This implies that the lateral boundary conditions, $v = 0$, $p = 0$, are invalid in a porous seabed of finite thickness. Thus, the numerical results of [39.29, 54] seem doubtful. In fact, this obstacle can be overcome by using the principle of repeatability [39.55], as suggested by the author [39.56].

Thomas developed a one-dimensional finite element method for a two-layered unsaturated seabed [39.57, 58]. His result agreed well with the analytical solutions of Yamamoto and Okusa. It also suggested that the stiffer sediment in the top layer dominated the response of the bottom layer in a two-layered seabed.

With a similar framework, a series of one-dimensional finite element models was developed for the wave-induced seabed response in a nonhomogeneous seabed [39.59–61]. The results of these numerical models agree well with previous two-dimensional experimental data and analytical solutions. In the models, the permeability and shear modulus are considered to vary with burial depth. *Jeng* and *Lin* [39.61] have also examined the influence of nonlinear wave components on the soil response. In addition, they examined the combined effect of cross-anisotropic soil behavior and nonhomogeneous soil characteristics on the wave-induced soil response [39.62].

The advantage of the above one-dimensional finite element models is the reduced computational time. However, such one-dimensional models cannot apply to the case with a structure. Thus, a two-dimensional finite element model was proposed [39.56] by employing the principle of repeatability [39.55].

u–p Model

Zienkiewicz et al. [39.4] proposed a simplified form of *Biot*'s poro-elastic theory [39.7]. In this model, the relative acceleration of fluid to solid is eliminated from the stress equilibrium equations, because the relative acceleration of fluid is apparently small for porous medium with a low permeability, and acted by low-frequency dynamic loading. Then, the governing equations (39.3) and (39.5)–(39.7) can be rewritten as

$$\frac{K_x}{K_z}\frac{\partial^2 p}{\partial x^2} + \frac{K_y}{K_z}\frac{\partial^2 p}{\partial y^2} + \frac{\partial^2 p}{\partial z^2} - \frac{\gamma_w n\beta}{K_z}\frac{\partial p}{\partial t} + \rho_f \frac{\partial^2 \epsilon_s}{\partial t^2} = \frac{\gamma_w}{K_z}\frac{\partial \epsilon_s}{\partial t}\,, \tag{39.8}$$

$$G\nabla^2 u + \frac{G}{1-2\mu}\frac{\partial \epsilon_s}{\partial x} = \frac{\partial p}{\partial x} + \rho\frac{\partial^2 u}{\partial t^2}\,, \tag{39.9}$$

$$G\nabla^2 v + \frac{G}{1-2\mu}\frac{\partial \epsilon_s}{\partial x} = \frac{\partial p}{\partial y} + \rho\frac{\partial^2 v}{\partial t^2}\,, \tag{39.10}$$

$$G\nabla^2 w + \frac{G}{1-2\mu}\frac{\partial \epsilon_s}{\partial x} = \frac{\partial p}{\partial z} + \rho\frac{\partial^2 w}{\partial t^2}\,, \tag{39.11}$$

where ρ_f is the density of the pore water; and $\rho = (1-n)\rho_s + \rho_f$ (where ρ_s is the density of solid).

Based on *Biot*'s poro-elastic theory [39.7, 63], *Zienkiewicz* et al. proposed a one-dimensional so-called *u-p* approximation for waves propagating over a porous medium [39.4]. Their approximation provided a simplified approximation for such a problem. Based on the numerical examples, Zienkiewicz concluded that neither the *u-p* approximation nor the dynamic approximation is required for the case of ocean waves over a seabed. However, in their examples, the compressive wave velocity under ocean wave loading was fixed at 1000 m/s, which is applicable during an earthquake, but may not be a reasonable value for most ocean environments. Thus, the conclusion drawn from the examples was in doubt.

Sakai et al. extended Mei and Foda's boundary-layer approximation to examine the effect of inertia and gravity force of the ocean waves on the seabed response, and verified it with results of their numerical model (finite element method) [39.64]. They concluded that the inertia term (acceleration) could be neglected under normal wave conditions except for breaking, whilst the gravity term can practically be ignored. Later, *Sakai* et al. modified the boundary-layer approximation to take into account the effect of wave-induced bottom shear stress which cannot be neglected in the surf zone [39.65]. In his paper, it was simply assumed

that the amplitude of the wave-induced bottom shear stresses was proportional to the wave pressure, without any phase lag. However, it has been reported that the wave-induced bottom shear stress has a phase lag of 45° to the wave-induced pressure at the seabed surface [39.66].

The one-dimensional *u-p* approximation [39.4] was further extended to two-dimensional waves over a porous seabed [39.5, 6]. The seabed was considered to be either infinite or finite thickness. It was found that the relative differences between the *u-p* approximation and the quasi-static solution may reach 17% of the amplitude of wave pressure.

In addition to the above analytical solutions, the finite element model (SWANDYNE II) was developed by *Chan* [39.67]. This model was initially developed for the earthquake-induced soil liquefaction and adopted for the wave-induced residual soil response around marine structure such as pipelines [39.68] and breakwaters [39.11]. In this model, the *u-p* approximation was used together with poro-elastoplastic constitutive model.

Full Dynamic Model

Based on the stress balance, momentum conservation and mass conservation, [39.7, 63] extended the above consolidation theory to dynamic theory by introducing the inertia terms of solid and fluid, to investigate the propagation of elastic waves in saturated porous medium. Then, *Biot* [39.69] proposed a set of general fully coupled governing equations for saturated porous medium with compressible fluid and incompressible soil particles for consolidation and dynamic problems, in which the effect of inertia terms of solid and fluid, and the interaction (drag force) between soil and pore fluid are taken into consideration. The stress equilibrium equation, momentum equilibrium equation and continuity equation, respectively, are (in the form of tensor)

$$\sigma_{ij,j} + \rho g_i = \rho \ddot{u}_{si} + \rho_f(\ddot{w}_{fi} + \dot{w}_{fj}\dot{w}_{fi,j}) \,, \tag{39.12}$$

$$-p_{,i} + \rho_f g_i = \rho_f \ddot{u}_{si} + \frac{\rho_f}{n}(\ddot{w}_{fi} + \dot{w}_{fj}\dot{w}_{fi,j}) + \frac{\gamma_w}{K_{ij}}\dot{w}_{fj} \,, \tag{39.13}$$

$$\dot{\epsilon}_{ii} + \dot{w}_{fi,i} + \frac{1}{Q}\dot{p} = 0 \,, \tag{39.14}$$

where u_s represents the soil displacement tensor, w_f denotes the related displacements between solid and pore fluid, and Q is defined as

$$\frac{1}{Q} = n\beta + \frac{1-n}{K_s} \,, \tag{39.15}$$

in which K_s is the bulk modulus of solid.

The major difference between dynamic analysis and *u-p* approximation has been the inclusion of acceleration due to pore fluid. With a similar framework to [39.4], *Jeng* and his co-workers investigated the effects of dynamic soil behavior on the wave-induced soil response through a two-dimensional analysis [39.8, 70]. Although their dynamic solution can provide a better prediction of the wave-induced soil response, their solutions were lengthy and difficult to be applied in engineering practice.

Based on the boundary-layer approximation, *Huang* and his co-workers further developed a series of analytical solutions for waves propagating over a soft poro-elastic bed [39.44, 71, 72]. Both linear and nonlinear wave loading were considered. Their approximation provided a simpler formulation, compared with the full closed-form solution.

Alternatively, on the basis of governing equations derived by *Mei* and *Foda* [39.43], *Yuhi* and *Ishida* [39.73] directly solved the boundary value problem, rather than using a boundary-layer approximation. The seabed was considered to be of infinite thickness. The characteristics of three waves within the soil column, including two compressive waves and one shear wave, was discussed in detail. This work has been later extended to the case of finite thickness [39.74].

Applicable Ranges of Biot's Models

To date, the full dynamic model, *u-p* approximation and Biot's consolidation equations have been widely used for investigating the quasi-static and dynamic problems. Basically, the full dynamic is applicable for all quasi-static and dynamic problems due to its relatively complete form to describe the mechanical behavior of porous medium, in which the acceleration of soil particles and pore fluid, and the relative displacement of pore fluid to soil particles are all included. However, some preconditions have been assumed when deriving the *u-p* approximation, and Biot's consolidation equations. Therefore, there are some limitations for applying the *u-p* approximation and Biot's consolidation equations to engineering problems.

Zienkiewicz et al. first made their attempt to investigate the applicable range of the *u-p* approximation and Biot's consolidation equations [39.4]. Based on a simple linear one-dimensional model, in which only one soil layer is contained, and subjected to periodic loading, the full dynamic form, *u-p* approximation, and Biot's consolidation equations are all analytically solved with the same boundary conditions. After comparing the solutions obtained from the three kinds of governing equations, a graph was provided to illustrate the applicable range of frequency and permeability for the *u-p* approximation, and Biot's consolidation equa-

tions. From the figure presented in [39.4, Fig. 3], it was concluded that the full dynamic form can be used in all zones, including I, II, and III. The *u-p* approximation can be used in zones I and II with the largest applicable frequency of approximately 1000 Hz. Biot's consolidation equations can only be used within zone I. It is only applicable for low-frequency and low-permeability problems.

It is noted that the above conclusion was obtained from the simplest one-dimensional soil layer; and the values of some property parameters of soil are assumed, such as $V_c = 1000\,\mathrm{m/s}$. As pointed out by *Jeng* and *Cha* [39.8], the velocity of wave in soil is significantly dependent on the saturation of soil and the water depth. The value of wave velocity used in [39.4] is only a special case. Therefore, the information coming from the figure is just a rough approximation, its applicability is limited.

Later, *Jeng* and *Cha* further analytically investigated the applicable range of the *u-p* approximation and Biot's consolidation equations for the two-dimensional (2-D) model of soil–wave interaction based on the full dynamic form [39.8]. Through solving the partial differential equations in the full dynamic form, they obtained the analytical solution of the pressure of the pore fluid, and the displacement of soil in a finite or an infinite seabed under wave loading. Then the analytical solutions based on the *u-p* form and Biot's consolidation equations are accordingly determined by ignoring the effects of the inertial term in the solution derived from the full dynamic form. In the above three analytical solutions, a linear ocean wave is used as the dynamic loading acting on the seabed. Due to the fact that the frequency of ocean wave is the magnitude of $O(10^{-1})$ Hz. It can definitely be clarified into the range of low frequency. Therefore, it is not surprising that the results from the *u-p* approximation and full dynamic form are always basically the same. A comparison has also been performed between the results obtained from the *full dynamic* form and from Biot's consolidation equations. Figure 39.1 shows the boundary line of applicable ranges for different permeabilities between the conditions under which *full dynamic* form or *u-p* approximation must be adopted and the conditions under which Biot's consolidation equations could be used. In Fig. 39.1, the abscissa Π_1 and the longitudinal coordinates Π_2 are defined as

$$\Pi_1 = \frac{K_z V_c^2 k^2}{\gamma_w \omega} \quad \text{and} \quad \Pi_2 = \frac{\rho_f \omega^2}{\left(\frac{G}{1-2\nu} + \frac{K_f}{n}\right) k^2}\,, \tag{39.16}$$

where K_z is the permeability coefficient, and k is the wave number. V_c^2 is the velocity of the compressibility

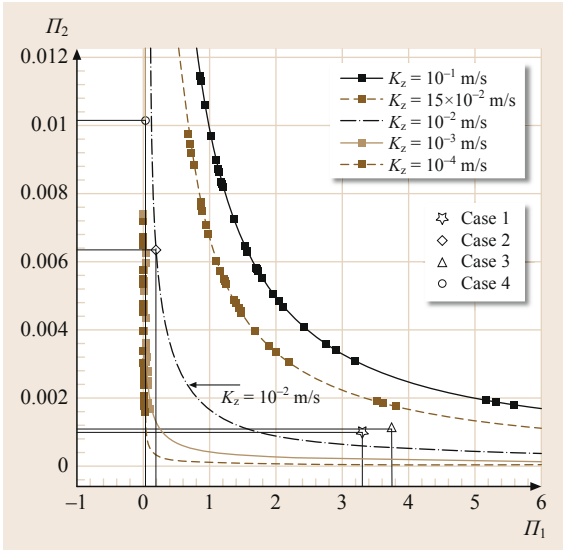


Fig. 39.1 The boundary line of applicable range for different permeabilities for full dynamic form and Biot's consolidation equations (after [39.8, 75])

wave in porous soil, and could be expressed as

$$V_c^2 = \frac{\frac{G}{1-2\nu} + \frac{K_f}{n}}{\rho_f}\,. \tag{39.17}$$

The parameter Π_1 is related to the permeability of soil, and Π_2 is related to the frequency of the ocean wave acting on the seabed. *Jeng* and *Cha* proposed an equation to describe the boundary line in the $\Pi_1 - \Pi_2$ nondimensional space [39.8]

$$\Pi_2 = C\Pi^m = 0.0298\left(\frac{K_z}{\Pi_1}\right)^{0.5356}\,. \tag{39.18}$$

For certain soil conditions, the full dynamic form or *u-p* approximation should be adopted if the point (Π_1, Π_2) is located above the boundary line, else a large computational error would result. The result obtained by [39.8] is more liable than that of [39.4] due to the fact that the soil model used by [39.4] is one dimensional, and the velocity of compressibility wave in the soil is assumed as 1000 m/s which is invalid for most soil layers.

Recently, *Ulker* et al. further investigated the applicable range of the full dynamic form, *u-p* approximation (partial dynamic), and Biot's consolidation equation (quasi static) [39.9, 76]. Firstly, a generalized analytical solution is developed for the wave-induced seabed response by adopting the similar method used by *Jeng* and *Cha* [39.8]. In the analytical solution, the full dynamic form is taken as the governing equation with the addi-

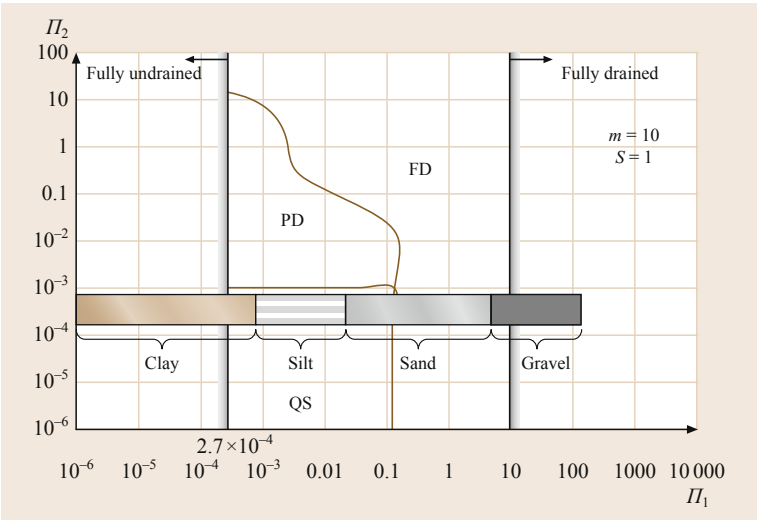


Fig. 39.2 Regions of applicability of the three types of governing equations for the wave-induced seabed response (after [39.9])

tion of the gravitational term, which was not considered by [39.8]. The solutions for the *u-p* approximation and Biot's consolidation equations are obtained by reducing the analytical solution from the full dynamic form. The linear wave and poro-elastic model are adopted as the external loading and the constitutive model of seabed soil. Through comparing the results of pore pressure, shear stress, and effective vertical stress obtained from the three types of governing equations, [39.9] proposed a graph (Fig. 39.2) to illustrate the applicable range of the three types of governing equations. In Figure 39.2, it is found that Biot's consolidation equations are accurate for clay soils; the *u-p* approximation is necessary for silty soils. For a sandy seabed, the formulation depends on the permeability and the loading frequency; the full dynamic model has to be used for gravels whose permeability is generally relatively a large magnitude.

In summary, the *u-p* approximation and Biot's consolidation equations are sufficient for the cases with low-frequency loading and small permeability soils; the full dynamic form has to be used for the cases with high-frequency loading and/or high-permeability soil. In this dissertation, the permeability of seabed is generally from 1.0×10^{-5} m/s to 1.0×10^{-4} m/s (maximum 1.0×10^{-2} m/s for parametric study); and the magnitude of dominant frequency is $O(10^{-1})$ Hz for ocean wave and $O(10^{1})$ Hz for earthquake loading. The *u-p* approximation is sufficient to describe the dynamic behavior of seabed foundation under marine structures.

39.1.3 Poro-Elastoplastic Models

All aforementioned poro-elastic models have been limited to small deformations, which is an idealized condition. However, large deformations will be a more important issue for engineering problems, especially under the action of a storm. In that case, poro-elastoplastic models are required to provide a better estimation of the soil response. Since a poro-elastoplastic model is more complicated than a poro-elastic model, especially in the problem of the wave–seabed interaction, to date, only a few investigations have been published.

Sekiguchi et al. may have been the first to derive an analytical solution for waves propagating over a poro-elastoplastic seabed with Laplace's transformation [39.77]. Their model clearly demonstrated the difference between poro-elastic and poro-elastoplastic models. Also, overall agreement between their model and centrifugal tests was reported. However, their model was based on the assumption of the seabed thickness being much less than the wavelength, which is questionable [39.78].

With the nonassociative bounding surface plastic constitutive model, *Yang* and *Poorooshasb* developed a numerical code to investigate the standing wave-induced soil response in a seabed [39.79]. Their model demonstrated the significant influence of permeability on the wave-induced seabed response. However, no comparison between elastic and their plastic models was made in their paper. Thus, the importance of plastic model was not observed.

Li et al. established a poro-elastoplastic model for wave propagation in saturated media, using the Drucker–Prager criterion to describe nonlinear constitutive behavior of pressure-dependent elastoplasticity for the solid skeleton [39.80]. Based on their model, they further investigated the stationary discontinuities and flutter instability of wave propagation in a saturated seabed, and concluded that the flutter instability may occur prior to stationary discontinuity only if the

derivative effective stress normal to the surface of discontinuity is compressive.

Recently, *Jeng* and *Ou* proposed a three-dimensional (3-D) poro-elastoplastic model for the wave-induced soil response around breakwater heads [39.11]. In the model, the poro-elastoplastic (PZ3) [39.67] was adopted to predict the liquefaction potential of the seabed in the vicinity of breakwater heads. Their study demonstrated the significant effects of large deformation on the prediction of liquefaction zone.

39.2 Mechanisms of Seabed Dynamics

In general, two mechanisms for wave-induced pore pressure have been observed in field measurements and laboratory experiments, as shown in Fig. 39.3. The first mechanism is resulted from the transient or oscillatory excess pore pressure and accompanied by attenuation of the amplitude and phase lag in the pore pressure changes [39.22, 26]. is particularly important for small-amplitude waves and it could only liquefy momentarily in the seabed under wave troughs [39.75]. The second mechanism is termed as the residual pore pressure, which is the build-up of excess pore pressure caused by the contraction of the soil under the action of cyclic loading [39.81, 82]. As reported in [39.75], the residual mechanism is more important for large wave loading.

As mentioned previously, the wave-induced pore pressure within marine sediments consists of two components: oscillatory ($\tilde{p}$) and residual ($\bar{p}$) mechanisms, as shown in Fig. 39.3, which can be expressed as

$$p = \tilde{p} + \bar{p}\,, \tag{39.19}$$

where $\tilde{p}$ represents the oscillatory pore pressure that leads to momentary liquefaction, while $\bar{p}$ represents the period-averaged pore pressure that leads to residual liquefaction, and is defined by

$$\bar{p} = \frac{1}{T}\int_{t}^{t+T} p\,\mathrm{d}t\,, \tag{39.20}$$

where T is the wave period and t is the time. Figure 39.4 illustrates the ranges that each mechanism dominate the process of the wave-induced soil response. In the figure, the scale factor (ϵ) is defined as

$$\epsilon = \frac{\bar{p}_{t\to\infty}}{|\tilde{p}|}\,, \tag{39.21}$$

where $\bar{p}_{t\to\infty}$ denotes the equilibrium residual pore pressure, that is, the pore pressure when time approaches to infinite, as shown in Fig. 39.3; and $|\tilde{p}|$ represents the amplitude of oscillatory pore pressure.

In Fig. 39.4, various coefficients of consolidation (c_v) are used to represent different types of soil. With given wave steepness (H/L) and relative water depth (d/L), we can identify the position of the case we considered. When the case is located above the curve of the given c_v, the residual mechanism will dominate the whole process. On the other hand, if it is located below the curve, the oscillatory mechanism will be more important than residual mechanism. As shown in the figure, residual mechanism will be more important in either shallow water depth or large waves.

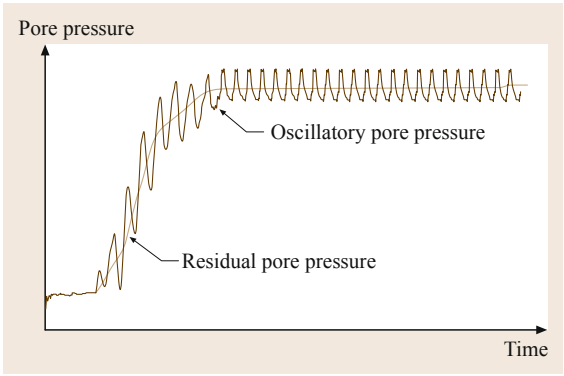


Fig. 39.3 Conceptual sketch of two different mechanisms of pore pressure (not in scale)

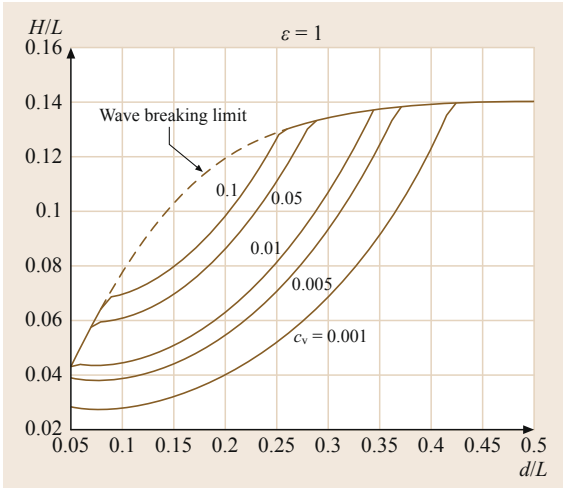


Fig. 39.4 Relationship of two mechanisms with various conefficient of consolidation (after [39.75])

In this section, we will outline the analytical solutions for both mechanisms and present several examples to show the effects of several important parameters. Then, we will present a new model that integrates both mechanisms to have a better understanding of the wave-induced soil response.

39.2.1 Oscillatory Soil Response

Numerous studies for wave-induced momentary seabed response, caused by oscillatory pore pressure, were completed in the past few decades. Among these, *Madsen* [39.26] and *Yamamoto* et al. [39.22] considered an infinite seabed with a similar framework. *Mei* and *Foda* [39.43] proposed a boundary-layer approximation to derive a rather simplified formulation for the wave-induced transient pore pressure, valid for coarse sand [39.45]. *Okusa* [39.28] further demonstrated the significant influence of the degree of saturation on the pore pressure and proposed the criteria of liquefaction. *Jeng* and his co-workers derived a series of analytical solutions for the oscillatory pore pressure within marine sediments with different considerations such as cross-anisotropic and nonhomogeneous soil layers, which were detailed in [39.75]. Here, we only outline the solution in an isotropic homogeneous seabed in two-dimensional.

Based on Biot's consolidation model, outlined in (39.3) and (39.5)–(39.7), we can have the analytical solution for the wave-induced socillatory pore pressure and soil displacements under 2-D progressive waves in an infinite seabed as prese [39.75, 83] as

$$\tilde{p} = \frac{p_o}{1-2\mu}\left[(1-2\mu-\lambda_*)C_1^\infty\, e^{kz} + \frac{\delta^2-k^2}{k}(1-\mu)C_2^\infty\, e^{\delta z}\right] e^{i(kx-\omega t)}, \quad (39.22)$$

$$\tilde{u} = \frac{ip_o}{2G}\left[(C_0^\infty + C_1^\infty z)\, e^{kz} + C_2^\infty\, e^{\delta z}\right] e^{i(kx-\omega t)}, \quad (39.23)$$

$$\tilde{w} = \frac{p_o}{2G}\left[\left(C_0^\infty - \frac{1+2\lambda_*}{k}C_1^\infty + C_1^\infty z\right) e^{kz} + \frac{\delta}{k}C_2^\infty e^{\delta z}\right] e^{i(kx-\omega t)}, \quad (39.24)$$

where the superscripts ∞ denote the coefficients in the solution for a porous seabed of infinite thickness.

Two parameters δ and λ_* are expressed as

$$\delta^2 = k^2\frac{K_x}{K_z} - \frac{i\omega\gamma_w}{K_z}\left(n'\beta + \frac{1-2\mu}{2G(1-\mu)}\right), \quad (39.25)$$

$$\lambda_* = \frac{(1-2\mu)\left[k^2(1-K_x/K_z) + \frac{i\omega\gamma_w n'\beta}{K_z}\right]}{k^2(1-K_x/K_z) + \frac{i\omega\gamma_w}{K_z}\left(n'\beta + \frac{1-2\mu}{G}\right)}. \quad (39.26)$$

In the above equations, the wave frequency ω ($= 2\pi/T$, where T is the wave period), wave number k ($= 2\pi/L$, where L is the wave length), and water depth (d) satisfy the dispersion equation

$$\omega^2 = gk\tanh kd\,, \quad (39.27)$$

and the amplitude of the dynamic wave pressure (p_o) is defined as

$$p_o = \frac{\gamma_w H}{2\cosh kd}\,, \quad (39.28)$$

in which H is the wave height.

The effective normal stresses (σ_x' and σ_z') and shear stress (τ_{xz}) can be expressed as

$$\tilde{\sigma}_x' = -p_o\left\{\left[(C_0^\infty + C_1^\infty z) + \frac{2\mu\lambda_*}{1-2\mu}C_1^\infty\right] e^{kz} + \left(k^2 - \frac{\mu(\delta^2-k^2)}{k(1-2\mu)}\right)C_2^\infty\, e^{\delta z}\right\} e^{i(kx-\omega t)}, \quad (39.29)$$

$$\tilde{\sigma}_z' = p_o\left\{\left(kC_0^\infty + C_1^\infty kz - \frac{2\lambda_*(1-\mu)}{1-2\mu}C_1^\infty\right) e^{kz} + \frac{1}{k(1-2\mu)}\left[\delta^2(1-\mu) - k^2\mu\right]C_2^\infty\, e^{\delta z}\right\} \times \cos nky\, e^{i(kx-\omega t)}, \quad (39.30)$$

$$\tilde{\tau}_{xz} = ip_o\left\{\left[kC_0^\infty + (kz-\lambda_*)C_1^\infty\right] e^{kz} + \delta C_2^\infty e^{\delta z}\right\} \times e^{i(kx-\omega t)}\,, \quad (39.31)$$

In Eqs. (39.22)–(39.31), the C_i^∞ coefficients are given by

$$C_0^\infty = \frac{-\lambda_*[\mu(\delta-k)^2 - \delta(\delta-2k)]}{k(\delta-k)(\delta-\delta\mu+k\mu+k\lambda_*)}\,, \quad (39.32)$$

$$C_1^\infty = \frac{\delta-\delta\mu+k\mu}{\delta-\delta\mu+k\mu+k\lambda_*}\,, \quad (39.33)$$

$$C_2^\infty = \frac{k\lambda_*}{(\delta-k)(\delta-\delta\mu+k\mu+k\lambda_*)}\,. \quad (39.34)$$

It is noted that the above solution is only for an infinite seabed. For more complicated cases, such as a seabed of finite thickness, a layered seabed, cross-anisotropic seabed, or nonhomogeneous seabed, readers can refer to [39.75].

39.2.2 Residual Soil Response

Based on the solution presented previously, (39.22)–(39.31), we have the amplitude of the oscillatory pore pressure ($|\tilde{p}|$) and shear stress ($|\tilde{\tau}_{xz}|$) for a saturated infinite seabed as

$$|\tilde{p}| = p_0\,\mathrm{e}^{-kz} \quad \text{and} \quad \tau_0 = |\tilde{\tau}_{xz}| = p_0 kz\,\mathrm{e}^{-kz}\,. \tag{39.35}$$

The residual pore pressure ($\bar{p}$) in a homogenous, isotropic soil can be derived from the one-dimensional Biot's consolidation equation [39.82]

$$\frac{\partial \bar{p}}{\partial t} = c_\mathrm{v}\frac{\partial^2 \bar{p}}{\partial z^2} + f\,, \tag{39.36}$$

where f is the mean accumulation pore pressure source term associated with the surface water waves [39.84]. A detailed derivation of (39.36) can be found in [39.82]. In (39.36), c_v is the coefficient of consolidation, given by

$$c_\mathrm{v} = \frac{2GK_z(1-\mu)}{\gamma_\mathrm{w}(1-2\mu)}\,. \tag{39.37}$$

Regarding the source term, the laboratory results of *de Alba* [39.85] suggested a nonlinear relationship between the development of pore water pressure and the number of load cycles as

$$\frac{u_\mathrm{g}}{\sigma_0'} = \frac{1}{2} + \frac{1}{\pi}\sin^{-1}\left[2\left(\frac{N}{N_\ell}\right)^{1/\theta} - 1\right], \tag{39.38}$$

where u_g is the pore pressure generation due to cyclic loading, σ_0' is the effect over burden, N is the number of cyclic loading, N_ℓ is the number of cycles to liquefaction, and θ is the shape factor suggested to be 0.7 [39.81].

To simplify the problem, a linear mechanism of pore pressure generation was proposed [39.81, 82]

$$\frac{u_\mathrm{g}}{\sigma_0'} = \frac{N}{N_\ell}\,, \tag{39.39}$$

where N is the number of wave cycle and N_ℓ is the number of cycles to liquefaction, which is a function of the cyclic shear stress ratio [39.81, 82],

$$N_\ell = \left(\frac{\tau_0}{\alpha_\mathrm{r}\sigma_0'}\right)^{-1/\beta_\mathrm{r}}\,, \tag{39.40}$$

where τ_0 is the maximum amplitude of wave-induced shear stress, and α_r and β_r are the functions of the soil type and relative density, respectively. The residual parameter (α_r and β_r) are related to the relative density of soil (D_r) [39.86]

$$\alpha_\mathrm{r} = 0.34D_\mathrm{r} + 0.084 \quad \text{and} \quad \beta_\mathrm{r} = 0.37D_\mathrm{r} - 0.46\,. \tag{39.41}$$

From (39.39), the source term of pore pressure generation can be expressed as

$$f = \frac{\partial u_\mathrm{g}}{\partial t} = \frac{\partial}{\partial t}\left(\sigma_0'\frac{N}{N_\ell}\right) = \frac{\partial}{\partial t}\left(\sigma_0'\frac{t/T}{N_\ell}\right) = \frac{\sigma_0'}{TN_\ell} = \frac{\sigma_0'}{T}\left(\frac{\tau_0}{\alpha_\mathrm{r}\sigma_0'}\right)^{-1/\beta_\mathrm{r}}\,, \tag{39.42}$$

where T is the wave period.

It is noted that the linear mechanism of pore pressure generation was first applied to the wave-induced pore pressure build-up in marine sediment by [39.81]. Since then, this relation has been commonly used in various approaches [39.82, 84, 87].

The shear stress in the source term depends on both wave and soil characteristics. Since the expression for the shear stress in a seabed of finite thickness is complicated, two simplified cases, shallow soil depth ($d/L \le 0.1$) and deep soil layer ($d/L \ge 0.3$) in addition to the finite soil model. We only present the analytical solutions for an infinite seabed, as detailed derivations are available in [39.75, Chap. 9].

For soil depth in the range $h/L \ge 0.3$, we consider the deep (infinite) depth approximation, the source term for deep soil can be expressed as

$$f = Az\,\mathrm{e}^{-\lambda z}\,, \tag{39.43}$$

$$\lambda = \frac{k}{\beta_\mathrm{r}} \quad \text{and}$$

$$A = \frac{\gamma'(1+2K_\mathrm{o})}{3T}\left(\frac{3p_0 k}{\alpha_\mathrm{r}(1+2K_\mathrm{o})\gamma'}\right)^{1/\beta_\mathrm{r}}\,, \tag{39.44}$$

where $\gamma' = \gamma_\mathrm{s} - \gamma_\mathrm{w}$ is the submerged weight of soil.

Then, the residual pore pressure is calculated using a Laplace transformation as

$$\bar{p} = \frac{2A}{c_\mathrm{v}\lambda^3}\left[1 - \left(\frac{\lambda z}{2} + 1\right)\mathrm{e}^{-\lambda z} - \frac{1}{\pi}\int_0^\infty \frac{\mathrm{e}^{-rc_\mathrm{v}\lambda^2 t}}{r(1+r)^2}\sin\left(\sqrt{r}\lambda z\right)\mathrm{d}r\right]. \tag{39.45}$$

The solution (39.45) is similar to the one proposed by *McDougal* et al. with corrections pointed out previously [39.84]. It is noted that the shear stress used in

the [39.84] solution was based on the assumption of an incompressible soil.

Sumer and *Fredsøe* also obtained an analytical solution, but in a different form [39.82]

$$\bar{p} = \frac{2}{\pi}\int_0^t\int_0^\infty\int_0^\infty \mathrm{e}^{-c_\mathrm{v}\xi^2(t-t')} \times \sin(\xi z) f(z') \sin(\xi z') \mathrm{d}z' \mathrm{d}\xi \mathrm{d}t' \,, \tag{39.46}$$

which is in their notation [39.82]. It is noted that both (39.45) and (39.46) are identical, as shown in [39.88].

For engineering practice, the most important task is to examine where liquefaction will occur and how deep it is. The criterion of residual liquefaction is

$$\frac{\bar{p}_{t\to\infty}}{\sigma_0'} = 1 \,, \tag{39.47}$$

which leads to

$$\bar{p}_{t\to\infty} = \frac{2A}{c_\mathrm{v}\lambda^3}\left[1-\left(\frac{\lambda z_\mathrm{L}}{2}+1\right)\exp(-\lambda z_\mathrm{L})\right] = \sigma_0' = \frac{(1+2K_0)}{3}\gamma' z_\mathrm{L} \,, \tag{39.48}$$

where z_L is the liquefaction depth.

Let

$$B = \frac{(1+2K_0)\gamma' c_\mathrm{v}\lambda^2}{6A} \,. \tag{39.49}$$

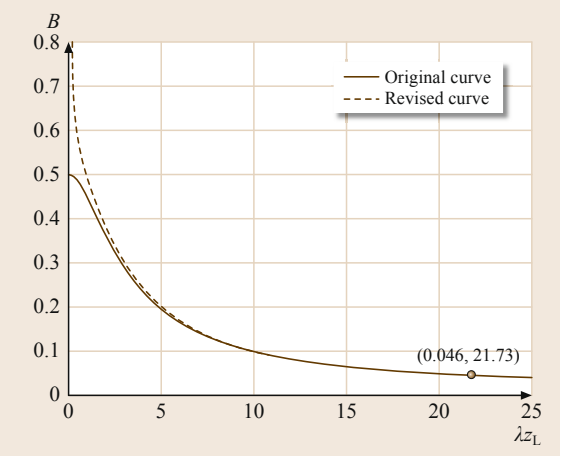


Fig. 39.5 Distribution of the maximum liquefied depth (z_L) and parameter B

Then the relationship of the maximum liquefied depth (z_L) and the parameter B can be plotted (see the original line in Fig. 39.5, called as the *J–S* (Jeng–Seymour) curve [39.88]). For engineering applications, given wave and soil conditions, we can determine the parameter B from (39.49). The maximum liquefied depth can then be easily determined from Fig. 39.5. It is noted that the relation presented in Fig. 39.5 is universal, applicable to all engineering conditions.

Herein, we further consider a special case assuming λz_L is small. With this assumption, we further expend $\exp(-\lambda z_\mathrm{L})$ with Taylor's expansion, and then using the first three terms, we have the solution of (39.48) as $z = 0$ or

$$z_\mathrm{L} = \frac{2}{\lambda}\sqrt{\frac{1}{2}-B} \,. \tag{39.50}$$

Note that (39.50) is only valid under the condition of $B \le 1/2$. It is noted that $B > 1/2$ will provide an unrealistic solution. Under such a condition, $z = 0$ will be the solution (i. e., no liquefaction occurs).

Recently, *Geremew* proposed a simplified model to directly adding oscillatory and residual components to examine the wave-induced liquefaction [39.89]. That is,

$$\frac{2A}{c_\mathrm{v}\lambda^3}\left[1-\left(\frac{\lambda z_\mathrm{L}}{2}+1\right)\mathrm{e}^{-\lambda z_\mathrm{L}}\right] + p_0\,\mathrm{e}^{-kz_\mathrm{L}} = \frac{(1+2K_0)}{3}\gamma' z_\mathrm{L} \,. \tag{39.51}$$

Unfortunately, numerous mistakes were found in their work, for example, the determination of the wavelength, and calculation of the liquefaction depth [39.90]. Based on (39.51), the revised *J–S* curve is illustrated in Fig. 39.5 together with the case presented in [39.89] ($B = 0.046$, $\lambda z_\mathrm{L} = 21.73$). Furthermore, based on the input data used in [39.89], the oscillatory component is only 1/70 of the residual component [39.90].

39.2.3 Integrated Model

In all aforementioned studies for wave-induced residual pore pressures in marine sediments, the source term was considered as the time-independent function because the maximum amplitude of the oscillatory shear stress was used. In fact, this source term was determined by the oscillatory shear stress which should be a time-dependent function. Furthermore, these approaches were one-dimensional (1-D) model, which may be insufficient to represent the real process. In this section, we re-define the source term as a time-dependent function with the instant oscillatory shear stress and re-derive the boundary value problem in 2-D.

New Model

In this study, the previous 1-D model can be re-derived to 2-D as

$$c_v \nabla^2 \bar{p} = \frac{\gamma_w c_v (1 + n\beta)}{K} \frac{\partial \bar{p}}{\partial t} + f(x, z, t) \,. \quad (39.52)$$

$$G\nabla^2 \bar{u} + \frac{G}{1-2\mu} \frac{\partial}{\partial x} \left(\frac{\partial \bar{u}}{\partial x} + \frac{\partial \bar{w}}{\partial z} \right) = \frac{\partial \bar{p}}{\partial x} \,, \quad (39.53)$$

$$G\nabla^2 \bar{w} + \frac{G}{1-2\mu} \frac{\partial}{\partial z} \left(\frac{\partial \bar{u}}{\partial x} + \frac{\partial \bar{w}}{\partial z} \right) = \frac{\partial \bar{p}}{\partial z} \,, \quad (39.54)$$

where the source term ($f(x, z, t)$) for the new model is defined as

$$f(t) = \frac{\partial u_g}{\partial t} = \frac{\sigma_0'}{T} \left(\frac{|\tilde{\tau}_{ins}(x, z, t)|}{\alpha_r \sigma_0'} \right)^{-1/\beta_r} \,. \quad (39.55)$$

It is noted that the governing equation used in the new model (39.52)–(39.54) integrated the mass conservation with momentum conservation and solved the wave-induced residual pore pressure and soil displacement together, which are different from the previous 1-D model [39.75, 82] in which only the pore pressure is solved and no other soil response information is mentioned.

In addition, the source term used in the previous models [39.75, 82] was generated by the maximum amplitude of transient shear stresses over a wave period, which is a simplified calculation method. While the instant absolute oscillatory shear stress is considered as the source of pore pressure generation in this new model, this new model will reflect the effects of instant oscillatory shear stresses on the pore pressure build-up, which will become a time-dependent function, as shown in (39.55).

To solve the two-dimensional governing equations (39.52)–(39.54) with the new source term, appropriate boundary conditions are required:

a) At the seabed surface ($z = 0$), the residual pore pressure is assumed as zero ($\bar{p} = 0$).
b) At the bottom of the porous seabed with finite thickness ($z = -h$), $\bar{u} = \bar{w} = \frac{\partial \bar{p}}{\partial z} = 0$.
c) Periodical lateral boundary condition is considered in the x direction.

Validations

The major differences between the present and previous models are the definition and numerical simulation method for the source term in (39.42) in which it is only a function of z and being time-independent in the maximum model, while it is a time-dependent function in the present instant model, (39.55).

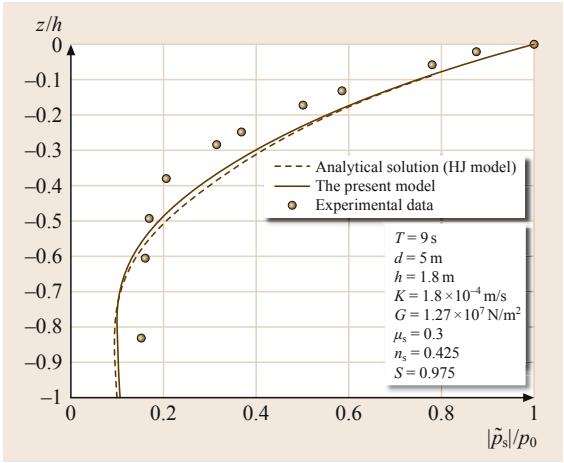


Fig. 39.6 Distribution of vertical distribution of maximum wave-induced oscillatory pore pressure versus relative soil depth for comparison between the present 2-D model (*line*) with 1-D experimental results (*points*) and Jeng's analytical solutions (*dashed line*)

To validate the present model, we compare the present numerical model with the previous experimental data for both oscillatory and residual mechanism, respectively. For the oscillatory mechanism, the numerical results for the maximum vertical oscillatory pore pressure ($\tilde{p}/p_0$) versus relative soil depth (z/h) are presented in Fig. 39.6. In the figure, both analytical solution [39.45] and experimental data [39.91] are included. As shown in the figure, the numerical results match well with both the analytical solution and the experimental data.

The previous experimental data reported in [39.86] are used to validate the residual mechanism of the present numerical model. The numerical results are observed in Fig. 39.7. In the figure, 1-D analytical solutions developed by [39.82, 92] are also included in the comparison. As shown in the figure, the overall trend of the residual pore pressure of the present 2-D instant model can capture the experimental data extremely well, and the value of the build-up pore pressure agrees quite well with the experimental data plotted in the figure, which provides much better prediction than the 2-D maximum model. There is a little difference of the trend of the build up pore pressure between the present instant 2-D model with the 1-D analytical model [39.82, 92], but the final residual pore pressure is acceptable. This comparison validates our new model and demonstrates the significant improvement of the prediction of the wave-induced residual pore pressures in marine sediment with the new definition of the source term in our new 2-D model.

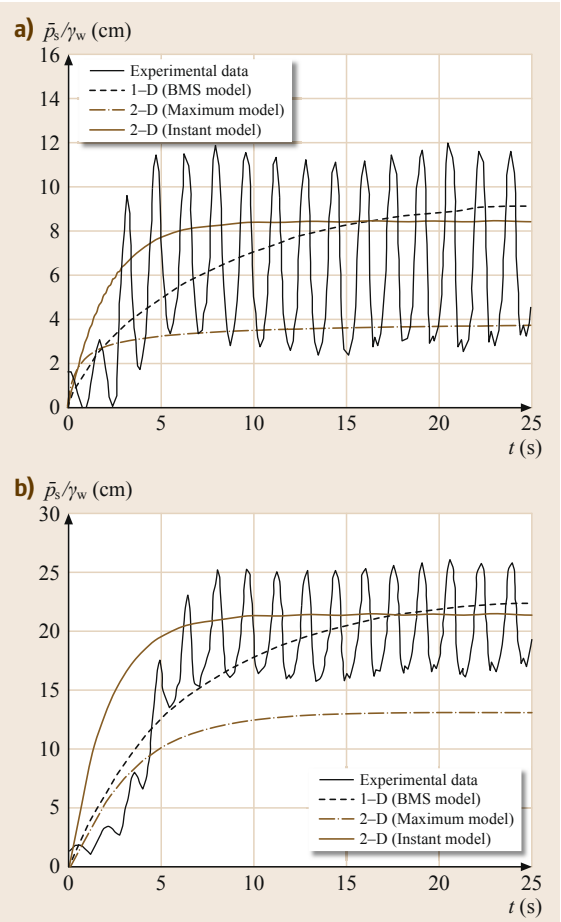


Fig. 39.7a,b Comparison of model results and experimental data (after [39.86]). **(a)** pore pressure at $(x, z) = (0, -0.085)$ m **(b)** pore pressure at $(x, z) = (0, -0.24)$ m. Input data: $H = 0.18$ m, $T = 1.6$ s, $d = 0.55$ m, $K = 1.5\times10^{-5}$ m/s, $n = 0.51$, $G = 1.92\times10^{6}$ N/m^2, $h = 0.4$ m, $\mu = 0.29$, $c_v = 0.0127$ m^2/s, $D_r = 0.28$

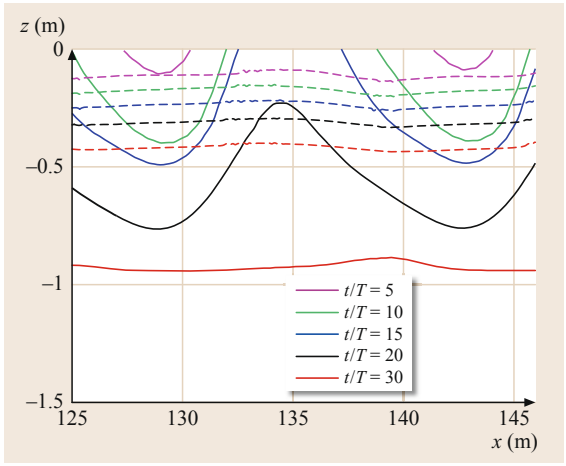


Fig. 39.8 Development of liquefaction zone from 2-D to 1-D for various types of soil with the instant model (*solid lines*) and the maximum model (*dashed lines*). Input data: $H = 2.8$ m, $T = 5$ s, $d = 4$ m, $K = 10^{-4}$ m/s, $n = 0.3$, $G = 3\times10^{6}$ N/m^2, $h = 20$ m, $\mu = 0.35$, $c_v = 0.1325$ m^2/s, $D_r = 0.54$, $\alpha_r = 0.246$ and $\beta_r = 0.165$

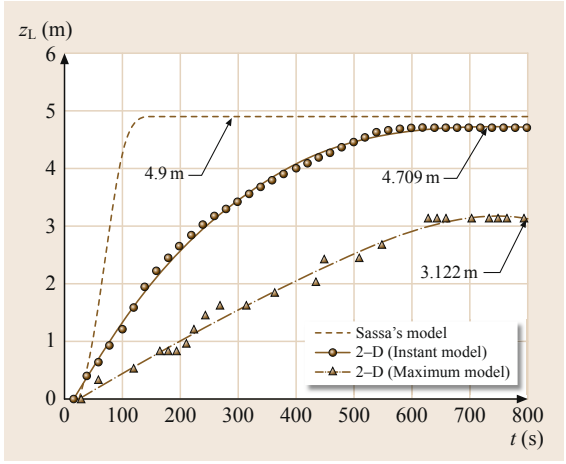


Fig. 39.9 Comparison of progress of wave-induced liquefied zone versus wave cycle (t/T) between Sassa's model (*dashed line*) and the present instant model (*line*) and maximum model (*dash-dotted line*). Input data: $H = 6.5$ m, $T = 10$ s, $d = 20$ m, $K = 1.5\times10^{-4}$ m/s, $n = 0.3$, $G = 2\times10^{5}$ N/m^2, $h = 20$ m, $\mu = 0.35$, $c_v = 0.0133$ m^2/s, $D_r = 0.4$

39.2.4 Wave-Induced Residual Liquefaction

The source term for the residual pore pressure generation is re-defined as two-dimensional and time-dependent function in our new model. This new feature will directly affect the pattern of the liquefaction zone. It is well known that liquefaction will occur when the excess pore pressure reaches the initial effective stress, $\bar{p}_s = \sigma_0'$. Figure 39.8 illustrated the variations of wave-induced residual liquefaction zones against the wave cycle (t/T) for various types of soil with different numerical model. As shown in the figure, for the 2-D instant model, the liquefaction occurs as a 2-D zone first and gradually moves down and behaves as a 2-D zone, after a certain wave cycles (e.g., 30 wave cycles in this example), the liquefaction zone changes from 2-D to 1-D. On the other hand, in the 2-D maximum model, the liquefaction zone always behaves as a 1-D zone and gradually moves down.

Sassa et al. proposed a 1-D elasto-plastic model, in which the source term was related to the rate of plastic volumetric deformation $\partial\epsilon^p/\partial t$, but the stress

ratio τ/σ'_{z0} was still estimated by the poro-elastic solutions [39.93]. *Liu* et al. extended Sassa et al.'s model by considering the effect of viscosity on the liquefied soil [39.94]. The comparison of the distribution of maximum liquefaction depth (z_L) versus time (t) between *Sassa*'s model [39.93] and the present instant and maximum 2-D model is presented in Fig. 39.9. It is clear from the results presented in Fig. 39.9 that the maximum liquefied depth will gradually move down and become a constant when the time will reach to a certain value, and the final liquefaction depth for the present instant model can match well with the results presented in *Sassa*'s elasto-plastic model [39.93], (the difference is only 3.9%) and provide a much better prediction than the 2-D maximum model. This preliminary results demonstrate the capacity of the present 2-D instant model in the prediction of progressive liquefaction.

39.3 Wave(Current)-Induced Soil Response in Marine Sediments

The co-existence of waves and currents in offshore area is a common physical phenomenon and their interaction is an important topic in the practices of coastal and ocean engineering. The presence of a current in a propagating wave will change the original characteristics of wave. For example, the following current will elongate the wave length, and the opposing current will shorten the wave length. In this study, to obtain more accurate results of seabed response under combined wave and current loadings, the third-order solution of wave–current interactions is used to determine the dynamic wave pressures acting on the seabed. On the problem of the third-order wave–current interaction, some works are available in the previous literatures. The interaction between the linear wave and uniform current has been studied by [39.95]. *Baddour* and *Song* investigated the interaction of linear wave and collinear current [39.96]. Later, they further extended this problem to third-order nonlinear wave and an uniform current [39.97]. However, this analytical solution for the interaction between the third-order nonlinear wave and uniform current is not right because the third-order terms are not completed. Recently, *Hsu* et al. gave out the complete form of analytical solution for the problem of the third-order nonlinear wave and uniform current interaction [39.98]. Later, *Jian* et al. develop an analytical solution for the interaction between gravity capillary short crested wave and a uniform current [39.99]. In this study, the analytical solution proposed by [39.98] is adopted to apply the wave pressure on porous seabed without marine structure.

Recently, a series of analytical solutions for the wave(current)-induced soil response in marine sediments were proposed based on various assumptions, for example, consolidation model [39.100], *u-p* approximation [39.101] and full dynamic model [39.102]. In addition to analytical solution, numerical models for the wave(current)-induced soil response were also proposed, based on consolidation model [39.103] and *u-p* approximation [39.104]. All these investigations demonstrated the significant effects of currents on the soil response in marine sediments.

39.3.1 Theoretical Formulations

Using the perturbation technique, *Hsu* et al. derived a third-order approximation of the dynamic pressure acting on the seabed which can be expressed as [39.98]

$$
\begin{aligned}
P_b(x,t) = {} & \frac{\rho_f g H}{2\cosh kd}\left(1-\frac{\omega_2 k^2 H^2}{2\,(U_0 k-\omega_0)}\right)\\
& \times\cos(kx-\omega t)\\
& +\frac{3\rho_f H^2}{8}\left(\frac{\omega_0(\omega_0-U_0k)}{2\sinh^4(kd)}-\frac{gk}{3\sinh 2kd}\right)\\
& \times\cos 2(kx-\omega t)\\
& +\frac{3\rho_f k H^3\omega_0(\omega_0-U_0k)}{512}\,\frac{(9-4\sinh^2(kd))}{\sinh^7 kd}\\
& \times\cos 3(kx-\omega t)\,,
\end{aligned}
\tag{39.56}
$$

where ρ_f is the density of sea water. When there is no current in wave ($U_0 = 0\,\mathrm{m/s}$), H is the wave height of the first-order wave, U_0 is the current velocity, g is the gravity, and the dispersion relation is given by

$$
\omega = \omega_0 + (kH)^2\omega_2\,, \tag{39.57}
$$

where $\omega_0 = U_0 k + \sqrt{gk\tanh kd}$ and

$$
\omega_2 = \frac{(9+8\sinh^2 kd+8\sinh^4 kd)}{64\sinh^4 kd}(\omega_0-U_0k)\,. \tag{39.58}
$$

In this study, the finite element model (SWANDYNE II), originally developed by [39.67] for the soil response under earthquake loading, is adopted for wave loading. The wave module (COBRAS, Cornell breaking wave and structure) to apply the wave loading to porous seabed was developed and integrated to

SWANDYNE II to form part of the numerical model PORO-WSSI II (POROus models for wave–seabed–structure interactions, version II). Details of the generalized FEM (finite element method) formulations are available in [39.11].

39.3.2 Effects of Currents

In this subsection, we compare the seabed response for the cases under waves loading with and without currents. Figures 39.10 a and 39.11a illustrate the vertical distributions of the seabed response under wave and following current ($U_0 = 1\,\mathrm{m/s}$) loading at $x = 125\,\mathrm{m}$ in both coarse and fine sand. Due to the fact that the absolute value of maximum dynamic response in seabed is greatly different under highly nonlinear wave loading, the maximum dynamic responses in seabed are compared, respectively, with their corresponding values when $U_0 = 0\,\mathrm{m/s}$. In the figures for maximum response, all seabed response variables are normalized by the maximum dynamic wave pressure along the seabed surface without current, that is, $(p_0)_{\max}$, given in (39.56) when $U_0 = 0\,\mathrm{m/s}$. It is clearly observed from Figs. 39.10 and 39.11 that the effect of ocean currents on the seabed response is significant in both coarse and fine sand. If a following current exists in the wave field, the magnitudes of the maximum seabed response, including pore pressure and effective stresses, are basically greater than that without currents. For the maximum shear stress in seabed, both cases are almost identical in the upper part of seabed. In the lower part of seabed, the magnitude of shear stress is greater when there is a following current. In coarse sand, the maximum relative difference between

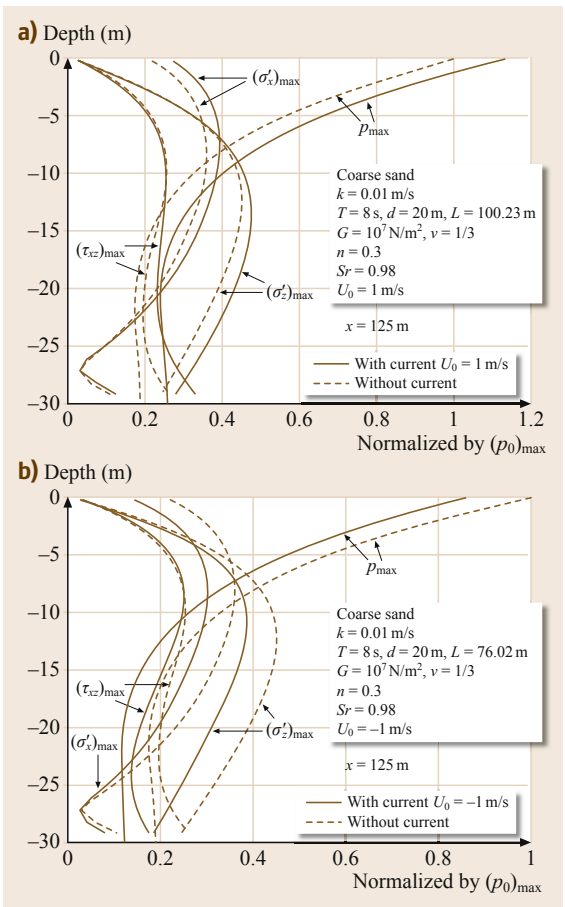


Fig. 39.10a,b Vertical distributions of seabed response under wave and current loading in coarse sand ($K = 10^{-2}\,\mathrm{m/s}$, $n = 0.3$). Input data: $H = 2\,\mathrm{m}$, $T = 8\,\mathrm{s}$, $d = 20\,\mathrm{m}$, $G = 10^7\,\mathrm{N/m^2}$, $h = 30\,\mathrm{m}$, $\mu = 1/3$, $S = 0.98$ **(a)** $U_0 = 1\,m/s$; **(b)** $U_0 = -1\,m/s$

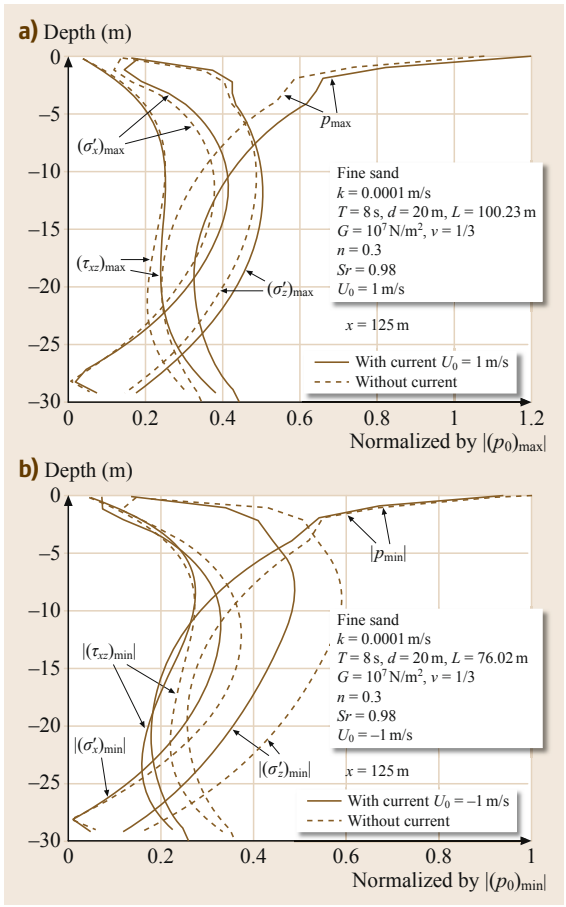


Fig. 39.11a,b Vertical distributions of seabed response under wave and current loading in fine sand ($K = 10 \times 10^{-4}\,\mathrm{m/s}$, $n = 0.2$). Input data: $H = 2\,\mathrm{m}$, $T = 8\,\mathrm{s}$, $d = 20\,\mathrm{m}$, $G = 10 \times 10^7\,\mathrm{N/m^2}$, $h = 30\,\mathrm{m}$, $\mu = 1/3$, $S = 0.98$ **(a)** $U_0 = 1\,m/s$; **(b)** $U_0 = -1\,m/s$

the two cases with/without currents can up to 15% for the maximum pore pressure, and 10% for $(\sigma_z')_{max}$, 5% for $(\sigma_x')_{max}$ and 10% for the shear stress $(\tau_{xz})_{max}$. It is noted that although the magnitude of the relative difference of seabed response between the two cases is not large, the absolute difference is huge because all quantities have been normalized by a great value $(p_0)_{max}$. All these results indicate that the seabed instability (such as liquefaction) is more likely to occur (will show in the latter section) if the ocean wave and following current co-exist simultaneously regardless of soil type.

Figures 39.10b and 39.11b further present the vertical distributions of the maximum seabed response under the nonlinear wave and opposing current loading at $x = 125$ m in coarse sand and fine sand when the current velocity $U_0 = -1$ m/s. It is also clearly observed that the effect of opposing current on the seabed response is also significant, as shown in Figs. 39.10b and 39.11b. However, the seabed response will be smaller than the case without current, which may reduce the potential of seabed instability.

39.3.3 Liquefaction of Seabed Under Combined Nonlinear Wave and Current Loading

It is well known that the porous seabed would liquefy under wave loading due to the build-up of excess pore pressure in seabed. In this study, to investigate the liquefaction properties in seabed under combined nonlinear wave and current loading, the liquefaction criterion proposed by [39.28] is adopted. It is expressed as

$$-(\gamma_s - \gamma_w)z \le \sigma_z' \,, \tag{39.59}$$

where the γ_s is the saturation unit weight of seabed soil, γ_w is the unit weight of water, and σ_z' is the wave-induced vertical dynamic effective stress. Actually, the liquefaction criterion, (39.59) means that the seabed will liquefy if the wave-induced vertical dynamic effective stress σ_z' (Noted: compressive stress is negative) is equal to or greater than original vertical effective stress $-(\gamma_s - \gamma_w)z$. In this study, the dynamic effective stresses are determined through the following three steps:

1. Calculating the consolidation state of seabed under the static water pressure.
2. Calculating the full effective stresses state of seabed under full water pressure including static pressure and the wave-induced dynamic pressure.
3. The dynamic effective stresses is determined by subtracting the effective stresses of consolidation state from the full effective stresses.

The transient liquefied zones move in seabed accompanying the movement of the third-order progressive wave. Therefore, there is no place which is always in the liquefied state if the elastic model is used for porous seabed. Figure 39.12 illustrates the variation process of liquefied depth in fine sand seabed in time domain at $x = 125$ m under nonlinear wave and current ($U_0 = -1$ m/s, 0 m/s and 1 m/s) loading. As shown in Fig. 39.12, the maximum liquefied depth in seabed is 1.16 m, 0.98 m, and 0.58 m when the velocity of current U_0 is 1 m/s, 0 m/s, and −1 m/s. Relative to the condition without current, the following current $U_0 = 1$ m/s makes the maximum liquefied depth increase 18%; while the opposing current makes the maximum liquefied depth decrease 57%. From this result, It is found that the following current makes the liquefied zone in seabed becomes larger than that when there is no current; and the opposing current is beneficial to prevent

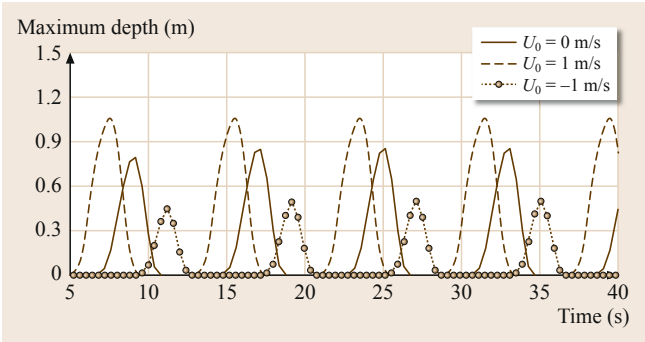


Fig. 39.12 The variation process of the liquefied depth in seabed (fine sand) at $x = 125$ m under nonlinear wave and different current loading. ($H = 3.0$ m, $d = 10$ m, $T = 8.0$ s)

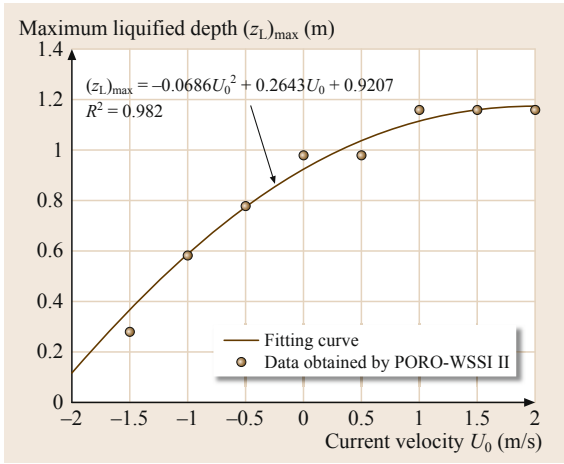


Fig. 39.13 The wave-current-induced maximum liquefied depth in seabed (fine sand) at $x = 125$ m versus different current velocity U_0, and the fitting curve ($H = 3.0$ m, $d = 10$ m, $T = 8.0$ s)

the seabed from liquefying. Figure 39.13 shows the relationship between the maximum liquefied depth and the current velocity. As illustrated in Fig. 39.13, the following current make the fine sand seabed more easy to get liquefied, the opposing current makes the fine sand seabed more difficult to be liquefied.

39.4 Seabed Stability Around Caisson Breakwaters

The phenomenon of wave–seabed–structure interactions (WSSI) has a major bearing on this issue and is central to the design of coastal structures such as breakwaters, pipelines and platforms. Numerous studies of wave-induced seabed response have been conducted since the 1970s, involving the investigations of pore pressures, effective stresses, and displacements [39.22, 26, 43, 45]. Most of them have based on *Biot*'s poro-elastic theory [39.3]. Later, the analytical solutions of dynamic response for a poro-elastic, isotropic seabed under linear wave loading have been developed, in which the inertia effects of the solid and/or pore fluid are included [39.5, 6, 8]. Recently, *Ulker* et al. further examined the applicable ranges of different approximations [39.9]. All these investigations have limited to the dynamic response of seabed under linear/nonlinear wave loading, without a marine structure.

There have been numerous investigations for the interactions of wave–seabed around marine structures, based on Biot's poro-elastic theory. Among these, *Mase* et al. developed a FEM numerical model to investigate the wave-induced pore water pressures and effective stresses under standing waves in a sand seabed and in the rubble mound foundation of a composite caisson-type breakwater based on Biot's consolidation equations [39.105]. Later, *Mizutani* et al. [39.106] and *Mostafa* et al. [39.107] developed a BEM (boundary element method)-FEM (finite element method) combination numerical model to investigate the wave–seabed–structure interaction. In their models, Poisson's equation is used to govern the irrotational wave field for an incompressible, inviscous fluid; and the poro-elastic Biot's consolidation equations are used to govern the porous seabed and structures. *Jeng* et al. developed a 2-D generalized FEM numerical model (GFEM-WSSI) to investigate the wave-induced pore pressure under a linear wave around a composite breakwater located at a finite, isotropic, and homogeneous seabed [39.108]. *Ulker* et al. investigated the dynamic response and instability of the seabed around a caisson breakwater under standing waves with a FEM numerical model [39.109], which is developed by considering the acceleration of relative displacements between solid and pore fluid particles. However, all these models have been based on potential flow theory for wave models.

In addition to the potential flow, models combining the Reynolds-averaged Navier–Stokes equations (RANS) for wave field and the volume-averaged Reynolds-averaged Navier–Stokes (VARANS) equations for the porous flow in porous medium have been developed to examine the phenomenon of wave propagation around marine structures [39.110–113]. The major advantage of using RANS and VARANS instead of potential flow for wave modeling is the capacity of capturing wave breaking, turbulence and detailed information of bottom boundary layers. In these models, the pressure, flow field in whole computational domain, and the flux at the interface between the porous seabed/marine structures and the seawater, are all continuous. However, the variations of effective stresses in the seabed and marine structures cannot be determined.

In this section, to improve the capability of modeling the mechanism of the WSSI problem, an integrated model (PORO-WSSI II) will be proposed. The VARANS equations are used for the description of wave motion and the porous flow in the seabed and marine structures. The dynamic Biot's theory is then used to describe the mechanical behavior of porous seabed. Furthermore, a one-way integrating algorithm, based on the radial point interpolation method, is adopted to link two models through the continuity of the pressure and velocity/flux at the interface between the seabed/marine structures and sea water.

39.4.1 Theoretical Models

The proposed integrated numerical model consists of two submodels: wave and seabed models. The wave model is used for generating waves and describing their propagation in a viscous fluid. The flow field inside and outside of porous media is determined by solving the VARANS equations [39.114], which are derived by integrating the RANS equations over the control volume. The influence of turbulence fluctuations on the mean flow is obtained by solving the volume-averaged $k-\epsilon$ turbulence model. In the VARANS equations, the interfacial forces between the fluid and solids have been modeled according to the extended Forchheimer relationship, in which both linear and nonlinear drag forces between the pore water and the skeleton of porous structures are included. More detailed information of the RANS and VARANS models are available in [39.114] and [39.115].

The seabed model, based on the dynamic Biot's equations (so-called *u-p* approximation) proposed by [39.4], is used to determine the seabed responses to the waves, including the pore pressure, soil displacements, and effective stresses. Unlike previous quasi-static soil behavior [39.3], the accelerations of the pore water and soil particles are considered in the present seabed model, but the relative displacements of pore fluid to soil particles are ignored. The finite element model (SWANDYNE II), originally developed by [39.67] for the soil dynamic response under earthquake loading, is adopted for the seabed response under wave loading. Details of the generalized FEM formulations are available in [39.11, 116].

A one-way integrating algorithm is adopted to integrate both models together. Unlike most previous investigations in which Poisson's equations or Laplace's equations, and the quasi-static Biot's consolidation equations are used, in this study, the VARANS equations and the dynamic Biot's equations are used to govern the wave motions and porous flows in the porous seabed and marine structures, and the dynamic mechanical behaviors of the seabed and marine structures. Due to the fact that the flow fields outside and inside the porous medium are coupled in the wave model, both the pressure and the flow velocity are continuous in the entire computational domain, especially at the interfaces between the seabed, marine structures, and the sea water. In the dynamic Biot's equations, the accelerations of solid and pore water are included.

In the process of integrating the VARANS equations and the dynamic Biot's equations, two types of mesh systems (matching mesh and nonmatching mesh) could be adopted in the numerical computations. In the numerical models proposed by [39.106] and [39.107], the matching mesh system is used, in which the same numbers of nodes are required along the sea floor because they have to share the same nodes along the surface of seabed, and marine structures. However, the size of elements in the fluid domain is generally much smaller than that in the solid domain. The ratio of the size of elements in the solid domain to that in fluid domain could vary between 5 and 20. Therefore, the nonmatching mesh system is used in this study. To integrate the wave and seabed models together at the interface with the nonmatching mesh system, a data exchange port between the VARANS equations and the dynamic Biot's equations is required, in which the radius point interpolation method proposed by [39.117] is adopted to implement the data exchange between the wave and soil models.

In the integrated model, to handle the interactions between wave, seabed, and marine structures, the Navier–Stokes equations for wave and porous flow, and the dynamic Biot's equations for the porous seabed have to be combined together through the continuity of pressure at the interface between the fluid domain and the porous medium. It is noted that there may be an argument of using Biot's poro-elastic model for the simulation of porous flows in rubble mound. However, due to the granular material properties, the pore pressures in the rubble mound predicted by Biot's model and VARANS model are close [39.116]. Furthermore, the main purpose of using Biot's model for rubble mound is to ensure the continuity of pressure field between the fluid domain and porous structures. Therefore, the present approach can provide a good estimation of pressure filed for engineering practice.

In the computation, the wave model is responsible for the simulation of wave propagation and the porous flow in porous structures (seabed, rubble mound, breakwater, etc.), and determines the pressure acting on the seabed and marine structures. Due to the fact that the VARANS equations are coupled at the interface between the fluid domain and the porous structures through the pressure and velocity/flux continuity, the pressure and the flow field are continuous in the whole computational domain. Meanwhile, the pressure/force acting on seabed and marine structures determined by the wave model is provided to the soil model through the data exchange port developed to calculate the dynamic response of seabed and marine structures, including the displacements, pore pressure, and the effective stresses. The proposed process is only valid to the cases with small deformation of seabed. *Liu* et al. proposed a coupled model for the interaction between wave and the pore water in seabed, in which the Navier–Stokes equation for wave motion on seabed, and Darcy's flow for the pore water in seabed are used [39.118]. The pressure and velocity continuity at their interface are implemented for the two models, namely it is a two-way coupling. However, the limitation of this method is that the effective stresses in seabed cannot be determined. It is noted here that although the VARANS equation and Biot's equation are only integrated together in this study, the flow field determined by the VARANS equation is continuous in the whole computational domain because the porous medium, such as rubble mound, seabed all are considered as porous structures in the wave model.

39.4.2 Verifications

To validate the proposed numerical model, four sets of previous experiments available in the literature are used. The experimental setup of these experiments is given in Fig. 39.14.

Lu conducted a series of laboratory experiments for the wave-induced dynamic response of sand bed in a wave flume, which is 60 m long, 1.5 m wide, and

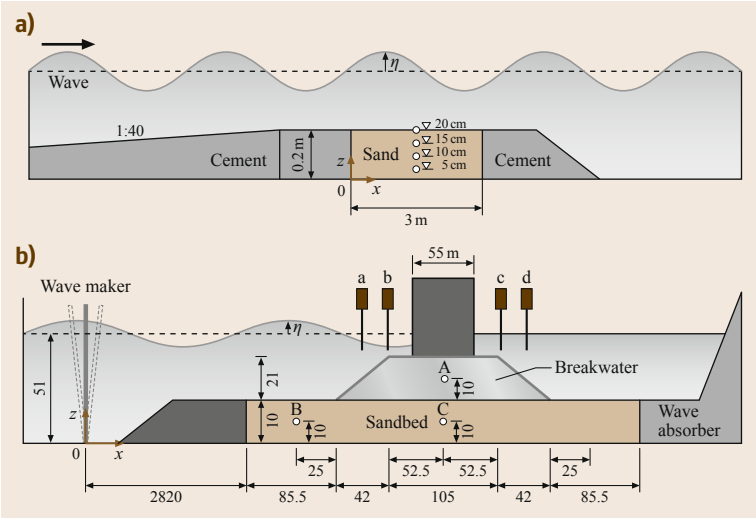


Fig. 39.14a,b Experiment setup of previous wave tank experiments for the validation of the present model: **(a)** *Lu*'s experiment – fifth-order wave and Cnoidal wave (after [39.119]), **(b)** *Mostafa* et al.'s experiment-Composite breakwater (after [39.107])

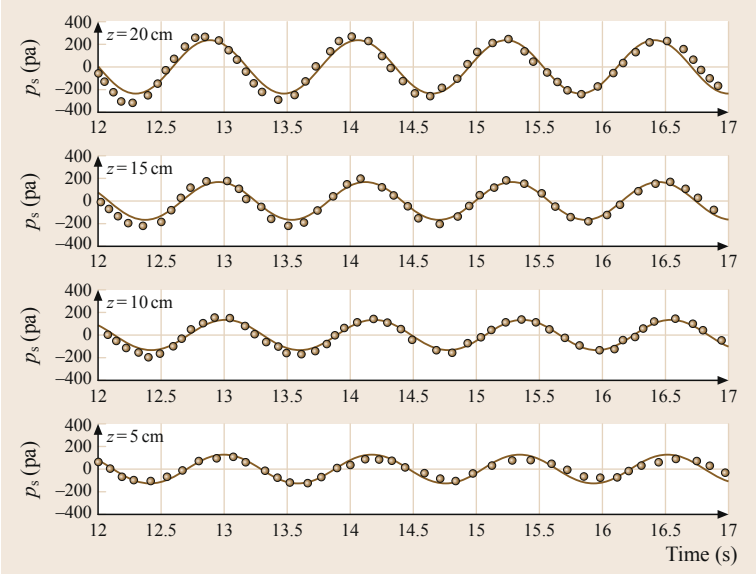


Fig. 39.15 Comparisons of wave-induced dynamic pore pressure on the midline of sand bed between the numerical results and the experimental data in Lu's experiments. —: numerical results; ○: experimental data. Input data: $H = 12\,\text{cm}$, $d = 40\,\text{cm}$, $T = 1.2\,\text{m}$, $G = 10^7\,\text{N/m}^2$, $\mu = 0.3$, $K_z = 10^{-3}\,\text{m/s}$

1.8 m high [39.119]. The waves generated in the wave flume include regular progressive waves and cnoidal waves. The periods of wave vary from 1.0 to 1.8 s and the wave heights vary from 8 to 16 cm. The experiment setup is shown in Fig. 39.14a. The pore pressure at the four points on the midline of sand bed is measured in experiments. In the present model, the fifth-order Stokes wave theory is adopted for the generation of the regular progressive wave ($H = 12\,\text{cm}$, $d = 0.4\,\text{m}$, $T = 1.2\,\text{s}$). As illustrated in Fig. 39.15, the numerical prediction of the wave-induced pore pressures overall agrees well with the experimental data [39.119].

Mostafa et al. conducted a series of experiments in the same wave flume to investigate the interaction between the wave, composite breakwater, and sand bed [39.107], as shown in Fig. 39.14b. In the experiments, a wooden box (the width is 55 cm) is placed on the breakwater to form a composite breakwater in the wave flume. Four wave gages are installed at points a, b, c, and d to monitor the wave profile; two of them are in front of the composite breakwater, while the other two are behind of the composite breakwater. Three pressure sensors are installed at points A, B, and C to record the pore pressure. Here, the wave model with wave maker of the second-order Stokes wave is adopted to simulate the generation, propagation, reflection, and interference of wave.

Table 39.1 Properties of seabed soil, rubble mound and caisson adopted in large-scale model

Medium	G [kN/m^2]	μ_s	k_s [m/s]	n	d_{50} [mm]	S_r
Seabed soil	1.0×10^5	0.33	0.0001	0.25	0.5	98%
Rubble mound	5.0×10^5	0.33	0.2	0.35	400	98%
Caisson	1.0×10^7	0.25	0.0	0.0	–	0%

The present model is adopted to simulate the interaction between the wave, composite breakwater, and sand bed. In computation, the sand bed and the rubble mound of the composite breakwater are treated as different porous structures in fluid domain; the wooden box is treated as impermeable structure in fluid domain in the wave model. The data exchange is implemented at the interface by the integrating algorithm at the interface between the solid domain (sand bed, rubble mound, and wooden box) and the fluid domain. It is noted that the buoyancy acting at the bottom of the wooden box applied by the pore water in the rubble mound has been considered in this case.

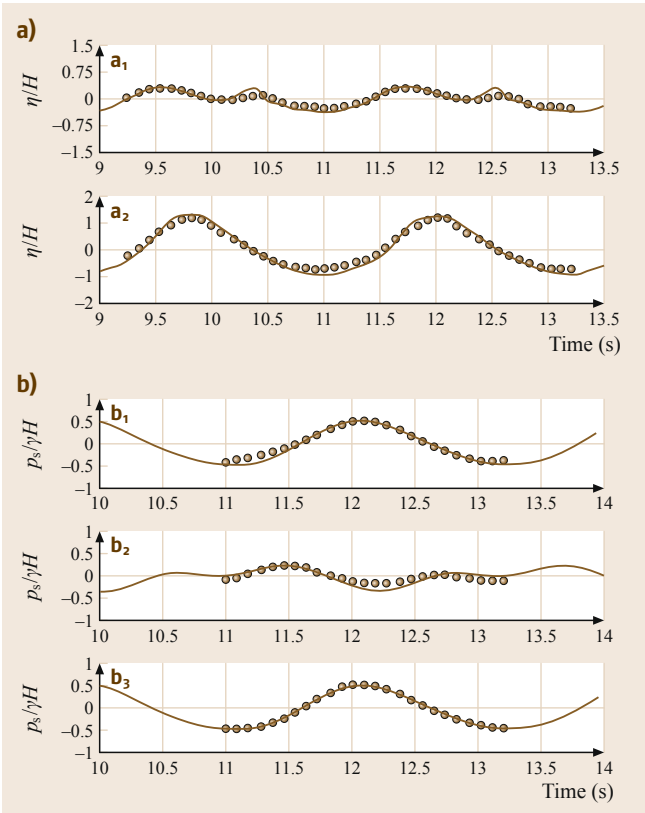


Fig. 39.16a,b Comparison of (a) wave profile and (b) pore water pressure in the seabed between the PORO-WSSI II model and the experimental data in *Mostafa* et al. (after [39.107]) for the wave profile. —: numerical results; ○: experimental data. Input data: $H = 5\,\text{cm}$, $d = 32\,\text{cm}$, $T = 2.2\,\text{m}$, $G = 5\times10^8\,\text{N/m}^2$, $\mu = 0.33$, $K_z = 2.3\times10^{-3}\,\text{m/s}$

The comparisons for the wave profile and the wave-induced dynamic pore pressure in sand bed and the rubble mound between the numerical results predicted by PORO-WSSI II and the experiment data are shown in Fig. 39.16. Due to the blocking effect of the impermeable wooden box, only little water can flow into and out the right side of the composite breakwater through the rubble mound. Therefore, the amplitude of wave behind the composite breakwater is very small. In Fig. 39.16a, only the wave profiles of points a and b are used to make the comparison between numerical results and experiment data. From the figure, it can be seen that the numerical results obtained by the present model agree well with the experiment data both for wave profile and wave-induced dynamic pore pressure.

39.4.3 Dynamic Response of Seabed Around Caisson Breakwaters

Once the initial consolidation state of seabed is determined under the static water pressure and weight of composite breakwater, it will be taken as the initial stress state when determining the dynamic response of seabed under dynamic ocean wave loading. In the model, the seabed and rubble mound are considered as porous medium, while the caisson is considered as impermeable structure in wave model. The full pressure acting on sea floor and composite breakwater is transmitted to the soil model that is used to predict the response of seabed and the composite breakwater. The predicted seabed responses induced by the full pressure are considered as full response, and the wave-induced dynamic response can be determined from the difference between full response and the initial consolidation state.

Figure 39.17 shows the distributions of dynamic effective stresses and pore pressure in the seabed and composite breakwater at $t = 73.6\,\text{s}$ under dynamic wave loading. The properties of seabed soil, rubble mound and caisson are listed in Table 39.1. According to the momentary liquefaction mechanism, the seabed soil will be liquefied when the total stresses become zero in the region, resulting in the failure of foundations in the vicinity of marine structures. On the right-hand side of composite breakwater, the effect of the ocean wave is limited in a range which does not exceed $x = 450\,\text{m}$ due to the blocking of breakwater. In the region far away from the composite breakwater, the effect of

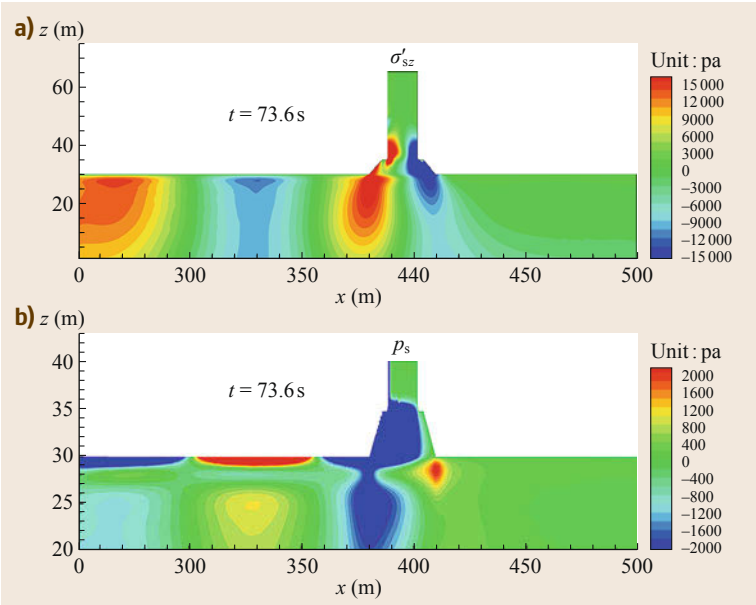


Fig. 39.17a,b The dynamic response of the seabed and composite breakwater under the ocean wave loading at $t = 73.6$ s. Wave characteristics: $T = 10$ s, $H = 3$ m, $d = 20$ m

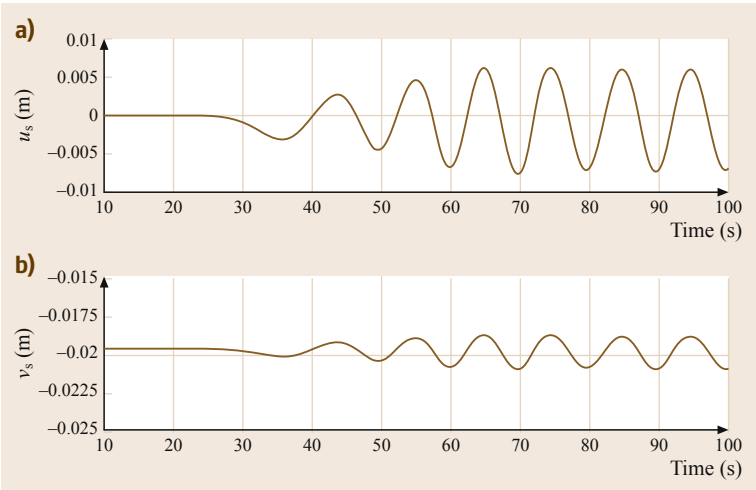


Fig. 39.18a,b The variation of horizontal and vertical displacements of the left corner of the impermeable caisson under ocean wave. Negative value of u_s means moving toward left, negative value of v_s means moving toward down **(a)** horizontal displacement; **(b)** vertical displacement

ocean wave basically disappears. At time $t = 73.6$ s, as shown in Fig. 39.17, the seabed near the rubble mound is likely to be liquefied when the wave trough is propagating on it, and the dynamic effective stresses and pore pressure are negative, which would lead to the collapse of the composite breakwater.

Before the ocean wave arrives at the structure, the composite breakwater gradually moves downward to the seabed due to its weight and reaches the initial consolidation state. Figure 39.18 shows the development of horizontal and vertical displacements of the upper left corner of the caisson. The results indicate that the structure vertically subsides about 19 mm during $0 < t < 25$ s. After the ocean wave reaches the caisson ($t > 25$ s), the caisson begins to vibrate vertically and horizontally with a small amplitude. After the full interaction of wave and structure ($t > 60$ s), the breakwater subjects to a periodical wave force and its induced vibration.

39.4.4 Wave-Induced Liquefaction Around Breakwaters

In this study, to investigate the liquefaction properties in a seabed under ocean wave loading, the liquefaction criterion proposed by [39.28] is used, which was given in (39.59). However, Eq. (39.59) is only applicable to the cases without a structure. For the cases with a ma-

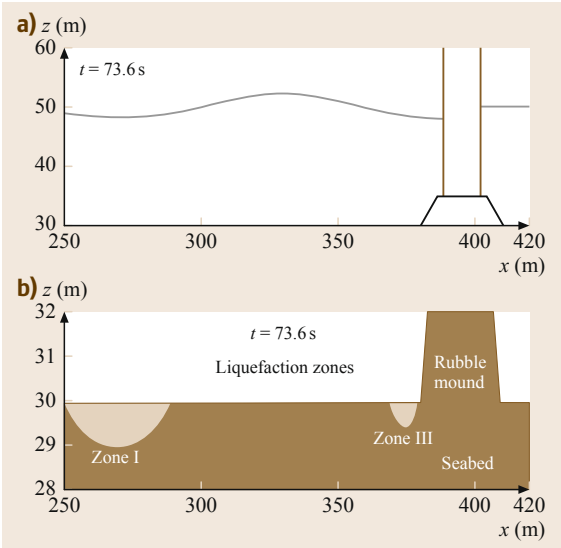


Fig. 39.19a,b The three liquefaction zones in seabed under ocean wave at time $t = 73.6$ sec **(a)** wave profile; **(b)** liquefaction zones

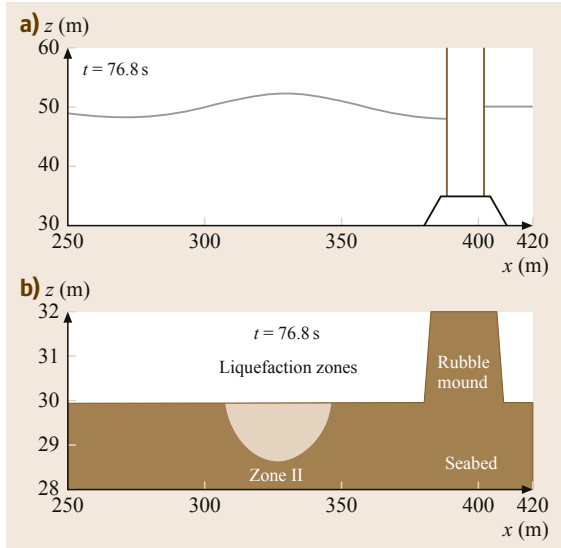


Fig. 39.20a,b The three liquefaction zones in seabed under ocean wave at time $t = 76.8$ sec **(a)** wave profile; **(b)** liquefaction zones

rine structure, it can be modified as

$$|(\sigma'_z)_{\text{initial}}| \le \sigma'_z \,, \tag{39.60}$$

where the $(\sigma'_z)_{\text{initial}}$ is the vertical effective stress at the initial consolidation state.

Figures 39.19 and 39.20 show the liquefaction zones in the seabed under the ocean wave loading at the time $t = 73.6$ s and $t = 76.8$ s, in which the modified liquefaction criterion (39.60) is adopted. As illustrated in the figures, there are two liquefaction zones in the region near the seabed surface at time $t = 73.6$ s; they are located at the range of $250 < x < 290$ m (Zone I) and $370 < x < 380$ m (Zone III), respectively. There is only one liquefaction zone in the region near the seabed surface at time $t = 76.8$ s, which is located at the range of $310 < x < 350$ m (Zone II). Zones II and III are very close to the foundation of breakwater, and they may have a large impact on the foundation stability. Therefore, we will further investigate the (depth, width, and area) development of these two liquefaction zones.

Figures 39.21 and 39.22 illustrate the variations of the liquefaction potential of Zones II and III under the ocean wave loading ($T = 10$ s, $H = 3$ m, $d = 20$ m), respectively. It can be seen from Fig. 39.21 that the liquefaction potential of Zone II is very small when the first wave trough passes through, but it increases largely during the second wave trough. After the interaction between wave and structure develops, the liquefaction depth, width, and area of Zone II further increase. The maximum liquefaction depth, width, and area are about 1.4 m, 41.0 m, and 38.5 m^2 (which occur at $t = 79$ s), respectively. As shown in Fig. 39.22, the liquefaction potential of Zone III has a similar development process as that of Zone II. The maximum liquefaction depth, width, and area of Zone III are about 0.46 m, 11.5 m and 3.85 m^2 (which occur at $t = 74$ s), respectively. The liquefaction potential of Zone III is much smaller than that of Zone II, which can be ascribed to that the weight of composite breakwater significantly increases the vertical effective stress of the initial consolidation state in Zone III.

Although the liquefaction potential in Zone III is small relative to that in Zone II, more attentions need to be paid to the Zone III as it is next to the foundation of composite breakwater. The soil liquefaction may lead to a collapse of composite breakwater. In engineering application, some methods, such as a replacement of the fine sand with gravel material, need to be adopted to protect the structure foundations.

The influence of the inertial terms associated with the accelerations of the pore water and soil particles on the wave-induced dynamic pore pressure and effective was discussed in the previous work [39.8, 9]. It is of interest to examine the effects of inertial term on the liquefaction potential. In Figures 39.21 and 39.22, the predictions of liquefaction zone by the conventional consolidation model are also included as dashed lines. As shown in the figure, the consolidation model [39.108] overestimates the liquefaction depth in Zone II, compared to the present dynamic model. However, the predictions of the consolidation model are slightly greater than that of the dynamic model.

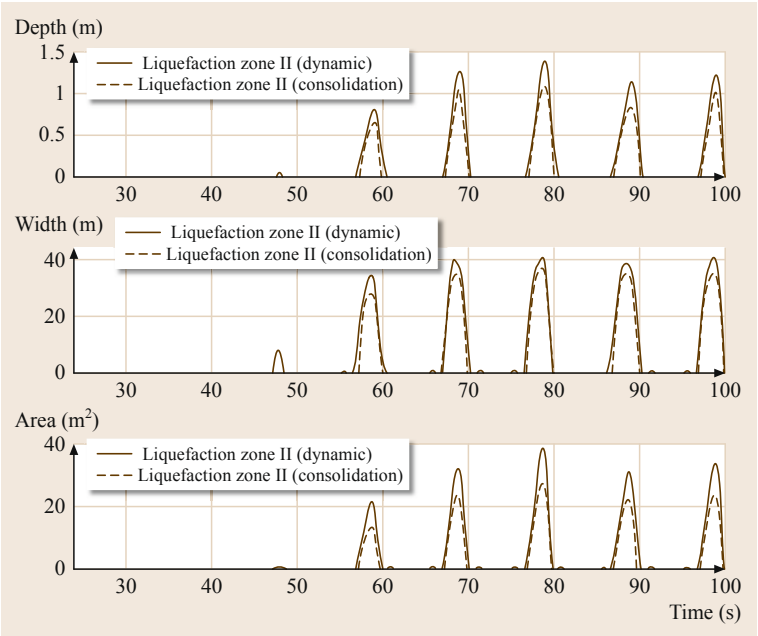


Fig. 39.21 The liquefaction properties (depth, width and area) in liquefaction Zone II

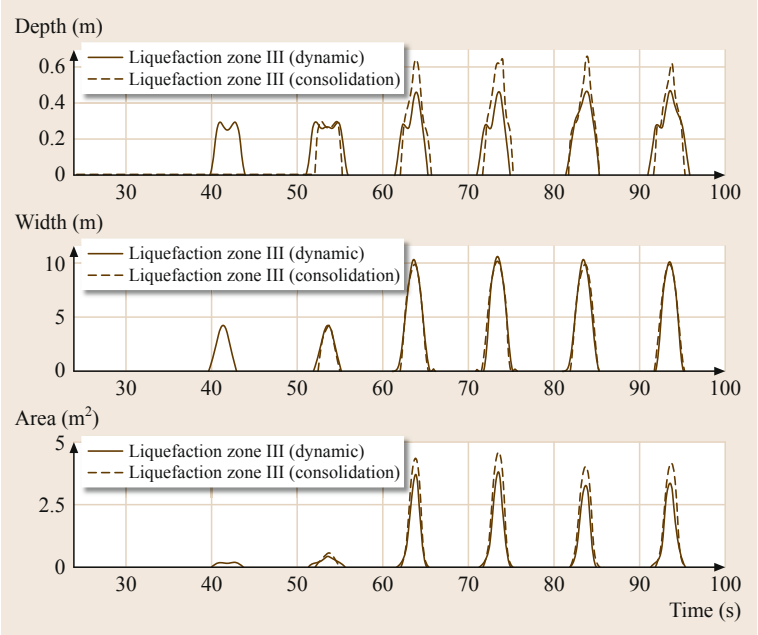


Fig. 39.22 The liquefaction properties (depth, width and area) in liquefaction Zone III

39.5 Remarks

In this chapter, an overview of the existing models for the wave-induced soil response around breakwaters was outlined. A new model integrating both oscillatory and residual mechanisms of the wave-induced soil response was also presented and validated with the existing experimental data. Then, the effects of currents on the

soil response were examined, followed by the application of the porous models to caisson-type breakwaters foundations.

In addition to the summary of the recent advances in the field of offshore geotechnics, the following areas can be further studied in the future:

1. Most existing models were based on the constitutive models (including plastic soil behavior) for onshore soils, rather than marine sediments. It is necessary to further develop the constitutive model for marine sediments.
2. Most previous studies for the wave–seabed–structure interactions have focused on the oscillatory mechanisms. It is desired to have an integrated model for both oscillatory and residual mechanisms for the cases around marine structures.
3. The existing models have been limited to uncoupled or weakly couple or one-way integrated approach, rather than fully coupling analysis, due to the complicate mechanism of wave–seabed–structure interactions. This area has a room for further developments.

References

39.1 J.A. Putnam: Loss of wave energy due to percolation in a permeable sea bottom, Trans. Am. Geophys. Union **30**(3), 349–356 (1949)

39.2 H. Nakamura, R. Onishi, H. Minamide: On the seepage in the seabed due to waves, Proc. 20th Coast. Eng. Conf. JSCE (1973) pp. 421–428

39.3 M.A. Biot: General theory of three-dimensional consolidation, J. Appl. Phys. **26**(2), 155–164 (1941)

39.4 O.C. Zienkiewicz, C.T. Chang, P. Bettess: Drained, undrained, consolidating and dynamic behaviour assumptions in soils, Géotechnique **30**(4), 385–395 (1980)

39.5 D.-S. Jeng, M.S. Rahman, T.L. Lee: Effects of inertia forces on wave-induced seabed response, Int. J. Offshore Polar Eng. **9**(4), 307–313 (1999)

39.6 D.-S. Jeng, M.S. Rahman: Effective stresses in a porous seabed of finite thickness: Inertia effects, Can. Geotech. J. **37**(4), 1388–1397 (2000)

39.7 M.A. Biot: Theory of propagation of elastic waves in a fluidsaturated porous solid, Part I: Low frequency range, J. Acoust. Soc. Am. **28**, 168–177 (1956)

39.8 D.-S. Jeng, D.H. Cha: Effects of dynamic soil behavior and wave non-linearity on the wave-induced pore pressure and effective stresses in porous seabed, Ocean Eng. **30**(16), 2065–2089 (2003)

39.9 M.B.C. Ulker, M.S. Rahman, D.-S. Jeng: Wave-induced response of seabed: Various formulations and their applicability, Appl. Ocean Res. **31**(1), 12–24 (2009)

39.10 S. Sassa, H. Sekiguchi: Analysis of wave-induced liquefaction of sand beds, Géotechnique **51**(2), 115–126 (2001)

39.11 D.-S. Jeng, J. Ou: 3D models for wave-induced pore pressure near breakwater heads, Acta Mech. **215**, 85–104 (2010)

39.12 R.O. Reid, K. Kajiura: On the damping of gravity waves over a permeable sea bed, Trans. Am. Geophys. Union **38**, 662–666 (1957)

39.13 J.F.A. Sleath: Wave-induced pressures in beds of sand, J. Hydraul. Div. ASCE **96**(2), 367–378 (1970)

39.14 P.L.F. Liu: Damping of water waves over porous bed, J. Hydraul. Div. ASCE **99**(12), 2263–2271 (1973)

39.15 P.L.F. Liu: On gravity waves propagated over a layered permeable bed, Coast. Eng. **1**, 135–148 (1977)

39.16 G. Dagan: The generalization of Darcy law for nonuniform flows, Water Resour. Res. **15**(1), 1–7 (1979)

39.17 P.L.F. Liu, R.A. Dalrymple: The damping of gravity water-waves due to percolation, Coast. Eng. **8**(1), 33–49 (1984)

39.18 S.R. Massel: Gravity waves propagated over permeable bottom, J. Waterw. Harb. Coast. Eng. ASCE **102**(2), 111–121 (1976)

39.19 W.W. Mallard, R.A. Dalrymple: Water waves propagating over a deformable bottom, Proc. 9th Annu. Offshore Technol. Conf. (1977) pp. 141–145

39.20 T.H. Dawson: Wave propagation over a deformable sea floor, Ocean Eng. **5**, 227–234 (1978)

39.21 H. Moshagen, A. Torum: Wave induced pressures in permeable seabeds, J. Waterw. Harb. Coast. Eng. Div. ASCE **101**(1), 49–57 (1975)

39.22 T. Yamamoto, H.L. Koning, H. Sellmeijer, E.V. Hijum: On the response of a poro-elastic bed to water waves, J. Fluid Mech. **87**(1), 193–206 (1978)

39.23 J.H. Prevost, O. Eide, K.H. Anderson: Discussion on 'Wave induced pressures in permeable seabeds' by Moshagen and Torum, J. Waterw. Harb. Coast. Eng. Div. ASCE **101**(1975), 464–465 (1975)

39.24 Z. Gu, H. Wang: Gravity waves over porous bottoms, Coast. Eng. **15**(5/6), 497–524 (1991)

39.25 T. Yamamoto: Wave induced instability seabed, Proc. ASCE Special Conf. Coast. Sediments (1977) pp. 898–913

39.26 O.S. Madsen: Wave-induced pore pressures and effective stresses in a porous bed, Géotechnique **28**(4), 377–393 (1978)

39.27 T. Yamamoto: Wave-induced pore pressures and effective stresses in inhomogeneous seabed foundations, Ocean Eng. **8**, 1–16 (1981)

39.28 S. Okusa: Wave-induced stress in unsaturated submarine sediments, Géotechnique **35**(4), 517–532 (1985)

39.29 B. Gatmiri: A simplified finite element analysis of wave-induced effective stress and pore pressures in permeable sea beds, Géotechnique **40**(1), 15–30 (1990)

39.30 D.-S. Jeng, J.R.C. Hsu: Wave-induced soil response in a nearly saturated seabed of finite thickness, Géotechnique **46**(3), 427–440 (1996)

39.31 M.S. Rahman, K. El-Zahaby, J. Booker: A semi-analytical method for the wave-induced seabed response, Int. J. Numer. Anal. Methods Geomech. **18**, 213–236 (1994)

39.32 E. Varley, B.R. Seymour: A method for obtaining exact solutions to partial differential equations with variable coefficients, Stud. Appl. Math. **78**, 183–225 (1988)

39.33 B.R. Seymour, D.-S. Jeng, J.R.C. Hsu: Transient soil response in a porous seabed with variable permeability, Ocean Engineering **23**(1), 27–46 (1996)

39.34 Y.S. Lin, D.-S. Jeng: The effect of variable permeability on the wave-induced seabed response, Ocean Eng. **24**(7), 623–643 (1997)

39.35 D.-S. Jeng, B.R. Seymour: Response in seabed of finite depth with variable permeability, J. Geotech. Geoenviron. Eng. ASCE **123**(10), 902–911 (1997)

39.36 D.-S. Jeng, B.R. Seymour: Wave-induced pore pressure and effective stresses in a porous seabed with variable permeability, J. Offshore Mech. Arct. Eng. ASME **119**(4), 226–233 (1997)

39.37 T. Kitano, H. Mase: Wave-induced porewater pressure in a seabed with inhomogeneous permeability, Ocean Eng. **28**, 279–296 (2001)

39.38 D.-S. Jeng: Wave-induced liquefaction potential in a cross-anisotropic seabed, J. Chin. Inst. Eng. **19**(1), 59–70 (1996)

39.39 D.-S. Jeng: Soil response in cross-anisotropic seabed due to standing waves, J. Geotech. Geoenviron. Eng. ASCE **123**(1), 9–19 (1997)

39.40 M. Yuhi, H. Ishida: Theoretical analysis of the response of a cross-anisotropic seabed to ocean surface waves, Proc. Jpn. Soc. Civil Eng. (JSCE) (1997) pp. 49–61

39.41 M. Yuhi, H. Ishida: Simplified solutions for wave-induced response of anisotropic seabed, J. Waterw. Harb. Coast. Eng. ASCE **128**(1), 46–50 (2002)

39.42 D.-S. Jeng: Discussion to 'Simplified solutions of wave-induced seabed response in anisotropic seabed' by Yuhi and Ishida, J. Waterw. Harb. Coast. Eng. ASCE **129**(3), 151–153 (2003)

39.43 C.C. Mei, M.A. Foda: Wave-induced response in a fluid-filled poro-elastic solid with a free surface – A boundary layer theory, Geophys. J. R. Astron. Soc. **66**, 597–631 (1981)

39.44 L.H. Huang, C.H. Song: Dynamic response of poro-plastic bed to water waves, J. Hydraul. Eng. ASCE **119**(9), 1003–1020 (1993)

39.45 J.R.C. Hsu, D.-S. Jeng: Wave-induced soil response in an unsaturated anisotropic seabed of finite thickness, Int. J. Numer. Anal. Methods Geomech. **18**(11), 785–807 (1994)

39.46 L.H. Huang, A.T. Chwang: Trapping and absorption of sound waves. II: A Sphere covered with a porous layer, Wave Motion **12**, 401–414 (1990)

39.47 T. Kitano, H. Mase: Boundary-layer theory for anisotropic seabed response to sea waves, J. Waterw. Harb. Coast. Eng. ASCE **125**(4), 187–194 (1999)

39.48 B.M. Sumer, N.S. Cheng: A random-walk model for pore pressure accumulation in marine soils, Proc. 9th Int. Offshore Polar Eng. Conf. (1999) pp. 521–528

39.49 W. Madga: Wave-induced uplift force acting on a submarine buried pipeline: Finite element formulation and verification of computations, Comput. Geotech. **19**(1), 47–73 (1996)

39.50 W. Madga: Wave-induced uplift force on a submarine pipeline buried in a compressible seabed, Ocean Eng. **24**(6), 551–576 (1997)

39.51 W. Madga: Wave-induced cyclic pore-pressure perturbation effects in hydrodynamic uplift force acting on submarine pipeline buried in seabed sediments, Coast. Eng. **39**, 243–272 (2000)

39.52 K. Zen, H. Yamazaki: Mechanism of wave-induced liquefaction and densification in seabed, Soils Found. **30**(4), 90–104 (1990)

39.53 K. Zen, H. Yamazaki: Oscillatory pore pressure and liquefaction in seabed induced by ocean waves, Soils Found. **30**(4), 147–161 (1990)

39.54 B. Gatmiri: Response of cross-anisotropic seabed to ocean waves, J. Geotech. Eng. ASCE **118**(9), 1295–1314 (1992)

39.55 O.C. Zienkiewicz, F.C. Scott: On the principle of repeatability and its application in analysis of turbine and pump impellers, Int. J. Numer. Methods Eng. **9**, 445–452 (1972)

39.56 Y.S. Lin, D.-S. Jeng: Effects of variable shear modulus on wave-induced seabed response, J. Chin. Inst. Eng. **24**(1), 109–115 (2000)

39.57 S.D. Thomas: A finite element model for the analysis of wave induced stresses, displacements and pore pressure in an unsaturated seabed. I: Theory, Comput. Geotech. **8**(1), 1–38 (1989)

39.58 S.D. Thomas: A finite element model for the analysis of wave induced stresses, displacements and pore pressure in an unsaturated seabed. II: Model verification, Comput. Geotech. **17**(1), 107–132 (1995)

39.59 D.-S. Jeng, Y.S. Lin: Finite element modelling for water waves – Soil interaction, Soil Dyn. Earthq. Eng. **15**(5), 283–300 (1996)

39.60 Y.S. Lin, D.-S. Jeng: Response of poro-elastic seabed to a 3-D wave system: A finite element analysis, Coast. Eng. Jpn. **39**(2), 165–183 (1996)

39.61 D.-S. Jeng, Y.S. Lin: Non-linear wave-induced response of porous seabed: A finite element analysis, Int. J. Numer. Anal. Methods Geomech. **21**(1), 15–42 (1997)

39.62 D.-S. Jeng, Y.S. Lin: Poroelastic analysis for wave-seabed interaction problem, Comput. Geotech. **26**(1), 43–64 (2000)

39.63 M.A. Biot: Theory of propagation of elastic waves in a fluidsaturated porous solid, Part II: High frequency range, J. Acoust. Soc. Am. **28**, 179–191 (1956)

39.64 T. Sakai, H. Mase, A. Matsumoto: Effects of inertia and gravity on seabed response to ocean waves. In: *Modelling Soil-Water-Structure Interactions*, ed. by P.A. Kolkman, J. Linderberg, K. Pilarczyk (A. A. Balkema, Rotterdam 1988)

39.65 T. Sakai, K. Hatanaka, H. Mase: Wave-induced effective stress in seabed and its momentary liquefaction, J. Waterw. Port Coast. Ocean Eng, ASCE **118**(2), 202–206 (1992)

39.66 K. Horikawa: *Nearshore Dynamics and Coastal Processes* (Univ. of Tokyo Press, Tokyo 1988)

39.67 A.H.C. Chan: A Unified Finite Element Solution to Static and Dynamic Problems of Geomechanics, Ph.D. Thesis (Univ. of Wales, Swansea 1988)

39.68 S.L. Dunn, P.L. Vun, A.H.C. Chan, J.S. Damgaard: Numerical modelling of wave-induced liquefaction around pipelines, J. Waterw. Port Coast. Ocean Eng. **132**, 276–288 (2006)

39.69 M.A. Biot: Mechanics of deformation and acoustic propagation in porous media, J. Appl. Phys. **33**(4), 1482–1498 (1962)

39.70 D.H. Cha, D.-S. Jeng, M.S. Rahman, H. Sekiguchi, K. Zen, H. Yamazaki: Effects of dynamic soil behaviour on the wave-induced seabed response, Int. J. Ocean Eng. Technol. **16**(5), 21–33 (2002)

39.71 T.W. Chen, L.H. Huang, C.H. Song: Dynamic response of poroelastic bed to nonlinear water waves, J. Eng. Mech. ASCE **123**(10), 1041–1049 (1997)

39.72 P.C. Hsieh, L.H. Huang, T.W. Wang: Dynamic response of soft poroelastic bed to linear water waves – A boundary layer approximation, Int. J. Numer. Anal. Methods Geomech. **25**, 651–674 (2001)

39.73 M. Yuhi, H. Ishida: Analytical solution for wave-induced seabed response in a soil-water two-phase mixture, Coast. Eng. J. **40**(4), 367–381 (1998)

39.74 D.-S. Jeng, T.L. Lee: Dynamic response of porous seabed to ocean waves, Comput. Geotech. **28**(2), 99–128 (2001)

39.75 D.-S. Jeng: *Porous Models for Wave-Seabed Interactions* (Springer, Heidelberg 2013)

39.76 M. Ulker, M.S. Rahman: Response of saturated and nearly saturated porous media: Different formulations and their applicability, Int. J. Numer. Anal. Methods Geomech. **33**(5), 633–664 (2009)

39.77 H. Sekiguchi, K. Kita, O. Okamoto: Response of poro-elastoplastic beds to standing waves, Soils Found. **35**(3), 31–42 (1995)

39.78 D.-S. Jeng: Discussion of 'Response of poro-elastic beds to standing waves' by Sekiguchi et al, Soil. Found. **37**(2), 139 (1997)

39.79 Q.S. Yang, H.B. Poorooshasb: Seabed response to wave loading, Proc. 7th Int. Offshore Polar Eng. Conf. (1997) pp. 689–695

39.80 X. Li, J. Zhang, H. Zhang: Instability of wave propagation in saturated poroelastoplastic media, Int. J. Numer. Anal. Methods Geomech. **26**, 563–578 (2002)

39.81 H.B. Seed, M.S. Rahman: Wave-induced pore pressure in relation to ocean floor stability of cohesionless soils, Mar. Geotechnol. **3**(2), 123–150 (1978)

39.82 B.M. Sumer, J. Fredsøe: *The Mechanics of Scour in the Marine Environment* (World Scientific, Singapore 2002)

39.83 J.R.C. Hsu, D.-S. Jeng, C.P. Tsai: Short-crested wave-induced soil response in a porous seabed of infinite thickness, Int. J. Numer. Anal. Methods Geomech. **17**(8), 553–576 (1993)

39.84 W.G. McDougal, Y.T. Tsai, P.L.F. Liu, E.C. Clukey: Wave-induced pore water pressure accumulation in marine soils, J. Offshore Mech. Arct. Eng. ASME **111**(1), 1–11 (1989)

39.85 P. de Alba, H.B. Seed, C.K. Chan: Sand liquefaction in large-scale simple shear tests, J. Geotech. Div. ASCE **102**, 909–928 (1976)

39.86 B.M. Sumer, V.S.O. Kirca, J. Frøsde: Experimental validation of a mathematical model for seabed liquefaction under waves, Int. J. Offshore Polar Eng. **22**, 133–141 (2012)

39.87 L. Cheng, B.M. Sumer, J. Fredsøe: Solution of pore pressure build up due to progressive waves, Int. J. Numer. Anal. Methods Geomech. **25**, 885–907 (2001)

39.88 D.-S. Jeng, B.R. Seymour: A simplified analytical approximation for pore-water pressure build-up in a porous seabed, J. Waterw. Port Coast. Ocean Eng, ASCE **133**(4), 309–312 (2007)

39.89 A.M. Geremew: Pore-water pressure development caused by wave-induced cyclic loading in deep porous formation, Int. J. Geomech. ASCE **13**(1), 65–68 (2013)

39.90 Z. Guo, D.-S. Jeng: Discussion of 'Pore-water pressure development caused by wave-induced cyclic loading in deep porous formation' by Geremew, Int. J. Geomech. **14**(2), 326–328 (2014)

39.91 B. Liu, D.-S. Jeng: Laboratory study for pore pressure in sandy bed under wave loading, Proc. 23rd Int. Offshore Polar Eng. Conf. (2013)

39.92 D.-S. Jeng, B.R. Seymour, J. Li: A new approximation for pore pressure accumulation in marine sediment due to water wave, Int. J. Numer. Anal. Methods Geomech. **31**(1), 53–69 (2007)

39.93 S. Sassa, H. Sekiguchi, J. Miyamamot: Analysis of progressive liquefaction as moving-boundary problem, Géotechnique **51**(10), 847–857 (2001)

39.94 Z. Liu, D.-S. Jeng, A.H. Chan, M.T. Luan: Wave-induced progressive liquefaction in a poro-elasto-plastic seabed: A two-layered model, Int. J. Numer. Anal. Methods Geomech. **33**(5), 591–610 (2009)

39.95 G.P. Thomas: Wave-current interactions: an experimental and numerical study. Part I. Linear waves, Appl. Math. Model. **110**, 457–474 (1981)

39.96 R.E. Baddour, S.W. Song: On the interaction between waves and currents, Ocean Eng. **17**(1/2), 1–21 (1990)

39.97 R.E. Baddour, S.W. Song: Interaction of higher-order water waves with uniform currents, Ocean Eng. **17**(6), 551–568 (1990)

39.98 H.C. Hsu, Y.Y. Chen, J.R.C. Hsu, W.J. Tseng: Nonlinear water waves on uniform current in Lagrangian coordinates, J. Nonlinear Math. Phys. **16**(1), 47–61 (2009)

39.99 Y.J. Jian, Q.Y. Zhu, J. Zhang, Y.F. Wang: Third order approximation to capillary gravity short crested

waves with uniform currents, Appl. Math. Model. **33**(4), 2035–2053 (2009)
39.100 Y. Zhang, D.-S. Jeng, F.P. Gao, J.S. Zhang: An analytical solution for response of a porous seabed to combined wave and current loading, Ocean Eng. **57**, 240–247 (2013)
39.101 B. Liu, D.-S. Jeng, J.-S. Zhang: Dynamic response of a porous seabed of finite depth due to combined wave and current loadings, J. Coast. Res. **30**(4), 765–776 (2014)
39.102 C.C. Liao, D.-S. Jeng: Wave (current)-induced soil response in marine sediments, Theor. Appl. Mech. Lett. **3**(1), 012002 (2013)
39.103 J.-S. Zhang, Y. Zhang, C. Zhang, D.-S. Jeng: Numerical modeling of seabed response to the combined wave-current loading, Int. J. Offshore Mech. Arct. Eng. ASME **135**(3), 031102 (2013)
39.104 J. Ye, D.-S. Jeng: Response of seabed to natural loading-waves and currents, J. Eng. Mech. ASCE **138**(6), 601–613 (2012)
39.105 H. Mase, T. Sakai, M. Sakamoto: Wave-induced porewater pressure and effective stresses around breakwater, Ocean Eng. **21**(4), 361–379 (1994)
39.106 N. Mizutani, A.M. Mostafa: Nonlinear wave-induced seabed instability around coastal structures, Coast. Eng. J. **40**(2), 131–160 (1998)
39.107 A.M. Mostafa, N. Mizutani, K. Iwata: Nonlinear wave, composite breakwater and seabed dynamic interaction, J. Waterw. Port Coast. Ocean Eng, ASCE **125**(2), 88–97 (1999)
39.108 D.-S. Jeng, D.H. Cha, Y.S. Lin, P.S. Hu: Wave-induced pore pressure around a composite breakwater, Ocean Eng. **28**(10), 1413–1432 (2001)
39.109 M. Ulker, M.S. Rahman, M.N. Guddati: Wave-induced dynamic response and instability of seabed around caisson breakwater, Ocean Eng. **37**(17/18), 1522–1545 (2010)
39.110 P.L.F. Liu, P. Lin, K.A. Chang, T. Sakakiyama: Numerical modelling of wave interaction with porous structures, J. Waterw. Port Coast. Ocean Eng, ASCE **125**(6), 322–330 (1999)
39.111 J.L. Lara, N. Garcia, I.J. Losada: RANS modeling applied to random wave interaction with submerged permeable structures, Coast. Eng. **53**, 395–417 (2006)
39.112 D.S. Hur, C.H. Kim, J.S. Yoon: Numerical study on the interaction among a nonlinear wave, composite breakwater and sandy seabed, Coast. Eng. **57**(10), 917–930 (2010)
39.113 S.H. Shao: Incompressible SPH flow model for wave interactions with porous media, Coast. Eng. **57**, 304–316 (2010)
39.114 T.J. Hsu, T. Sakakiyama, P.L.F. Liu: A numerical model for wave motions and turbulence flows in front of a composite breakwater, Coast. Eng. **46**, 25–50 (2002)
39.115 P. Lin, P.L.F. Liu: A numerical study of breaking waves in the surf zone, J. Fluid Mech. **359**, 239–264 (1998)
39.116 D.-S. Jeng, J.H. Ye, J.-S. Zhang, P.L.F. Liu: An integrated model for the wave-induced seabed response around marine structures: Model verifications and applications, Coast. Eng. **72**, 1–19 (2013)
39.117 J.G. Wang, B. Zhang, T. Nogami: Wave-induced seabed response analysis by radial point interpolation meshless method, Ocean Eng. **31**(1), 21–42 (2004)
39.118 P.L.F. Liu, Y.S. Park, J.L. Lara: Long-wave-induced flows in an unsaturated permeable seabed, J. Fluid Mech. **586**, 323–345 (2007)
39.119 H.B. Lu: The Research on Pore Water Pressure Response to Waves in Sandy Seabed, Ph.D. Thesis (Changsha Univ. Science and Technology, Changsha Hunan 2005)

40. Buoy Technology

Andrew Hamilton

Small buoy and mooring systems have a long history and many uses ranging from navigation and marking buoys, meteorological monitoring systems, oceanographic measurement systems, wave-measurement buoys, and tsunami warning systems. Many of these designs are very mature and have benefited from decades of field experience. Others are more recent and rely heavily on synthetic materials, modern electronics, and analytical and computer analysis techniques to create high-performance systems that operate in extreme depths, boundary currents, remote locations, enable a power and data connection to submerged instrumentation, or have predictable responses to allow accurate measurement of the sea state. The material presented here applies to systems that are less than approximately 5 m in diameter and have masses less than 5000 kg or so. This chapter outlines the various choices facing designers of these systems, presents approaches to the design of standard systems, presents the material properties of commonly used materials, and describes the analysis techniques and tools typically applied to systems that must meet specialized functional requirements. The chapter concludes with diagrams of several mooring systems that have been successfully fielded.

40.1 **Buoy and Mooring Types and Design Considerations** 937
40.1.1 Buoys 938
40.1.2 Moorings 940
40.1.3 Failures and Hazards 942

40.2 **Buoy and Mooring System Components** 942
40.2.1 Strength Members 942
40.2.2 Terminations and Strain Relief 946
40.2.3 Buoys and Floats 947
40.2.4 Hardware 947
40.2.5 Anchors 948
40.2.6 Acoustic Releases 949

40.3 **Analysis Techniques** 949
40.3.1 Environmental Conditions 950
40.3.2 Static Analysis 950
40.3.3 Frequency Domain Analysis 952
40.3.4 Nonlinear Analysis 954

40.4 **Example Designs** 956
40.4.1 Chain-Catenary Mooring 959
40.4.2 Auto-Detection Mooring 959
40.4.3 Deep-Water Mooring with Inductively Linked Instrumentation 961

References 961

40.1 Buoy and Mooring Types and Design Considerations

Because of the rich set of trade-offs that must be managed in a successful design, buoy/mooring systems tend to be very customized to each particular application. Differences in deployment depth, sea conditions, current conditions, service intervals, system payload requirements, deployment options, and budget constraints must all be considered. Long-term moorings are inherently high-fatigue systems, with hundreds of thousands of tension and motion cycles per month. Systems deployed for long periods of time with minimal maintenance often fail in unexpected ways that are hard to anticipate ahead of time. For these reasons, experience with what has worked before is very important, and adapting existing successful designs to new applications is generally a good idea where possible.

When designing a mooring system, the size and shape of the surface expression is determined by the requirements of its application and by the environmental conditions it will experience. Properties of the surface expression will impact the rest of the mooring system directly and should be identified early in the design process if possible. Typically the buoy size is set by payload requirements, including the mass and size of any instrumentation, batteries, solar/wind arrays, radio antenna height, and stability requirements. The smallest sized buoy system that meets the require-

ments is usually the most economical solution. The cost of materials, deployment, recovery, and maintenance increases with buoy size, sometimes dramatically as larger mooring components, ships, cranes, and winches are required. A mooring design often proceeds by first identifying the payload requirements and subsequently designing a buoy that can house that payload and be sufficiently stable for operation and maintenance.

Typical installations consist of a floating buoy element held in place by a mooring system. These two components interact with one another significantly enough that the behavior of the entire system should be considered during design. In spite of this coupling, however, it is useful to consider the behavior of the buoy alone without the mooring attached for several reasons. The buoy's behavior without the mooring's influence is generally well approximated by linear analysis, and important frequency domain characteristics such as resonance behaviors can be examined. Additionally, it is possible for the buoy to become separated from the mooring during installation, service, or recovery. Ensuring that the buoy alone has sufficient stability to stay upright dramatically increases the chances of recovery. Additionally, drifting buoys are commonly used for measuring sea-surface velocities and therefore have no mooring loads. However, these devices typically have drogues that affect the buoy dynamics [40.1].

Section 40.1.1 outlines the dynamics of various kinds of small freely floating buoys. Later sections describe more sophisticated ways to analyze the coupled buoy/mooring system that generally rely on time-domain computer simulations due to the nonlinear nature of the mooring dynamics. These techniques have the advantage of providing a good estimate of a particular system's behavior but suffer from the fact that the entire design space must be traversed with individual simulations, making general conclusions about a system's behavior difficult to determine. For instance, for a given sea-state, a mooring system may behave very differently in large currents compared to slack conditions, and each condition must be analyzed separately.

In mooring a buoy, there are a wide variety of mooring system types from which to choose from. Again the selection of the mooring is subject to a large number of considerations including the installation depth, tidal variations, seafloor conditions, seafloor impact restrictions, expected current profiles and sea conditions, system lifetime, accuracy requirements for the buoy location, and sometimes requirements to provide a power or data connection to the sea-floor or elements along the mooring string. The equipment available for deployment and recovery (ships, winches, and cranes) also dictates the options for the mooring system, and it is not uncommon for a system to be designed for a particular class of deployment vessel. Section 40.1.2 outlines the range of mooring options available to the system designer and discusses the capabilities and limitations of each.

40.1.1 Buoys

Discus and toroidal type buoys have shallow drafts and large water-plane areas, resulting in systems that have low drag-to-buoyancy ratios and are subsequently relatively easy to moor. A potentially undesirable characteristic of these kinds of buoys is that they tend to heave and pitch so that they follow the sea surface almost perfectly. This results in pitch behaviors that may interfere with desired measurements and can introduce some wear and fatigue issues into the connection between the surface buoy and the mooring. Additionally, when large enough in diameter these buoys are stable upside down as well as right-side up, so once capsized the restoring moment of the mooring can be insufficient to right the system; they tend to stay capsized. Figure 40.1a presents an estimate of the heave and pitch response of this kind of buoy. These buoys have been widely used by the National Oceanic and Atmospheric Administration (NOAA) National Data Buoy Center for meteorological measurements. That organization has also deployed much larger discus buoys of up to 10 m in diameter [40.2].

Surface buoys of intermediate aspect ratios are often the result of compromises that require the diameter of a system to be small for handling, and the payload or buoyancy of the system to be large enough that the draft must be increased. The result is a draft that is of the same order of magnitude as the diameter. A sphere-shaped buoy naturally falls into this category and is a common choice for marker buoys because of the availability and simplicity of steel spheres. Figure 40.1b illustrates the typical response of a system of this type. As seen in the estimated heave and pitch response curves, this type of buoy will pitch less than the discus buoys but will still follow the wave-surface in heave. What is evident, however, is the presence of a moderate resonance peak that is not present in the low aspect-ratio discus buoy.

For the comparison presented in Fig. 40.1, the initial stability (metacentric height) of the buoys analyzed has been chosen to be the same in each case to highlight the effects of the different geometry. Typically, freely floating discus buoys will be much more stable than intermediate aspect ratio buoys due to the large water-plane area. In the case of moored systems, however, the mooring-line tension contributes significantly to the stability of an intermediate aspect ratio buoy because of the larger distance between the connection point and the

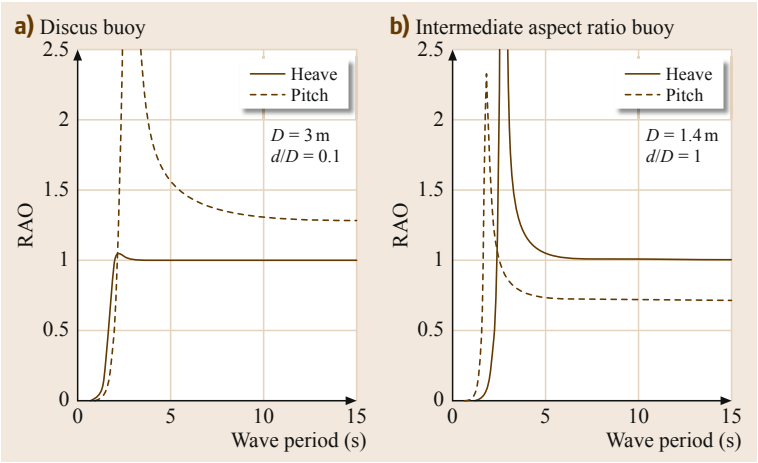


Fig. 40.1a,b Comparison of response amplitude operator (RAO) characteristics for buoys of different aspect ratio: **(a)** discus buoy, **(b)** intermediate aspect ratio buoy (d = draft, D = diameter

waterplane. Therefore, the resulting stability of the two types of buoy are often comparable.

These figures make it clear that small buoy systems are wave followers in almost all cases. Higher aspect ratio buoys have a pronounced heave resonance, located at a frequency below where most of the energy is present in open-ocean waves. The pitch response is broader and depends strongly on the buoys stability and pitch moment of inertia. Therefore, to determine the exact pitch response of a given system the RAO curves need to be computed for the particular buoy design in question; Sect. 40.3.3 presents the needed data to perform these calculations for cylindrically shaped buoys.

Spar buoys are at the other extreme and have large drafts and small water-plane areas. If made large enough the motions of these buoys can be almost totally decoupled from the sea-surface, but systems with that behavior are generally larger than would be considered for the purposes highlighted in this section. In small and intermediate sizes, the spar shape can have some applications. They will heave and be over topped by waves occasionally, but can be engineered for minimum pitch and roll. Figure 40.2 illustrates the expected heave response of three different sizes of spar buoy. As shown, these systems have a very pronounced resonance response, and only the largest systems can be designed to place this resonance at wave periods longer than typical ocean waves.

Regardless of aspect ratio, most buoys are designed to be axisymmetric. There have been boat-shaped systems such as the US Navy's NOMAD buoy that were designed in the 1960s and deployed for decades. These are moored from one end and rotate to face into the wind and waves. In large seas, surfing and slamming of these systems can become an issue. Relatively little recent work has been done with these shapes, and the

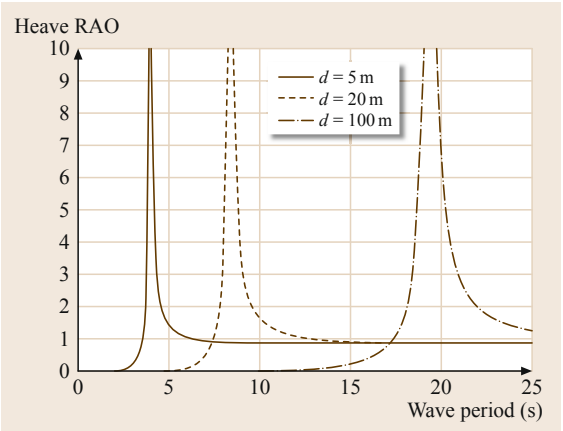


Fig. 40.2 Comparison of response amplitude operator (RAO) characteristics for spar buoys of different lengths (d = draft)

vast majority of systems designed and built today have axisymmetric hulls. Even with axisymmetric hulls the above-water instrumentation and equipment can introduce rotation due to wind loading. A design decision must be made to keep the system as axisymmetric as possible or to install vanes to consistently orient the buoy to the wind. In the latter choice, the prevailing weather conditions must be considered: in many locations the wind rotates consistently with passing weather systems, and care needs to be taken to keep from introducing more rotations than the mooring system can tolerate, or include swivels in the design.

Although buoy systems traditionally have sufficient buoyancy to ensure that the buoy remains on the sea surface in the most severe sea-state expected, systems are also sometimes designed and built with minimal reserve buoyancy. These stay on the surface and operational in moderate sea states but begin to be pulled under water

when faced with large waves. These systems must necessarily be designed to withstand submergence which can be costly, but the mooring size and cost can usually be reduced, and very shallow water deployments in which the system must survive breaking waves become possible. This design approach drives these systems to the middle of the aspect-ratio range but the techniques applied here to estimate buoy-only response curves are not applicable in conditions that lead to the buoy submerging. This is because once the buoy submerges, the water-plane area that provides the restoring force is removed, which introduces a significant nonlinearity into the analysis that is not captured by the linear analysis. Therefore, the occasionally submerging nature of these buoys introduces analysis and modeling challenges that are described in the computer modeling section below.

40.1.2 Moorings

There is a wide variety of mooring types available, and the selection for a given application involves evaluating the trade-off between many factors including water depth, buoy station-keeping performance, deployment strategy, cost factors, and special requirements like the need to provide a power and/or data path between the surface mooring and submerged instrumentation. In almost all cases, a key consideration in the mooring design is the amount of compliance provided to allow the surface buoy to move in response to waves and sea-surface level changes due to tides. There are three main techniques to provide the required compliance: a *chain-catenary* mooring in which part of the mooring string simply lies on the seafloor; a *semi-taut* design that relies on elasticity of the mooring components themselves to stretch when loaded; and an *inverse-catenary* technique in which the mooring line is induced to assume an *S* shape in the water column by weighting and floating different sections of the mooring. These three types of moorings are illustrated in Fig. 40.3. In many cases hybrid designs make use of more than one of these techniques.

Chain-catenary moorings are often the simplest to design and install. This type of mooring has many advantages such as a tolerance to water-depth uncertainty or changes, the use of durable low-cost materials, and a range of anchor options. One downside of this approach is that the design becomes more challenging in deeper water depths. In depths above a few hundred meters, the horizontal forcing of the mooring string and buoy due to currents and winds begins to overwhelm the practical amount of chain that can be placed on the seafloor. Even modest amounts of current lift the chain and pull directly on the anchor. This effectively eliminates the compliance effect. Chain-catenary moorings also have relatively large watch-circles and will have a significant impact on the seafloor. Thus, they are inappropriate in marine-protected areas or other locations where the impact to the seafloor needs to be minimized. Finally, although the wires and chains used in these types of moorings are usually very durable and resistant to damage from fishing operations or ship traffic, the constant wearing of the chain on the seabed can be an issue for some seabed types and may not be suitable at all for rocky seabeds that have a high likelihood of snagging the mooring chain and eliminating all compliance.

Semi-taut moorings rely on the elasticity of the mooring elements themselves to provide the needed compliance. Typically these designs are used in deep-water installations where a relatively small amount of elasticity in the mooring materials translates to a significant extension across the loads seen during service. For reasons of elasticity and in-water weight, these moorings are typically constructed of rope materials such as nylon, polyester, polypropylene, and polyolefin. Although potentially elastic enough, steel wire-rope simply becomes too heavy in long lengths, driving up the system design into a spiral of becoming larger than can be supported by the types of buoy systems considered in this chapter. Semi-taut mooring designs do not include any chain or material on the seafloor aside from the anchor, so they have a minimal seafloor impact and are good candidates for including acoustic release devices. As an exception to the deep-water application of these moorings, some shallow-water moorings utilize rubber-based hoses and cords to provide extreme elasticity to a mooring over a short length. Semi-taut moorings typically are well behaved in that they include a lot of elasticity, and the tension in the mooring string rarely goes to zero. This limits snap loading that can be present in the other types of mooring systems. In all cases of this type of mooring, good knowledge of the water depth at the deployment site is required, and careful attention must be paid to the design and fabrication of the mooring string. In particular, accurately measuring long lengths of ropes is a nontrivial practical issue that has significant bearing on implementation of these systems.

Inverse-catenary moorings do not rely entirely on the elasticity of the mooring string, but instead introduce some compliance by creating an *S-curve* in the geometry of the mooring string. This is accomplished by floating part of the mooring string above the anchor and weighting part of the mooring. There are many ways to do this including the use of ropes of different densities. Polypropylene, Spectra, and polyolefin are all less dense than water while polyester and nylon are

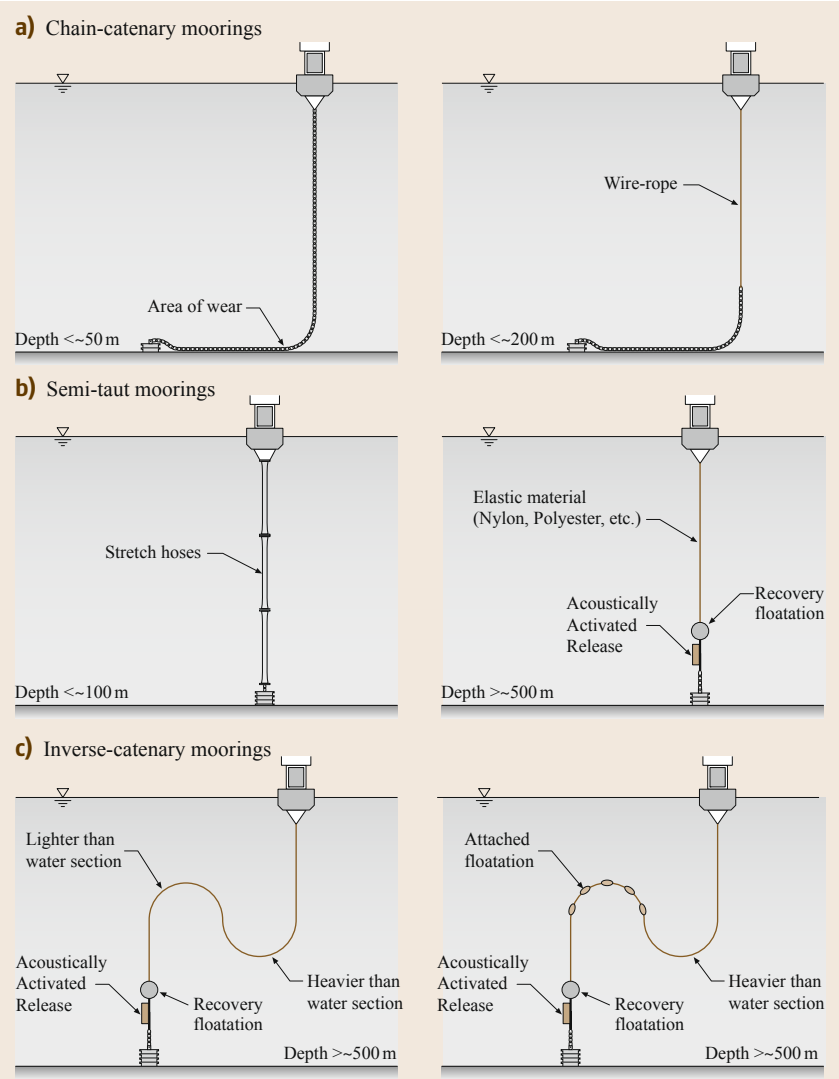


Fig. 40.3a–c Three mooring types typical in small buoy systems: **(a)** chain-catenary mooring, **(b)** semi-taut moorings, **(c)** inverse-catenary moorings

more dense. In other cases when the mooring application requires wire-rope or other heavier-than-water materials (as may be the case if electrical conductors need to be included), direct attachment of flotation to the mooring string may be required. Unlike semi-taut moorings, this type of mooring tends to be poorly behaved in large sea-states. Low-tension conditions at some point in the mooring string are often unavoidable in low-current regimes. In the presence of high currents, the geometric compliance tends to be minimized by the stretching out of mooring string, which can happen in surprisingly low currents for deep moorings. In all cases, abrupt changes in the material properties (stiffness, elasticity, mass) should be minimized as unusual dynamics can be excited at such locations, resulting in highly repetitive motions and premature failure. Techniques of specialty ropes that blend the density of the material gradually from one region to another have been employed, as well as extruded flotation jackets over wire-rope; these solutions are typically highly customized and expensive to fabricate.

Moorings are often also characterized by *scope*, the ratio of mooring length to water depth. Semi-taut moorings have the lowest scope, greater than 1 when deployed, but less than 1 when the mooring materials are measured in a relaxed state. Inverse-catenary moorings are usually designed with a scope between approximately 1.1 and 1.6. Chain-catenary moorings typically require higher scopes for deeper depths; the important consideration in this case is that at least part of the mooring chain remains on the seafloor so that the anchor does not experience vertical loads.

40.1.3 Failures and Hazards

Mooring system failure modes are numerous, but the most common are vandalism, fatigue, corrosion, and fish-bite. Vandalism and interference from commercial traffic is a common cause of the loss of service of a mooring system. Problems include collisions, fishing line and net entanglement, fish operations utilizing buoys as fish-aggregating devices, and theft. These issues depend on location and cannot always be eliminated, but mitigation techniques include armoring rope strength members and making buoys difficult to board [40.3]. Fatigue and corrosion failures are also the causes of mooring failures and careful engineering must be performed to anticipate the possible fatigues issues that can arise [40.4]. Failures due to fish-bite also occur and are hard to anticipate or design for [40.5].

40.2 Buoy and Mooring System Components

A wide variety of materials are used in oceanographic buoys, often depending on local availability and past practices of the designer and builder. This section outlines the specifics of various materials and components typically used and provides tables of approximate physical characteristics for these materials. Table 40.1 compares the fundamental properties of the materials used in oceanographic mooring systems, and more detailed tables throughout this section show the particular characteristics of components made of these materials in a specified construction. In all cases, these tables are meant to guide design, but the specifics of the actual components used should always be checked as there will be some variability around these examples.

40.2.1 Strength Members

Strength members fall into four broad categories: chain, wire-rope, synthetic ropes, and specialty stretch elements that are typically made of rubber compounds. Regardless of the material that the strength members are made of, an important characteristic to consider is the torque and rotation behavior of these elements. Because of the constantly oscillating tensions, even a small tendency to rotate as the element is put under tension can add up to significant rotation after hundreds or thousands of meters. In addition to imparting a torque on the ends, the induced torque and rotation can cause strength members to *hockle* when unloaded, resulting in a tight bend radius that can lead to failure when the load is reapplied. In less severe cases, too much torque-imbalance can cause self-wear of the various strands or components of a strength member where they move past one another. For these reasons, constructions that are inherently *torque-balanced* are primarily used.

Chain

Chain is usually included in mooring systems for its weight and durability, so it is often not necessary to use high-performance chain. Chain grades (30, 43, 70) represent the modulus of the steel in 1E6 psi units. Strengths higher than grade 43 should be avoided if galvanized to avoid hydrogen-embrittlement problems that can occur when high-strength steels are galvanized.

Chain selection is usually made based on size, weight, and availability. Table 40.2 presents chain

Table 40.1 Properties of common materials used in ocean buoy systems (after [40.6–9])

Material	Density		Modulus		Tensile strength	
	[kg/m^3]	[lb/in^3]	[Gpa]	[ksi]	[MPa]	[ksi]
Improved plow steel	8050	0.291	200	29 000	1517	220
Extra-improved plow steel	8050	0.291	200	29 000	1793	260
A36 structural steel	8050	0.291	200	29 000	500	73
316 stainless steel	7970	0.288	195	28 300	550	80
17-4 PH stainless steel	7800	0.282	196	28 300	1030	100
Nitronic 50 Stainless Steel	7880	0.285	190	28 000	827	120
UHMW-HDPE[a]	980	0.035	0.725	105	50	7.3
Nylon-66 fibers	1140	0.041	5.5	800	965	140
Polyester fibers	1380	0.050	13.8	2000	1172	170
Spectra/Dyneema fibers	980	0.035	117	17 000	2586	375
Kevlar-29 fibers	1440	0.052	70	10 200	2923	424
Kevlar-49 fibers	1440	0.052	112	16 300	2999	435

[a] Ultra high molecular weight-high density poly-ethylene

Table 40.2 Properties of common strength members (after [40.7, 10–12])

Material	Diameter [m]	Mass [kg/m]	Density [kg/m^3]	Wet weight in seawater [N/m]	Elastic stiffness (E A) [N]	Breaking strength [N]
Chain (grade 43)						
3/8″ (10 mm)	–	2.028	8050	17.363	52 810 000	72 100
1/2″ (13 mm)	–	3.430	8050	29.364	93 890 000	122 800
5/8″	–	5.503	8050	47.109	146 700 000	173 500
3/4″	–	7.949	8050	68.047	211 260 000	269 600
Wire–rope						
1/4″ 3X19 jacketed	0.008	0.192	3925	1.381	4 300 000	29 600
5/16″ 3X19 jacketed	0.011	0.313	3236	2.087	7 000 000	44 000
3/8″ 3X19 jacketed	0.013	0.447	3532	3.109	10 000 000	61 800
7/16″ 3X19 jacketed	0.014	0.585	3640	4.101	13 000 000	83 600
Nylon rope						
1/2″ 12-strand	0.013	0.110	1140	0.109	139 000	48 000
3/4″ 8-strand	0.019	0.203	1140	0.201	209 000	72 100
7/8″ 8-strand	0.022	0.285	1140	0.282	284 000	97 900
1″ 8-strand	0.025	0.353	1140	0.350	348 000	120 100
Polyester rope						
3/8″ 12-strand	0.010	0.064	1380	0.162	270 000	27 000
1/2″ 12-strand	0.013	0.149	1380	0.376	580 000	58 000
5/8″ 12-strand	0.016	0.220	1380	0.555	830 000	83 000
3/4″ 12-strand	0.019	0.285	1380	0.719	1 100 000	110 000
1″ 12-strand	0.025	0.516	1380	1.302	1 980 000	198 000
Polypropylene rope						
3/8″ 12-strand	0.010	0.040	940	−0.035	168 000	16 000
1/2″ 12-strand	0.013	0.070	940	−0.062	305 000	29 000
5/8″ 12-strand	0.016	0.116	940	−0.103	505 000	48 000
3/4″ 12-strand	0.019	0.144	940	−0.128	611 000	58 000
1″ 12-strand	0.025	0.312	940	−0.277	1 084 000	103 000
Spectra/Dyneema						
1/4″ 12-strand	0.006	0.024	980	−0.011	1 086 000	38 000
3/8″ 12-strand	0.010	0.054	980	−0.024	2 486 000	87 000
1/2″ 12-strand	0.013	0.095	980	−0.043	4 314 000	151 000
5/8″ 12-strand	0.016	0.152	980	−0.068	6 714 000	235 000
Vectran rope						
1/4″ 12-strand	0.006	0.033	1400	0.087	1 077 000	42 000
3/8″ 12-strand	0.010	0.068	1400	0.179	2 205 000	86 000
1/2″ 12-strand	0.013	0.131	1400	0.344	4 000 000	156 000
5/8″ 12-strand	0.016	0.208	1400	0.547	6 256 000	244 000

weights and strengths for several common sizes used in small mooring systems. Long-link chains are popular also because they are easier to shackle to, especially mid-span. Detailed information on standards for chain used in offshore applications can be found in [40.13].

Wire-Ropes

For wire-ropes used in small mooring systems, the dominant construction for mooring use is a 3X19 arrangement with three twisted main strands, each consisting of 19 wires each. The contra-helical winding of the 19 galvanized steel strands relative to the 3 main strands results in a design that is balanced and rotates little under load. This design has proven itself in thousands of applications and is the usual starting point for wire-rope in oceanographic mooring applications. Most commonly, wire-rope used in mooring systems has a polyethylene or polypropylene jacket extruded over the wire.

In addition to improving durability, the jacket provides a surface to clamp instrumentation to. For low-bandwidth communications, the jacketing forms a current loop through the wire-rope, with the conductive seawater as a return path. By coupling signals to this loop data can be transmitted up or down the wire between instruments and the surface buoy, or between instruments. This is the most common communication technique for retrieving data from submerged instrumentation and is

dramatically less expensive than including a dedicated copper or optical path in the mooring cable.

Synthetic Ropes

Although durable, wire-rope sometimes cannot be used because of its weight (e.g., the self-weight of 5000 m of 3X19 wire-rope introduces a load at the top, which is 80% of its working load), or because the mooring design requires a more elastic element. In these cases synthetic ropes are used. There are a wide variety of material options and construction types, typically characterized into two groups, high-performance synthetic fibers, and industrial fibers. High-performance fibers are defined as having a tenacity of greater than 15 g/denier [40.14], with strengths similar to steel wires, much less weight, and slightly lower modulus. Industrial fibers typically have about half the strength of the high-performance fibers. The strength penalty manifests itself in larger sizes that increases drag, but the higher elasticity and lower cost of these fibers mean they are widely used in mooring systems.

Table 40.1 highlights the material properties of the various fibers described below. Table 40.2 includes the properties of typical ropes manufactured from these materials.

High-Performance Synthetic Fibers.

Kevlar: Kevlar fiber was introduced in 1970 by Dupont and was the first high-performance fiber with a strength close to steel. It is only a little heavier than water with a specific gravity of 1.44, its stretch at break is about 4%. Creep of this fiber is very low which makes it useful for electro-optical-mechanical cables, but it is easily damaged by self-abrasion in repetitive cycling applications such as mooring cables. A specific failure mode known as kink-band formation dramatically weakens the fiber and occurs when the fiber undergoes compression. For this reason, it is a poor choice for surface mooring designs that include low-tension bending in which the inside of the bend allows the fibers to go into compression. Due to its relatively low cost, it is an appropriate choice for subsurface moorings that include electro-optical components and is often used in remotely operated vehicle (ROV) cables where the cable only rarely experiences significant loads.

Vectran: Vectran is a trade name for a liquid crystal polymer fiber introduced by Celanese in the mid 1980s. Its density, strength, and modulus are approximately equal to Kevlar, and the fiber has much less tendency to form kink-bands, making it well suited as a strength material in cables and ropes for surface moorings.

Spectra and Dyneema: These fibers are made from ultrahigh molecular weight polyethylene, and therefore have a density less than water and float (specific gravity = 0.98). With low weight, high strength, and high resistance to self-abrasion, this material is well suited to mooring applications in which a light, stiff strength element is needed. Traditionally the limitation of this fiber is a tendency to creep under sustained load, making it unsuitable for use in electro-optical-mechanical cables. Manufacturing improvements seem to be slowly eliminating this characteristic so this could change in the future.

Zylon: Dow Chemical developed this fiber in 1990 and marketed it under the trade name PBO. It is now manufactured and marketed as Zylon. It is an extremely strong fiber, about 50% stronger than Kevlar and Vectran and has a similar specific gravity of 1.53. Costs remain high, and it is therefore restricted to very specialized applications and has not yet found its way into mooring applications. If costs come down in the future, it is a good candidate for electro-optical-mechanical cables.

Industrial Fibers.

Nylon: Nylon was developed in the 1930s by DuPont and is a very stretchy fiber with elongations of 15–30% at break and has a low specific gravity (1.14). This light, inexpensive, elastic material is a good candidate for semi-taut mooring designs in deep water. Typical construction is a torque balanced 8 strand *Woods Hole* lay [40.11]. It does absorb water readily, which changes its length over long distances so care must be taken in specifying and fabricating moorings using this material.

Polyester: Polyester is the heaviest of the industrial fibers with a specific gravity of 1.38. Stretch at breaking loads is 12–18% but only 3–5% at working load. There is a tendency for some initial stretch at first loading so care is required in specifying and fabricating semi-taut moorings.

Polypropylene: Polypropylene has a low specific gravity (0.91) and floats readily in sea-water. This makes it a common choice for the lower section of an inverse-catenary mooring design. Because polypropylene has strength and stretch behavior similar to polyester, inverse-catenary moorings benefit from compliance created from both the geometry of the mooring and the material stretch under load. Polypropylene does exhibit some creep behavior so is inappropriate for use in center-strength-member electro-optical-mechanical cables, described below.

Electro-Optical-Mechanical Elements

In cases where electrical power needs to be transmitted along the mooring string, or if high data-rate communications along the mooring are required, it can be necessary to include copper or optical-fiber elements in the strength member. This introduces a number of requirements on the behavior of the strength member that affect the component's design and construction, as well as the mooring system's design as a whole. In particular, the total strain of copper wires and optical fibers must be kept very low, below 0.5% at the peak tensions experienced. Although this amount of strain is below the yield point of copper, it is not below the fatigue point, and after many cycles the copper strands will disintegrate. Optical fibers themselves, on the other hand, are fairly elastic, but cannot tolerate exposure to water for long-term use. The fibers are therefore typically encased in a very small diameter stainless-steel tube which acts as a one atmosphere pressure housing for the fiber. However, 0.5% is about the yield point of this tube and once yielded, relaxation of the cable causes the steel tube to kink and fail. There are a range of approaches to housing optical fibers for long-term at-sea use, but all rely on techniques that limit the allowable strain of the assembly.

There are two cable designs that address this low strain requirement. Traditionally, electromechanical cables are of a center conductor core. In this design contra-helical layers of the strength member (high-performance synthetic material or steel) are spiraled around a core consisting of the electrical and/or optical conductors (Fig. 40.4). The cable manufacturing industry has extensive experience with this construction technique from acoustic tow cables and remotely operated vehicle tethers. Unfortunately, this arrangement does little to alleviate the strain in the conductors relative to the strain in the entire cable. For mooring applications, the cables tend to become heavily overbuilt in terms of ultimate strength in order to limit the conductor strains.

An alternative approach is a center-strength-member construction in which the conductors are spiraled around a strength member (Fig. 40.5). In this design, the larger pitch diameter of the conductors provides an opportunity to introduce considerably less strain to the conductors relative to the overall strain of the cable. If the pitch of the spiraled conductors is matched to the diameter reduction of the center strength member under tension, low strains in the conductors can be realized in a cable that maintains high elasticity compared to center conductor core approaches [40.15]. While conceptually appealing, there have been challenges to implementing this type of design related to the terminations, in which a transition from high elasticity in the cable to zero elasticity in the termination must be managed. Although the designs aim for zero strain in the conductors, it is generally impossible to achieve this in practice, and even small amounts of stress in the outer helix of copper can compromise the torque balance of the cable. Compared to center-conductor core cables there is much less experience building this type of design in the cable industry, and some trial and error should be expected. Additionally, in this design approach the conductors are on the outside of the cable, in harm's way, and must be suitably protected during installation and maintenance.

For passing electrical conductors through short lengths that do not require compliance, techniques involving chains encased in polyurethane with the electrical conductors have proven reliable [40.16]. The high axial stiffness of the chain relieves all significant strain from the conductors, and the polyurethane stiffens the assembly to reduce strain due to bending. This technique is typically too heavy for extended lengths, but is useful for transition elements and below surface buoys.

Elastic Hoses and Cords

In cases where the traditional methods of providing mooring compliance are insufficient, specialty elements that stretch hundreds of percent at working loads can be included. There are two main applications of this. The first are very shallow moorings in which a chain-catenary mooring is not an option. The other are deep-water moorings that require a stiff electro-optical-

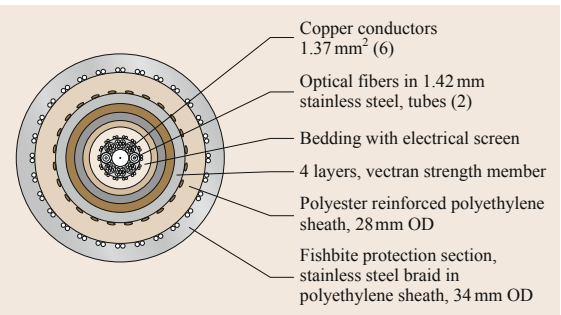


Fig. 40.4 Center conductor cable

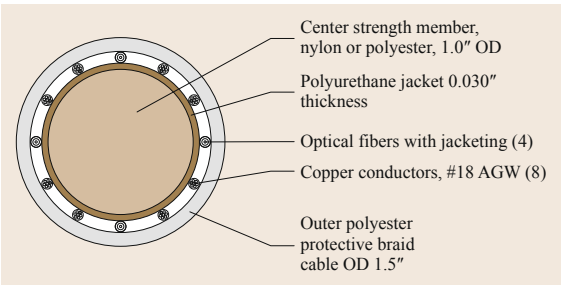


Fig. 40.5 Center strength member cable

Fig. 40.6 WHOI stretch hose shown on deck before deployment (photo courtesy of Woods Hole Oceanographic Institution (WHOI))

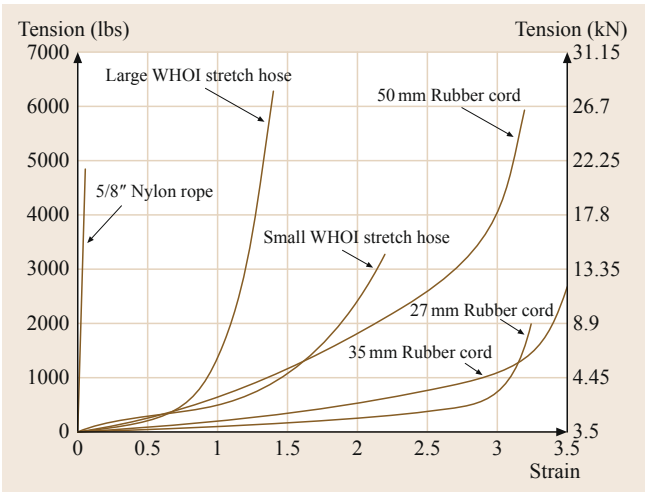


Fig. 40.7 Stress–strain behavior of various typical elastic elements, nylon rope shown for comparison

mechanical cable that does not supply enough compliance even in an inverse-catenary arrangement.

In cases where it is desired that subsurface instrumentation be isolated from surface-buoy-induced wave-motions, the high elasticity of these elements can perform this role provided that the subsurface instrumentation has sufficient mass to avoid being moved around. In the typical arrangement of buoy, stretch-hose, and then instrumentation, a trade-off appears between the tensions in the stretch hose and the amount of motion isolation achieved. This trade-off is dictated by the instrumentation mass: a very massive subsea element will be well isolated by the stretch-hose, but the tensions and elongation of the stretch hose will have a large amplitude.

The two main constructions of these kinds of elements are solid rubber cords with an external extension limiter and a more sophisticated reinforced rubber hose that includes nylon cords that gradually begin to take the load and reduce the elasticity of the element as the hose is stretched. In the latter case, it is possible to include electrical and optical conductors in the hose-wall layout at a neutral stretch angle, providing a durable power and data connection in an elastic component. This design has been used quite successfully. In addition to providing a conductor path, the durability of the rubber hose stands up well to the often difficult environment below the surface buoy [40.17] (Fig. 40.6).

Table 40.2 outlines typical mechanical properties of these stretch-hose elements. The elastic behavior of these components is far from linear however, and an estimate of the stress–strain curve needs to be included in any mooring modeling work that is performed. Figure 40.7 presents the stress–strain curve for a variety of typical elastic elements and shows the typically very soft behavior at low tensions and the gradual stiffening as the component is stretched. In all cases, these elements are designed to stiffen up as they extend and the stress strain curve becomes nearly vertical beyond what is shown. The WHOI stretch hoses accomplish this by including nylon cords in the rubber hose lay-up that bear more and more load at higher extensions. The rubber cords shown in [40.18] include an external safety line that takes over at extensions greater than about $\Delta L/L = 3$.

40.2.2 Terminations and Strain Relief

Terminations of wire-rope are typically a swaged type in which a metal socket is deformed onto and around the wire-rope using appropriate dies. A less common approach is a poured socket in which the wire-strands are spread out into a cone-shaped socket and a hard polyurethane resin is poured into the socket to lock the wires in place. In this approach, careful de-greasing and preparation of the wires and socket is required to develop the full strength, which can be nearly 100% of the breaking strength of the wire-rope. For synthetic ropes, splicing techniques specific to each material and rope construction are used. Termination of electro-optical-mechanical cables depends highly on the construction of the cable, center-conductor-core cables usually have the strength member arranged in contra-helical layers and in this case poured sockets are used similar to those used for wire-ropes. The center-strength-member cables that have been built have had a braided core which is terminated in a splice after the conductors are peeled off the outside. In these designs, it can be challenging to manage the transition from the elastic portion of the cable to the necessarily rigid termination.

In all cases, bending strain reliefs are very important elements of mooring systems. Abrupt changes in material properties along the mooring string can lead to ex-

acerbated motions and repetitive strains. Polyurethane bending strain reliefs can be designed to maintain a minimum bend radius cross the range of expected bending moments. Standard and proven solutions exist for common wire-rope terminations, but are typically engineered specifically for custom EOM cables. Where shackles and chains are included, the requirements on the bending strain reliefs are less stringent because these elements are incapable of supporting significant moments.

40.2.3 Buoys and Floats

Surface Buoys

For surface buoys, there are two main construction techniques: a closed-cell foam construction to which required towers, bridles, and instrumentation are attached; and a hull-like construction of steel, aluminum, fiberglass, or molded polyethylene. A common foam material is a foamed ionomer resin that is a thermoplastic polymer produced by DuPont [40.19]. This material is a closed-cell foam with a density of less than 10% that of water. Although very tough and durable, this material does have some compressibility, so it is only useful on or near the surface. Tests of this material's compressibility and water absorption have been done, and a relatively high variability due to construction technique was found [40.20].

Hull-like constructions for surface buoys can provide significant interior volumes for instrumentation and payload but suffer catastrophic failure in the case of a hull breach. Nevertheless, many buoys have been built and deployed in this manner; with suitable coatings, maintenance, and corrosion protection, steel and aluminum buoy hulls have been very reliable but expensive to maintain. Fiberglass hulls are less common, but provide higher performance from the point of view of weight-to-buoyancy ratio. Hybrid approaches have also been pursued with success for several decades in which lightweight nonstructural expanding foam is poured into a hull cavity.

There are few standards in buoy design aside from the relatively few organizations that deploy and maintain many buoys such as weather service buoys and navigation buoys. Typically, buoy designs are customized for specific applications and designer preferences. Critical parameters that must be considered in all designs such as stability, reserve buoyancy, and resonant behavior are discussed below.

Subsurface Flotation

For subsurface flotation such as that used to support an inverse-catenary mooring design, or provide recovery buoyancy for the anchor end of a mooring string, the critical parameters are material mean density and depth rating. The two primary types include pressure vessels that maintain a low-pressure interior volume and syntactic foam materials that typically consist of microspheres as flotation, bound together with an epoxy or polyester resin. For flotation purposes, evacuated volumes are almost always spherical in nature because of the efficiency of that shape as a pressure vessel. Materials used range from plastics for shallow-water applications, steel for moderate-depth applications, and glass-spheres for very deep applications. It should be remembered that in all cases, a submerged evacuated volume can represent an enormous amount of potential energy that can be released in an implosive manner in the case of failure. In one case of an imploding carbon-fiber electronics housing on an ROV at a depth of 1000 m, nearly all of the surrounding aluminum pressure vessels were deformed to the point at which the O-ring mating surfaces were out of specification and the housings needed to be replaced. For this reason, manned submersibles typically will not operate around this type of flotation, and this needs to be considered in mooring applications that include servicing by ROV or manned submersible.

An alternative to evacuated volumes for subsurface flotation is syntactic foam. This material is typically lower performance (higher density at a given depth rating) than evacuated glass spheres but can be engineered to the exact depth requirement of the system, can be formed or machined into just about any shape, and does not present an implosion hazard.

Table 40.3 compares the characteristics of a range of typical flotation used in small mooring systems.

40.2.4 Hardware

Hardware in small mooring systems usually consists of galvanized low-carbon anchor-bolt shackles, pearlinks, and rings. Table 40.4 contains a summary of typical mechanical properties for these components. Stainless steel is often used when low magnetic influence is required, but stress and crevice corrosion issues need to be considered. Fatigue issues are also important in specifying hardware components. For steel, the fatigue limit is about 50% of the ultimate tensile strength; if all stresses in a component are kept below this level, essentially infinite life is achieved. Manufacturers of shackles and similar components often specify a breaking strength and a safe working load. When the component is intended for overhead lifting, this safe working load includes a five times safety factor, putting the stresses in the component well below the fatigue limit. So adhering to this type of working load specification is one way to address fatigue. For custom elements such as instrument cages that also act as a load path,

Table 40.3 Properties of common flotation used in mooring systems (after [40.21–24])

Description	Diamete [m]	Mass [kg]	Mean density [kg/m^3]	Wet weight [N]	Depth rating [m]
Glass spheres					
12″ Sphere	0.3048	6.0	404.7	−98.0	7000
13″ Sphere	0.3302	9.1	482.7	−102.0	9000
17″ Sphere	0.4318	17.7	419.9	−249.2	6700
Steel spheres					
30″ Sphere	0.76	75	326	−15.7	380
37″ Sphere	0.94	137	315	−30.3	302
40.5″ Sphere	1.03	186	325	−39.3	410
49″ Sphere	1.24	488	489	−52.5	1075
Plastic floats					
10″ (25.4 cm)	0.254	3.63	423	−48.3	1800
9″ (23.0 cm)	0.23	2.31	363	−39.2	1800
6″ (16.0 cm)	0.16	0.39	182	−15.4	400
5″ (12.5 cm)	0.125	0.275	269	−6.2	600

Table 40.4 Properties of common mooring components (after [40.25])

Component	Inside length [m]	Mass [kg]	Wet weight in seawater [N]	Breaking strength [N]
Anchor shackles (galv. steel)				
3/8″	0.0366	0.15	1.3	53 400
1/2″	0.0478	0.36	3.1	106 800
5/8″	0.0605	0.62	5.3	173 550
3/4″	0.0715	1.23	10.5	253 650
1″	0.0955	2.28	19.5	453 900
Pear links (galv. steel)				
3/8″	0.0572	0.1	0.9	43 790
1/2″	0.0762	0.25	2.1	70 490
5/8″	0.0955	0.48	4.1	101 990
3/4″	0.114	0.85	7.3	145 250

more careful fatigue engineering should be performed with attention to stress concentrations and weld heat affected zones. In cases where weight is critical, heat treating, shot peening, and other surface treatments may be appropriate.

40.2.5 Anchors

Dead-weight drag embedment anchors are the most common types of anchor systems used in small buoy and mooring systems. In all cases the nature of the mooring design and conditions on the seafloor affect the anchor selection.

Deadweight anchors rely on their own weight to stay in contact with the seafloor, and the resulting friction to resist horizontal motions. Used train wheels are a popular source of steel anchor material, and when stacked and welded assume a mushroom shape which can significantly increase the anchors' ability to resist movement when the mushroom shape becomes buried. Purpose built steel and concrete pyramid anchors are also available and behave similarly. Key features to dig into the seafloor are sometimes included to increase a deadweight anchors capacity in cases where it is unlikely to become buried. The amount of lateral resistance a given anchor will provide depends highly on the nature of the seafloor. When this is unknown as is often the case for buoys deployed in remote locations, it is necessary to include significant factors of safety in determining the anchor weight.

Chain-catenary moorings with sufficient chain on the seafloor can often use embedment anchors of moderate size. Because these moorings are almost always single-point moorings, the nature of the weather and current conditions affect the suitability of embedment anchors. Changing load directions can pull the anchor from the seafloor. However, the light weight of this type of anchor makes them attractive for ease of deployment and recovery.

The numerical mooring analysis techniques described below in Sect. 40.3.4 usually result in a time history of the anchor loads and direction in various sea conditions. This is a useful information in sizing any anchor system and is a key output of most analyses.

Table 40.5 Properties of common acoustic release devices used in mooring systems (after [40.26–28])

Release	Length [m]	Mass [kg]	Wet weight in seawater [N]	Release load rating [N]	Lifting load rating [N]	Depth rating [m]
Benthos						
875-TD	0.53	4.5	14	1766	4464	500
867-A	0.63	9	36	2207	4464	305
866-A	0.66	18	128	21 582	21 582	2000
865A	0.96	33	245	49 050	49 050	12 000
865-H	1.26	154	1265	221 706	260 946	12 000
Sonardyne						
ORT (Type 7409)	0.70	22	167	16 677	12 508	2000
DORT (Type 7710)	0.70	22	167	16 677	12 508	6000
EdgeTech						
SPORT MFE	0.63	4.2	9	2453	6377	400
SPORT LF	0.72	4.72	13	2453	6377	400
PORT MFE	0.69	9.1	35	2453	9810	3500
PORT LF-SD	0.72	7.2	25	2453	9810	3500
PORT LF	0.66	11.3	47	3434	9810	3500
8242 XS	0.95	36	275	53 955	73 575	6000

More and more attention is being paid to the effect on the seafloor of mooring systems and many marine sanctuaries require trawl-resistant anchor designs for anchors that are left behind, or are mandating complete recovery of anchors which introduces additional design challenges.

40.2.6 Acoustic Releases

An important component in many mooring systems is an acoustically activated release located near the anchor. This device allows the mooring string to be released from the anchor for recovery, significantly reducing the lifting capacity required on the vessel. This technique leaves the anchor behind on the seafloor. In this scenario, recovery buoyancy above the acoustic release is required, which can unfortunately impact the anchor size requirements since this buoyancy works directly against the weight of the anchor during service. In some cases, an acoustically triggered line-pack with flotation is also included in the system, allowing the anchor to be recovered separately from the mooring string. Table 40.5 includes the properties and specifications of a number of commonly used acoustic release systems.

To reduce the ships lifting capacity requirements, typically moorings with acoustic releases are deployed using an *anchor-last* method in which the anchor is pushed off the ship's deck and allowed to free-fall to the seafloor after the buoy and mooring string have been placed in the water. The load on the mooring string is necessarily less than the anchor weight during this operation. The rapid deceleration of the anchor when it hits the seafloor raises the possibility that an acoustic release or other component just above the anchor could hit the anchor and be damaged. In practice, the mooring string during an anchor-last deployment has a high drag-to-inertia ratio and comes to a rest very quickly after the anchor touches down.

40.3 Analysis Techniques

Once the buoy size and shape, installation depth, mooring type, and environmental conditions for a buoy system have been identified, analysis of the system's behavior can proceed. Typically three levels of complexity are considered. The first is the system's static response to wind and current loading, and buoy pitch/roll stability. This analysis of the buoy and proposed mooring design helps determine if the system has sufficient buoyancy and stability for operation.

The second is a frequency domain analysis of the buoy's response to wave-forcing. The dynamics of the freely floating buoy are much simpler to analyze than the coupled buoy/mooring system, and some important insights can be gained. Although the buoy dynamics are modified by the existence of the mooring line, the buoy-only analysis provides a good general guideline to the buoy's behavior and a starting point for analyzing the more complicated buoy/mooring system.

This dynamic behavior is important to consider in the design so that the final system responds safely to wave forcing and does not exhibit resonance behaviors that could cause damage. Linearized frequency domain analysis provides an estimate of these behaviors but becomes progressively less useful in deep-water systems in which the behaviors of the mooring system begin to dominate the buoy dynamics.

Finally, nonlinear analysis of the coupled mooring–buoy system provides the most accurate prediction of system behavior in any given wind, wave, and current condition. These systems are nonlinear, however, so it is required that this analysis be performed through numerical solution of the governing differential equations. And as a nonlinear system, each environmental condition must be examined individually. This can be time-consuming and expensive to cover the entire range of possible conditions a system may see in service. For instance, a system may behave very differently in large waves and low currents, versus the same large waves in high currents. As such, a given mooring system analysis typically includes a matrix of analysis in low-, medium-, and high-current environments and in moderate, rough, and severe wave conditions. Typically it is assumed that wind forcing and wave forcing are correlated. Even this matrix of nine conditions still leaves a lot of ground unexamined; for instance, in some locations, currents run against the prevailing storm conditions at some times, and with the waves at other times. Therefore, knowledge of the deployment location or the mooring system's sensitivities impacts which conditions are analyzed.

A range of numerical tools have been developed to do this and are described below. Both open source and commercially produced solutions are available. These tools should be applied whenever a buoy and mooring design strays too far from existing proven designs, or when the design requirements of the system require that safety margins in stability and peak loads be reduced. Although it is rare to be able to validate the results of these tools with field measurements, this has been done occasionally [40.29, 30] and decent agreement has been found. In any case numerical analysis can provide good comparative analysis between variations on a given design, and sensitivity analyses are often performed to optimize a given design.

In all cases, analysis provides guidance on the loads that a particular system will experience in service. The factor of safety that is then applied to these estimates for design purposes is highly variable and depends upon the confidence the designer has in the model inputs (environmental conditions), the fatigue strength characteristics of the materials used, the need to reuse components for subsequent deployments, the planned lifetime of the system, the amount of control the designer has over the operation and maintenance of the system, and the tolerance for additional weight and cost in the system.

40.3.1 Environmental Conditions

Characterization of the ocean environment is covered in detail in Chap. 35 and those techniques are used directly in the buoy/mooring analysis described below. In general, three environmental factors are important: wind, waves, and currents. The wind climatology at any given spot in the ocean is probably the most well characterized due to the existence of scatterometer satellite measurements that have been ongoing for several decades [40.31, 32]. The wave environment also needs to be known; ideally the wave climatology can be estimated from time-series measurements of nearby wave-measurement buoys. This requires that such a system is in place, and that a sufficient time history is present to reliably estimate the return period of severe conditions. When these measurements are not available, this data can be estimated from global wave models such as WAVEWATCH III [40.33]. Commercial companies exist that will aggregate available data and provide the expected wave climatology at a particular location, which can be very convenient. Current behaviors at a particular location are the most problematic in that little data typically exists, and the current profile can have significant impact on a mooring system's behavior. Time histories of global surface current measurements made by satellites over many years are available [40.34], but these measurements are limited to the movement of the surface water. Knowledge of the subsurface current profile typically requires direct measurements using Acoustic Doppler Current Profiling (ADCP) technology or discrete current sensors [40.35] and is often not available. In the worst case, the designer of a system must make assumptions about the current profile throughout the water column and it is typical to analyze a range of possibilities; a *slab* current condition in which the entire water column is moving at the same speed, this is only likely to occur in practice in shallow conditions. Other possibilities are a linearly decaying current from the surface to the bottom, or to some depth in the water column. Uncertainty in the current profile magnitude and shape is often the largest uncertainty that a mooring designer faces.

40.3.2 Static Analysis

There are two main elements of the static analysis of a mooring system. The first is the initial stability of the surface buoy itself in which the weight and bal-

ance of the system is considered and the stability of the system is examined. These characteristics can be computed analytically from knowledge of the shape and mass distribution of the buoy. The second is the mooring strings behavior in a range of constant wind and current loading and this must be computed numerically in most cases. This result includes a computation of the mooring strings position and tensions, along with the current and wind loading that would cause the buoy to submerge.

Stability

For buoys that are always on the sea surface, stability in heave is provided by the presence of the water plane. Deeper submergence of the buoy creates an increase in buoyancy so the buoy will rest at a stable equilibrium at all times. The static draft of a freely floating buoy is found by simply equating the weight of the buoy to the buoyancy force, which increases as the buoy is further submerged. Additional vertical forces such as the mooring-line tension will change the draft. Buoy systems are typically designed with a buoyancy reserve of 100% or more; that is, the force required to submerge the buoy entirely is greater than twice the weight of the buoy plus the minimum static mooring load. The amount of reserve buoyancy implemented in a system depends on the mooring type, and expected variation in static loads (winds and currents), peak dynamic loading, and the consequences of having the buoy submerge in extreme conditions.

It is useful to compute the force required to increase the buoy draft. This parameter gives an indication of buoy sensitivity to added payload or weight gain due to bio-fouling. For a cylindrical wall-sided buoy this is

$$\frac{\delta F}{\delta d} = A_{\mathrm{wp}}\rho \mathrm{g} = \frac{\rho \mathrm{g}\pi D^2}{4} \qquad \left(\frac{F}{L}\right) . \tag{40.1}$$

For pitch and roll stability, the buoy is characterized by the well-known metacentric height which includes the righting moment contribution from the water-plane area as a modification of the distance between the center of gravity height above the keel $\overline{\mathrm{KG}}$ and the center of buoyancy height above the keel $\overline{\mathrm{KB}}$. See Fig. 40.8. It is the contribution of the water plane that allows floating objects to remain right-side up in spite of having a center of gravity above the waterline. The metacentric height $\overline{\mathrm{GM}}$ is computed as follows for the case of a freely floating buoy without mooring-line tension,

$$\overline{\mathrm{GM}} = \overline{\mathrm{KB}} - \overline{\mathrm{KG}} + \frac{I_{\mathrm{wp}}}{\nabla} , \tag{40.2}$$

where I_{wp} is the second moment of area of the buoy water plane and ∇ is the buoy displacement. When $\overline{\mathrm{GM}}$ is negative, the system will not be stable in pitch/roll and will heel or capsize in even the calmest seas. For wall-sided buoys of cylindrical geometry, this expression becomes

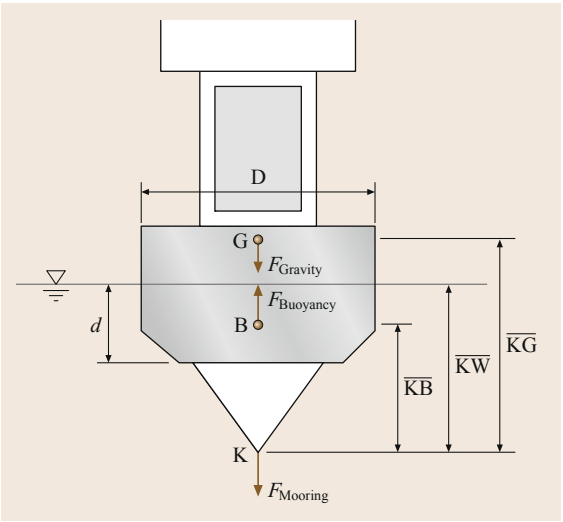


Fig. 40.8 Buoy schematic showing important dimensions for stability calculations

$$\overline{\mathrm{GM}} = \frac{d}{2} + \frac{D^2}{16d} - \overline{\mathrm{KG}} , \tag{40.3}$$

where d and D is the buoy draft and diameter, respectively.

A first-order correction to this expression can be made that incorporates the effect of the mooring-line tension by assuming that the mooring line exerts a vertical force at the keel of the buoy. For this purpose the magnitude of the mooring-line tension can be selected as the weight of the mooring string supported by the buoy. This is the minimum average tension the mooring line can exert. Any current or wind loading increases the tension from there (and introduces a lateral force and heeling moment to the buoy). The inclusion of this force (T) introduces a righting moment to the buoy and increases $\overline{\mathrm{KB}}$ but decreases the righting moment contribution of the water-plane area by increasing the buoy displacement. For wall-sided buoys, the modified meta-centric height expression in terms of $\overline{\mathrm{GM}}$ from the freely floating case is

$$\overline{\mathrm{GM}}' = \overline{\mathrm{KB}} + \frac{\delta\nabla}{2A_{\mathrm{wp}}} + \frac{I_{\mathrm{wp}} - \overline{\mathrm{KG}}\nabla}{\nabla + \delta\nabla} , \tag{40.4}$$

where A_{wp} is the buoy water-plane area and $\delta\nabla$ is the displacement increase due to the mooring-line tension ($\delta\nabla = T/\rho \mathrm{g}$).

This expression can be expanded for small values of $\delta\nabla/\nabla$ to provide insight into how the static mooring-line tensions will affect the buoy stability,

$$\overline{\mathrm{GM}}' \approx \overline{\mathrm{GM}} + \frac{\delta\nabla}{2A_{\mathrm{wp}}} + \frac{\delta\nabla}{\nabla}\left[\overline{\mathrm{KG}} - \frac{I_{\mathrm{wp}}}{\nabla}\right] . \quad (40.5)$$

From this it is clear that the mooring-line tensions have a larger stabilizing effect when the distance from the attachment point is far from the center of gravity (large $\overline{\mathrm{KG}}$), and when the buoy has a small waterplane area and waterplane moment of area. In other words, mooring-line tensions provide a relatively larger stabilizing effect to intermediate aspect ratio buoys than to discus buoys.

Mooring-Static Tensions and Draft

Prediction of the mean position (draft and heel) of a buoy based on current and wind loading is a nonlinear problem and so is less straightforward. Analytical solutions for simplified catenary situations exist but typically numerical solutions are used to predict this behavior. The numerical tools described below include facilities to compute the buoy systems response to constant wind and current loading, and this is usually the first step of the dynamic analysis to provide an initial condition for the dynamic simulation. The numerical simulations can also include point loading in the mooring string and other discontinuities in the material properties and loadings, which is typically not possible in analytical solutions.

40.3.3 Frequency Domain Analysis

The dynamics of the buoy itself can be predicted by equating the sum of forces acting on the buoy to the product of the acceleration and inertia. The simplest case of this analysis relies on assumptions that all forces vary linearly with wave-height and period, which is generally true for long-period waves. The appropriateness of this assumption deteriorates as the wavelengths get shorter (and collapses completely when the waves begin to break), but useful insights from this type of analysis are still possible. Again the two relevant modes of motion are the heave behavior and the pitch motions. This section presents an analysis of the buoys linearized response in the frequency domain and is most relevant to the freely floating buoy case. Non-linear numerical analysis techniques that predict the behavior of the coupled buoy–mooring system are described in the next section.

Response Amplitude

Figure 40.1 shows response amplitude operators for a range of mooring aspect ratios in the freely floating condition. These are examples of typical small moorings and highlight the important feature of resonant behavior due to the interaction of the buoys inertia and hydrostatic restoring force. For high-aspect ratio buoys the heave resonance can be dramatic; for lower aspect ratios this resonance is less pronounced but does mark the frequency at which the buoy motions transition from following the sea-surface to not reacting to the wave excitation. This section outlines how these response curves are generated and provides tools to compute the freely floating response of simple buoy shapes.

For linearized frequency domain analysis, the fundamental assumption is that all motions and forces oscillate at the same frequency as the incoming waves, and that a system's behavior in the presence of many wave frequencies is the sum of its behavior in each individual frequency. Under these assumptions, the differential equations of motion become algebraic equations that are simply solved. To accomplish this, each varying quantity is represented by a complex amplitude which represents the amplitude and phase of the sinusoidal motion. For instance, the complex response amplitude in heave is

$$\begin{aligned} z_3(\omega, t) &= Re\left[Z_3 \mathrm{e}^{\mathrm{i}\omega t}\right] \\ &= Re[Z_3]\cos(\omega t) - Im[Z_3]\sin(\omega t) \\ &= |Z_3|\cos\left(\omega t + \arctan\left(\frac{Im[Z_3]}{Re[Z_3]}\right)\right) . \end{aligned} \quad (40.6)$$

For heave, the solution of the equations of motion is

$$Z_3(\omega) = \frac{X_3(\omega)}{-\omega^2(a_{33}(\omega) + M) + \mathrm{i}\omega b_{33}(\omega) + \rho g A_{\mathrm{wp}}} , \quad (40.7)$$

where X_3 is the complex amplitude of the wave exciting force acting on the buoy, a_{33} is the heave added mass, b_{33} is the heave wave damping, M is the buoys mass, and $\rho g A_{\mathrm{wp}}$ is the buoyancy restoring force, written for cylindrical wall-sided buoys in (40.1). The hydrodynamic coefficients (X_3, a_{33}, and b_{33}) depend upon the geometry of the submerged part of the buoy and are functions of ω, while $\rho g A_{\mathrm{wp}}$ and M are constants for a given buoy and independent of frequency [40.36].

For pitch motions, the mass distribution within the buoy becomes important and is incorporated into the solution for the pitch dynamics in the form of the moment of inertia of the buoy in pitch I_5,

$$Z_5(\omega) = \frac{X_5(\omega)}{-\omega^2(a_{55}(\omega) + I_5) + \mathrm{i}\omega b_{55}(\omega) + mg\overline{\overline{\mathrm{GM}}}} . \quad (40.8)$$

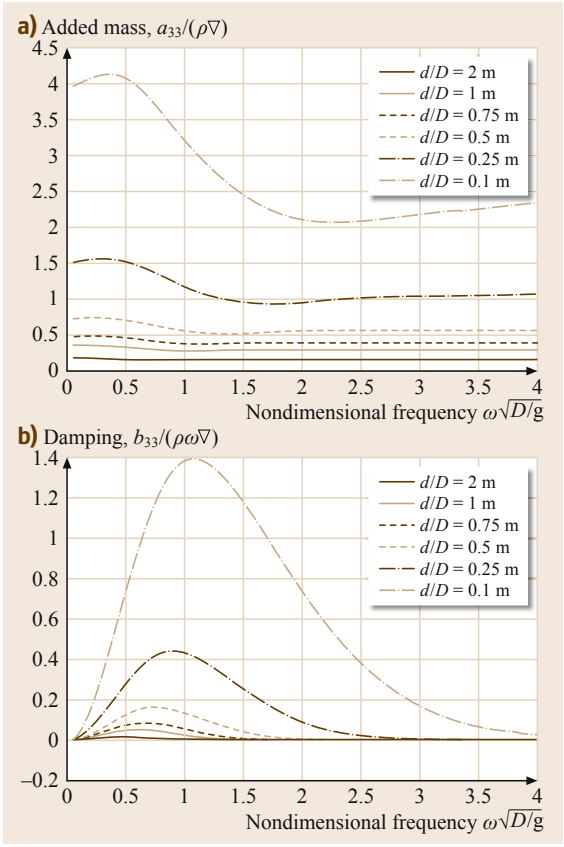


Fig. 40.9a,b Nondimensionalized heave hydrodynamics for cylindrical buoys

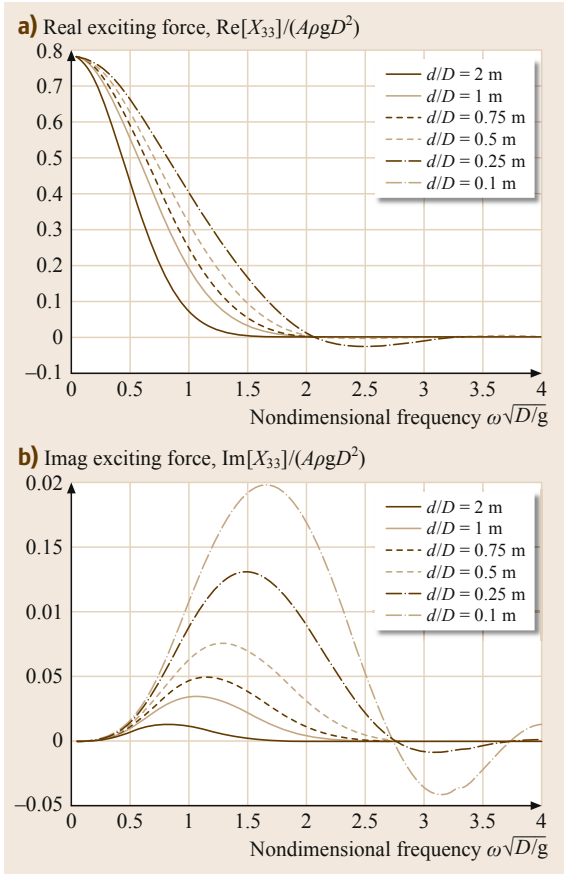


Fig. 40.10a,b Nondimensionalized heave wave-exciting forces for cylindrical buoys

For pitch motions, the hydrostatic restoring force is characterized by the metacentric height ($\overline{\mathrm{GM}}$).

If the hydrodynamic coefficients of the buoy are known, along with the mass and size of the buoy, the motions can be fully characterized using (40.7) and (40.8). The response amplitude operator is the magnitude of Z_3 per unit amplitude of the incoming waves A. For pitch, the RAO is typically defined as the magnitude of the pitch angle relative to the slope of the incoming waves ($\frac{|Z_5|}{A\omega^2}$).

Computing the hydrodynamic coefficients can be done analytically for simple shapes [40.37, 38] or numerically for more complicated shapes [40.39]. For cylindrical buoy shapes of various aspect ratios, the necessary hydrodynamic coefficients are presented in Figs. 40.9–40.12. In order to reduce the number of coefficients presented, frequency has been nondimensionalized by the buoy diameter and gravity ($\sqrt{D/g}$). For convenience, Fig. 40.13 can be used to relate nondimensional frequency to wave period for a range of buoy sizes.

These results are valid for freely floating bodies, but a first-order correction that includes the mean mooring tension is simple to generate by using hydrodynamic coefficients that represent the resulting deeper draft. For the pitch result the metacentric height also needs to be replaced with the modified metacentric height $\overline{\mathrm{GM}}'$ from (40.5). Note that in applying this correction, the buoy mass M and moment of inertia I are unchanged from the freely floating case, even though the buoy has a larger displacement.

Resonance

Resonance in heave and pitch occur when the denominator of (40.7) and (40.8) are at a minimum, for example,

$$\omega_{\mathrm{n}}^{\mathrm{H}} = \sqrt{\frac{\rho g A_{\mathrm{wp}}}{(M + a_{33})}} \qquad (40.9)$$

and,

$$\omega_{\mathrm{n}}^{\mathrm{P}} = \sqrt{\frac{mg\overline{\mathrm{GM}}}{(I_{55} + a_{55})}}\,. \qquad (40.10)$$

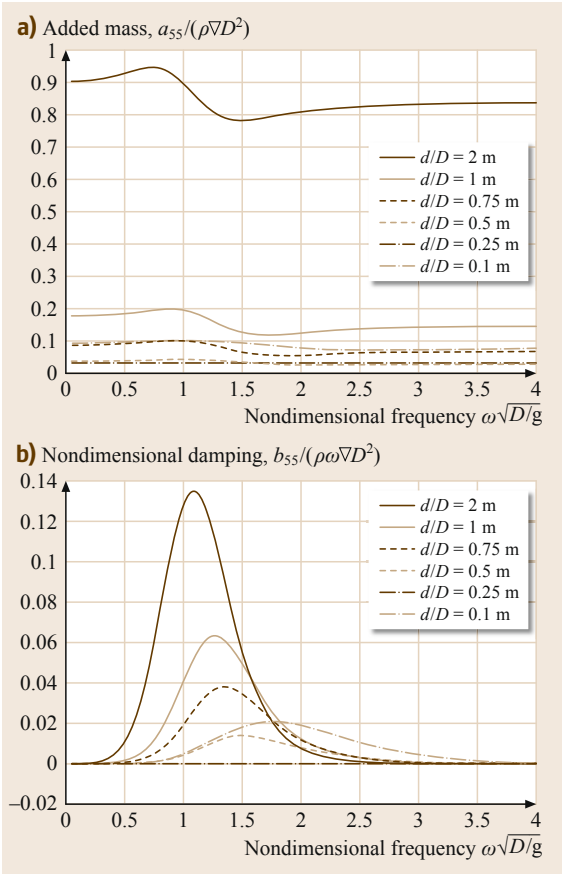


Fig. 40.11a,b Nondimensionalized pitch hydrodynamics for cylindrical buoys

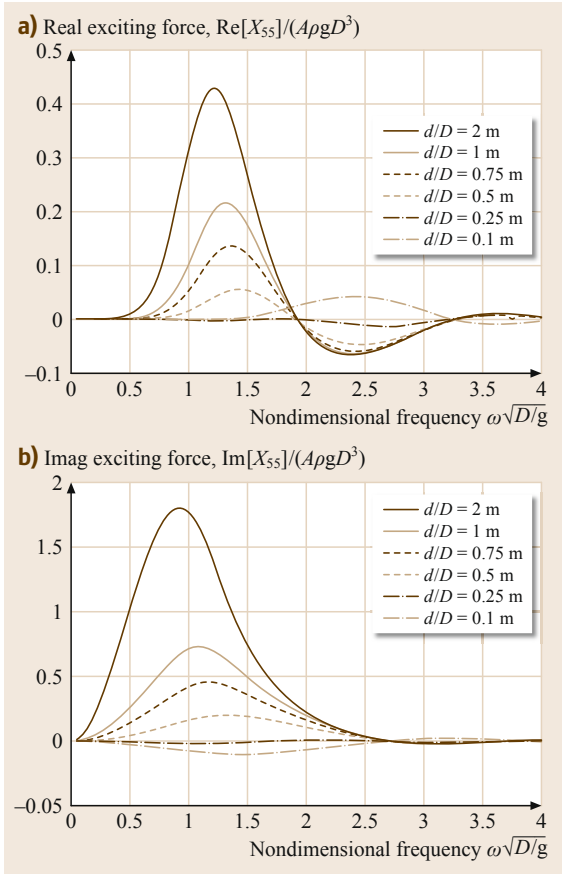


Fig. 40.12a,b Nondimensionalized pitch wave-exciting forces for cylindrical buoys

The magnitude of the resonance depends on the wave-exciting force and amount of damping at the resonance frequency.

$$\left|Z_3(\omega_n^H)\right| = \left|\frac{AX_3}{\omega_n^H b_{33}(\omega_n^H)}\right| , \tag{40.11}$$

$$\left|Z_5(\omega_n^P)\right| = \left|\frac{AX_5}{\omega_n^P b_{55}(\omega_n^P)}\right| . \tag{40.12}$$

The expressions for the natural frequency are implicit in frequency and depend on the hydrodynamic coefficients' dependence on frequency. Figure 40.14 presents the heave resonance frequency and amplitude for cylindrical buoy shapes of various aspect ratios, computed based on the hydrodynamic coefficients presented in Figs. 40.9 and 40.10.

40.3.4 Nonlinear Analysis

Although frequency-domain analysis provides some insight into the behavior of small buoy systems, the presence of the mooring generally cannot be ignored when predicting system behaviors in specific conditions. Figure 40.15 illustrates the system characteristics

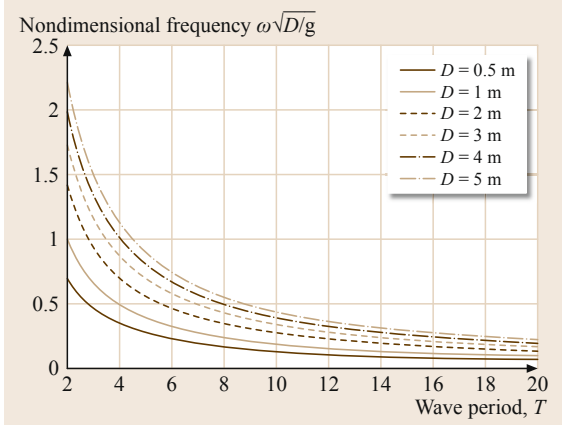


Fig. 40.13 Nondimensionalization wave frequency for different size buoy diameters

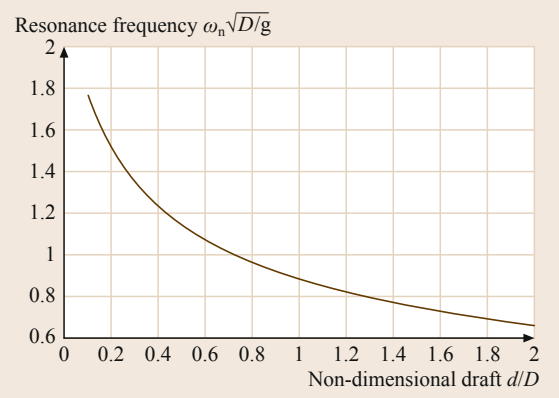


Fig. 40.14 Heave resonance-frequency of freely floating cylindrical buoys

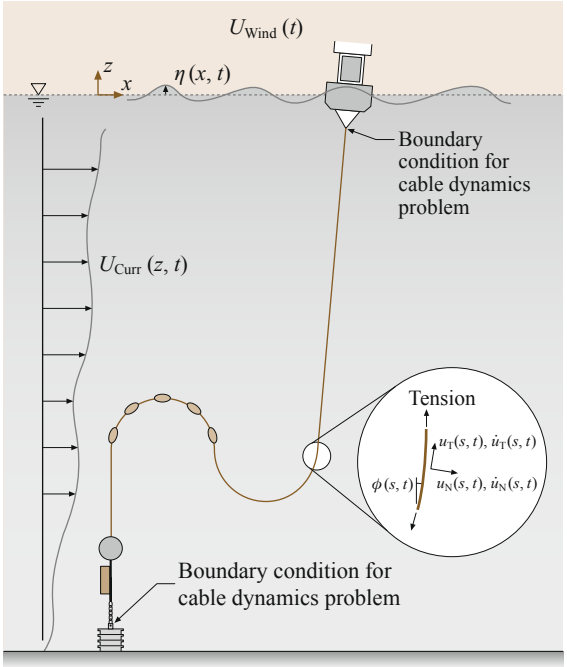


Fig. 40.15 Schematic of nonlinear mooring dynamics problem

and forcing inputs that are important. The three main environmental inputs are wind and wave forcing on the buoy and current forcing on the mooring. As in the frequency-domain analysis, the buoy is represented by its size, mass, and mass distribution. Additionally, a relation between the wind speed and resulting forces on the buoy is required (surface area and drag coefficient). The mooring characteristics that are required for this type of analysis are the mooring cable mass per unit length, hydrodynamic behavior, and constitutive relation that relates stress and strain. The hydrodynamic behavior is usually characterized as drag (force proportional to velocity squared) and added-mass (force proportional to acceleration) of the cable in the normal and tangential directions. The constitutive relations for the material are the stress/strain curve axially, and the moment/curvature relationship in bending. These are most simply represented as linear relations (EA and EI for linear constitutive relations axially and in bending). In some cases, such as the stretch hoses described above linear representation is inappropriate and the nonlinear relation such as those plotted in Fig. 40.7 should be used. Because the solution is being performed numerically, these nonlinearities (and the discontinuity that results when transitioning from material to another) does not introduce any fundamental problem.

The equations of motion governing the behavior of the mooring result from an application of Newton's second law ($F = ma$) in which the external forces are balanced with the inertial forces for each part of the system. A continuous representation of the equations of motion can be derived by applying this concept to each infinitesimal element of the system [40.40]. For solution, a discretized version of these equations is generated by utilizing finite difference techniques. In addition to providing a technique for numerical solution of the underlying continuous equations of motion, this approach can also accommodate lumped parameter representations of some elements. For instance, a common strategy is to represent the strength members as a finite-difference representation of the cable-dynamic equations, and then include the effect of items attached to the cable by simply increasing the mass, drag, weight, and buoyancy at appropriate nodes, such as where flotation or instruments are attached. In all cases, the unknowns that are solved for are the time histories along the cable of strain, shear force, inclination to vertical, and curvature. From these unknowns all quantities of interest can be found such as tension in the cable from the stress/strain relationship, and mooring line location from an integration of inclination, curvature and strain.

In the general problem the three main forces acting on the buoy and mooring (wind, wave, and currents) are time varying, requiring boundary conditions at the cables ends, and an initial condition for the state of the entire mooring. The boundary conditions can be simply provided as known points for each end versus time (such as at the anchor), or by coupling to the solution of the buoy dynamics equations of motion which are solved simultaneously. The initial condition is the solution of a version of the equations of motion in which the time-dependent terms are set to zero, and the wind, wave, and current forcing are constant (the only constant wave-forcing that is physically plausible is calm water). In addition to providing an initial condition for

the time-dependent solution, the static solution is very useful in its own right and provides the designer important information about the mooring geometry, static loads, and buoy draft as a function of currents and winds.

The mass, size, and hydrodynamic characteristics of all elements in the mooring system must be known or estimated. For cables and ropes, drag and added-mass coefficients equivalent to an infinitely long cylinder are appropriate for the normal motion. For the tangential direction the drag and added mass are much lower, but can be estimated to reflect the degree of smoothness the particular element possesses. For other components such as acoustic releases and flotation, tabulated data for drag coefficients of various items can be used [40.41].

Numerical Solution

The equations of motion described above can be reduced to a coupled set of nonlinear first-order differential equations. Solution of these equations can be performed in a number of ways, and all approaches are subject to the stability and accuracy concerns related to numerical analysis of such systems [40.42]. In some cases, discontinuities in the cable properties can cause problems due to limited precision in the numerical solution, so the computer codes must have strategies for dealing with these through adaptive grid or time-step refinement.

A variety of computer codes have been written to solve these problems, two examples are WHOI-Cable [40.40] and Orcaflex [40.43]. Although these two codes solve the cable dynamics equations motion in a similar way, the treatment of the buoy boundary condition differs.

WHOI-Cable focuses on the dynamics of the cable itself and simplifies the buoy-end boundary condition by assuming that the buoy is a pure wave-follower in heave, but is free to move laterally on the sea surface subject to the forces acting upon the buoy (mooring loads, hydrodynamic forces, and wind loading). For buoys of intermediate to low aspect ratio that have sufficient reserve buoyancy to always remain on the surface, this is a good assumption and simplifies the analysis. The user does not need to specify the mass distribution in the buoy, or determine the wave hydrodynamic coefficients for the floating body. The program does compute the buoy draft based upon the hydrostatic restoring force for the static solution, but then simply assigns the buoys vertical velocity to be the same as the wave's vertical velocity for the dynamic analysis. For this reason, this approach does not give a good idea of the time-varying buoy draft that results from wave forcing. While simple, this limitation precludes the accurate analysis of buoys with significant wave-hydrodynamic forces, or those that are designed to become submerged in large waves.

Orcaflex, on the other hand, has the capability to solve the coupled buoy/cable equations of motion. This requires the computation of the wave-hydrodynamic coefficients for the particular surface buoy under consideration. WAMIT is the tool most commonly used for this, although for simple shapes data from the literature is available (or Figs. 40.9–40.12 could be used for cylindrical buoys). The inclusion of the buoy equations of motion in the solution removes the assumption that the buoy is a wave-follower and more complicated situations including submerging buoys or wave-energy extracting buoys can be considered. In extreme cases of submerging buoys, the assumptions inherent in linearized wave hydrodynamics also break down and other (often heuristic) assumptions about the buoys behavior must be made.

These techniques typically do not have the capability to resolve vortex-induced vibrations and the resulting increase in drag on a mooring line. To do so would require solving for the local flow field around the rope or cable in conjunction with the cable motions. Although examining the Strouhal number for a given cable diameter and current speed does give some idea of the expected vortex-shedding frequency, it is not straightforward to identify the natural frequency of the cable in question. Practically speaking, vortex-induced vibration and the resulting *strumming* are commonly observed in high current regimes or during anchor-last deployment operations in which the mooring string is moving laterally through the water. The best that can be done for analysis is to include a multiplier on the drag coefficient of the cables.

40.4 Example Designs

As mentioned in the introduction, small buoy mooring systems have a long history and existing successful designs can provide an excellent starting point for designing new systems. This section provides descriptions of several systems that employ the components and features described above.

The literature also contains many examples of successful mooring design. In particular, the US Office of Naval Research/Marine Technology Society has initiated and sponsored a bi-annual Buoy Workshop, starting in 1996, intended to facilitate the exchange of technical information about small buoy/mooring sys-

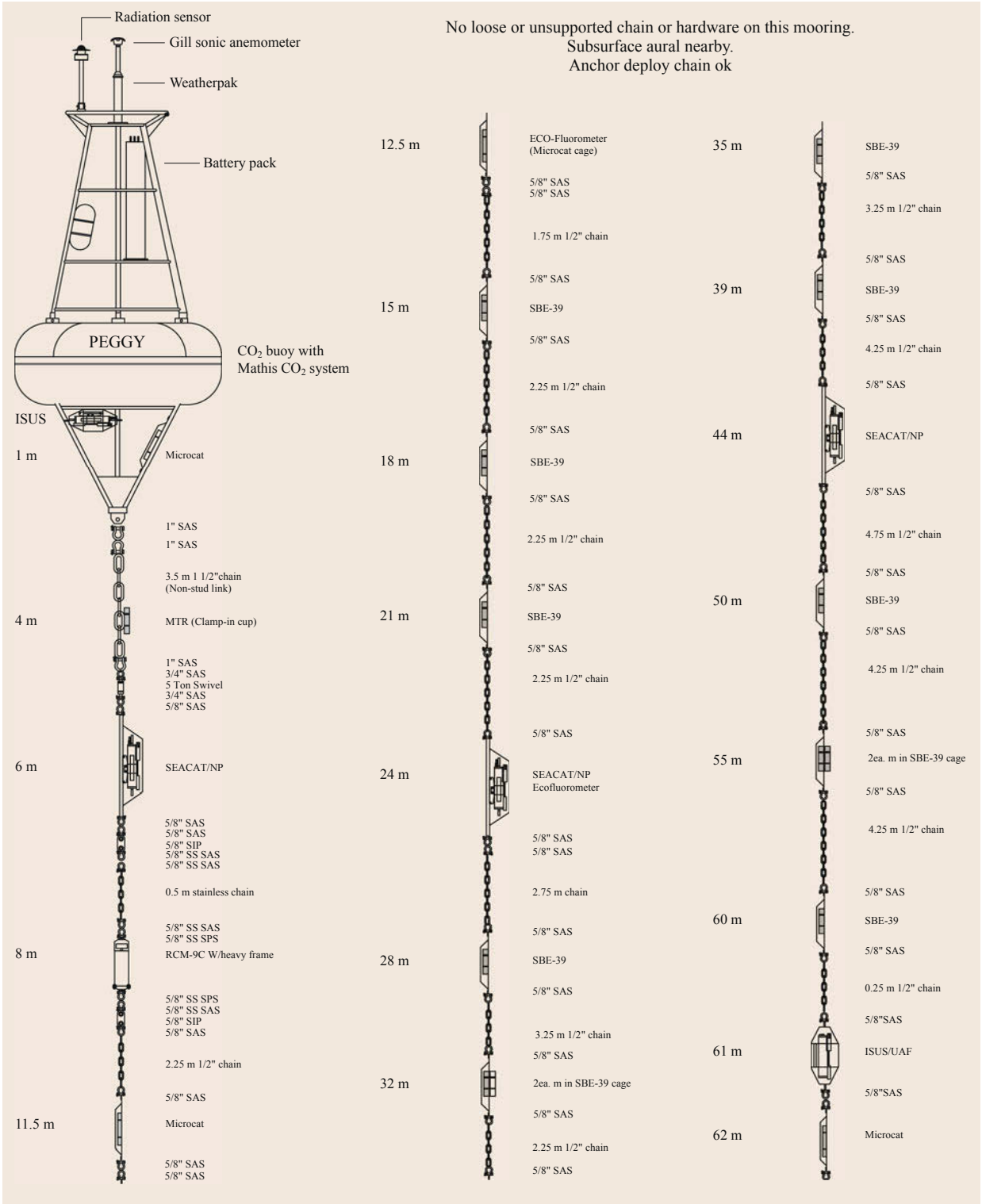
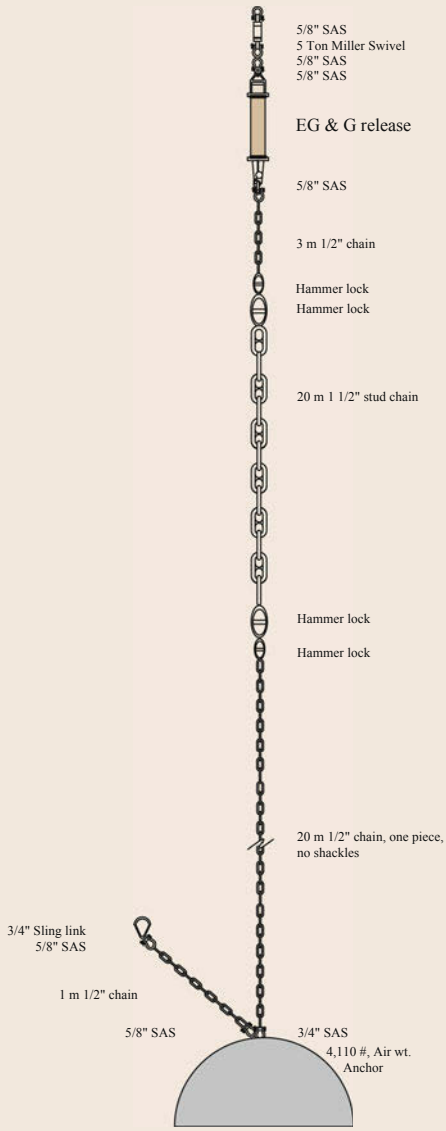


Fig. 40.16 Chain-catenary mooring with instrumentation throughout mooring string (diagram courtesy of NOAA-PMEL (Pacific Marine Environmental Laboratory))



Scope 1:1.5		
Surface instruments		
Inst	Ser #	
Weatherpak		
Gill sonic		
Linkquest mod		
Dcu iridium		
Eppley radiation		
Subsurface	Instruments	
Depth (m)	Inst.	Ser #
1	Microcat	
1	ISUS	
4	MTR	
6	SEACAT/NP-no.pump	
8	RCM-9/C	
12	Microcat	
12.5	Eco-Fluor	
15	SBE-39	
18	SBE-39	
21	SBE-39	
24	SEACAT/NP-pump	
24	Ecofluor	
28	SBE-39	
32	MTR	
32	MTR	
35	SBE-39	
39	SBE-39	
44	SEACAT/NP	
50	SBE-39	
55	MTR	
55	MTR	
60	SBE-39	
61	ISUS	UAF
62	Microcat	
	Ac.rel.	
72	Bottom	

NOAA-PMEL-FOCI
7600 Sandpoint Way NE
Seattle, Wa. 98115
(206) 526-6180

Mooring:	**14BSM-2A**
Location:	**Bering Sea**
Drawn by: Rick Miller	Date: 6 Dec. 2013
Approved by:	Date:

Fig. 40.16 (continued)

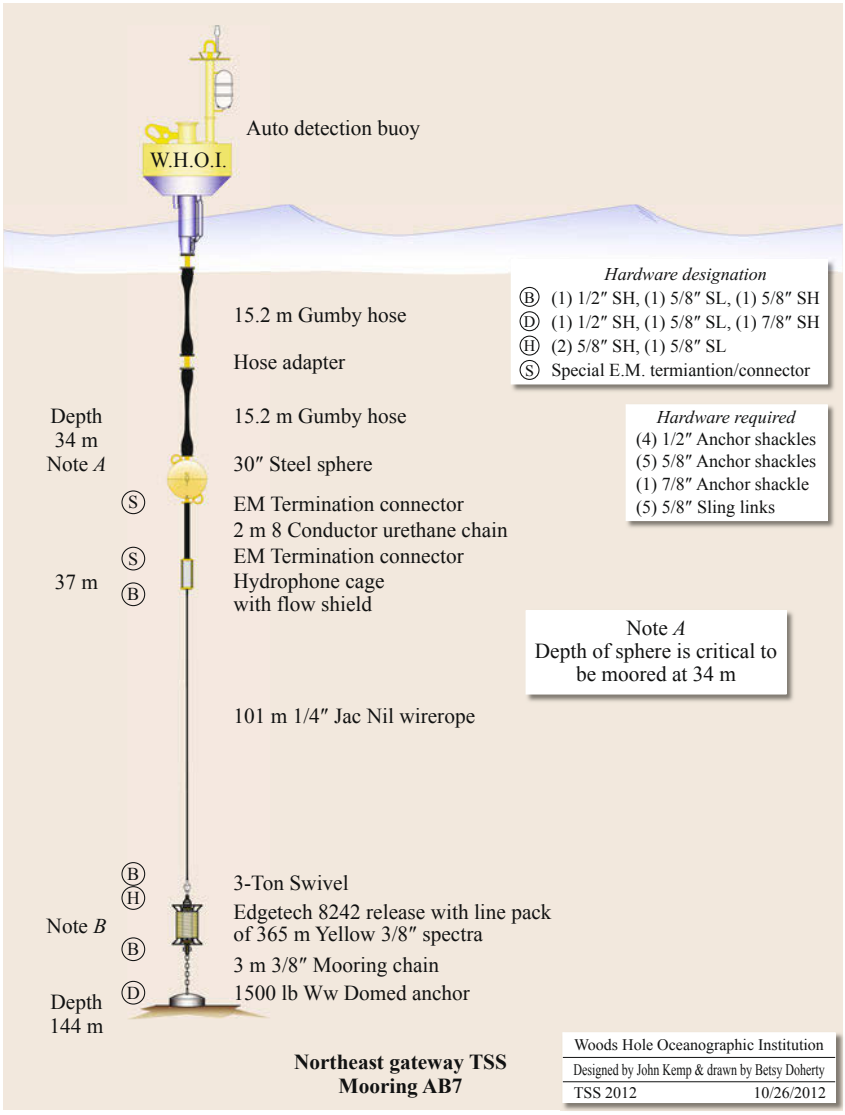


Fig. 40.17 Woods Hole whale listening mooring with two stretch-hoses with electrical conductors (diagram courtesy of Woods Hole Oceanographic Institution)

tems. The proceedings from these meetings are an excellent source of solutions to the practical problems encountered in designing, building, and fielding these systems [40.44].

40.4.1 Chain-Catenary Mooring

Figure 40.16 illustrates a chain-catenary mooring design used in shallow water (72 m). The mooring has a scope of 1.5 and the section of chain that transitions off the seafloor is considerably heavier than the section which is connected to the anchor. The additional size in this section address increased wear and provides tension to the mooring string above. Instrumentation throughout the mooring string is battery operated and self-logging.

40.4.2 Auto-Detection Mooring

Figure 40.17 shows a semi-taut mooring design that utilizes a highly elastic element below the buoy to provide enough compliance in the design in relatively shallow water (144 m). A key requirement for this system is that it includes a quiet place to mount a hydrophone for passive acoustic applications such as the detection of whales in the area. Matching the inertia of the submerged sphere and the stretch-hose elasticity isolates much of the surface motion from the

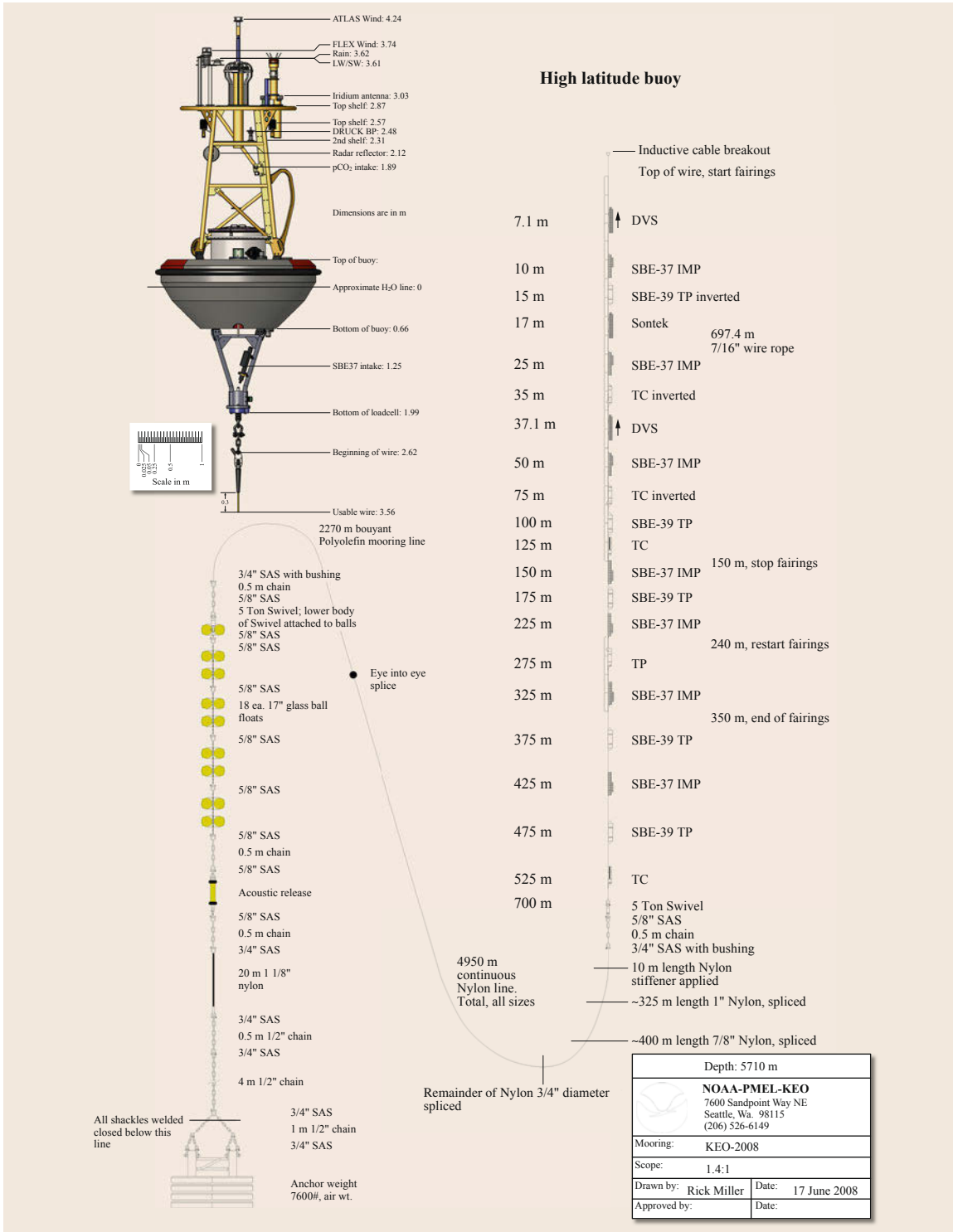


Fig. 40.18 Deep water inverse-catenary mooring utilizing nylon, polyolefin, and wire-rope strength materials (diagram courtesy of NOAA-PMEL)

lower part of the mooring, limiting flow-induced noise at the hydrophone. Additionally, the connections near the hydrophone and below the buoy are bolted flange connections, eliminating noise created by shackle connections in waves. Electrical conductors are helixed into the walls of the stretch hose at sufficient angle to reduce conductor strain to acceptable levels; these provide a power and data path between the buoy and the submerged hydrophone. This mooring was developed by WHOI [40.45, 46] and has been successfully deployed for many years; the stretch hoses have proven to be quite robust and are routinely reused for multiple year-long deployments without mechanical or electrical failure.

40.4.3 Deep-Water Mooring with Inductively Linked Instrumentation

Figure 40.18 illustrates a common oceanographic mooring design in deep water (5710 m). The inverse-catenary behavior is achieved by utilizing alternating sections of polyolefin and nylon rope and the mooring has a scope of 1.4. The buoy is a toroidal design, the behavior of this geometry is similar to that of a discus buoy due to its relatively large water-plane second moment of area to displacement ratio. The upper section of the mooring is jacketed wire-rope and includes inductively coupled instrumentation to monitor conditions throughout the water column. This system uses the seawater as a return-path for an conductive loop through the wire-rope. Data is inductively coupled onto the wire-rope at the top and at each instrument. Acoustic releases above the anchor facility recovery of this mooring which is deployed in an *anchor-last* fashion. Seventeen inch glass spheres above the release provide recovery buoyancy.

This mooring design is the result of iteration in which a failure of the nylon-rope section in a previous deployment of a similar system was addressed. A detailed report on the failure analysis and mitigation can be found in [40.29].

References

40.1 R. Lumpkin, S.A. Grodsky, L. Centurioni, M. Rio, J.A. Carton, D. Lee: Removing spurious low-frequency variability in drifter velocities, J. Atmos. Ocean. Technol. **30**, 353–360 (2013)

40.2 D.J. Maxwell, T. Mettlach, B. Taft, C. Teng: The 2010 national data buoy center (NDBC) mooring workshop, Proc. OCEANS (2010)

40.3 C. Teng, S. Cucullu, S. McArthur, C. Kohler, B. Burnett, L. Bernard: Vandalism of data buoys, Mar. Weather Log **54**(1) (2010) http://www.vos.noaa.gov/MWL/apr_10/vandalism.shtml

40.4 M. Grosenbaugh: Designing oceanographic surface moorings to withstand fatigue, J. Atmos. Ocean. Technol. **12**, 1101–1109 (1995)

40.5 H. Berteaux, B. Prindle, D. May: The deep sea moorings fishbite problem, Proc. OCEANS (1987)

40.6 American Iron and Steel Institute: *Steel Products Manual: Wire and Rods, Carbon Steel* (AISI, Washington DC 1975)

40.7 Samson: Industrial Catalog, http://www.samsonropecatalogs.com/home/100239.pdf (2010)

40.8 DuPont: Kevlar Technical Guide, http://www.dupont.com/content/dam/dupont/products-and-services/fabrics-fibers-and-nonwovens/fibers/documents/Kevlar_Technical_Guide.pdf

40.9 Nitronic50 Data, http://www.matweb.com/search/datasheet.aspx?matguid=5bc866b641534bbb921b0ce16255233a&ckck=1

40.10 Nilspin 3X19 Wire-Rope DataSheet, http://unionrope.com/Resource_/RopeProduct/1245/Nilspin-Product-Brochure.pdf

40.11 Buccaneer: Oceanographic Rope DataSheet, http://www.bucrope.com/Oceanographic-Rope.html

40.12 Peerless: Peerless/ACCO Chain Catalog, http://www.peerlesschain.com/catalogs/catalog-2010/files/per073_catalog_6_2_10%20full.pdf

40.13 Det Norske Veritas: Offshore Mooring Chain Standard DNV-OS-E302 (Høvik 2008)

40.14 The Cordage Institute: *Fibers for Cable, Rope and Twine (CI-2003)* (The Cordage Institute, Wayne 2005)

40.15 W. Paul: *Hose Elements for Buoy Moorings: Design, Fabrication and Mechanical Properties*, Woods Hole Oceanographic Institute Tech. Rep. WHOI-2004-06 (2004)

40.16 J. Kemp, D. Peters: WHOI's experience with electromechanical chains on surface moorings, ONR/MTS Buoy Workshop (2004)

40.17 D. Frye, J.A. Hamilton, M. Grosenbaugh, W. Paul, M. Chaffey: Deepwater mooring designs for ocean observatory science, Mar. Technol. Soc. J. **38**(2), 7–20 (2004)

40.18 DataWell: DataWell Rubber Cord Brochure, retrieved from http://www.datawell.nl/Portals/0/Documents/Brochures/datawell_brochure_rubbercords_b-21-02.pdf

40.19 DuPont: Surlyn 8660 Datasheet, retrieved from http://www.dupont.com/content/dam/assets/products-and-services/packaging-materials-solutions/assets/surlyn_8660_.pdf

40.20 H. Berteaux, A. Bocconcelli, M. Gould, S. Kery: *Testing and Evaluation of SURLYN Foam and SPECTRA*

Fiber Ropes for Buoy Systems Applications, WHOI Tech. Rep. 88-32 (1988)
40.21 McLane Steel Flotation, http://www.mclanelabs.com/master_page/product-type/mooring-products/steel-flotation
40.22 McLane Glass Ball Flotation Datasheet, http://www.mclanelabs.com/sites/default/files/sub_page_files/McLane-G2200-Datasheet_0.pdf
40.23 Benthos: Glass Sphere DataSheet, http://teledynebenthos.com/product/flotation_instrument_housings/flotation-glass-spheres
40.24 Atlantic Floats: Product List, http://www.atlanticfloats.com/frames_product.htm
40.25 The Crosby Group: Catalog, http://www.thecrosbygroup.com/html/default.htm
40.26 Benthos: Acoustic Release Product Selection DataSheet, http://teledynebenthos.com/product_dashboard/acoustic_releases
40.27 Sonardyne: Acoustic Release Product Selection, http://sonardyne.com/products/release-a-actuation/acoustic-release-transponders.html
40.28 EdgeTech Acoustic Transponding Releases, http://www.edgetech.com/pdfs/ut/Acoustic-Comparison.pdf
40.29 N. Lawrence-Slavas, C. Meinig, H. Milburn: *KEO Mooring Engineering Analysis*, NOAA Tech. Memorandum OAR PMEL-130, 2006)
40.30 J.A. Hamilton, M. Chaffey: Use of an electro-optical-mechanical mooring cable for oceanographic buoys: Modeling and validation, Proc. OMAE (2005) pp. 1–7
40.31 C.L. Wu, Y. Liu, K.H. Kellog, K.S. Pak, R.L. Glenister: Design and calibration of the SeaWinds scatterometer, IEEE Trans. Aerosp. Electron. Syst. **39**, 94–109 (2003)
40.32 M.H. Freilich, B.A. Vanhoff: The accuracy of preliminary WindSat vector wind measurements: Comparisons with NDBC buoys and QuikSCAT, IEEE Trans. Geosci. Remote Sens. **44**, 622–637 (2006)
40.33 H.L.Tolman: User manual and system documentation of WAVEWATCH III^{TM} version 3.14, Tech. Note, MMAB Contribution(276) (2009)
40.34 California Institute of Technology: Ocean Surface Current Analysis (OSCAR) Third Degree Resolution User's Handbook, http://apdrc.soest.hawaii.edu/doc/osarthirdguide.pdf (2009)
40.35 W.J. Emory, R.E. Thompson: *Data Analysis Methods in Physical Oceanography* (Elsevier, Amsterdam 2001)
40.36 J.N. Newman: *Marine Hydrodynamics* (MIT Press, Cambridge 1997)
40.37 A. Hulme: The wave forces acting on a floating hemisphere undergoing forced periodic oscillations, J. Fluid Mech. **121**, 443–463 (1982)
40.38 R.W. Yeung: Added mass and damping of a vertical cylinder in finite-depth waters, Appl. Ocean Res. **3**(3), 119–133 (1981)
40.39 C.-H. Lee: *WAMIT Theory Manual*, MIT Report 95-2 (MIT, Cambridge 1995)
40.40 J.I. Gobat, M.A. Grosenbaugh, M.S. Triantafyllou: *WHOI Cable: Time domain numerical simulation of moored and towed oceanographic systems*, WHOI Tech. Rep. 97-15 (Woods Hole Oceanographic Institution, Woods Hole 1997)
40.41 S.F. Hoerner: *Fluid-Dynamic Drag: Theoretical, Experimental and Statistical Information* (Hoerner Fluid Dynamic, Bakersfield 1965)
40.42 C.W. Gear: *Numerical Initial Value Problems in Ordinary Differential Equations* (Prentice-Hall, Upper Saddle River 1971)
40.43 Orcina Ltd.: OrcaFlex Manual (Version 9.8a), http://www.orcina.com/SoftwareProducts/OrcaFlex/Documentation/OrcaFlex.pdf
40.44 ONR/MTS Buoy Workshop, http://www.whoi.edu/buoyworkshop/index.html
40.45 E. Spaulding, M. Robbins, T. Calupca, C.W. Clark, C. Tremblay, A. Waack, J. Kemp, K. Newhall: An autonomous, near-real-time buoy system for automatic detection of North Atlantic right whale calls, Proc. Meet. Acoust. **6**, 010001 (2009)
40.46 EOM Offshore LLC: EOM Auto-Detection Mooring Brochure, http://eomoffshore.com/pdf/EOMautodetectmooring.pdf

41. Liquefied Natural Gas Carriers

Krish P. Thiagarajan, Robert Seah

Global energy demand is rising as the world's population is growing. Various forecasts predict that by 2050, there could be 2.5 billion more humans than today, who would use twice the energy consumed today. Energy companies are examining all possible sources of energy as part of their widening portfolio.

Natural gas (NG) is an abundant resource that is still relatively a newer source for exploitation. With improving technology, NG can be an alternative fuel to generate electricity with operating costs comparable to coal or nuclear energy sources. Liquefaction of NG is an important part of the process, as liquefied natural gas (LNG) is compact and can be transported with special-purpose carrier vessels. Large offshore gas fields call for floating LNG (FLNG) terminals, which are capital and technology intensive. Various aspects of FLNG technology are being researched and developed currently. Notwithstanding this, companies are developing and deploying FLNG solutions based on past experience and proprietary technology. This chapter of the handbook deals with various aspects of LNG, design of carriers and especially the cargo holds, environmental and design challenges of these carriers and FLNG.

41.1 **Types of LNG Carriers** 963
41.1.1 LNG Carrier Containment Systems 964

41.2 **Thermodynamics of LNG** 968
41.2.1 Properties of LNG 968
41.2.2 LNG Stages 969
41.2.3 LNG Onshore Storage Considerations 969

41.3 **Environmental Challenges** 970
41.3.1 Open Ocean Voyages 970
41.3.2 Boil-Off 971

41.4 **Fluid Structure Interaction of LNG Systems** 971
41.4.1 Wave-Induced Motions 971
41.4.2 Sloshing Impact and Consequences 972
41.4.3 Scaling 974
41.4.4 Multiple Body Interaction in Shallow Water 977

41.5 **Design Methodologies of LNG Containment Systems** 978
41.5.1 Comparative Analysis 978
41.5.2 Model Testing 980

References 983

41.1 Types of LNG Carriers

Liquefied natural gas (LNG) is transported at a cryogenic temperature of −163 °C with an ullage pressure slightly above atmospheric. This requires that the vessel's LNG containment system should be designed to provide thermal isolation of the cargo to restrict boil-off and to provide thermal protection for hull steel against embrittlement by the cargo. Of course, with the cargo being a flammable product, the containment system also needs to isolate the contents from the atmosphere. These basic requirements make the maritime transport of LNG not a straightforward matter.

The first LNG vessel carried cargo from Lake Charles, Louisiana, to the Canvey Island, United Kingdom, in 1959, demonstrating the viability of maritime transport of LNG. This vessel, *The Methane Pioneer*, was a converted diesel powered cargo vessel and at 103 m (338 ft) length overall (LOA), had a capacity of just 5000 (5 K) m^3 of LNG carried in five independent aluminum prismatic tanks. As a prototype vessel, she made just seven trips and was renamed as the Aristotle in 1967. Interestingly, this vessel is noted to have imported LNG back into the United States when its cargo was offloaded in Boston in 1969. Based on the experience garnered by the Methane Pioneer, the first true purpose built commercial LNG carrier (LNGC), the Methane Princess and its sister vessel, the Methane

Fig. 41.1 Moss-type LNG carrier and portside ship-to-ship (STS) cargo transfer operations (courtesy of Hoegh LNG)

Progress, were constructed. These vessels gave birth to the LNG trade industry by delivering the first shipments from Algeria to UK in 1964. In contrast to the Methane Pioneer, the Methane Princess had a capacity of $27.4 \times 10^3\,m^3$ of LNG in nine Conch-type independent prismatic aluminum tanks.

Today, the largest LNG carriers are of the Qmax class, by Qatargas, with LOA of 345 m and a capacity of $266 \times 10^3\,m^3$. This is followed by the Qflex class of vessels with a LOA of 315 m and a capacity of $216 \times 10^3\,m^3$ of LNG at 100% fill. Standard LNG carriers have an average capacity of about $160 \times 10^3\,m^3$ and are likely to utilize a membrane containment system from Gaztransport and Technigaz (GTT). While there have been many developments and designs for the containment system in the prior decades since the Methane Pioneer, the vast majority of new and operating vessels use one of the following four containment system types: IHI-SPB self-supporting prismatic tanks, Kvaerner–Moss self-supporting spherical tanks (e.g., Fig.41.1), GTT's Technigaz Mark III membrane tanks, and GTT's Gaz Transport NO96 membrane tanks.

41.1.1 LNG Carrier Containment Systems

The purpose of the containment system is to prevent release of the cargo into the environment, minimize thermal ingress to reduce cargo boil-off as well as to protect the hull from the cryogenic temperatures of the cargo. All LNG carriers built on or after July 1986 are subject to the International Maritime Organization (IMO), international code for the construction, and equipment of ships carrying liquefied gases in bulk (IGC code). The code has several tank type definitions but categorizes the majority of LNG containment systems into just two main categories: membrane tank systems and independent self-supporting tanks.

Membrane Tank Systems

Membrane tank systems such as shown in Fig. 41.2 consist of two containment barriers, each formed by a metal membrane backed by an insulation layer. The ullage pressure in the tanks is maintained slightly above atmospheric pressure with the LNG at its phase-change temperature at this pressure. Boil-off is controlled by the thermal insulation and thus penetrations through the insulation have to be minimized. The two insulation spaces are maintained with a nitrogen gas environment which is monitored for the presence of natural gas as an indicator for loss of containment. In contrast to the self-supporting type, the membrane systems rely on the inner hull to provide structural strength and support to the primary and secondary barriers.

The popular NO96 system (Fig. 41.3) utilizes a 0.7 mm thick Invar film, a nickel alloy with very low thermal expansion rate, as the containment membrane. Plywood boxes, typically filled with perlite, act as the thermal insulation component as well as to provide support for the membrane and to transfer membrane stresses to the hull. The plywood boxes are reinforced with internal ribs to provide additional strength. Total insulation thickness is approximately 0.5 m. Two layers of the membrane–insulation combination are used to provide the primary and secondary containment required by the IGC code.

The Mark III system (Fig. 41.4) is similar to the NO96 system but uses a 1.2 mm thick stainless steel membrane atop a reinforced polyurethane (PU) foam insulation layer. Due to the thermal expansion properties of stainless steel, the Mark III primary barrier

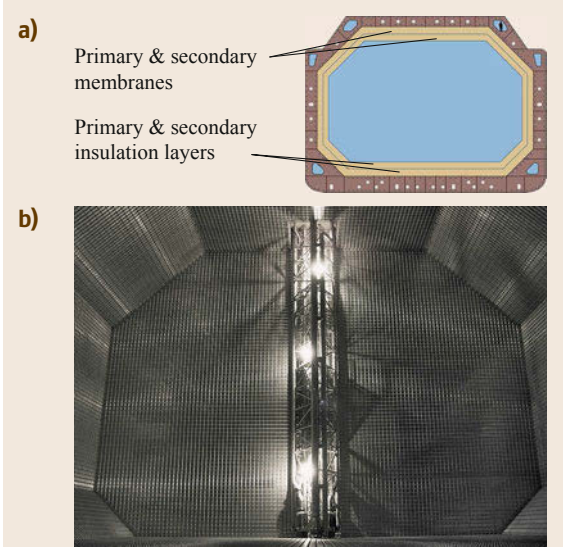


Fig. 41.2 (a) Membrane schematic and (b) pump tower @ GTT

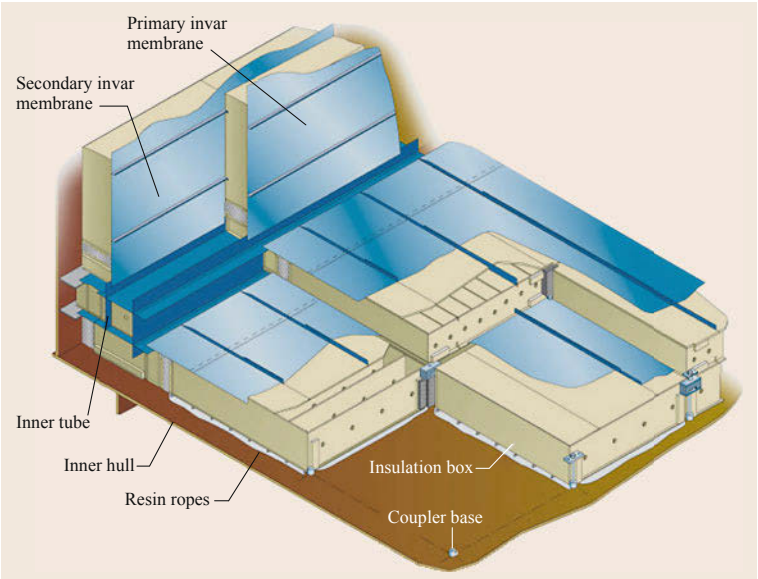


Fig. 41.3 NO96 containment system schematic (after [41.1])

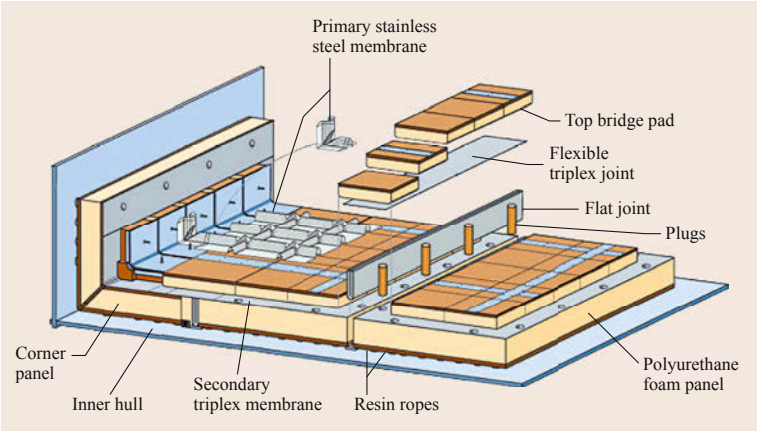


Fig. 41.4 Mark III containment system schematic (after [41.1])

incorporates a *waffle-like* system of corrugations to absorb the thermal expansion stresses. Insulation thickness for the Mark III system is approximately 0.3 m. For secondary containment, a different membrane, atop PU foam, composed of a composite laminate called Triplex is used instead.

GTT France also offers the CS1 membrane system (Fig. 41.5) which combines the features of the NO96 and Mark III system. The CS1 uses a primary membrane of Invar and the Triplex secondary composite membrane. Reinforced PU foam is used as the insulation layer with a total thickness of approximately 0.3 m.

Advantages. GTT's membrane containment systems, mostly NO96, have experienced a rapid rise in market share to become the dominant containment system deployed in new build carriers. One major advantage over competing designs is its cost efficiency. As the membrane systems rely on the hull for the structural strength, coupled with their compact design, they are able to maximize the storage capacity available for a given hull form. In addition, the costs of the membrane systems are relatively inexpensive, making the membrane systems the most cost-efficient containment system thus far. These systems also have a modular construction methodology enabling flexibility in construction and design. The low main deck profile not only reduces the windage area but also makes the membrane systems particular appealing for FLNG that require deck space for topside equipment.

Other operational advantages of the membrane design include quick cool down periods (low thermal capacity) during loading and hence quicker turnaround. Finally, GTT have licensed their design to a large

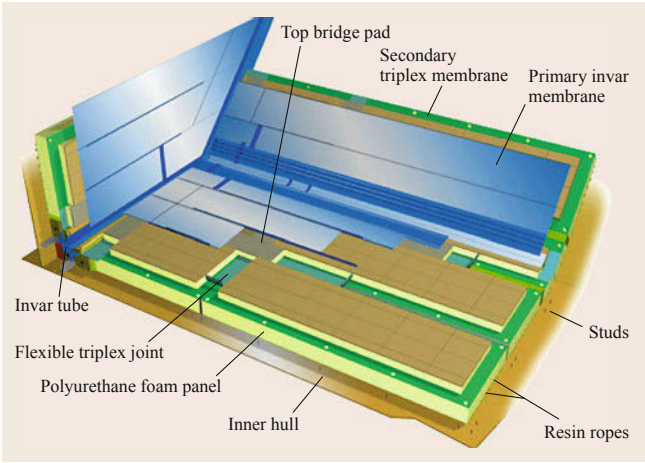


Fig. 41.5 CS1 containment system schematic (after [41.1])

number of shipyards, which has commercial flexibility advantages.

Disadvantages. The primary disadvantage of the membrane system is its apparent vulnerability to LNG sloshing induced damage. As the main load-bearing component is the hull itself, it is theoretically possible for the high-impact pressures induced by sloshing to locally damage the insulation or even breach the containment barrier. To mitigate this, the membrane tanks incorporate several design features. In Fig. 41.2, there are corner chamfers in the top 30% and bottom 10% of the tank height. The chamfering mitigates the impact of the LNG on the walls of the tanks at fill levels above 70% and below 10%. Consequently, the capacity of the tanks is reduced and most class societies require that the cargo level be either above 70% or below 10%. This fill restriction places limits on the ability of the LNG carrier to support spot market demands although this is somewhat mitigated by being able to shift cargo from tank to tank. The vulnerability to sloshing also places environmental limits on the vessels. Another mitigation measure is to reinforce the plywood insulation boxes for the NO96 system by increasing the number of internal plywood ribs in each box. This can be done selectively at tank locations most prone to sloshing impact damage, such as the tank ceiling for cost efficiency.

As the membrane system is attached to the inner hull of the vessel, the installation of the containment system is serialized with respect to the construction of the hull, that is, the installation of the containment systems can only commence after the inner hull is completed. Also, as a consequence of not having a void space between the tanks and the hull, inspection, and maintenance are inhibited.

Independent Self-Supporting Tanks

Independent self-supporting tanks, as the category name implies, are containment systems which are not part of the ship's hull and thus do not contribute to the strength of the hull. This category is further subdivided into types A, B, and C. Type A tanks are designed according to recognized standards or class rules and require a full secondary barrier. Examples of a type A tank are the aforementioned conch tanks. Type B tanks are designed using model tests and refined analysis methods and only require a partial secondary barrier. IHI-SPB and Kvaerner–Moss tanks are examples of type B tanks. Both type A and type B tanks have a design vapor pressure of less than 0.7 bar. In contrast, type C tanks are generally pressure vessels which require no additional secondary barriers.

The Moss–Rosenberg spherical LNG containment system (Fig. 41.6) was developed by Kvaerner in 1971. It consists of a steel tank cover, a PU foam insulated spherical aluminum (or 9% Ni steel) tank shell structure (primary barrier), a steel support skirt just below the equator of the sphere, a pipe tower, and a drip pan with splash walls (partial secondary barrier) below the shell. The tank loads, including weight, thermal and contraction stresses are transferred to the ship hull via the support skirt and hence the hull supports which the tanks sit on. The carrier is dual hulled and equipped with side ballast tanks. LNG leak detection is enabled by monitoring the atmosphere in the void space between the tank and hull as well as by temperature sensors monitoring the drip pan.

Advantages – Spherical

One significant advantage of the spherical tanks over the membrane tanks is the relatively low level of sloshing loads and hence do not have fill level restrictions or sloshing-related environmental restrictions that the membrane tanks are subjected to. It is hence favorable

Fig. 41.6 Moss–Rosenberg spherical LNG tank

for spot market LNG transactions. Also, in contrast to the membrane tanks, the spherical tanks can be built independently and later integrated with the hull resulting in flexibility in construction scheduling. As mentioned above, the resulting void space between the hull and tanks is monitored for the presence of methane for leak detection but also serves to provide access to the tank exterior (and hull interior) for inspection and repair, providing a greater sense of integrity. In the event of major damage, individual tanks can be replaced. Finally, due to the separation between the tank and hull, the tanks have a measure of protection in the event of a collision or grounding.

Disadvantages – Spherical

The disadvantage of the spherical tank is its space efficiency compared to prismatic tank forms. Further, due to the additional structural steel required for spherical tanks, these tanks are relatively heavy and costly when compared to membrane tanks and are thought to have led to its marked decrease in market share from its dominant position for LNG carriers several decades ago. The relatively higher structural weight per unit of cargo not only implies a greater unit cost but also incurs higher toll fees at the Suez canal which calculate fees according to the vessel's gross tonnage. The spherical geometry also leads to higher tank heights with increased wind loading. This is exacerbated when trying to layout process equipment around or above the tanks, as in an FLNG application. An increase in wind loading may give rise to seakeeping and resistance issues for LNGCs and stability or process efficiency issues for FLNGs.

In 1983, Ishikawajima–Harima heavy industries (IHI) introduced their SPB LNG tanks, first installed in two 87 000 m^3 LNGCs in 1993. SPB stands for self-supporting, prismatic IMO type B independent tank containment system, and an example is shown in Fig. 41.7. Like the Moss–Rosenberg spherical tanks it does not rely on the ship structure for strength. Instead, SPB tanks are distinctive in that they are constructed of Al-alloy (or 9% Ni steel) stiffened plate structures and are subdivided by a centerline liquid-tight bulkhead and swash bulkhead, giving the containment system relative high strength compared to the other systems. Each tank is insulated by PU foam and transfers loads via reinforced plywood supports to the inner bottom hull. In contrast to both membrane and Moss–Rosenberg systems, there is no pump tower or column. Instead, fill and discharge lines are supported by the girders and bulkheads.

Advantages – SPB

The SPB tanks share many of the advantages with both prismatic tanks and spherical tanks. Since SPB tanks are independent, like the Moss–Rosenberg tanks, they can be constructed in parallel with the hull as well. Note that even though these tanks are commonly referred to as *prismatic*, in actuality they conform closely to the hull form and the first bow tank is usually wedge shaped. This makes them particularly space efficient. Further, due to their strength there is no requirement for chamfered tank tops, and SPB tanks are known for their flush deck profile and low wind loading. This also makes them particularly suitable to layout process equipment atop the tanks, especially for FLNG applications and advantageous for maneuvering. Sloshing is not a concern for SPB tanks due to their stiffened plate structure. The internal stiffeners act as baffles to limit the extent and coherence of sloshing loads within the tanks. Thus, SPB tanks do not have fill restrictions and the associated environmental limits, similar to the Moss–Rosenberg tanks, albeit for different reasons.

Disadvantages – SPB

The main drawback of the IHI-SPB containment system is its cost which is higher than both the Moss–Rosenberg and the membrane systems. Also, there are only two yards with the capability to build the SPB system as compared to 25 for membrane systems. These

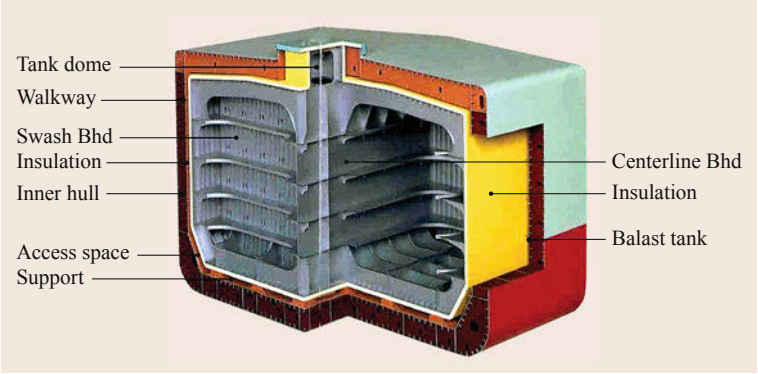


Fig. 41.7 IHI-SPB containment system

are likely the reasons that, at the time of this writing, there are only two LNG carriers in service using the IHI-SPB containment system. However, interest in the IHI-SPB containment system has recently grown as the industry moves toward floating LNG facilities. It turns out that the increased cost of the SPB system is relatively small compared to overall development cost for an FLNG facility, especially in light of the increased strength, reliability, and overall operability compared to the cheaper membrane systems. As mentioned previously, the SPB system also has the advantage of a flat deck to aid in the layout of LNG process trains.

41.2 Thermodynamics of LNG

The actual composition of LNG, while being predominantly methane (CH_4), varies based on its origins as well as the sales specifications it has to satisfy. Thus, there is variability in its properties and the values stated in the following sections are representative of a *typical* composition.

41.2.1 Properties of LNG

Once properly processed, LNG is a colorless, odorless, nontoxic, and noncorrosive liquid with a boiling point of −162 °C (111.15 K) at standard atmospheric pressure. It has a density of 450 kg m^{-3} and is approximately 1/600th the volume of the gaseous phase. The flammability limits are between 5% and 15%. The specific energy is approximately 50 MJ kg^{-1}, thus one reason for LNG's relative popularity is its efficient energy density being 22.5 kJ m^{-3} (22.5 MJ L^{-1}) in comparison to compressed natural gas (CNG) at 9 kJ m^{-3} (9 MJ L^{-1}). Heat of vaporization for regasification is 213 kJ kg^{-1} at standard atmospheric pressure.

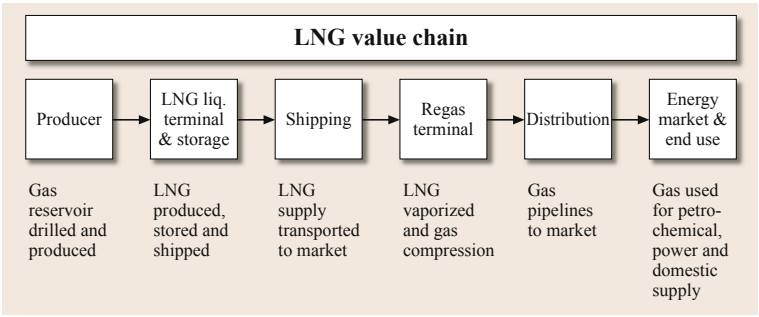


Fig. 41.8 LNG value chain

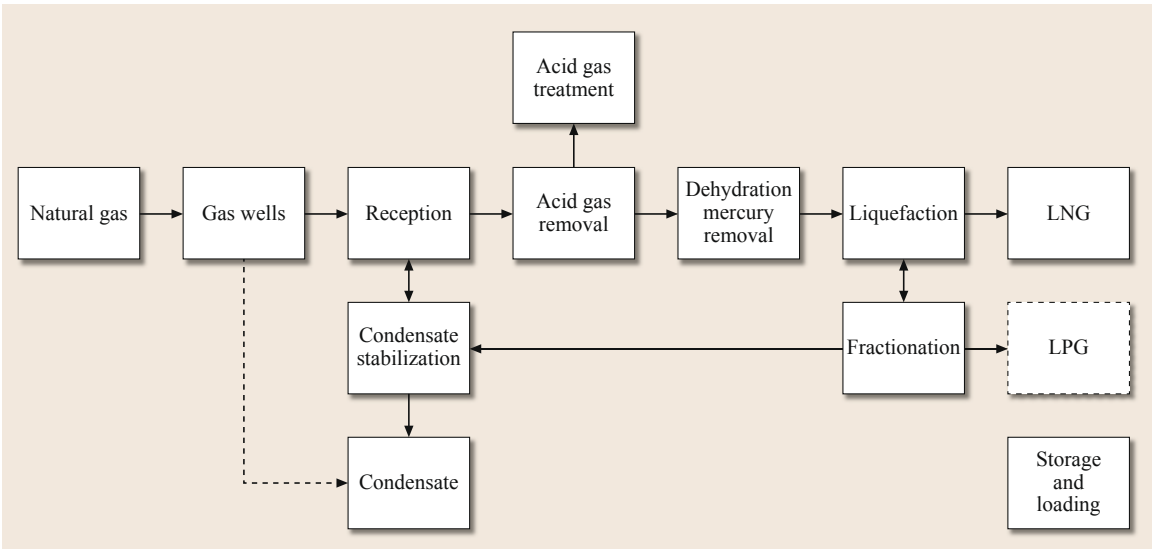


Fig. 41.9 Liquefied natural gas production (after [41.2])

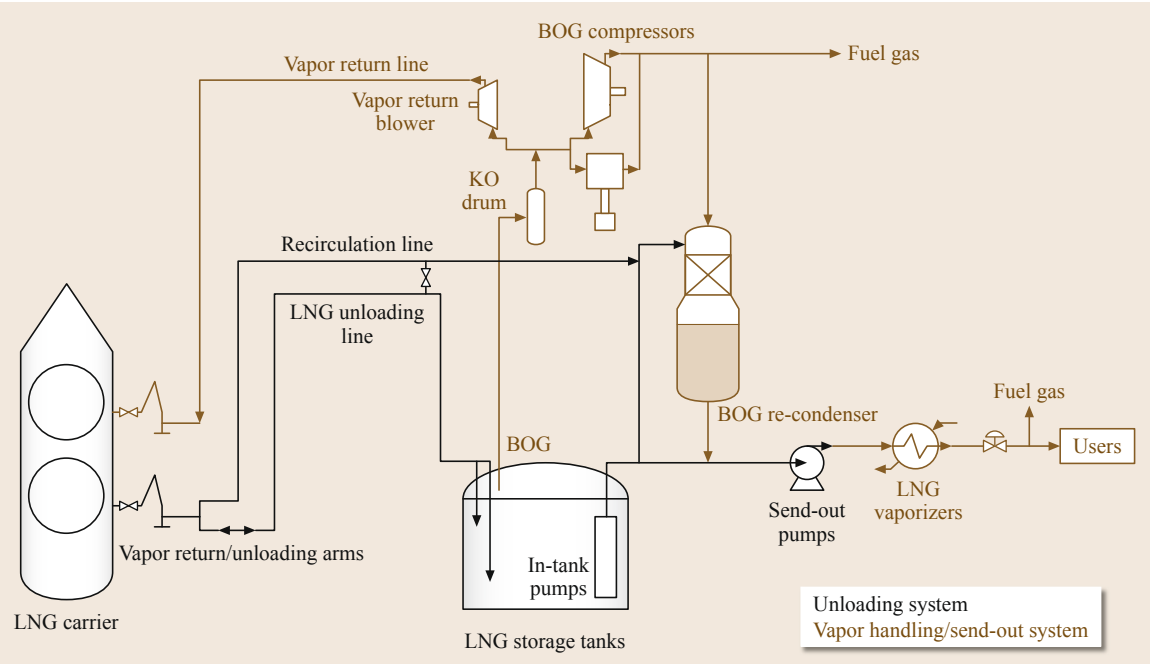


Fig. 41.10 LNG offloading line diagram

41.2.2 LNG Stages

The primary stages in the LNG value chain include gas treatment, liquefaction, storage, transport, offloading and regasification (Fig. 41.8).

During the gas treatment stage, the gas is brought to sales specification including heating value control by removal of natural gas liquids (NGL), dehydration, and CO_2 removal. A typical flow chart of the treatment and production process is shown in Fig. 41.9. Due to strict limits on toxic hydrogen sulfide (H_2S), sulfur, and mercury content, these contaminants have to be removed at this stage as well, lest they be released into the end user environment. Mercury, in particular, also has a corrosive effect on aluminum which is used extensively in the liquefaction (cryogenic) processes.

After processing, the sweet dry gas is ready for liquefaction. While there are numerous LNG liquefaction processes, the majority of the LNG plants use the Air Products propane precooled mixed refrigerant (C3MR) process, the ConocoPhilips' optimized cascade process (CoPOC) or the Linde multi fluid cascade (MFC) process, with C3MR being the most common.

Typically, the produced LNG is pumped into storage at this stage, either for transport or for later introduction into the retail gas pipeline. As most LNG plants are shore-based, the storage tanks are also onshore. Exceptions include floating LNG (FLNG) vessels and floating regasification and storage units (FSRU) which utilize the containment systems described above.

41.2.3 LNG Onshore Storage Considerations

The storage of LNG, being a cryogenic fluid, is by nature a complex but necessary function in the production of LNG. LNG is stored in specialty insulated tanks, either at the LNG plant prior to transport, during transport on LNGC or at the receiving terminal prior to regasification into a distribution pipeline. LNG is stored in a boiling state (162 °C) at close to atmospheric pressure. One particular feature of LNG storage and transfer is the handling of boil-off gas (BOG), (Fig. 41.10) As LNG is stored in a boiling state, heat ingress into the tanks causes a small proportion of the liquid LNG to boil off. The rate of BOG generation is determined by the rate of heat ingress into the tank through the insulation and tank penetrations such as the pump tower and tank supports. LNG is also used as the cryogenic agent to cool down the storage tanks prior to transfer of cargo, a process that may take more than 10 h. Additionally, LNG is continuously circulated through the transfer lines to maintain the cryogenic temperature. As a result, the LNG flashes to gaseous phase, generating BOG. BOG is generally recovered and either re-condensed or used as a local fuel source and its handling requires special consideration. BOG

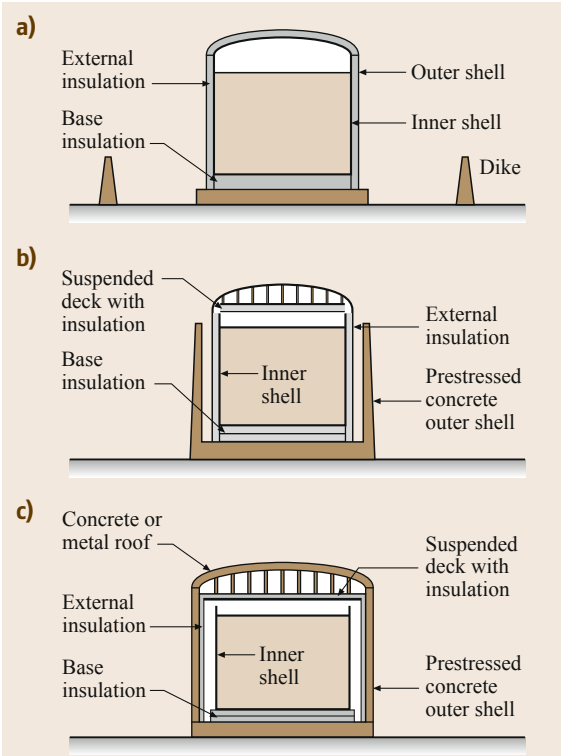


Fig. 41.11a–c Onshore LNG containment systems: **(a)** single containment; **(b)** double containment; **(c)** full containment

is also supplied via return lines to the storage tanks to maintain the near atmospheric pressure within the tanks.

Onshore LNG tanks (Fig. 41.11) feature both a primary and a secondary containment system. The latter is designed to contain the LNG liquid in the event of a failure in the primary containment system but may not always be designed to contain the vapor phase as well. The primary containment system (or inner tank) is generally constructed of a 9% nickel steel alloy and insulated using perlite. A greater degree of insulation is required on the base of the tank to insulate the tank from the ground. The base or slab may also contain heating elements to prevent the ground from freezing. The secondary containment system consists of a pre-stressed concrete outer shell or even an earthen wall.

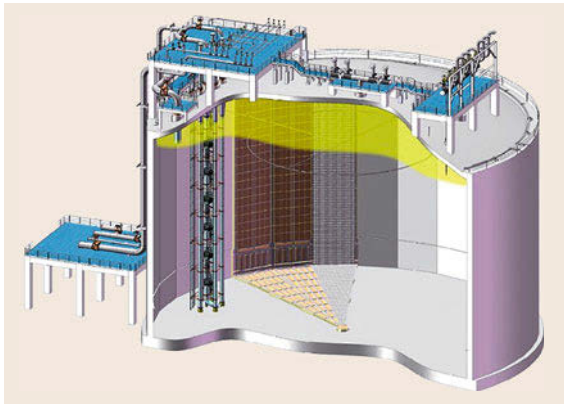


Fig. 41.12 Onshore LNG membrane containment system (after [41.1])

In contrast to the above where both primary and secondary containment systems are self-supported, membrane tanks (Fig. 41.12), similar to those used in LNG carriers, have a primary containment system which derives its structural strength from the prestressed concrete shell. Primary containment is provided by a stainless steel membrane mounted on a plywood-foam insulating panel to transfer the loads to the concrete shell. A secondary membrane barrier is typically also included. This system is typical of in-ground LNG tanks.

41.3 Environmental Challenges

41.3.1 Open Ocean Voyages

An LNGC during voyage in an open ocean encounters various sea-state conditions that can affect its motion response and maneuverability. Traditionally, LNGC were planned for port to port voyages and hence the tanks were either full or empty. The potentially dangerous internal free surface condition was thus avoided. As is well known, internal free surfaces can compromise safety and reduce vessel stability. For example, *Halkyard* et al. [41.3] say that the metacentric height of a vessel is reduced by the net sum of the water plane stiffness lost due to internal surfaces. This is given by

$$\mathrm{GM(new)} = \mathrm{GM} - \sum \frac{\rho_c}{\rho} \frac{i_{xx}}{\nabla} \,. \qquad (41.1)$$

Here the original metacentric height GM is reduced by the net effect of water plane stiffness (i_{xx}) lost due to internal surfaces. ρ_c and ρ are the densities of the cargo and ambient sea water, respectively, and ∇ is the displaced volume of the vessel.

Recently, spot market trading of LNG has become advantageous and economically viable for operators. This however, requires an LNGC to ply from one location to another collecting liquid cargo, thus operating

with one or more of its tanks partially full. While this causes a decrease in stability as discussed above, it also can result in potentially large sloshing loads on internal panels of LNC cargo holds. The internal surface of these holds is not designed for large impact loads, and hence dents and possible damage to the insulation can occur. A leak in insulation can result in loss of liquid cargo being released as vapor. A rapid increase in temperature can result in rapid expansion of cargo leading to possibility of explosion.

While traveling in waves, it is important to recall that the waves seen by a ship is altered by its so-called encounter effect. The frequency of excitation for a ship is given by ω_e, and is related to the actual wave frequency ω, by the formula

$$\omega_e = \omega - \frac{U^2}{g} \sin\beta \,, \tag{41.2}$$

where U is the speed of the ship and β is the angle of approach of a wave system. Thus at various partially filled conditions, significant sloshing may be experienced if the natural sloshing frequencies are matched with the encounter frequency seen by the ship. Since the latter is a function of ship speed, sloshing motions may be experienced at certain portions of the voyage only.

When an LNGC is loading or offloading at an exposed terminal, potential for sloshing again exists if the LNGC encounters a changing wind or wave pattern. Standard operating practice normally includes provision for emergency operations in case of adverse weather. Another situation conducive for sloshing relates to emergency disconnect during loading/offloading of an LNGC with an offshore liquefaction plant. In all these cases, potential sloshing loads can arise.

Yet another structural consideration caused during offloading relates to low-cycle high-stress fatigue. *Wang* et al. [41.4] discuss the fatigue stresses caused on welded joints in bulkheads separating two tanks one of which is full while the other one is empty. Duty cycles associated with such operations can produce stresses that exceed the yield strength of the weld. The frequency of these loads is low and hence in the plastic range and not covered in a standard *S–N* curve. The article suggests a calculation procedure based on the so-called Neuber's rule.

41.3.2 Boil-Off

Typical fill conditions of LNG tanks are close to 98% of volume. The small amount of atmosphere allows for vapor accumulation from the LNG. As LNG is close to its boiling temperature, some amount of LNG is converted into vapor, causing a boil-off condition. The boil-off rate is quite an important safety concern and is stipulated among others by IMO at 0.12% for 125 000 m^3 storage. The boil-off rate ($\dot{B}$) can be simply characterized by the following equation

$$\dot{B} = \frac{q}{h\rho_c V} \,. \tag{41.3}$$

Here q denotes the heat influx into an LNG tank of volume V from the enclosure, h and ρ_c are the latent heat coefficient and density of the LNG, respectively. *Zakaria* et al. [41.5] use computational fluid dynamics (CFD) to simulate the heat transfer across a membrane tank. The tank is considered to be one of the four tanks that make up a 160 K m^3 ship. The boil-off is found to be approximately 0.155%/(kg day) when the ship is operating in seawater at 0 °C, and ambient air at 5 °C.

41.4 Fluid Structure Interaction of LNG Systems

41.4.1 Wave-Induced Motions

An LNG carrier, like any other ship with forward speed will encounter waves during its voyage and hence should have good seakeeping characteristics. This is even more important if any of the tanks is partially filled with liquid cargo. During the loading and unloading stages, the LNG carrier behaves like a moored vessel and here again wave-induced motions are important.

A floating vessel's response to oncoming waves is characterized by its response amplitude operator (RAO). The RAO is strictly a complex quantity, comprising of amplitude and phase information. One can define the amplitude of an RAO as [41.6]

$$|\mathrm{RAO}|_i = \frac{|\eta_i|}{A} \,, \tag{41.4}$$

where η_i denotes the motion being considered, with the subscript following standard notation, that is, $i = 1, 2, 3$ denoting the three translations surge, sway, and heave, respectively, and $i = 4, 5, 6$ denoting the three rotations, roll, pitch, and yaw, respectively, A is the wave amplitude. In typical offshore computations, numerical methods based on potential flow assumptions are used. Then viscous damping is artificially added, which alters the resonant peaks in the RAO.

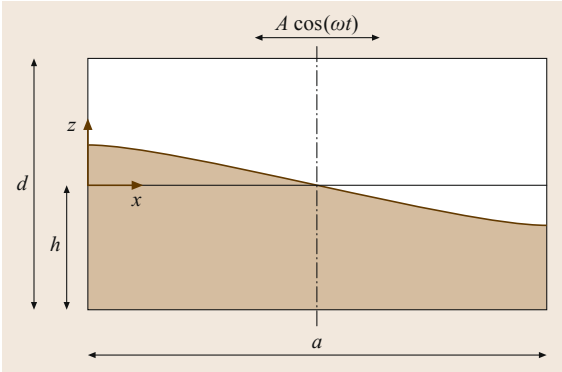


Fig. 41.13 Layout of rectangular tank subject to sloshing motions (after [41.7])

In a classical naval architectural sense, liquid cargo is treated hydrostatically and hence its effect on stability is evaluated. When tanks are partially filled then dynamic effects can be important. *Thiagarajan* et al. [41.7] suggest that in a linear diffraction problem, it is sufficient to consider the linear transfer function of sloshing, which introduces the first level of dynamics into the problem. For a tank of width a filled up to height h, subject to translational oscillation of frequency ω and amplitude A, the dynamic pressure on the walls is given in terms of vertical coordinate z ($z = 0$ is the free surface) by

$$p(z,t) = \sum_{n=1}^{\infty} \frac{2\rho a A(1-\cos n\pi)}{(n\pi)^2} \frac{\cosh\left(\frac{n\pi(z+h)}{a}\right)}{\cosh\left(\frac{n\pi h}{a}\right)} \times \frac{\omega^4}{(\omega^2-\omega_n^2)} \cos\omega t \,. \tag{41.5}$$

Here ω_n is the sloshing natural frequency, given for the n-th eigenmode by

$$\omega_n^2 = g\frac{n\pi}{a}\tanh\left(\frac{n\pi h}{a}\right) \,. \tag{41.6}$$

The pressure may be integrated to provide a force to be used in global response calculations.

A different linear computational approach to what was described above was used by *Chen* [41.8]. His computation employed the same Green's function to solve the flow inside the tanks as well as the flow external to the ship hull. A small amount of wall reflection was employed inside the tank, thus simulating damping. When motion was close to resonance then significant nonlinear effects occurred, requiring extra numerical treatment. *Chen* [41.8] presented numerical results at no damping ($\epsilon = 0$) and at 2% ($\epsilon = 0.02$) for an LNG carrier with two tanks, and compared it with model tests at 1 : 50 scale. The prototype LNGC (Fig. 41.14a) was 274 m long, 44.2 m wide, and was loaded to a draft of 11.58 m. The two tanks were of lengths ranging from 41 to 47 m, and width 39.1 m. The tanks were filled with cargo of 10 m in height. The motion RAO's of this LNGC are shown in Fig. 41.14b–d. The uncoupled fundamental natural frequency of sloshing in each tank is 0.72 rad s^{-1} based on (41.6). Resonant sloshing motion in the tank causes a dip in the sway RAO of the LNGC (Fig. 41.14b). The roll RAO of a floater is normally single peaked. However, in the presence of an internal surface, two peaks are visible, on either side of the sloshing frequency of the two tank system. This may be thought to be analogous to anti-roll tanks, which are designed to reduce the peak roll motion of a vessel. The effects of inertia of the vessel are offset by the restoring effect of the liquid movement resulting in a cancelation. On the other hand, the inertia of the vessel and the liquid cargo individually provide the two peaks on either side. The heave RAO shown in this figure is typical of floating vessels, and is unaffected by partially filled cargo tanks.

41.4.2 Sloshing Impact and Consequences

Sloshing is associated with uncontrollable motion of liquids within the walls of a containment structure. While normally sloshing loads are handled by structural elements supporting the inner walls, a membrane-type LNG tank encounters problems because structural members supporting the insulation material do not have sufficient strength. This has resulted in surface deformation [41.9] or for a chronology of sloshing related events one may refer to *Gavory* and *de Seze* [41.10]. The consequences of even a small failure of the insulation of an LNG system can be significant, if LNG is allowed to evaporate into NG which can escape to adjacent regions.

Impact loads due to sloshing are significant when the liquid is excited in and around its natural frequencies. These frequencies (41.6) are highly geometry and fill level dependent. Figure 41.15 shows the variation of first and third sloshing frequencies with fill level in the tank (shown as an aspect ratio of fill height to tank width). The even harmonics are not significant since there is cancellation of energy from both ends of the tank [41.7].

Considerable studies have focused on measuring and modeling sloshing impact pressures. For example, one can refer to the sloshing symposia organized as part of the international offshore and polar engineering (ISOPE) conferences from 2010 to 2014. Of note here is the SLOSHEL joint industry project [41.11], where significant experimental work has been con-

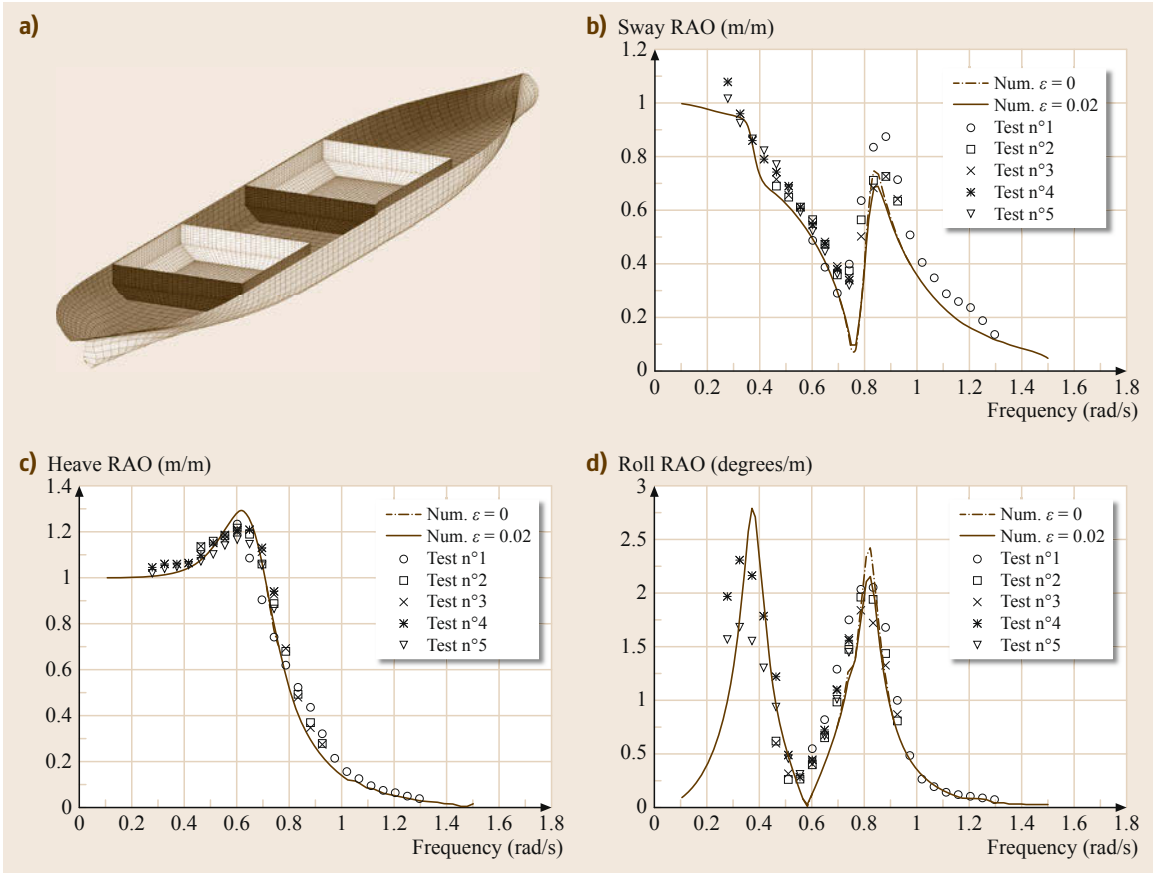


Fig. 41.14 **(a)** Mesh of LNGC and two tanks (after [41.8]), **(b)** sway RAO [$\mathrm{m\,m^{-1}}$] in beam sea, **(c)** heave RAO in beam sea, and **(d)** roll RAO in beam sea. All x-axes: frequency [$\mathrm{rad\,s^{-1}}$]

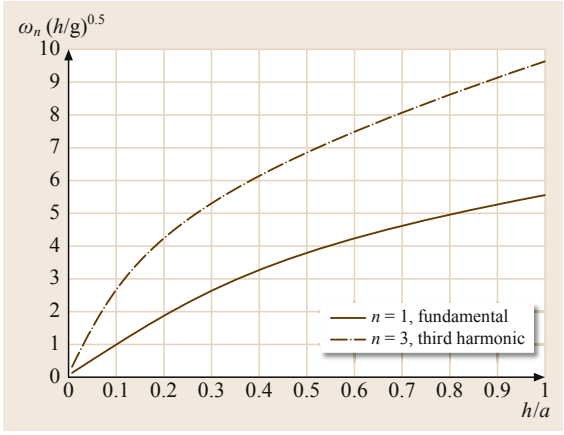


Fig. 41.15 Sloshing natural frequencies as functions of tank aspect ratio

ducted on the nature of the impact pressures. Sloshing pressures are highly stochastic. Even for systematic sway motion of the tank, there is significant variability of sloshing pressure from cycle to cycle. For example, Fig. 41.16 shows a time series of the pressure measured during an experiment. Such pressure signatures seen in this figure are typical of many reported papers.

While the magnitude of impact is of primary consideration, it is also important to note the time scale within which the impact occurs. Details of the formation and disintegration of the pressure peak is shown in Fig. 41.17 taken from *Pistani* and *Thiagarajan* [41.12]. The temporal quantities like rise time and decay time are also stochastic, however, with much smaller standard deviation than the impact magnitude.

High pressure is normally evidenced in the top corners of a rectangular tank, and hence it is common practice to provide a chamfer to the tanks (Fig. 41.2). Experiments show reduction in peak pressure when the corners are chamfered. For example, *Pistani* et al. [41.13] compare the pressures measured during an experiment at fill equal to 30% of tank height. There is notable decrease in pressure due to chamfer.

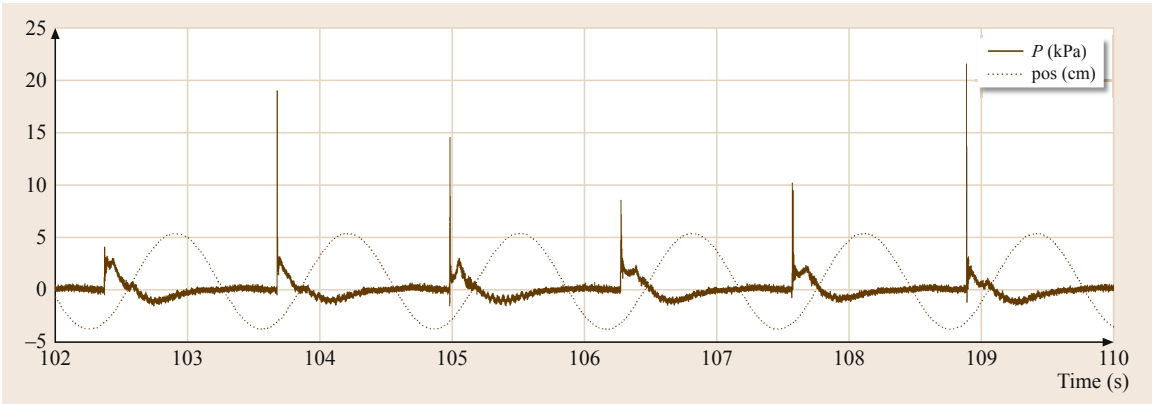


Fig. 41.16 Pressure trace from a transducer (in kPa) due to sloshing from a sway motion of a tank (pos in cm) (after [41.12])

Conventional wisdom [41.10] suggests that reducing the dimensions of the tank can reduce the effects of sloshing, and this principle has been adopted in SPB tanks (Sec. 41.1.1, *Independent Self-Supporting Tanks*) and in developments such as Shell's Prelude, where two rows of tanks are employed. Nevertheless, the accepted design practice [41.14] is to conduct model tests to ascertain the level of sloshing impact pressures and incorporate those in the design.

41.4.3 Scaling

An important pathway to developing robust full-scale design is via scaled model tests in a laboratory. Of various factors that affect design, three aspects have been tested quite rigorously in an ocean engineering laboratory:

1. Liquid sloshing
2. Coupled motions between a floater
3. Liquid cargo and vessels moored side by side.

The starting point of a scaling discussion is dimensional analysis. *Bass* et al. [41.15] have presented an excellent overview of the factors affecting the scaling of sloshing loads. Let us consider the sloshing base case – a rectangular tank in sway motion. The tank length is L, and its filling height is h. Table 41.1 summarizes all important parameters, and also points to the complexity of scaling all the parameters in a single experiment.

One can identify several nondimensional parameters from this list of parameters and some of them are expressed in Table 41.2. *Bass* et al. [41.15] have obtained the following relation for the impact pressure (P_w), which also shows resemblance to the list in Table 41.2:

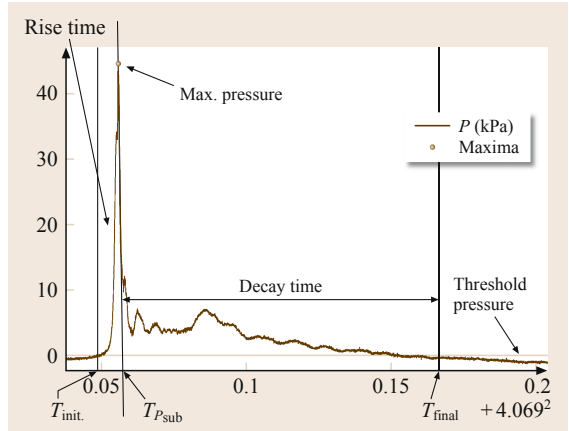


Fig. 41.17 Definitions of pressure impact during sloshing (after [41.12])

$$\frac{P_w}{\rho_l g L} = f'\left[\frac{\rho_l g^{\frac{1}{2}} L^{\frac{3}{2}}}{\mu_l}, \frac{P_u}{\rho g L}, \frac{E_l}{\rho g L}, \frac{P_u - P_v}{\rho_l g L}, \frac{\rho_g}{\rho_l}, m_0, \frac{\rho_l g L^2}{\sigma_l}, \frac{\rho_l C_l (t_w - t_{sat})}{\rho_v h_{fg}}, \text{wall properties, geometry}\right] . \tag{41.7}$$

Yung et al. [41.16] defined a parameter interaction index (Ψ), to characterize the effect of ambient vapor on the magnitude of impact. By introducing the pressure–density relationship for a polytropic process ($\frac{p}{\rho^\kappa}$ = constant) into the dimensionless unsteady Bernoulli equation, the interaction index is obtained as

$$\Psi = \frac{\rho_g}{\rho_l}\frac{\kappa - 1}{\kappa} . \tag{41.8}$$

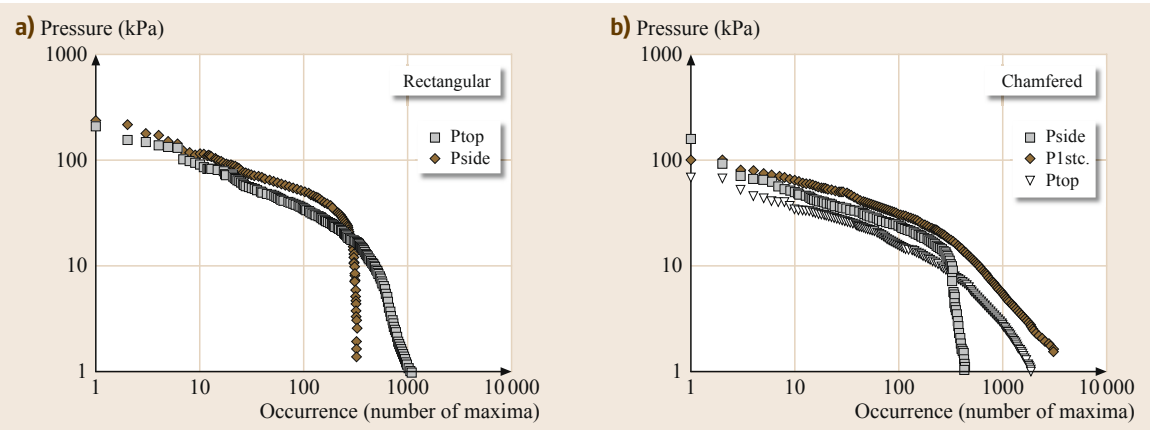


Fig. 41.18 Maximum of the pressures impacts for rectangular and chamfered tanks from *Pistani* et al. [41.13], using transducers on side wall, first corner of chamfer and top (ceiling) fill level here is 30% of tank height

Table 41.1 Key parameters affecting a sloshing experiment

Geometric parameters	**Symbol**	**SI units**
Length of the tank	L	m
Tank height	T	m
Water depth	h	m
Wall inclination	α	rad
Kinematic parameters		
Acceleration	g	$\mathrm{m\,s^2}$
Angular frequency	ω	$\mathrm{rad\,s^{-1}}$
Amplitude of oscillation	A	$\mathrm{m\,s^{-1}}$
Speed of sound	C	$\mathrm{m\,s^{-1}}$
Dynamic parameters		
Ullage pressure	P_u	$\mathrm{N\,m^{-2}}$
Bulk modulus of liquid	E_l	$\mathrm{N\,m^{-2}}$
Bulk modulus of vapor	E_v	$\mathrm{N\,m^{-2}}$
Ullage density	ρ_u	$\mathrm{kg\,m^{-3}}$
Liquid density	ρ_l	$\mathrm{kg\,m^{-3}}$
Viscosity of the liquid	μ_l	Pas
Wall pressure	P_w	$\mathrm{N\,m^{-2}}$
Wall elasticity	E	$\mathrm{N\,m^{-2}}$
Vapor pressure of liquid	P_l	$\mathrm{N\,m^{-2}}$
Thermodynamic parameters		
Ullage temperature	T_u	K
Liquid temperature	T_L	K
Thermal diffusivity of liquid	α_l	$\mathrm{m^2\,s^{-1}}$
Heat transfer coefficient of liquid	h_l	$\mathrm{W/(m^2\,K)}$
Thermal conductivity of liquid	k_l	$\mathrm{W/(m\,K)}$
Coefficient of expansion of liquid	β	$\mathrm{K^{-1}}$
Specific heat of liquid at constant pressure	C_p	$\mathrm{J/(kg\,K)}$
Specific heat of liquid at constant volume	C_v	$\mathrm{J/(kg\,K)}$
Surface tension	σ	$\mathrm{N\,m^{-1}}$

Table 41.2 Some important nondimensional numbers governing the physics of sloshing

Dimensionless number	**Definition**
Froude number	$\frac{A^2\omega^2}{gL}$
Euler number	$\frac{P_\mathrm{u}}{A^2\omega^2\rho_\mathrm{l}}$
Compressibility number of ullage gas	$\frac{\omega A}{\sqrt{\frac{E_\mathrm{v}}{\rho_\mathrm{v}}}}$
Density ratio	$\frac{\rho_\mathrm{l}}{\rho_\mathrm{u}}$
Reynolds number	$\frac{\rho_\mathrm{l}\omega AL}{\mu_\mathrm{l}}$
Prandtl number	$\frac{\mu C_\mathrm{p}}{K}$
Weber number	$\frac{\rho_\mathrm{l}A^2\omega^2L}{\sigma}$
Bond number	$\frac{(\rho_\mathrm{l}-\rho_\mathrm{u})L^2\mathrm{g}}{\sigma}$
Mach number	$\frac{L\omega}{C}$
Cavitation number	$\frac{P_\mathrm{u}-P_\mathrm{v}}{0.5A^2\omega^2\rho_\mathrm{l}}$

Here κ is the polytropic index of the ullage gas whose density is ρ_g. The liquid's density is given as ρ_l. The interaction index captures the condensation thermodynamics when gases are trapped in the flow causing entrapped pockets. Neglecting condensation can lead to exaggerated dynamics of entrapped air, such as springing oscillations. They argue that complete dynamic similitude is possible for the sloshing problem if the five parameters: Euler number, Froude number, Reynolds number, Weber number, and the interaction index are identical.

More commonly used tandem offloading method employs transfer from the stern of one vessel (e.g., FLNG) to the bow of a shuttle tanker. Hoses can be

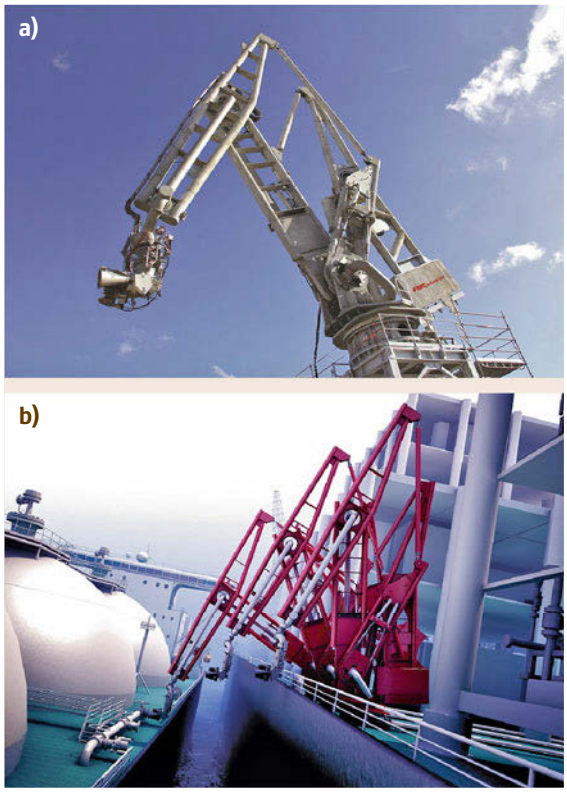


Fig. 41.19a,b Offshore loading arm (courtesy of FMC Technologies) **(a)** demonstrating flex joint technology, and **(b)** attached to an LNGC for cargo transfer

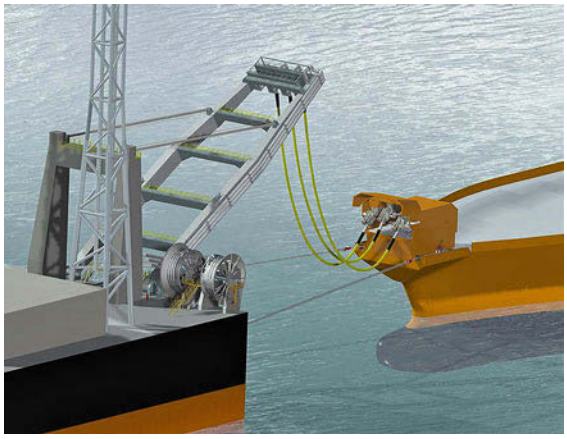


Fig. 41.20 Aerial hose tandem offloading, offshore cryogenic transfer system (courtesy of OneSubsea)

aerial, floating, submerged, or rigid arm arrangements (Fig. 41.20). For tandem offloading, relative motions can be large when conducted in diverse seastates. Recent Joint industry projects (JIP) have resulted in development of customized 8″ and 16″ cryogenic hoses

Table 41.3 Final dimensions and stiffness properties of a hose (after [41.17])

Inner diameter	16″ (406 mm)
External diameter	27″ (686 mm)
Weight – empty in air	88 kg m^{-1}
Weight – full of LNG	160 kg m^{-1}
Storage bending radius	4.5 m
Operational bending radius	10.0 m

for transfer of LNG. These are multi-sheathed pipes capable of handling fluids at −163 °C, large volumes at nominal pressures of 3–10 atm. Sample segments have been tested extensively in dynamic and load conditions. *Cox* et al. [41.17] provide the final dimensions and stiffness properties of a hose presented in Table 41.3.

One of the challenges with this hose is to balance the working weight with the length stipulated by the bending radius. Other JIPs have reported 16″ internal diameter (ID) pipe development with vacuum insulation technology to enable cryogenic transfer [41.18]. Several guides exist to assist with development of offloading system design and operation. *DNV* [41.19] lists the following:

- OCIMF/SIGTTO (Oil Companies Internatinal Marine Forum/Society of Interntional Gas Tanker and Terminal Operators):
  - Ship to ship transfer guide (liquefied gases)
  - LNG STS transfer guidelines
  - Standardization of manifold for refrigerated liquefied gas carriers (LNG).
- Euronorm:
  - EN1474-1 installation and equipment for liquefied natural gas. Design and testing of marine transfer systems. Design and testing of transfer arms.
  - EN1474-2 requirements for flexible hoses.
  - EN1474-3 risk-based qualification of offshore transfer systems.

Cryogenic cargo transfer poses challenges compared to conventional fuels. Some of the challenges [41.19] relate to safety of containment of cryogenic products, ramp-up and shut down operations, and ability to transfer LNG and vapor return within the limited time window of operation. Side by side offloading is currently applied to calm sea conditions, where relative motions between vessels are small. The vessels are separated by fenders, whose diameters are in the range of 5 m [41.19]. Common transfer hoses are aerial, or supported by rigid arms with swivel connections. An example of the flexjoint technology is the Chiksan loading arm developed by FMC, (Fig. 41.19) [41.20] is deployable. The arm uses a hydraulic quick connect

disconnect system, with a constant tension guide line connecting the arm and the manifold. The system can accommodate up to 4 m heave and 30 m drift between two vessels.

41.4.4 Multiple Body Interaction in Shallow Water

When two structures are floating in proximity, important hydrodynamic interactions can occur. Some of the more obvious ones are related to phase and magnitude difference in motions among the structures. When horizontal plane motions are out of phase, then station-keeping problems such as the structures drifting close to each other can arise. Alternatively, structures drifting apart can raise the tension in connecting hoses or guy lines to critical levels. In some cases, snap loads can occur. Phase difference in vertical plane motions such as heave and pitch can result in difficulties in cargo and personnel transfer.

Less obvious effects caused by proximal structures relate to changes in hydrodynamic pressure distributions, resulting in changes in added mass and damping coefficients. Furthermore, when structures are held close to each other *gap* problems such as resonant fluid motions can occur. Based on experimental observations, *Buchner* et al. [41.22] note that head sea waves can get amplified in the gap region. This can result in spikes in RAO and quadratic transfer function (QTF) of individual structures. Figure 41.21 from *Buchner* et al. [41.22] shows a large spike in heave RAO of a shuttle tanker when moored alongside a floating production storage and offloading (FPSO) (vessel). Due to motion coupling, a spike at the same frequency is noticed in pitch as well.

Simulations of gap problems can result in unrealistic free surface motions, which can be considered as inherent drawbacks of the state-of-the-art diffraction methods. These methods rely on discretized integral equation formulations that have limitations on the minimum spatial dimensions which can be handled without significantly increased discrete surface elements or panels. See [41.23] for an excellent overview of early works on this subject. Several authors have followed *Chen* [41.24] in imposing an artificial lid on the gap, which is equivalent to imposing an artificial damping on the free-surface motions at the gap. The boundary condition on the free surface at the gap is corrected by imposing a damping term (defined by ε).

$$-\omega^2\phi_s + g\frac{\partial\phi_s}{\partial z} - i\varepsilon\omega\phi_s = 0\,, \tag{41.9}$$

where φ_s is the scattering or radiation potential and ω is the wave frequency. The difficulty is in knowing the

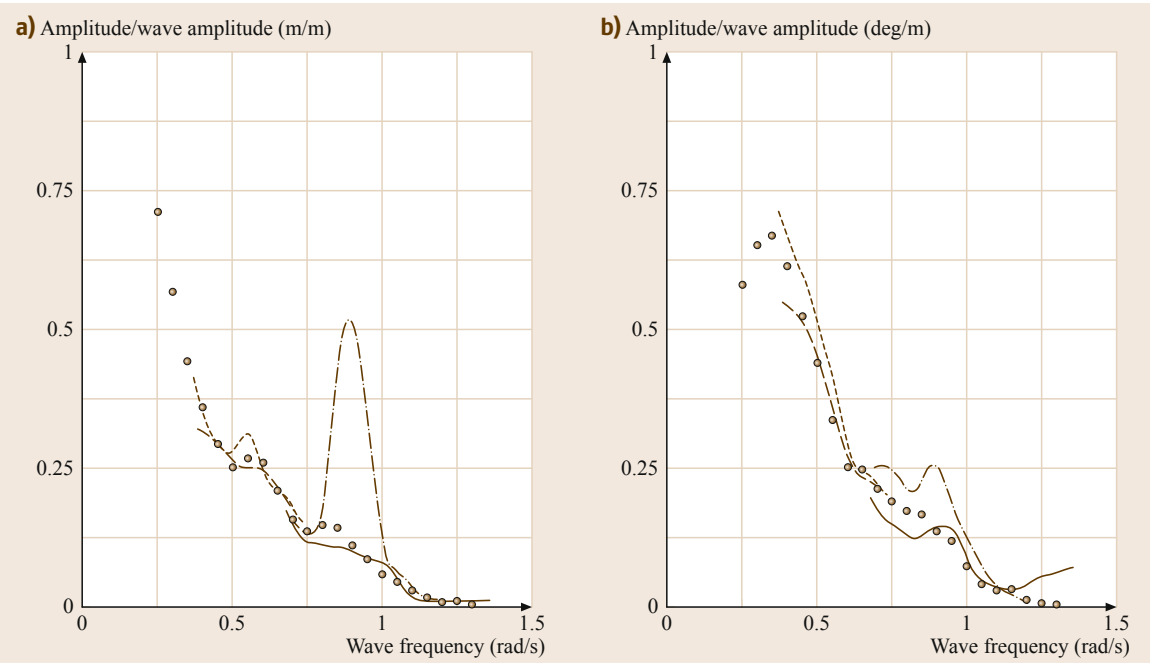


Fig. 41.21 Heave and pitch RAO for a shuttle tanker moored alongside a LNG FPSO (after [41.21]). Model test data at sea (*solid line*) and swell (*dashed line*) frequency domain analysis (*circles*), and time domain analysis in sea and swell (*dashed-dot line* and *long-dash line*)

exact value of damping to be imposed. In fact, *Pauw* et al. [41.25] show from comparison of experiments and simulations that no single value of damping can be universally applied. The optimal value has to be chosen based on the influence on the second-order drift forces. On the other hand, *Lu* and *Chen* [41.26] suggest that a damping value of 0.4 gives good agreement with linear transfer functions obtained from experiments and CFD.

The situation gets more complicated in shallow water, as the effect of the bottom further complicates the flow around the structures, resulting in changed viscous damping. For example *Buchner* et al. [41.22] show through experiments and simulations that the damping due to bilge keels of an LNGC is altered when placed in proximity to a concrete gravity structure, like a dock.

41.5 Design Methodologies of LNG Containment Systems

The design of LNG carriers and floating LNG facilities require additional unique analysis techniques that general cargo carriers are unencumbered by, due to the complex nonlinear physics involved in LNG sloshing. The crux of the issue, if not already clearly evident from the previous section, is the prediction of LNG sloshing loads from smaller scale model tests. Traditional scaling laws such as Froude, Reynolds, and Euler do not perform satisfactorily. In particular, if Froude scaling were to be applied, predicted prototype sloshing impact pressures would be extremely large and do not correlate to the rate of containment damage seen in the LNGC service histories. Such scaling laws taken in isolation obviously do not take into account thermodynamic and fluid–structure interactions. Further, micro and macro bubble dynamics play an important cushioning effect, an accounting of which cannot be complete without consideration of phase-change and cavitation, due to the fact that LNG is transported in a constant state of boiling.

In order to circumvent some of these complexities, the technique used to certify the design of new membrane LNG carriers by class societies is known as *comparative analysis* which relies on the past performance of a reference carrier to predict the performance of a new carrier based on tank sloshing tests. The method proved adequate but was not conducive to the progression of carrier size and thus the largest carriers that could be built for a while were roughly around 138 000 m^3 in capacity. More recent Qflex and Qmax carriers, with capacities of 216 and 260 k m^3, were designed using enabling research, performed by ExxonMobil corporation, which represents a significant leap forward in the fundamental understanding of the sloshing processes.

In contrast to the membrane LNG tanks, the design requirements for LNG carriers with independent tanks are much less onerous, largely due to the reduced sloshing risk or innate strength of the containment system. For instance, whereas ABS has several publications relating to the design guidance for membrane LNG carriers, in the case of their publication [41.14] on building and classing liquefied gas carriers with independent tanks, sloshing is briefly discussed in a subsection related to dynamic loading criteria. The remaining portion of this section will deal largely with assessment of the membrane type containment systems.

41.5.1 Comparative Analysis

Comparative analysis is a technique of assessing the performance of a new-build LNG carrier using membrane containment systems by scaling the sloshing loads from a reference carrier to the new design. The reference carrier is selected to have a known service history in which no damage to the containment systems is to have occurred. In essence, the new carrier is designed to have a higher load resistance than the reference carrier and since the reference carrier did not suffer any damage, it is assumed that the new carrier will be less likely to. Logically, this technique works best when the new design does not differ significantly from the reference design. Therefore, significant changes in geometry and/or capacity are impeded using this technique. The major classification societies have published guidance documents regarding this analysis methodology, such as *ABS* [41.27], *BV* [41.28], and *DNV* (Det norske Veritas) [41.19]. Figure 41.22 shows how comparative analysis is utilized in ABS' current strength assessment procedure for membrane-type LNG containment system.

The two main steps in comparative analysis are the determination of sloshing design loads and the containment strength assessment for both the new and reference carrier. Traditionally, the determination of sloshing design loads is performed solely via small-scale model testing. With modern advancements in computing, CFD is increasingly used to validate and augment the model tests. However, due to the statistical nature of sloshing loads, the computational burden to completely replace model testing makes this an impossibility at the time of this writing. In addition, while CFD is fairly accurate at predicting the spatial distribution of sloshing loads, the actual peak pressure is

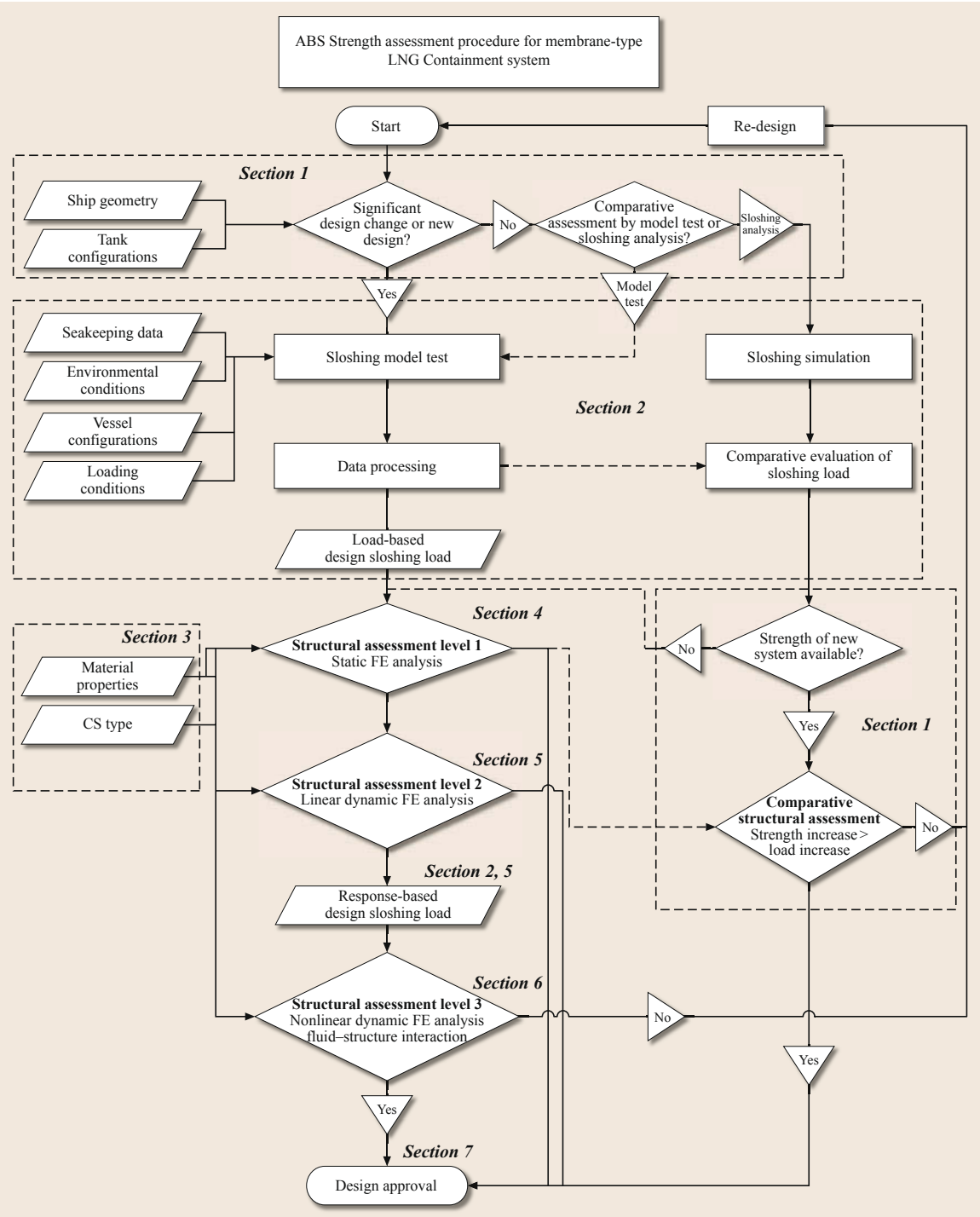


Fig. 41.22 American Bureau of Shipping (ABS) strength assessment procedure for membrane-type LNG containment system

difficult to capture accurately and depends largely on the knowledge of the CFD practitioner. Model testing too is not without its challenges with regard to the special requirements that sloshing testing imposes. This will be covered in a later section. At this point, it is important to note that the reference and new service history needs to be accounted for, including ship motion characteristics and service metocean environment, especially if comparative analysis is being used to qualify an older tanker for new service.

Sloshing impact is a dynamic process and thus the dynamic structural capacities for the different failure modes must be assessed, for the containment system on both the reference carrier and the new design. This is generally carried out using dynamic finite element analyses and needs to account for the different design features and material properties of the membrane CCS type (NO96, MkIII, CS1, etc.). Detailed procedures can be found in the guidance documents mentioned above, as well as simplified methods that may be applicable.

Once the sloshing design loads and the containment system capacities have been determined for both the systems, the comparative analysis can be performed. Here we examine the stated procedure from *BV* [41.28] which is carried out in two stages. In the first stage as shown in Fig. 41.23, a scaling factor λ for the sloshing loads is determined.

$$\lambda = \min\left(\frac{C_{\text{ref}}}{P_{\text{ref}}}\right) , \tag{41.10}$$

where C_{ref} is the full-scale dynamic capacity of the containment system (versus surface) used in the reference vessel and P_{ref} is the design sloshing loads for the reference vessel obtained from the model tests. λ is chosen such that the scaled design load curve is tangent to the capacity curve.

In step 2, the load-scaling factor λ is applied to the design sloshing load curve of the new design, P_{target}. The new design is deemed acceptable if the capacity exceeds the scaled sloshing loads after incorporating a safety factor (SF) (Fig. 41.24). That is, if

$$C_{\text{target}} \geq \text{SF}\lambda P_{\text{target}} . \tag{41.11}$$

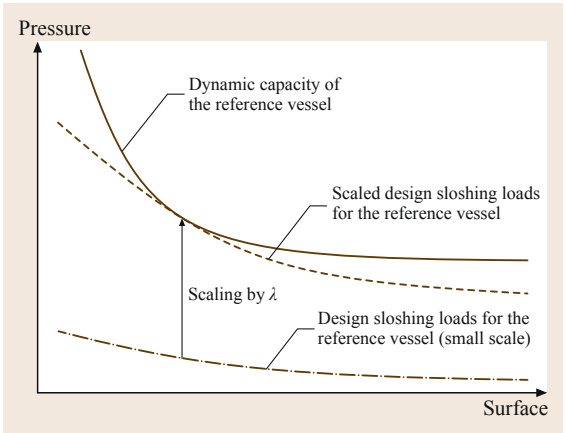


Fig. 41.23 Scaling of design sloshing loads for reference vessel

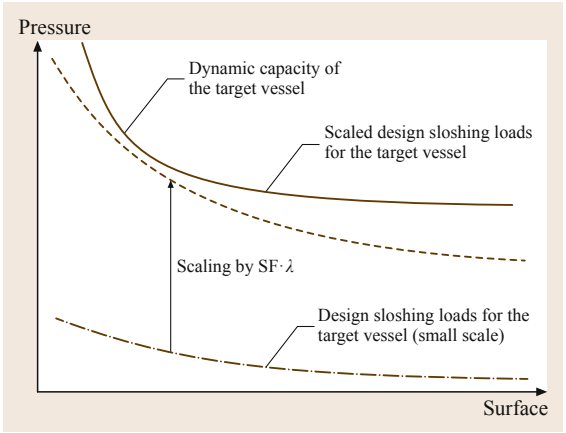


Fig. 41.24 Scaling of design sloshing loads for target vessel

41.5.2 Model Testing

Sloshing model tests are the favored technique to obtain sloshing loads. A tank geometrically similar to the specific cargo tank being studied is mounted atop a 6 degree of freedom platform like a hexapod or Stewart platform, (Fig. 41.25). The motion of the platform is a scaled synthetic time history of the vessel motion over an expected extreme seastate. Clusters of pressure sensors are located at positions where extreme pressures are expected. The traces from the pressure sensors are post processed to determine rise times, peak pressures, and decay times. A statistical model is built up of the sloshing impact pressures and used to determine the design sloshing impact loads for the containment system.

Tank

It is essential that the sloshing test tank should be an accurate representation of the actual cargo tank. A typical LNG carrier with a membrane containment system will have several prismatic tanks and a wedge-shape tank at the bow of the carrier. The sloshing impact types experienced by the latter are different from the prismatic tanks and are induced by a different mode of motion as well. Regarding the geometric scale of the test, it is

Fig. 41.25 Sloshing test using hexapod (courtesy of SYMETRIE)

recommended that the model scale should be as close as possible to the prototype scale in order to minimize scale effects. Obviously, there are practical limitations to the size of the model test tank such as the laboratory space available but more importantly the power required to excite the tank accurately according to the specified motion. Test tanks are generally designed to be transparent or with view ports to observe physical sloshing. However, steel or aluminium may be used to provide the necessary strength in the case of large tanks or a pressurized/depressurized tank.

Fluid Selection and Ullage Pressures

The typical gas–liquid combination used in sloshing tests is air–water. Further, the pressure in the tank ullage space is uncontrolled, that is, at atmospheric pressure. However, it is known that gas bubble cushioning has a significant effect on the impact pressures, with gas–liquid density ratio being an important parameter. The density ratio for air–water is significantly different from LNG–methane. Another phenomenon that may affect the cushioning effect is the bubble collapse as the methane condenses back into the bulk liquid. To try to account for some of these effects, various alternative gas–liquid combinations have suggested. Some combinations that can be found in the literature include SF_6-water, air-boiling water, and CO_2-carbonated water. Pressurized and depressurized ullage spaces have also been investigated.

Sensors, Sensor Distribution, and Data Acquisition

As Fig. 41.26 shows, a sloshing tank is generally outfitted with several arrays of pressure sensors. It is generally recommended that individual pressure sensors have as small a sensing surface as possible in order to capture a wide range of pressure versus loaded area. Depending on the geometric scale, even a sensor with a 3 mm diameter sensor may represent a large area in prototype scale. Sensors should also have an appropriate dynamic range and response, as model scale sloshing pressures increase in milliseconds and recorded peak pressures are at several bar which is typical. It is important that the sensor performance is prequalified prior to the sloshing test. For example, most off-the-shelf pressure sensors are designed for single-phase applications and the sensor might experience thermal shock each time it is wetted as a wave impacts it, leading to a spurious drop in pressure trace which is not physical in nature. Wedge drop tests are sometimes used to qualify the sensors. Special care should be paid in mounting the sensor flush with the inner surface of the tank as any protrusion or cavity will affect the fluid flow around the sensor and hence the pressure trace itself.

Peak sloshing pressures are generally expected at several locations within the tank, especially at corners and knuckle points. Peak pressures are also inversely proportional to the size of the loading area, that is, very high pressures can be recorded over a small area but are smaller when averaged over a larger area. It is important to investigate both area extremes as smaller areas

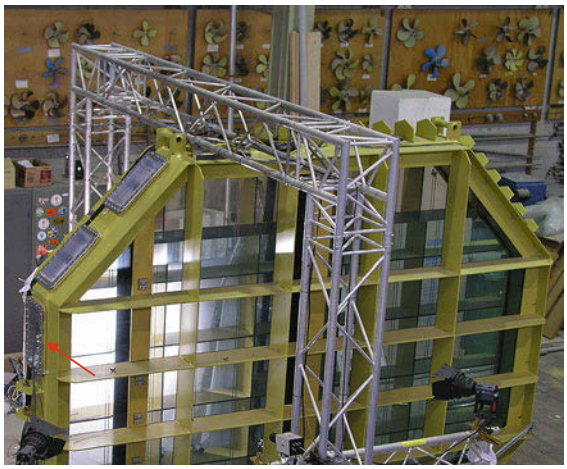


Fig. 41.26 Pressure sensor array on two-dimensional (2-D) sloshing tank (photo courtesy of MARIN)

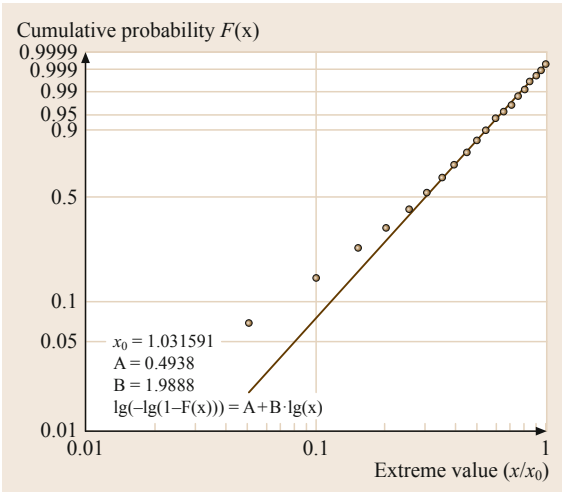


Fig. 41.27 Typical Weibull tail fitting

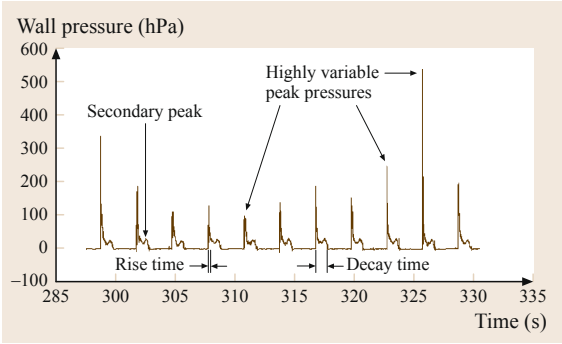


Fig. 41.28 Characteristic sloshing impact pressure trace

require a larger *local* pressure to induce a structural response and the converse is true for larger areas. For this reason the sensors are deployed in arrays of 3×3 or larger. The relationship between peak pressure and loaded area, such as that shown in the above section, can then be easily investigated by averaging over different combinations of pressure sensors.

As mentioned above, model scale sloshing impacts have rise-times in milliseconds, therefore the data sampling rate is recommended to be at least 20 kHz in order to accurately capture the peak of each sloshing impact. Such a high sampling rate in turn requires a specialized high-fidelity data acquisition system. Due to the large number of sensors and high sampling rate, the data storage strategy requires careful consideration. Some techniques that have been employed are to compute and store the data statistics only and to discard the actual pressure time trace. Triggering methods have also been used to store only data in a window centered around each impact. This is a very efficient method as much of the pressure trace show little sloshing activity. Of

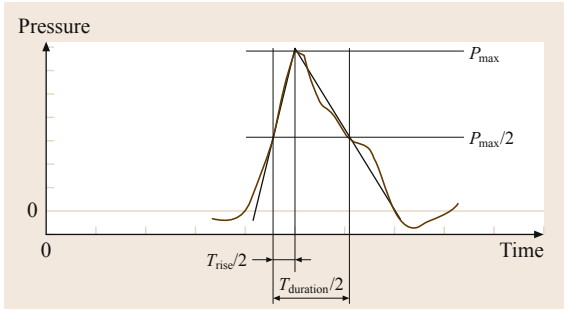


Fig. 41.29 Typical sloshing impact pressure trace with idealized form

course, if one wishes to correlate the pressure trace with the motion of the tank that is usually sampled at a much lower rate, a scheme to synchronize the low- and high-speed data acquisition has to be devised.

Post-Processing

Numerous individual runs are performed in a sloshing model test campaign. In general, the expected service environment of the carrier is examined to determine the conditions resulting in motions that will induce the most extreme sloshing loads. Each condition is repeated several times and at different fill levels to ensure that the sloshing loads for that condition can be appropriately modelled statistically. Typically, each sloshing run is conducted for 3–5 h prototype time to generate enough sloshing impacts in order to determine an appropriate fit using an extreme value statistical model for the impact pressures. The Weibull distribution (Fig. 41.27) is typically used, although other models such as the generalized Pareto may also be used. Once the extreme value distribution has been fitted, the distribution can be used to predict the expected maximum sloshing pressure for the tested condition at the appropriate probability level.

The individual sloshing traces (Fig. 41.28) may also be processed into an idealized representation for ease of data processing. The most common representation is the triangular impulse shape. The description described here is the *ABS* [41.27] formulation. Figure 41.29 shows a typical impact recorded in the pressure trace. The sloshing impact is characterized by a peak pressure P_{max}, a short rise time T_{rise}, typically only several milliseconds in model scale and the decay time which is measured in tens of milliseconds. As mentioned previously, the sampling rate needs to be sufficiently high to accurately capture the profile of the sloshing impact time history. For each sloshing impact the idealized triangular representation is thus determined from the P_{max} and the time it takes for the pressure to rise from $P_{max}/2$ to P_{max} and from the time it takes to decay from P_{max} to $P_{max}/2$.

References

41.1 Gaztransport and Technigaz: http://www.GTT.fr

41.2 International Finance Corp.: *LNG Liquefied Natural Gas Facilities* (World Bank, Washington 2007)

41.3 J.E. Halkyard, J. Filson, P. Hawkey, K.P. Thiagarajan: Floating structure design. In: *Handbook of Offshore Engineering*, ed. by S.K. Chakrabarti (Elsevier, Amsterdam 2005)

41.4 X. Wang, J.K. Kang, Y. Kim, P.H. Wirsching: Low cycle fatigue analysis of marine structures, Proc. 25th Int. Conf. Offshore Mech. Arct. Eng. (2006), Paper 92268

41.5 M.S. Zakaria, K. Osman, M.N. Musa: Boil-off gas formation inside large scale liquefied natural gas (LNG) tank based on specific parameters, Appl. Mech. Mater. **229–231**, 690–694 (2012)

41.6 S.K. Chakrabarti: *Hydrodynamics of Offshore Structures* (Springer, New York 1990)

41.7 K.P. Thiagarajan, D. Rakshit, N. Repalle: The air–water sloshing problem: Fundamental analysis and parametric studies on excitation and fill levels, Ocean Eng. **38**, 498–508 (2011)

41.8 X.-B. Chen: Offshore hydrodynamics and applications, IES Journal A Civ. Struct. Eng. **4**(3), 124–142 (2011)

41.9 B. Kayal, C.-F. Berthon: Analytical approach to predict sloshing severity in LNG membrane tanks based on optimized series of model tests, Proc. 23rd Int. Offshore Polar Eng. Conf., Vol. 3 (2013) pp. 313–319

41.10 T. Gavory, P.E.D. Seze: Sloshing in membrane LNG carriers and its consequences from a designer's perspective, Proc. 19th Int. Offshore Polar Eng. Conf., Vol. 3 (2009) pp. 13–20

41.11 L. Brosset, Z. Mravak, M. Kaminski, S. Collins, T. Finnigan: Overview of Sloshel project, Proc. 19th Int. Offshore Polar Eng. Conf. (2009), ISOPE-1-09-037

41.12 F. Pistani, K.P. Thiagarajan: Experimental measurements and data analysis of the impact pressures in a sloshing experiment, Ocean Eng. **52**, 60–74 (2012)

41.13 F. Pistani, K. Thiagarajan, D. Roddier, T. Finnigan: Comparison of sloshing impacts for rectangular and chamfered LNG tanks, Proc. ASME and 30th Int. Conf. Ocean Offshore Arct. Eng. (2011), Paper 49452

41.14 American Bureau of Shipping: *Guide for Building and Classing Liquefied Gas Carriers with Independent Tanks* (American Bureau of Shipping, Houston 2014)

41.15 R.L. Bass, E.B. Bowles, R.W. Trudell, J. Navickas, J.C. Peck, N. Yoshimura, S. Endo, B.F.M. Pots: Modeling criteria for scaled lng sloshing experiments, J. Fluids Eng. **107**(2), 272–280 (1985)

41.16 T.-W. Yung, R.E. Sandström, H. He, M.K. Minta: On the physics of vapor/liquid interaction during impact on solids, J. Ship Res. **54**(3), 174–183 (2010)

41.17 P.J.C. Cox, J.-M. Gerez, J.-P. Biaggi: Cryogenic flexible for offshore LNG transfer, Proc. Offshore Technol. Conf. (2003), Paper OTC 15400

41.18 G.F. Clauss, F. Sprenger, D. Testa: *Dynamics of Offshore LNG Transfer Systems in Harsh Seas*, Vol. 104 (Jahrbuch der Schiffbautechnischen Gesellschaft e.V., Hamburg 2010)

41.19 Det Norske Veritas: *Classification Notes No. 30.9 Sloshing Analysis of LNG Membrane Tanks* (Det Norske Veritas, Høvik 2006)

41.20 C. Pashalis: Latest developments in offshore FMC loading systems, LNG Journal **July/August**, 20–21 (2004)

41.21 B. Buchner, A. van Dijk, J.D. Wilde: Numerical multiple-body simulations of side-by-side mooring to an FPSO, Proc. 11th Int. Offshore Polar Eng. Conf. (2001), Paper 2001-JSC-286

41.22 B. Buchner, G. de Boer, J. de Wilde: The interaction effects of mooring in close proximity of other structures, Proc. 14th Int. Soc. Offshore Polar Eng. (2004), Paper No. JSC-364

41.23 R.H.M. Huijsmans, J. Pinkster, J. de Wilde: Diffraction and radiation of waves around side-by-side moored vessels, Proc. 11th Int. Offshore Polar Eng. Conf. (2001)

41.24 X.-B. Chen: Hydrodynamic analysis for offshore LNG terminals, Proc. 2nd Workshop Appl. Offshore Hydrodyn. (2005)

41.25 W.H. Pauw, R.H.M. Huijsmans, A. Voogt: Advances in the hydrodynamics of side-by-side moored vessels, Proc. 26th Int. Conf. Offshore Mech. Arct. Eng. (2007), OMAE2007-29374

41.26 L. Lu, X.-B. Chen: Dissipation in the gap resonance between two bodies, Proc. 27th Int. Workshop Water Waves Float. Bodies (2012)

41.27 American Bureau of Shipping: *Guidance Notes on Strength Assessment of Membrane-Type LNG Containment Systems Under Sloshing Loads* (American Bureau of Shipping, Houston 2009)

41.28 Bureau Veritas: *Strength Assessment of LNG Membrane Tanks under Sloshing Loads Guidance Note N1564 DTR006* (Bureau Veritas, Neuilly Sur Seine 2011)

42. Salvage Operations

Michael S. Dean

Salvage engineering is a broad-based, interdisciplinary field. Salvage engineers must have a firm grasp on the principles of naval architecture to be able to assess the strength and stability of a damaged vessel. A working knowledge of strength of materials, mechanics, dynamics, and structures is requisite, as is some understanding of soil mechanics, fluid dynamics, coastal processes, safety engineering, and the theory and practice associated with rigging systems design and operation, pumping operations, compressed air system design and operation, metals fabrication, industrial processes, and explosive use. An effective salvage engineer will also be familiar with ship operations, deck seamanship, machinery operation, diving, and oil-spill remediation. That said, an often quoted anecdote states that:

> *Ship salvage is a science of vague assumptions based on debatable figures from inconclusive instruments, performed with equipment of problematic accuracy by persons of doubtful reliability and of questionable mentality.*

The referenced publications contain thousands of pages of information developed from decades of experience and operations. Anyone undertaking marine salvage operations would be well advised to spend time within these publications to leverage the wealth of that knowledge. The material presented throughout this chapter is from a set of the United States Navy's Salvage Technical Manuals (Vol. 1–3) [42.1–3] published by Commander, Naval Sea Systems Command, Office of the Director of Ocean Engineering, Supervisor of Salvage and Diving.

42.1 **The Casualty and Response** 986
42.1.1 Contract Types 987
42.1.2 Salvage Personnel 987
42.1.3 The Salvage Engineer's Role 988

42.2 **Introduction to Salvage Engineering** 988
42.2.1 Offshore and Coastal 989
42.2.2 Harbor and Inshore 989
42.2.3 Cargo and Equipment Recovery 989
42.2.4 Pollution and Hazmat 990
42.2.5 Wreck Removal 990
42.2.6 Clearance 990

42.3 **Data, Surveys, and Planning** 991
42.3.1 Essential Data and Ship Information 991
42.3.2 Salvage Survey 992

42.4 **Types of Operations** 996
42.4.1 Free Floating 997
42.4.2 Grounded Ships 999
42.4.3 Sunken Ships 1001

42.A **Appendix: Dewatering** 1003
42.A.1 Dewatering Tools and Methods 1003
42.A.2 Compressed Air Dewatering 1003
42.A.3 Recoverable Buoyancy 1004
42.A.4 Vents and Standpipes 1004
42.A.5 Expedient Blowing Fittings 1005
42.A.6 Removing Liquids Other than Floodwater 1005
42.A.7 Combined Use of Air and Pumping 1005
42.A.8 Induced Buoyancy 1006
42.A.9 Expedient Buoyancy Recovery 1006
42.A.10 Cast-in-Place Foam 1007

42.B **Appendix: Common Formulas, Calculations, and References** 1008
42.B.1 Trim and Stability 1008
42.B.2 Hull Strength in Salvage 1011
42.B.3 Lifting and Heavy Rigging 1013
42.B.4 Compressed Air 1016
42.B.5 Pumping 1018
42.B.6 Petroleum, Oils, and Lubricants 1024

42.C **Appendix: De-Beaching** 1028
42.C.1 Refloating Stranded Ships 1028
42.C.2 Moving Stranded Ships 1029
42.C.3 Reducing Ground Reaction 1029
42.C.4 Reducing Friction 1029
42.C.5 Increasing Water Depth 1030
42.C.6 Temporary Reduction 1030
42.C.7 Applying Forces 1031
42.C.8 Tug Bollard Pull 1031
42.C.9 High-Capacity Winches 1032
42.C.10 Ground Tackle 1032
42.C.11 Heaving System Location 1032

42.C.12 Underwater Excavation 1033
42.C.13 Scouring 1034
42.C.14 Air Lifting 1035
42.C.15 Jetting 1035
42.C.16 Dredging 1035
42.C.17 Mechanical Dredges 1036
42.C.18 Hydraulic Dredges 1036
42.C.19 Salvage Dredging 1036
42.D **Appendix: Refloating** 1037
42.D.1 Stability of Sunken Ships 1037
42.D.2 Main Deck Above the Surface 1037
42.D.3 Main Deck Partially Above the Surface 1037
42.D.4 Main Deck Submerged 1037
42.D.5 Longitudinal Stability 1038
42.D.6 Longitudinal Position of Center of Gravity (LCG) 1039
42.D.7 Longitudinal Position of the Center of Buoyancy (LCB) 1039
42.D.8 Longitudinal Center of Flotation (LCF) 1039
42.D.9 Longitudinal Metacenter (ML) 1039
42.D.10 Longitudinal Metacentric Radius (BML) 1040
42.D.11 Height of the Longitudinal Metacenter (KML) 1040
42.D.12 Longitudinal Metacentric Height (GML) 1040
42.D.13 Keeping the Ship Upright 1040
42.D.14 Strength of Sunken Ships 1042
42.D.15 On-the-Side Refloating 1042
42.D.16 Upside-Down Refloating 1043
42.D.17 Righting Capsized Ships 1044
42.D.18 Initial Calculations 1044
42.D.19 Righting Methods 1045
42.D.20 Variable Factors 1046
42.D.21 Offshore Operations 1047
42.D.22 Righting Plans 1047
42.D.23 Calculations 1047
42.D.24 Headframes 1048
42.D.25 Types of Headframes 1048
42.D.26 Connection of Pulling Wire 1050
42.D.27 Connection to Headframes 1050
42.D.28 Connection to the Hull 1050
42.D.29 Padeyes 1050
42.D.30 Special Bollards 1051
42.D.31 Chain Pigtails 1051
42.D.32 Location of Both Hauling and Lifting Points on Hull 1051
42.D.33 Headframe Attachment Position 1051
42.D.34 Floating Cranes and Sheer Legs . 1052
42.D.35 Small-Capacity Buoyant Lift Devices 1053
42.D.36 Hauling System Anchorages 1053
42.D.37 Shore-Based Anchorages 1053
42.D.38 Weight Removal 1054
42.D.39 Weight Removal 1055
42.D.40 Weight Removal Methods 1055
42.E **Appendix: Wrecking in Place** 1055
42.E.1 Wrecking-in-Place Methods and Techniques 1055
42.E.2 Manual Cutting 1055
42.E.3 Mechanical Demolition 1055
42.E.4 Explosive Sectioning 1056
42.E.5 Burying or Settling Wrecks 1056
42.E.6 Problems in Wrecking in Place ... 1056
42.E.7 Environmental Regulation and Constraint 1056
42.E.8 Manual Cutting 1056
42.E.9 Underwater Manual Cutting 1057
42.E.10 Underwater Cutting Processes ... 1057
42.E.11 Surface Manual Cutting 1057
42.E.12 Lifting 1057
42.E.13 Mechanical Demolition 1058
42.E.14 Chain Cutting 1058
42.E.15 Advantages and Disadvantages of Chain Cutting 1058
42.E.16 Cutting Chains 1059
42.E.17 Preparation for Chain Cutting 1059
42.E.18 Chain-Cutting Operations 1060
42.E.19 Improvising Chain-Cutting Systems 1060
42.E.20 Wreck Grabs 1061
42.E.21 Wreck Punches and Chisels 1062
42.E.22 Explosive Cutting 1063
42.E.23 Burial, Flattening, and Reduction of Wrecks 1063
42.E.24 Burial of Wrecks 1063
42.E.25 Flattening Wrecks 1064
42.E.26 Ship Settling 1064
42.E.27 Ship Flattening 1065
42.E.28 Cutting Down Wrecks 1065
References 1065

42.1 The Casualty and Response

More often than not, initial assumptions must be made regarding the characterization of the casualty. Initial reports often contain conflicting information and the salvage team must initiate action for a rapid salvage assessment without having all the facts. Insurance coverage may very well frame the salvage response, affect the salvage contract type and therefore the composition of the salvage team. A cast

of characters: ship owners, underwriters, classification societies, salvage contractors, and regulators will all play a role in any significant salvage operation. Early salvage engineering assessments of structural integrity and stability can often determine the options available to the salvage team, the contract type preferred by the owners and underwriters and the actions of regulators in response to the developed salvage plan.

42.1.1 Contract Types

The type of contract can have a significant effect on the pace and success of the operation. Salvage and wreck removal operations invariably carry a degree of financial risk to the contractor and will affect the demands and expectations of the salvage officer/master on the personnel involved. There are several types of contracts used for salvage and wreck removal:

- *Open Form Contracts*: There are several open form contracts used in commercial salvage especially in cases of offshore casualties of flooding and towing to a port of refuge. These contracts are agreement that the salvor will undertake the salvage operation using his best endeavors with the compensation being settled by arbitration if the salvor is successful. If the salvors are not successful, they receive no compensation, thus the common term No Cure–No Pay for these types of contracts. This contract type will call on the salvage engineer to deliver the technical analysis very quickly and positively as time is of the essence. There is no room for second guessing in situations of emergency salvage and towing.
- *Fixed Price No Cure-No Pay Contracts*: This type of contract is frequently seen in salvage, particularly in wreck removals. Typically, wreck removal is a slower paced operation for the salvage engineer as there is generally time for planning trade-offs and survey opportunities.
- *Fixed Price Contracts*: Fixed price contracts are sometimes used in salvage and wreck removal. In this type of contract, the risk to the contractor is high, and it can be expected that a *risk factor* of up to 100% will be applied to the bid price. This type of contract can lead to significant pressures on the salvage engineer to expedite work and increase the pace of analysis if costs are becoming out of control and progress is not being made.
- *Time and Materials*: The time and materials contract is particularly applicable where there are a large number of unknowns that cannot be accounted for or estimated accurately in salvage planning. This type of contract has the disadvantage of not presenting much incentive to the contractor to finish his work expeditiously and cheaply. These contracts will typically provide the salvage engineer with a lot of analysis work sorting out the trade-offs in the salvage planning.
- *Cost Plus Fee*: Like time and materials contracts, costs plus fee contracts are useful when the salvor's costs are difficult to predict. Fees that are based on a percentage of the costs encourage the contractor to run his costs up, but incentive fees based on performance give the contractor an inducement to reduce costs and maximize his profits – advantageous to all. The salvage engineer can expect cost trade-off analyses.

42.1.2 Salvage Personnel

The salvage officer/master is typically the recognized leader of the salvage effort and the one on-scene person who directs the activity of the salvage team. Salvage team personnel include operations personnel who are expert in rigging and crane operations, fabrication and welding, mechanical and industrial repair and maintenance trades, boat operators, hydraulics and air systems technicians, remotely operated vehicle (ROV) operators and divers who along with the salvage engineering personnel must work together in a cohesive team that takes advantage of the strengths each discipline brings to the project. Shoreside there are expert facilitators and logistics personnel whose job is to procure and deliver equipment. Fitting the team together and getting the most out of each member requires skill on the part of the leadership, and maturity and cooperation from all the team members:

- *Salvage Officer/Salvage Master*: The salvage officer/master typically is a naval officer or master mariner (United States Coast Guard (USCG) licensed) who is fully at home in command of ships and tugs at sea. They must be expert in all of the operational skills required to conduct ship operations and towing. The salvage officer, if not himself a diver, should have a comprehensive knowledge of diving safety, the diving environment, environmental limitations on divers, and decompression practices. Salvage officers must keep the divers' safety paramount; diving safety must never be compromised. The job of salvage officer/master calls for extraordinary leadership skills.
- *Salvage Engineer*: Salvage engineers usually are technical advisors reporting directly to the salvage officer/master and are seldom in direct charge of working a salvage crew. The salvage engineer's

main role is to put numbers and dimensions to the emerging salvage plan and to analyze physical forces of ship buoyancy and stability, strength of material, pumping systems, flow rates, in fact everything involved with implementing salvage techniques. They must have excellent communication skills and an uncanny ability to guess what the salvage officer/master will need in the way of analysis and information to move from a plan concept to execution by the salvage crew. Ideally, they will combine the skills of seasoned naval architects and ocean engineers with that of marine engineers who are fully familiar with all shipboard systems. Salvage engineers are typically degreed Naval Architects, licensed Marine Engineers, and/or a P.E. with specialty training in ocean engineering. It is a great asset to the salvage engineer to have been to dive school.

Salvage teams are typically made up of personnel with years of experience in industrial trades including machinery operations, rigging and crane operations, hydraulics, pumping, and compressed air systems. They must be fully at home in a life at sea. Experience in fabrication and welding is essential on virtually every salvage operation. Salvage operations work around the clock when it is practical to do so and there must be sufficient people with the right skills to make multiple shifts effective. There must also be enough supervisors to supervise multiple shifts effectively. There must be strong supervision on all shifts. The strongest supervisors should be assigned to the night shifts.

- *Divers*: Divers and tenders are an integral and very special part of almost every salvage crew. They carry out a wide variety of mechanical work requiring fine judgment underwater – an environment that is hostile and hazardous. Because of the medium in which they work and the constraints it places on them, divers require much more time to carry out a task than surface workers require for the same task. A good salvage diver requires a high degree of patience, persistence, and a desire to see the job through.

42.1.3 The Salvage Engineer's Role

Salvage engineers bring to a salvage operation a deep technical background, often not balanced by a commensurate amount of experience. Other members of the team may have considerable experience and knowledge of what actions worked on other jobs, but may not have the technical depth to understand why and how they did work, how close they were to disaster, or how much overkill they had built in. Self-styled practical people with extensive experience, inherent common sense, and knowledge of how to do things, and theoretical engineers make a powerful team. They must take the time to listen to one another, understand one another's reasoning, and develop mutual respect.

Salvage engineers may be placed in various positions within a salvage response. They may be advisors to the owners or underwriters where they will analyze a salvage plan to assess its feasibility; they may be part of the salvors team where they will assist the salvage masters in plan development; they may be part of the regulators' team where their analysis will identify plan risk. Regardless of where they fit in the organization, engineers should take the time and make the effort to seek the advice and ideas of the people who will build and install their designs or implement a salvage plan. Minor design changes that do not affect the function of a design may make its construction and installation considerably more effective. Engineers who take the time to listen to and discuss the salvage crew's ideas will find the crew coming to them freely with valuable suggestions. The salvage engineer must not exhibit *pride of authorship* and must be prepared to absorb practical criticism to adapt salvage techniques to the particular job at hand.

A very real hazard exists when engineers and practical people develop a dichotomy and fail to communicate. Salvage operations have been severely disrupted when practical people have modified engineers' designs, unintentionally destroying their strength or integrity. Engineers must get out on deck and down in the holds, observe what is going on, see where the trouble spots are, and make an effort to understand the practical people.

42.2 Introduction to Salvage Engineering

There are no *standard* procedures or set plans in salvage operations; there are only well-known techniques that are employed in series or in parallel to accomplish the salvage goal. The salvage engineer and his team are expected to be conversant and analytically competent in the entire array of techniques used to effect salvage. The salvage operation will be incremental with the first steps generally taken to stabilize the situation (extinguish the fire, stop the flooding, pin the grounded ship, etc.), and then move on to gain control over the vessel by shifting weights, ballasting, patching, etc., and only then free the ship. A good salvage plan based on tried and true techniques will not ensure success, but an operation that is not well thought out and organized in the early

stages has little chance of success. Adequate planning is impossible without a thorough and detailed survey of the casualty and the salvage site. This is seldom accomplished before the initial salvage plan has met the reality of the operation. Often the second casualty of a salvage operation is the salvage plan (the first being the casualty itself). The salvage engineer must approach each casualty armed with a comprehensive knowledge of the general physical principles at play in the type of salvage casualty (grounding, stranding, capsize, flooding and list, collision, etc.) and in whatever setting (Offshore, Harbor, Coastal, etc.) and the best (if limited) information available. To demonstrate the workability of the salvage plan and the sequence of techniques involved, the salvage engineer must evaluate the position and condition of the ship, understand her difficulties, and conceptualize the work and methods to accomplish the aims of the operation.

Salvage is conducted to recover useable or valuable objects from the sea, or to eliminate a pollution hazard. A ship is salvaged because there is a reason for its recovery or removal – not simply because it is wrecked. The goals of a salvage operation influence virtually every decision throughout the operation. Military salvage is constrained by economy; yet, cost is only one of the several factors affecting the decision to undertake a given salvage operation. Tactical, strategic, political, environmental, and other factors also affect the decision. Although salvage engineers seldom decide whether a salvage operation will be undertaken, their evaluation of the technical requirements, difficulties, and probable cost strongly influence those decisions.

A number of authorities have attempted to expound definitive categorizations of salvage and related work, variously based on location, operational environment, or the type of casualty; no classification system can adequately describe all aspects of marine salvage. A brief examination of the different types of salvage work, however, can illustrate how varying conditions can constrain or facilitate operations.

42.2.1 Offshore and Coastal

Coastwise salvage is the refloating of ships stranded in exposed coastal waters. The fate of grounding casualties exposed to surf, swell, currents, and weather will deteriorate with time, often rapidly. A simple grounding on a coastal bar can lead to broaching and with the effects of wind and sea end up high and dry in a matter of hours, often before tug assistance can be brought to bear. Windows of opportunity created by abnormally high tides or fair weather may close abruptly and not reopen for weeks or months, while the casualty deteriorates. Often a simple grounding will evolve into a wreck in place operation. With offshore and ocean salvage, assistance must be prompt and effective to preserve the value of ship and cargo, making offshore salvage perhaps the most difficult type of salvage. Offshore salvage is undertaken from especially constructed and outfitted salvage vessels and tugs. Portable equipment must be relatively compact and lightweight, as it often must be transported by a helicopter or a small boat, and positioned by manual labor. Conditions are often not suitable for the employment of floating cranes, construction tenders, dredges, and similar floating plant which are usually designed for operation in sheltered waters. Furthermore, such equipment may not be immediately available, because of prior commitments or long transit time from the nearest port. Swift, but accurate and thorough surveys are required to ascertain the casualty's condition and to provide input for an engineering analysis to support a timely salvage plan.

A vessel damaged but still freely afloat may require many different services. The most useful service may be to assist the ship's own damage control effort. Measures taken to assist a damage control effort are, for the most part, planned responses to likely situations; any engineering or specific planning must be quick and extemporaneous. After the situation has stabilized, the salvage effort enters a phase where more detailed planning and engineering may be required.

42.2.2 Harbor and Inshore

Salvage of ships stranded or sunk in sheltered waters is termed as harbor or inshore salvage. Casualties in harbors or other sheltered waters are not usually subject to the rapid deterioration characteristic of offshore salvage; unless they obstruct important channels or facilities, time is not a driving factor. There is usually time for thorough surveys and planning, and to arrange for equipment and services from port facilities, widely dispersed salvage units, or contractors. As floating cranes can generally be employed, portable equipment is not as constrained by size and weight as in offshore salvage. Carpenters, stevedores, and general labor can be obtained from local labor pools.

42.2.3 Cargo and Equipment Recovery

In some cases, a casualty may not be worth saving, but her cargo or equipment may have a high priority for salvage, such as high-value cargos, critical war materials, sensitive military items, machinery, weapons mounts, etc. For example, the battleships ARIZONA and UTAH, sunk at Pearl Harbor, were never raised, but most of their guns were removed for use ashore or on new-construction combatants. When planning for cargo

and equipment removal, the integrity and stability of the hull must be evaluated to protect the environment. Machinery removal will often introduce machinery lubricants into the hull and or surrounding waters, which must be addressed in the salvage pollution plan. When there is no free communication with the sea, the hull can provide an excellent envelope to capture fuel and lubricant, and employing internal oil skimming equipment is a crude measure to take before attempting to move or section the hull for removal. There are often pressures to remove cargo prior to refloating a stranded ship. Generally, cargo removal for lightering is inefficient as compared to bunker or seawater ballast removal.

42.2.4 Pollution and Hazmat

Whereas early salvors' focus was entirely on the recovery of ship and cargo, today's environmental awareness dictates that all salvage is aimed at protecting the environment. There are an increasing number of salvage operations that do not involve the recovery of the vessel itself, either intact or in pieces. Many of these operations involve the recovery of oil, hazmat, and other materials that are a threat to the marine environment. Public and Government awareness of our environment has given way to the development of laws and obligations of the responsible party to clean remains of shipwrecks of all pollutants. In addition to the containment of pollutants, salvors must often work to protect sensitive marine ecosystems (e.g., coral reefs), where salvage work is often carried out.

Pollution Prevention and Control

Despite the time-critical nature of many salvage operations, in all but in some cases of military emergency the salvor's attention must be given to the prevention and control of pollution from the casualty. Pollution control and abatement efforts will affect the Salvage engineer's approach to vessel recovery and refloat tasks, for example, in cases where oil tanks have been holed and exposed to the sea it may be necessary to access the tank from the deck to strip the oil from the tank before attacking other operations. This type of action is fraught with hazards and the salvage engineer who is assisting in the planning (pumping flow rates, suction head, viscosity, etc.) the work must be cognizant of how risks of explosion are to be safely mitigated. The salvage engineer must be familiar with the applicable regional and local oil and hazardous substance (OHS) spill contingency plans; salvage plans must address control and cleanup of spills that have already occurred, as well as prevention of further pollution. The salvage engineer gets involved with pollution control in many salvage scenarios including cases where oil and hazardous materials (HAZMATs) are in free communication with the sea (ship is holed), or liquid cargos are being lightered or transferred from a casualty to an attending vessel.

Hazardous Materials

HAZMAT such as flammable or explosives, corrosive, reactive, or toxic cargos are an increasingly important part of any marine salvage operation. Operations involving chemical tankers, barges, or naval combatant vessels have inherent HAZMAT risks. Increasing amounts of HAZMAT are transported to treatment/disposal sites, leading directly to more HAZMAT salvage situations. Also, HAZMAT encountered in the salvage operation can stop or delay work for fear of hazarding workers or the environment. The presence of recognized HAZMAT may prohibit open ocean wreck disposal or scrap sale. In many cases, proper precautions by the salvage team can minimize the impact on the salvage effort. In such cases, HAZMAT removal contractors must be incorporated into the salvage plan.

Ecosystem Protection

Rapid removal of a vessel from its compromised position is generally the best method of mitigating environmental damage. The longer a vessel remains in extremis, the more likely the surrounding ecosystem will be adversely affected by leaking fuel and lubricants, hazardous cargos, or simply mechanical action of the hull on the bottom. The expediency of removal must however be weighed against the damage inflicted on the environment by the salvage process. In evaluating processes and equipment, engineers must consider the effects, for example, of excavators operating in sensitive marshlands or wire ropes and anchors abrading nearby corals.

42.2.5 Wreck Removal

Salvage forces may be called upon to remove hazardous or unsightly wrecks that have little salvage value. Wrecks are refloated or removed by the most feasible or suitable methods available to the salvors, without regard for the salvage value of the wreck. All that is desired is the wreck to be removed; salvors may cut the wreck into easily handled sections or refloat and remove the hulk in one piece, based only on their evaluation of which technique is easier. If the wreck is to be sunk in deep water, however, cast-in-place foam would be a poor choice as a refloating method.

42.2.6 Clearance

Harbor and waterway clearance involves the coordinated removal or salvage of a number of wrecks.

Obstructions may include vessels of all sizes and descriptions that have stranded or sunk with varying degrees of damage as a result of storm, collision, fire, combat, sabotage, or intentional wrecking by retreating enemy forces. In addition, vehicles, rail cars, port equipment, collapsed bridges and piers, and other materials may also require removal.

Clearance work is similar to wreck removal in that concerns over the preservation of the value of obstructing vessels or hardware generally give way to the requirement to restore the waterway to useful operation. Harbor clearance is often part of a larger port rehabilitation effort that may involve elements of the Army, Navy, or Marine Corps engineer branches; the Army Transportation Corps; the Coast Guard; Army and Navy Supply organizations; or civilian contractors. Mine, ordnance, or booby-trap clearance may be necessary in military harbors, combat-obstructed waterways, or ports intentionally wrecked by enemy forces. Various Federal, state, and local agencies will be involved in peacetime clearance operations.

42.3 Data, Surveys, and Planning

42.3.1 Essential Data and Ship Information

The conditions at a salvage site are seldom fully defined in the beginning and often are not ever fully defined during a salvage operation. Conditions change with time and weather. The salvage engineer's information gathering should begin immediately on the notice of involvement and well before the salvage team begins to arrive on scene, and should continue until the operation is complete.

Intact Ship Information

Intact ship information is available in the documents as follows:

Curves of form (displacement and other curves) and/or hydrostatic tables containing as much of the following as possible:

- Displacement in saltwater
- Vertical center of buoyancy (VCB or KB)
- Longitudinal position of the center of buoyancy (LCB)
- Tons per inch submersions (TPI)
- Moment to change trim one inch (MT1)
- Height of transverse metacenter above keel (KMt)
- Longitudinal metacentric radius (BML)
- Areas of waterplane (Awp).

Drawings and other reference material include:

- Trim and stability booklet
- Damage control manual
- Lines drawing, body plan and table of offsets
- Bonjean curves
- Booklet of general arrangements
- Capacity plan
- Docking plan
- Inboard and outboard profiles.

Structural drawings include:

- Midship section
- Shell expansion
- Scantling plans and sections
- Structural inertia and section modulus tables by frame or station
- Weight distribution curve.

Computer model of the vessel from ship owner, classification societies, etc.

Relevant information includes the ship's daily logs to include:

- Ship's log's – draft, trim, and list conditions
- Cargo loadout and tank soundings
- Speed and heading at the time of the casualty.

Many factors influencing a salvage operation vary in relative importance – not all apply to every case. Information required and its significance depends on the type of casualty. Salvage engineers must understand the factors influencing the salvage operation. They must evaluate the relative priority and influence of each factor on the choice of salvage methods. Demands for additional information arise as the salvage plan develops and situations change. It is imperative that the salvor engineer be intimately familiar with the vessel's condition, particularly whether the locations of cargo and liquids aboard the casualty are a result of damage or normal stowage.

Many salvage operations are time-critical – particularly stranding and afloat casualties. There is not always time to wait until conditions are optimum, or until all information has been gathered. Salvage engineers must obtain the best information they can in the time available, estimate information they cannot obtain, perform the necessary analyses to assist in formulating

the salvage plan, and proceed with an analysis that is sufficiently flexible to change as more information is acquired.

42.3.2 Salvage Survey

In a salvage survey, information about a casualty is gathered by inspecting the ship and the surrounding area. The salvage engineer should be prepared to accompany the salvage crew on any and all surveys. The primary purpose of the survey is to gather information for the development of the salvage plan. Information gathering is a dynamic process that is never truly complete. It begins when the first salvor arriving at the salvage site takes stock of the situation. If communications can be established with the casualty's crew or other observers, or documentary information about the ship or area is available, information gathering should begin before salvors arrive on scene, and continue throughout the operation. Full or partial surveys should be repeated as often as necessary for salvors to stay abreast of changing conditions.

The keys to a good survey are the verification of observations and the organization and presentation of the information collected.

A survey will report only observations. It is the salvage engineer's task to interpret the conditions observed to determine their relevance to the condition of the ship. Like a survey, the interpretation of the results must be an ongoing process that continues throughout the operation and is constantly revised as the survey is refined.

Survey Breakdown

The salvage survey breaks down into several interdependent surveys (preliminary, detailed hull, topside, interior, diving, hydrographic, site, safety, and other surveys). How these surveys interrelate, and the extent and importance of each, depends on the type of casualty. Since stranding salvage is the most familiar and possibly the most frequent type of salvage, surveys are first presented as they pertain to strandings. Differences between stranding surveys and those for other types of casualties will be discussed afterward.

Preliminary Survey

The preliminary survey verifies information received from the casualty itself, ships in company, owners, or other observers. All reports should be checked because initial observations may no longer be relevant, or information important to salvage engineer may have been overlooked. A *desktop survey* should be conducted to assemble as much documented information as possible about the vessel, its contents, and the bathymetry of the salvage site. The documented information aids initial evaluation of the situation and provides starting points for detailed surveys. At least, the preliminary survey should:

- Verify the casualty's changes in drafts, trim, and list over several tide cycles.
- Verify bottom soundings reported by the crew – corrected to tidal datum.
- Determine if the vessel is merely aground, or stranded high and dry.
- Verify the casualty's position and attitude on nautical charts.
- Determine if shipboard equipment and systems are functioning.
- Determine if the ship is lively and moves with the tide.
- Verify tank soundings, flooding extent, structural damage.
- Determine if there is progressive flooding.

Aerial Observations

It may be extremely useful for key salvage personnel to fly over the salvage site to observe changes in color and other indications of shoal water. Because of its ability to move slowly or hover, a helicopter is the best platform for this type of aerial survey. Aerial observation provides a view of the overall site (including the shore). Potential beach gear locations, mooring sites, locations ashore for deadmen or highlines, lightering arrangements, retraction channels, etc., can be selected from the air. The suitability of these sites can then be confirmed by a detailed survey.

Initial Information

The preliminary survey should begin before salvage forces arrive and should incorporate reports from the ship's crew. Merchant ship information can be obtained from the owners or their agents. Information on many merchant vessels is also available through the USCG Headquarters, the National Cargo Bureau, and classification society registers. Aerial or satellite reconnaissance of the stranding site can also provide basic information about the casualty. The casualty or ships in company can provide a great deal of information about the casualty and site. Salvors can transmit portions of the standard survey form for completion. Post-stranding drafts should be taken as soon as possible after the incident. Early information forms a basis for preliminary planning and initial estimates of the effort, time, and assets required for the salvage. The salvage assets dispatched to the scene are determined by the information available at mobilization. Whether or not the salvage forces have means to stabilize the casualty immediately can make the ultimate difference between success and

failure. The casualty may deteriorate beyond the point of salvage, or fair weather windows may be lost while waiting for additional equipment or forces. It is better to arrive with too much equipment than too little, but salvage forces and assets are not unlimited. Information obtained before salvors arrive on scene can identify items requiring further investigation so surveys can be focused for the maximum return. Information to be requested from the salvage site includes:

- An accurate position of the stranding site giving latitude and longitude, along with applicable chart numbers, and means of fixing the position.
- Date and time of stranding.
- Drafts on sailing from the last port and estimated at time of stranding.
- Fuel, water, ammunition, and other weights expended or transferred since sailing.
- Post-stranding drafts forward, amidships, and aft, with the state of tide and the time taken.
- Soundings along the entire length of the ship, corrected to the datum of the chart of the area.
- Course and speed at time of stranding and ship's heading after stranding, with details of changes.
- Liveliness of the ship.
- Weather conditions, including wind direction and velocity, current weather, and forecast weather at the site.
- Sea and current conditions, including the direction and height of seas and swells; state of tide, weather, and sea conditions at time of tranding.
- Tide range and reference station.
- Extent and type of damage to the ship.
- Location of grounding points and estimated ground reaction.
- Type of seafloor at the site.
- Status of ship's machinery.
- Status of liquid loading (cargo, fuel, fresh water, ballast).
- Ship's cargo list or manifest, including amount and location of known HAZMAT.
- Assistance available on scene.

On Scene

The detailed on-scene survey refines the preliminary survey and collects the specific information listed in the salvage survey form. The detailed survey consists of five surveys:

- Topside
- Interior hull (including machinery)
- Diving and exterior hull
- Hydrographic
- Vessel attitude.

Topside Survey. The topside survey gathers information about the exterior of the ship above the weather decks. Particular items of concern are:

- Type, location, safe working load, and operating condition of all deck machinery and cargo gear
- Location and estimated safe working load of tug and beach gear attachment points, including working space for pulling devices
- Location and estimated weight of top hamper and superstructure if it appears that topside weight must be removed
- Operating condition of the ship's boats.

Interior Hull Survey. The interior hull survey gathers information about the interior of the ship and its contents. The interior survey includes:

- Examining in detail the condition and contents of every space below the main deck.
- Taking soundings or ullages of all spaces containing liquids.
- Determining the condition of installed drainage system piping, valves, pumps, and other associated equipment.
- Determining the location and operating condition of all cargo and ballast pumps and the arrangement of associated piping and manifolds.
- Determining the location and condition of all cargo and stores and obvious hazards, such as flammables and chemicals.
- Determining the location, weight, volume, and class of all ammunition magazines and the operating status of the magazine sprinkler systems and location of their controls.
- Determining the location of all structural damage, that is, holes, tears, cracks, weeping seams, panting bulkheads, etc.
- Determining the location, type, and estimated weight of loose or displaced cargo, equipment, or permanent/solid ballast that could be removed.
- Investigating items of special interest, such as open decks where the beach gear can be rigged, or identifying locations for necessary openings.
- Determining the availability and location of material that may be useful in salvage.
- Determining the location and size of any cross-connections for liquid tanks that could be closed or left open.

Machinery. Machinery status and condition are of great interest to the salvage engineer. The availability of electrical power, compressed air, deck machinery, pumps, and other equipment can greatly simplify the

task of salvaging the vessel. Operational propulsion machinery can assist the refloating effort and control the casualty once free of the beach. The value of the machinery plant may be a significant portion of the casualty's total value for either commercial or military considerations. Specialized cargos, such as refrigerated goods, may require certain equipment to prevent degradation or hazard formation. Proper operation of ventilation systems can prevent the accumulation of dangerous gases. In some cases, the machinery plant and other equipment may be the object of the salvage operations, with salvage of the hull conducted only to provide a convenient barge for transporting the machinery. Flooding can damage machinery severely, especially if the water level rises and falls, exposing saltwater-drenched machinery to air, or if adequate water damage protection measures are not taken when the space is dewatered.

Diving or ROV Exterior Hull Survey. The diving or ROV survey examines the underwater portions of the ship's hull and the exterior portion of the hull below the main deck. This survey may be conducted by a dive team, or in some cases using an ROV to view the sunken or stranded vessel. Underwater surveys may not be strictly required for all groundings; in some cases, the hull condition can be adequately determined from internal surveys. Diving surveys provide more accurate information concerning the hull and bottom conditions and should be conducted unless time constraints, lack of divers, or sea conditions prevent them. If a diving or ROV survey is not conducted or is severely limited, for whatever reasons, the interior survey must be unusually comprehensive to compensate for the lack of detailed hull and bottom condition information; conclusions about the bottom may have to be drawn from topside observations alone. The diving or ROV survey determines or checks for the following:

- Amount of the hull in contact with the seafloor and a description of the points of contact.
- Existence and location of pinnacles.
- Existence and location of impalements.
- Location and size of all cracks, tears, holes, and indentations in the underwater portion of the hull and in the portion between the waterline and the weather deck.
- Condition of all sea suctions, valves, and fittings, and whether or not they are clear.
- Condition and operability of all underwater appendages, including bilge keels, sensors, stabilizers, rudders, shafting and bearings, and propellers.
- Signs of leaks or escaping fuels, pollutants, or liquids.
- Type of seafloor soil and the presence, location, and extent of scouring or buildup.

Underwater Video. Whenever possible, video should be used for underwater surveys. Video – particularly low-light-level video – has greater sensitivity and can record more detail than the diver's eye. Videotapes can be reviewed repeatedly at the convenience of the viewers; tapes from earlier surveys can be compared with the most recent versions to measure changes. Technical personnel who are not divers can observe underwater conditions through video. Videotapes of areas of the hull in contact with the seafloor, underwater damage, and hull appendages and openings are particularly valuable. Black-and-white video generally provides better detail resolution and low-light performance, but color video may be preferable for general surveys because many seafloor features and damage to hull areas are highlighted by color contrasts, such as coral heads, rock formations, scraped paint, marine growth, etc.

Hydrographic Survey. The hydrographic survey determines the condition of the sea and seafloor in the area where operations will be taking place. Included in the hydrographic survey are:

- Comparison of the observed tides with the predicted tidal information.
- Determination of the strength, period, and times of local currents, and the durations of high and low water slack and their relationship to the times of high and low tide.
- Periodic observations of the sea and swell height, period and direction of seas, and their impact on the salvage operation.
- Soundings all around the ship, in areas where beach gear or moorings will be laid, or where salvage ships or other vessels will operate.
- A seafloor profile chart of the beach gear area to assist in design of the beach gear legs.

Vessel Attitude.

Draft Readings. Accurate stranded drafts are often difficult to obtain, but are essential because they form the basis for many salvage calculations. For instance, the ground reaction calculation is dependent on drafts; ground reaction influences much of the operation. Efforts to obtain accurate grounded drafts are well spent. In swells, draft readings taken from a boat are the most accurate. When the hull girder is significantly deflected, it may be necessary to obtain drafts at several locations along the ship's length to accurately determine buoyancy distribution for strength and ground reaction calculations. If drafts at stations between draft marks

are required (or the draft marks are obliterated), freeboard to the main deck or other convenient deck can be measured with a weighted line or chain. The freeboard is then subtracted from the height of the deck above the keel, as taken from a profile plan or tabulated data. As the sea surface is seldom calm, drafts should be determined by averaging at least three readings at each location.

Whenever drafts are taken, the time, date, and state of the tide are recorded and drafts are reduced to the tide datum; normally that datum is the same as that of charts of the local area. Reduction to the tide datum is required because the draft readings of a stranded ship vary with the tide. When ground reaction at the tide datum is known, the effect of tide on ground reaction can be determined.

Determining Casualty Movement. Early in the salvage operation, it should be determined if the ship is moving. Observation of widely separated navigation ranges – preferably one in line with the casualty and one on the beam – is the fastest way to detect casualty movement. Navigational ranges can be established on natural landmarks, or, if the depth of water permits, range poles can be set up and observed to see if the ship falls off the range. Alternatively, bearings to fixed, easily identifiable objects should be recorded at regular short intervals. When no movement is observed for some time, the interval between readings may be increased. A large pendulum-type clinometer should be set up on the casualty in an easily observed location to show changes in list (or roll, if the ship is lively).

Hull Deflection. Hull deflection can be related to hull stress and bending moment. Deflection can be determined by plotting drafts. It can also be determined by measuring heights relative to a horizontal datum with a surveyor's transit, theodolite, or siting level, or from a taut wire stretched between fore and aft points. The line or curve of deck or other measured surface plotted from the measurements is compared to ship's drawings to determine the amount and direction of hull deflection.

Hull Strain Readings. If the hull is lifting to the rising tide or swells, the hull deflection should be checked. Dial indicators installed between frames measure strain in adjacent longitudinal hull members. Increases in hull strain indicate an increase in hull stresses. Sudden increases in bending stresses may indicate that hull failure is possible, and that changes in loading must be made quickly to reduce stresses. Dial indicators can also show if the hull girder is suffering permanent strain. In many instances, the hull will flex measurably in response to water-level fluctuations or the passage of waves. Flexing will alternately increase and decrease longitudinal strain. In extreme cases, flexing will cause strain reversals, from elongation to compression and back again. If the strain is elastic, the range of dial indicator readings will remain fairly constant and an approximate midrange point can be determined after a few cycles. A steady shift in the midrange point in either the positive or negative direction that cannot be tied to observed water-level fluctuations indicates a change in the overall hull girder deflection. The change may be either increased or decreased elastic deflection resulting from changes in total bending moment, or permanent deformation, indicating that stresses have exceeded the yield limit in that part of the hull girder. If stresses have exceeded the yield limit, the hull is in danger of catastrophic failure. Hull members should be examined for signs of yielding; changes in hull girder loading should be carefully reviewed to determine the cause of the shift in deflection levels. In the same way, dial indicators can be set up between side-shell longitudinals to indicate vertical shear stresses and between deck girders to show horizontal shear.

Damage Reports. Damage reports prepared by the ship's crew can assist in prioritizing areas to be surveyed. Particular attention should be paid to secondary damage, such as abnormal bulkhead flexing, cracked seams, hatches and doors that no longer close, cracking or flaking paint, or other signs of stress or hull distortion. As these conditions may indicate more serious damage, their causes should be determined. The diving team should be briefed on the location and type of all damage found inside the hull, so that they can check for underwater damage in those areas. Frame numbers, spray- painted on the hull, help divers and boat crews to orient themselves along the length of the ship. Weighted lines, hung over the side and into the water at specified locations (frame numbers) help divers orient themselves underwater and locate damage relative to known points. Spreading damage, such as lengthening or widening cracks, increasing or decreasing dimensions of openings and clearances, increasing areas of deformation, or areas of paint cracking and flaking, is especially important because it indicates progressive weakening of the hull or increasing stress.

Tank Soundings. All tanks should be sounded frequently and the soundings compared to those taken before and since the stranding. On most tankers, the contents of cargo tanks are tabulated against ullages (the height of the empty space above the liquid, usually measured from the lip of the expansion trunk or ullage port), rather than soundings (liquid depth). Fuel

and other oil tanks should be checked for the presence of water with indicator paste, thief samples, or by opening the tanks. Some tankers are equipped with ultrasonic sounding devices that can locate the oil surface as well as the oil–water interface. Use of the casualty's tank capacity tables will be greatly simplified if survey crews measure liquid surfaces and interfaces from the same datum used in the tables, that is, soundings or ullages. It may be necessary to brief survey crews used to taking soundings on the correct method of measuring ullages.

Tides. Tides may vary, both in height and time, from those predicted in the tide tables. A tide gage should be set up, and regular readings taken and compared to predicted tides. Local mariners can often provide the best information about tides and currents at the stranding site.

Salvage Plan and Updates

Analysis of the data gathered by the surveys is not of full value until it is used to formulate a salvage plan. The salvage plan lists the work to be done, the salvage techniques to be employed, and matches them with the resources available, schedules it, sets forth the responsibilities of individuals and organizations, and provides a vehicle for the coordination of all salvage efforts to meet target dates and times. Development of the salvage plan begins when the initial information about the casualty is received and continues with updates throughout the operation. The salvage engineer's contribution to the salvage plan is to quantify the information to assist the salvage officer/master in communicating the salvage plan to officials and the salvage crew. The salvage engineer can expect to provide the salvage officer/master with updated calculations at any time during the operation. Often the updates will be required early in the AM local time for briefings. They can be very comprehensive updates and may require a long sleepless night. By way of example in the case of a grounding the engineer may have to provide calculations that include:

1. Ground reaction
2. Freeing force
3. Location of the neutral loading point, if applicable
4. Stability – aground and afloat
5. Strength of the hull girder, damaged areas, attachment points, and rigging
6. A summary of the rationale for the selection of specific retraction and refloating techniques.

42.4 Types of Operations

Salvage operations generally involve three basic casualty circumstances: free floating, grounded, or sunken. A free-floating vessel, whether intact or damaged, partially flooded, or even capsized will have some measure of longitudinal or transverse stability along with a waterplane as well as be lively to the sea. Tidal changes will not affect the casualty's buoyancy or stability. In contrast a grounded vessel, whether just touching ground and trimming or listing to its ground reaction (aground), or stranded and working to the current, surf, wind, and tide is being subject to forces that her designers had never intended and had never built into the design. High structural loads can hole the vessel, and the action of the sea can turn the grounding into a stranding, leaving the casualty high and dry with virtually no residual buoyancy no matter what the state of tide. In an exposed coastal or reef situation, the stranded vessel's predicament can progress from a simple grounding to a hopeless shipwreck in a matter of days. Sunken vessels have insufficient waterplane to float with the tide and are not going anywhere. Depending upon the sunken vessel's angle of repose and extent of ground reaction, the sunken vessel's plight will deteriorate with scouring and an accumulation of silt to add to the submerged weight of the vessel and her cargo. In a harbor or protected water setting, however, sunken vessel have remained for years before being refloated or wrecked in place. Sunken vessels may be grouped according to residual buoyancy characteristics as follows:

- Completely submerged (upright or on its side) or capsized (upside-down) with no residual buoyancy
- Submerged with a small portion of the exposed hull above the water surface
- Partially submerged with a substantial portion of the hull above the surface.

Sunken ships may start to refloat before all of the buoyancy lost in sinking has been recovered. At liftoff from the bottom residual water remaining on board can cause dangerous free surface and problems with stability, list, and local and overall hull girder strength. The salvage of a sunken ship requires not only recovering enough buoyancy to refloat the ship, but also distributing buoyancy to obtain suitable conditions of stability, trim, and strength.

The fundamental perils and physics of each type of casualty will dictate the techniques and equipment to undertake the salvage operation, as well as the tools and

methodology used by the salvage officer/master to control the situation. These tools and methodology selected will invariably involve salvage engineering to quantify the effects of various approaches.

42.4.1 Free Floating

A freely floating vessel, whether upright, or subject to extreme list and trim (or even capsized) will be in a state of stable equilibrium on the water surface. Naturally, as progressive flooding and loss of buoyancy continue for a holed vessel, the margin of stability will deteriorate and a stable situation may become unstable. In a state-of-stable equilibrium, the centers of buoyancy and gravity will be in longitudinal and transverse alignment even at excessive angles of trim and list and metacentric height (GM) will be positive. And despite these large angles of trim and list, or even in a capsize situation, there will be a waterline, with waterplane inertia and an instantaneous metacenter even if the vessel is upside-down and riding on a pocket of air. Free-surface effects of flooded spaces or partially filled tanks (slack) will have a huge impact on the state of trim and list, as well as the dynamic range of stability and the period of roll and pitch, and will contribute greatly as to how quickly the situation may become hopeless and the vessel founders and sinks. By definition, a submerged surface vessel is one without any waterplane and it is a rare case indeed if the fully submerged surface vessel does not touchdown on the bottom.

Battle or collisions damage, even the effects of loss of watertight integrity may lead to the ultimate loss of the vessel due to progressive interior flooding (settling), downflooding (after bulkhead deck submergence), or loss of stability due to a massive loss of waterplane between transverse watertight bulkheads invariably to be followed by capsize or diving when the flooding is largely symmetric between bulkheads. In addition to these effects, the hull girder strength may be compromised and bending and torsional loads may cause the ship to break up. Salvage operations on a free-floating vessel will generally involve continuing the damage control measures taken by the ship's crew at the earliest stages of casualty response, and if the vessel is unable to transit on her own power towing will be involved. Free-floating casualties as a result of fire, explosion, collision, or a massive shift of cargo weight are a particularly difficult form of casualty. For the purposes of this chapter on operations, it will be assumed that the damage control measures are being performed by the ship's crew at the time the salvage engineer's involvement begins. It is also assumed that the ship's condition will be one of excess trim and lists with massive internal flooding, such that the use of ship's standard intact stability data (Trim and Stability Booklet or Cross Curves) are of limited value.

Free Floating – Excessive Trim and List

Excessive list and trim will generally be characterized by loss of buoyancy hence freeboard coupled with unusually great trim or list angles (i. e., drafts and angles not found in the vessel's Cargo Loading Manual or Trim and Stability Booklet). The loss of the margin of stability will be a result of impaired dynamic stability and will usually be sensed by a very long period of roll (lolling) about an equilibrium angle of list, or even upright and apparently in stable equilibrium. The first concern in such situations is to determine the margin of stability and to analyze the consequences of various techniques to improve seaworthiness.

Many of the operational techniques used to improve seaworthiness involve discharge or transfer of liquid loads combined with efforts to minimize or neutralize the effects of large areas of liquid free surface. In cases where the ship's bilge and ballast or cargo transfer systems are operational, the liquid load may be modified to improve overall stability by raising the metacentric height, as in the case of ballasting. Unfortunately, adding ballast will rob the casualty of freeboard and introduce transient free surface effects during the transfer process. There are always trade-offs. For areas of the vessel that are flooded (as in the case of cargo holds), it may be necessary to patch hull or bulkhead penetrations and use salvage pumps to discharge the flooding water. Counterflooding from the sea is another technique often considered and one which will have a trade-off between the good effects of helping to correct off-center weight but will have a negative effect on transient free surface and a loss of freeboard. It is better to shift liquid cargo internally. In any event, techniques such as patching and de-watering will be essential to restoring seaworthiness where there are large areas of free surface.

Development of a lightering plan is essential to establish the sequence of liquid cargo and fuel as well as predict the effects of transient free surface on the vessel's margin of stability. Lightering can also include the discharge of weights from the deck, as in the case of discharging cargo containers from a disabled vessel.

On the structural side of the casualty analysis estimating the residual strength of the vessel's hull girder, and calculating the still water or wave bending moments on the hull from the changes in hull buoyancy and distribution of flooding water are necessary to prevent further structural damage and associated flooding. Structural calculations are often necessary when the use of compressed air is used to displace water as the loading on the boundary structure, such as the underside of the main deck, can be excessive.

Without benefit of a validated computer model of the hull, it may be necessary to extrapolate or estimate quantities, such as the displacement, transverse center of buoyancy (TCB) and longitudinal center of buoyancy (LCB) of the vessel from line drawings and bonjean curves. There is often a serious question to consider as to which method of analysis is best and quickest to use in doing hand calculations in the form of either the added weight or lost buoyancy method to quickly gauge the effect of various *what-if* scenarios. Free floating damaged casualties in peril of capsize or breakup will tax the naval architect/salvage engineer's best judgment and computational/estimating skill. In situations like this, intentionally grounding the vessel (beaching), if that option exists, may be prudent. Even then there will be questions to resolve such as of how to put the vessel aground (speed and heading) and at what state of tide so as to minimize capsize forces and distribute the ground reaction forces on the hull plating. Once aground, analyzing methods of stabilizing the situation and then refloating the intentionally grounded vessel must be carried out quickly. Stranding is generally considered an unintentional grounding of a vessel, as discussed in Sect. 42.4.2.

Appendix 42.A describes dewatering tools and methodology, and Appendix 42.B contains methods and provides common formulas, calculations, and references to refer to when developing salvage plans for a free-floating vessel.

Free Floating – Capsized

Experience has shown that ships freely floating upside down and exhibiting a freeboard roughly equivalent to, or only slightly greater than the height of the double bottom are generally quite stable. Similarly, full-bodied ships or barges floating at heel angles of 130–150° with the turn of their bilge a few feet out of the water are generally quite stable. These ships are often found to be riding on pockets of trapped air, or have great quantities of cargo that are buoyant in seawater. If the vertical center of buoyancy (COB) lies directly above the center of gravity (COG), the vessel will be in stable equilibrium. For capsized vessels, free floating or submerged it is always important to have B lie above G. Similarly, when the longitudinal centers of gravity and buoyancy align the vessel will stabilize in trim. It is quite possible for a vessel to be in stable transverse equilibrium yet unstable longitudinally and trim excessively before equilibrium is established. In this state, she may drift or possibly be towed with either her bow or stern dragging until fetching up and grounding as she nears shallow water. At this time, an additional set of forces will come into play as a result of the ground reaction. This phenomenon is discussed under the Grounding section.

Depending upon the vessel's subdivision, both transverse and longitudinal, the fluid interface at the free surface will affect the vessel's rise in its COG. Free surface reduces righting arm in floating capsized vessels in the same manner as it affects surface vessels floating right side up. The opposing shifts of liquid and air can be considered to cause a rise in the COG (added weight method) or a lowering of the COB (lost buoyancy method).

In most capsize circumstances, dive assets will be required for access and subsea entry into the vessel for survey work, to introduce compressed air into compartments from subsea, or to do the necessary rigging for a crane lift. This is a dangerous work, made even more difficult in the open ocean and the transfer from larger vessels to dive tenders. It is generally better to shift the capsized vessel into protected waters wherever possible. In harbor clearance operations, shifting the floating capsized vessel to shallow water can be the prelude to wrecking in place. Grounding the freely floating capsized vessel in a harbor, especially near a pier can often be the first step toward parbuckling the vessel into a righted and sunk position, from which the operation can become one of refloating a vessel that is stranded.

Salvors will often use compressed air to enhance the freeboard or trim of a floating capsized vessel and to recover buoyancy. Calculating the volume, pressure, and flow rate of air to be introduced and its likely effect on trim and stability is essential. Fortunately, in the case of a floating vessel, the issue of expanding air as the vessel rises in the water column is not important as it is when refloating a capsized ship from the bottom. The air pressure at equilibrium for a given increase in buoyancy will have an effect on the structural loading of shell or inner bottom plating as the air pressure required to dewater a space increases with hydrostatic head. In addition to using compressed air, it is possible to induce buoyancy and stability by using pontoons, foam, and other buoyant objects in the vessel's holds or even floating at the vessel's waterline outside of the hull.

On some salvage jobs involving a capsized vessel afloat, it will be possible to parbuckle the vessel to an upright position by flipping it or lifting it to a deck barge from its capsized state using heavy lift assets. This is a challenging job at sea because ship motions will affect loads on shear leg cranes and the rigging can be a nightmare for the lift or rotation. In addition, the balance of forces among the crane's load, the instantaneous center of residual buoyancy, and the casualty's flooded weight can become dramatically out of control as the rotation over-tops and the crane load goes slack. The suddenly upright vessel may skate directly toward the crane barge and collide. If possible, it is better to tow the vessel to sheltered water to complete the task and

simplify the rigging demands to control motion. With a thorough knowledge of heavy lift and rigging loads saddles, padeyes, and other rigging attachments must be designed, fabricated, and installed. In addition, it is not unheard of for the capsized vessel, especially if floating with her bilge high, to be pre-rotated by cutting holes through the shell plating into the low-side compartment, and pre-ballasting a space to shift the TCG to the low side. This operation is dependent on a large surplus of buoyancy and requires detailed calculations to move step by step through the process.

42.4.2 Grounded Ships

All grounding casualties have one essential feature in common: part, if not all, of the vessel's weight is supported by ground forces, commonly called the ground reaction. The distribution can range from a vessel simply touching down (aground) on a bar, reef, or ledge at her bow or bilge and lively with the tide and sea, to the stranded situation where there is little or no motion throughout the tide cycle. A stranded vessel that is high and dry will have virtually the entirety of its weight support by a uniformly distributed load or ground reaction (like a vessel in dry-dock), or in some cases it may be close to refloating if the buoyancy of the displaced volume of surrounding water can be made to exceed the ship's weight, for example, by lightering cargo or liquid loads to reduce the tons aground. In the aground case, trim and stability will join longitudinal strength considerations to dictate the course of action needed to refloat the casualty. When a vessel is stranded high and dry, its trim and stability characteristics are not uppermost in the salvor's thinking, which is generally how to stabilize the deteriorating situation and how to refloat. However, when the vessel is aground on what can be considered a pinnacle, or a single point, the ground reaction can have a significant effect on the vessel's trim, stability, and hull girder strength. In almost all cases of grounded ships in an exposed situation, there is the utmost urgency because the vessel is in a dangerous position and time will be of the essence with an exposure to wind, sea, tides, and currents combining to broach the vessel and drive it further ashore.

The type of grounding casualty and the extent of damage, as well as the conditions of exposure will have a huge impact on the salvage operation. In addition, the steps taken to initially stabilize the situation, for example, by pinning the vessel and taking steps to prevent broaching by deployment of anchors, where possible, will affect the best course of action and the options available. This in turn will dictate the techniques used to refloat and the computational tasks in quantifying refloat options. The fundamental goal in the partnership between the ship's crew in its Damage Control process and the salvors who arrive on scene is the safe refloating of the stranded ship, meaning the goals are to save lives, the ship, its cargo, and conduct pollution abatement. The reality is sometimes a steady sequence of ineffective steps to reduce the tons aground, or improve residual buoyancy leading to a deterioration of options and finally broaching. The sequence of unfortunate events from grounding to wrecking in place can easily follow well-established perils of the sea, grounding to broaching, to stranding, to being left high and dry, and finally to shipwreck. In the process of trying to beat the forces of nature, the salvage engineer may be called upon to exhaust his entire bag of engineering tricks.

The three phases of a grounding salvage operation are:

1. The stabilization phase – when steps are taken first by ship's crew and then salvors to prevent further damage to the vessel and keep the ship from being driven further around, or broaching.
2. The refloating phase – when the salvage plan is being executed, often in the form of a sequence of techniques applied to achieve piecemeal goals leading to extraction from the situation.
3. The post-refloating phase – when the ship returns to the situation of being a free-floating casualty.

Aground

The first successful steps taken to save the ship aground often begin with the master's actions within the first hours of grounding. These steps include the immediate notification of the ship's condition to the authorities, determination of the vessel's watertight integrity, and the survey of compartment flooding, perimeter soundings of the seafloor and draft readings, and an accurate determination of the vessel's liquid-loading conditions for consumables and fuel. Ship's personnel and remote salvors and salvage engineers can then begin to determine the tons aground. The master must begin the steps to pin the vessel to stabilize the situation if there is any indication that she will broach in the current and seaway, and resist the temptation to jettison weights or back down if the vessel's hull has been breached. If the assessment indicates that there is no flooding and the ground reaction calculation suggests that the vessel may be retracted at the next high tide through a revision to the vessel's cargo, fuel, and ballast loadout then the proper course of action will be to pin the vessel hard in the interim and develop the sequence of internal ship cargo, ballast, and fuel transfers to be completed during the tide cycle. The calculations involved in this analysis are derived from fundamentals of naval architecture augmented for grounding. If the

vessel is aground on a single pinnacle in water of sufficient surrounding depth, it will pivot about that point as the tide changes and may be treated as a statically determinate problem by equilibrating residual buoyancy and centers with the known ship weight and centers (from pre-stranding ship displacement and centers). The sequencing of shipboard cargo, fuel, or ballast transfer steps require the vessel to stay aground through the low water cycle until the extraction process is ready to start on the rising tide. At this time, the liquids used to pin the vessel will be discharged (ballast) thus lowering the ground reaction or shifted (fuel or ballast) to shift the LCG and trim the vessel to lower the ground reaction. A successful extraction like this will rely heavily on correctly estimating the freeing force, ground reaction, and the bottom's coefficient of friction. It also helps to have tugs assisting to augment ship's power. If the ship cannot be extracted at full power on the next tide the engines should be placed on standby and the vessel should be re-pinned to prevent being driven further ashore while awaiting the arrival of salvage tugs and salvors to take charge of the extraction operation. In the interim whatever assistance should be used to stabilize the ship in a position to safely ride the currents without broaching. Time is not on the side of a vessel aground.

Stranded and Beached

From the salvage engineer's perspective, the conditions of a stranding are seldom fully defined in the beginning and often are not completely defined during the salvage operation. Salvage of a stranded ship is time-critical because weather and sea conditions always worsen the situation. In addition, unlike the case of simple pinnacle groundings (with one with two pinnacle supports) a stranded situation will invariably involve a statically indeterminate set of engineering calculations (somewhat similar to dry-docking calculations with many of the blocks removed or of differing elastic properties). The appropriate soil mechanics characteristics of the seafloor, reef, or shelf must be quickly identified. There is no time to wait until conditions are optimum or until all information is known. Obtaining the best information possible, estimating unobtainable information, analyzing the sequence of activity, and developing quantitative information without delay are often the norm. There are many methods for refloating stranded ships; none is correct in every circumstance. There is no simple formula. Salvors are limited only by their knowledge and imagination. This section addresses conditions that affect stranded ships and what methodologies are typically used to refloat.

The initial salvage step for a stranded refloat operation is to stabilize the vessel by deploying a system of ground legs to control the vessel swing and keep her in the correct direction to facilitate extraction. The anchoring system will consist of the anchor itself (drag embedment, pile, deadweight, etc.), mooring line (chain, wire or synthetic rope), which together form the ground leg, and pulling system for tensioning (linear cable pullers, winches, purchases, etc.). Tugs are used to deploy the ground legs and can apply pulling forces directly on the casualty with their application of power measured in terms of bollard pull. These forces are typically applied in the plane of the casualty to achieve the following effects: retracting directly off the beach, swinging the casualty into a more favorable position, inducing wrenching motion to break suction or reduce ground friction, holding the casualty in position, as in a swift current, for example, or controlling the casualty's motion during the refloat evolution (checking lines).

Once stabilized or under control relative to her heading with the beach, reef, or shore the casualty will be lightered as much as possible to minimize the ground reaction and reduce the buoyancy needed for refloat. In addition, it will be necessary to safely remove all fuel and other pollutants that are potentially at risk if the vessel begins to break up, or is holed during the extraction phase of the refloat process. In compartments where cargo oil or fuel tanks have been holed and exposed to the sea, salvage pumps may be needed to strip the oils from the tank top and avoid the water bottom that would be sucked into the vessel's own cargo and fuel oil transfer system. Based upon hydrostatic balance between the head of oil and the casualty's draft at the holed tank's location, the quantity of oil in the tank at equilibrium can be estimated. Opening an inerted space is fraught with danger. Sometimes the lightering operation will be to shore instead of to a standby barge or vessel. Calculating estimated operation times for the transfers and evaluating how the ground reaction will change as a result is required to finalizing the lightering plan. As the tide changes, updates will be made to determine the buoyancy deficit for refloat. When possible, much of the deck cargo will be removed and predictions will have to be made to determine what the trim, list, and stability characteristic of the vessel will be at refloat. This will factor into the extraction process and depth of water requirements. If repeated efforts in cargo lightering are unsuccessful in refloat at high water, it may become necessary to start to dig the vessel out.

Difficult refloat operations may often require underwater excavation techniques to scour, jet, airlift or dredge seafloor material to reduce friction, modify the ground reaction, pass messengers, or create a channel to extract the vessel into deeper water. These techniques may use a tug's propeller wash to scour or a crew of divers with jet and airlift tools. Suction dredging is an

important tool to cut trenches alongside the stranded vessel, especially in high and dry situations, to set up pockets of deeper water where the vessel will settle, list or be pulled into a new position to increase buoyancy and thereby decrease ground reactions and eliminate friction. Each change in the angle of repose presents a new distribution of ground reactions and hull local and girder loadings. Continuous longitudinal strength calculations to analyze the hull girder sheer, bending, and torsional loads for each successive stage of the operation, so as not to break the vessel, must be undertaken. In addition, high spots in the contact points along the hull can create local loads sufficient to hole the vessel. Both unsupported and grillage panels of bottom plating must be analyzed to understand local loading. Sometime soil collapse may cause the casualty to roll suddenly and excessively to a new equilibrium position requiring an update of all of the structural loading and equilibrium calculations.

Appendix 42.C describes de-beaching tools and methodology and Appendix 42.B contains methods and provides common formulas, calculations, and references to refer to when developing salvage plans for a grounded vessel.

42.4.3 Sunken Ships

Stranded and Sunk – Refloat

A partially submerged stranded and sunken ship will typically be internally flooded in at least some very large spaces or tanks between transverse and longitudinal bulkheads. If the internal flooding rises and falls with the tide, the ship is said to be tidal, which is a sure sign that there is free communication with the sea in the area under survey, for example, in hold areas where the watertight integrity of the side shell is compromised as in the case of a collision followed by parallel sinkage without capsize in a harbor, port facility, or shipping channel. A similar phenomenon can happen when a vessel strands and tears open its shell plating into holds or tanks. If the space is not tidal and a differential level exists between the tidal waterline and internal flooding, residual buoyancy can be obtained by judicious pumping of accessible spaces (dewatering) that actually have a watertight envelope. The refloat evolution is undertaken by first establishing the tons aground for the sunken vessel and then developing a plan to remove enough weight or add enough additional buoyancy to float the vessel. Weight removal can be accomplished by dewatering spaces by pumping, use of compressed air in certain spaces to displace water (and being careful to be sure the air bubble cannot escape suddenly if the vessel rolls), lightering of cargo from holds or superstructure to lower both the weight and vertical distance from the keel to the center of gravity (KG). Any submerged space, like a forepeak tank, that is dewatered by displacing flooding water with air needs to be checked for residual strength as it in effect becomes a cofferdam habitat subject to hydrostatic head. The increased buoyancy developed as the air expands as the vessel rises in the water column must be calculated into the total buoyancy equation. Underwater cutting is often used to provide a path for the displaced water to exit and limit the expansion of air as the casualty rises in the water column. Hot taps can be employed as well to create a path for flooding water to discharge. Often the challenge of refloating a stranded and sunken vessel is one of the patching holes in the ship's water tight envelope (side shell, inner bottom, or internal longitudinal and transverse bulkheads). Divers are essential in completing this mission. At the conclusion of all of these preparatory steps, the key determination in planning the refloating operation is to predict at what stage of preparation and tide liftoff will take place.

For a vessel completely submerged, the challenge is even greater. Not only will all operations aboard the submerged vessel require diver interface, but also sunken ship stability in the water column is harder to control as there will be no waterplane. Dewatering can sometime be accomplished by fabricating surface piercing cofferdam bulkheads. Pontoons, side shell cofferdams, and sponsons can be attached to the hull to created buoyancy and provide some measure of stability at liftoff. The stability of an upright sunken ship at liftoff depends primarily upon whether the main deck is above water, partially submerged (i. e., there is limited waterplane), or completely submerged.

As the ship is raised, various methods are used to prevent it from capsizing while it passes through ranges where it is inherently unstable or develops instability from free surface. The most common method is to refloat one end of the vessel while keeping the other end firmly in contact with the bottom. The contact with the ground prevents the ship from taking on a dangerous list or capsizing. In deep water, keeping one end of the ship in contact with the ground and limiting the rise of the other end prevent extreme trim. Before the grounded end is raised, free surface in the floating end is reduced by dewatering. Spaces low in the ship, such as double bottom tanks, may be flooded and pressed up to both eliminate free surface and to lower the COG and increase the metacentric height. At the same time, topside weights are removed and other portable weights moved as low in the ship as possible to lower the COG. With free surface minimized from the floating part of the ship and the COG as low as practicable, the grounded end may be raised. These methods are not always adequate; careful stability calculations with a detailed

consideration of free surface should be made before attempting to raise the grounded end. Large free surfaces can be broken up by repairing damaged bulkheads and by building temporary bulkheads within flooded spaces. The work of building temporary bulkheads is considerably reduced if the bulkhead is built with high-pressure concrete pumped into simple forms. Both bulkhead reinforcements and temporary bulkheads should be built wider at the base than at the top to assist in lowering the COG. A force may be applied to the ship being raised to produce a moment that counters an upsetting moment. This is done by attaching barge or ship mounted cranes or tackles rigged to apply a vertical force near the side of the ship to steady her list. When the ship begins to list, a force is applied to counter the list and bring the ship back to the upright position. Pontoons, cofferdams, barges, or lift craft may be rigged to provide a force that counters a heeling moment and keeps the ship upright. The pontoons must be rigged tightly to the ship so that when the ship begins to heel, she will also attempt to submerge the pontoon on the low side. The additional buoyancy of the submerging pontoon, coupled with the loss of buoyancy from the pontoon on the high side, creates a moment that rotates the ship back to an upright position. Tightly rigged pontoons and lift craft not only provide an uprighting force from their buoyancy but also act as an increased waterplane of a system composed of the ship and pontoons. The waterplane of the pontoons increases the metacentric height and overall stability of the ship–pontoon system. This advantage is gained only when the ship and pontoons are so tightly rigged that they function as a unit. If the ship is free to render in a cradle formed by the rigging, the pontoons provide only buoyancy. Ships sunk in harbors have been kept upright by rigging purchases from their mastheads to anchoring points ashore. If this method is attempted, the ship should be held securely by mooring lines, as there will be a tendency for the ship to kick out from under the strain of the purchases on the masts.

One of the most secure methods of controlling the ship is to rig beach gear and haul the ship into shallower water as it is lightened or lifted. The beach gear is kept under constant heavy tension so that the ship moves into shallower water in constant contact with the bottom until it has reached a location where it may safely be dried out and refloated. Beach gear rigged ashore may be hauled with winches, linear pullers, or heavy vehicles. Tugs may be used to help move and direct the ship. Ashore, heavy tracked vehicles may also be used to haul lines for positioning the ship. Keeping a ship in contact with the bottom, either at one end or all along her length, also assists in controlling trim and preventing the loss of longitudinal stability caused by water rushing to the low end. Once the ship is refloated and stable, it can be handled as a free-floating casualty as discussed in Sect. 42.4.1.

Appendix 42.A describes dewatering tools, Appendix 42.D describes refloating tools and methodology and Appendix 42.B provides common formulas, calculations and references to refer to when developing salvage plans to refloat a stranded and sunken vessel.

Capsized and Sunk

Influences on capsized ships arise from two principal sources: condition in which the ship has capsized and sank and the environmental and physical working conditions. As often as not, ships will capsize as they sink in the water column, and they come to rest at a severe angle. Ships sunk in this condition present a more complex problem than those sunk upright. An entirely new dimension is added to the salvage operation because the vessel is usually righted before refloating. Ships are righted by developing a moment of force to overcome the moment of weight that holds the ship in her capsized position. With the exception of wreck removal there are four methods of handling a capsized and sunken ship:

- Refloating the ship as it lies and moving it to another location (Sect. 42.4.3)
- Wrecking in place or otherwise disposing of it in situ (Sect. 42.4.3)
- Rotating the ship until it is completely upside down and refloating the upside-down ship (Sect. 42.4.1)
- Rotating the ship until it is upright in place, then refloating it (Sect. 42.4.3).

Wreck Removal

Ships seldom capsize and sink in ways favorable to salvors, rarely allowing a convenient and comparatively straightforward righting operation. Statistically, it is as probable that a ship may capsize:

- Effectively blocking an entrance channel or canal
- In the middle of a wide river or navigable waterway where she is a traffic hazard
- In comparatively deep water, but where she is still a dangerous obstruction to traffic
- In the middle of a harbor or bay some distance from the nearest shoreline or harbor installation.

Relocate. The ship that capsizes and sinks effectively blocking a channel or canal creates a difficult situation because working methods often are reversed. Normally, salvors right a sunken ship before trying to refloat her. A capsized ship that blocks or seriously obstructs an important channel or canal is treated differently. As an obstruction to maritime operations, such a wreck must be moved quickly. Sometimes the most effective method of moving the ship is to refloat it on its side,

then haul or tow her away from the waterway. If the ship cannot be refloated completely, it may be practical to lighten or lift one end at a time, slewing or hauling the ship sideways to clear the channel. Under these circumstances, salvors have altered sunken ships' situations to their advantage by:

- Moving the ship, refloated on its side, to a more convenient righting location
- Swinging the refloated ship to face in the best direction for parbuckling
- Refloating the ship in such a way that it rotates, so that it is totally upside down and can be moved away immediately after refloating.

Without heavy lifting units that make wrecking in place an alternative, refloating first and righting later has many advantages. Where heavy lifting units are available, an analysis of both plans may show that refloating and moving away have both cost and time advantages over wrecking in place.

Wreck in Place. When a ship cannot be parbuckled into an upright position or a refloat and tow option is just not practical as in the case of a stranded vessel that simply cannot be refloated, then wrecking in place is the only remaining option. The wreck in place evolution will typically require heavy lift capabilities (shear leg cranes, derrick barges) along with jackup platforms configured to render chain or cutting wire to sever the hull much like a cheese slicer. Chopper beams are sometimes used, especially in shallower water where the wreck may have been beached some time before. The wreck in place evolution is driven by cutting, rigging, and lifting and the salvage engineer will be required to calculate weights and centers for portions of the ship that are to be severed from the hull girder, or superstructure, rigged for lifting successfully. There will be intensive diving operations often with tunneling operations to sling under the vessel. In situ welding, surface or subsea will be required along with detailed calculations and strength analysis for padeyes. The naval architecture and marine engineering disciplines will be of particular use in identifying internal structure that will need to be severed before taking up the load. This can include piping systems, box girders, and other massive structural subassemblies. Sections of the ship up to several thousand tons may be handled in a single pick using tandem cranes and then loaded on deck barges.

At times the wreck in place operation will resemble the shipbreaker techniques used to build down the superstructures and lighten the ship's weight, and then using linear pullers anchored to the shore to drag the hull ashore compartment by compartment and sequentially chop the hull. This is not a very sophisticated method but in the right circumstances it can be a very cost-effective technique to get the job done. One important note about wreck in place concerns the necessity to control pollution releases of petroleum, oil and lubricants (POL) and other contaminates aboard the vessel.

Appendix 42.E describes tools and methodology to refer to when developing a wreck in place salvage plan.

42.A Appendix: Dewatering

42.A.1 Dewatering Tools and Methods

1. Patching and sealing
   - Materials – steel, aluminum, wood, concrete, composites, glass-reinforced plastic (GRP)
2. Pumping:
   - Pump types
3. Compressed air:
   - Displacement method
4. Combined air and pump
5. Induced buoyancy.

Removing floodwater, or *dewatering*, recovers buoyancy lost by flooding, as opposed to increasing buoyancy by removing cargo, stores, or other weights that were in the ship before the casualty. Floodwater is removed by pumping, compressed air, or by *induced buoyancy* methods that displace water with buoyancy objects. Leakage into spaces to be pumped must be reduced to less than available pumping capacity by patching and/or securing open valves and fittings. Submerged compartments can be pumped through cofferdams that extend the boundaries of the space above the waterline. The tops and sides of spaces to be dewatered with compressed air must be made airtight. Spaces dewatered with induced buoyancy need not be made watertight, but other preparations may be necessary. Each dewatering method has advantages and disadvantages.

42.A.2 Compressed Air Dewatering

Buoyancy can be recovered by forcing water out of flooded spaces with compressed air. Compressed air dewatering is used when pumping is not practical or the construction of the casualty particularly suits it to use

of compressed air:

- Dewatering large tanks, holds, or machinery spaces that have bottom damage.
- Dewatering double-bottom or deep tanks.
- Dewatering cargo tanks in tankers and bulk liquid tanks in other ships.
- Dewatering holds filled with cargo that prevents placing of pump suctions or lowering of pumps to reduce suction lift.
- Dewatering sunken ships submerged too deeply for the practical use of cofferdams and pumps.
- Regaining sufficient buoyancy to refloat capsized vessels upside down or on their side.
- Recovering buoyancy in flooded submarine compartments and ballast tanks.

Because the tops and sides – as well as the bottoms – of liquid tanks are watertight and airtight, buoyancy can be recovered from flooded doublebottom or deep tanks without patching and little other preparation, so long as damage is confined to the bottom and lower side plating. Buoyancy quickly recovered from flooded tanks may be sufficient to refloat stranded ships, or prevent damaged ships from sinking.

Despite its advantages, compressed air dewatering has a number of disadvantages:

- The preliminary work to prepare most types of ships for blowing with compressed air is much greater than for pumping. Extensive patching and/or diving work may be required to obtain a sufficient degree of airtightness. An accepted rule of thumb is that compressed air will escape four to six times faster than water through a given opening. Considerable strengthening of decks and hatches may be required before dewatering with compressed air. Compressed air exerts its upward force on the tops of compartments, from the inside. A ship's structure is, in general, designed to withstand sea pressure from below and outside the hull.
- Compartments can be overpressurized if vents are not properly sized or become obstructed, or blowing operations are not carefully monitored.
- Ships raised on compressed air leave the bottom suddenly and can be difficult to control during ascent as the compressed air expands.
- Because of the inefficiencies of pumping compressible fluids, the dewatering capacity of a compressor is much less than that of a pump of equal weight.

The basic requirements for compressed air dewatering are a blowing connection to which an air hose is attached, a gage to monitor compartment pressure, and a route for the water to escape. The blowing connection can consist of an air fitting attached to a vent pipe or sounding tube, or permanently installed in the deck above the compartment to be dewatered. The diameter of the air fitting should equal or exceed the air hose diameter. An air fitting can also be mounted on a plate sized to fit a manhole, scuttle, or butterworth fitting. Figure 42.1 shows a typical salvage blowing fitting. As an alternative, air hoses can be led through hull openings below the waterline. A valve to control air pressure can be installed on the air fitting or in the supply hose if a relief or dump valve is installed upstream in the system. The pressure gage should be installed on the air fitting plate or other *direct access* to the space. If the gage is installed in the air supply line, air flow must be secured to get an accurate reading of the pressure inside the space (this means the gage must be *downstream* of the supply valve).

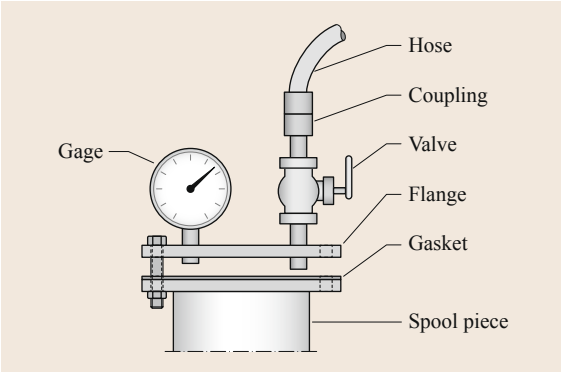


Fig. 42.1 Blowing fitting

42.A.3 Recoverable Buoyancy

The buoyancy that can be recovered from any compartment is directly proportional to the volume of the compartment that can be blown down without air leaking from the compartment. If a compartment has damage only at the bottom, nearly all the buoyancy can be recovered. If damage extends up the side of the compartment, only the volume above the highest extent of the damage represents recoverable buoyancy.

42.A.4 Vents and Standpipes

In a compartment with holed bottom plating, water is forced out through the openings in the bottom of the space as compressed air fills the compartment from the top down. A space flooded without free communication to the sea can be dewatered by use of a standpipe, or by cutting openings in the hull as shown in Fig. 42.2. The volume of water expelled, and thus the buoyancy regained, can be varied by adjusting the length of the standpipe or height of the hull openings.

Compressed air volume, and therefore added buoyancy, can also be limited by securing blowing before the water level reaches the vent holes or standpipe. This method is acceptable for strandings, afloat casualties, and ships sunk to shallow depths. By controlling the amount of buoyancy recovered in different spaces, trim, list, and draft can be controlled as desired.

For deeply submerged ships or objects, partly dewatering a space with compressed air is very hazardous and should not be attempted; as the casualty rises, the air expands, adding buoyancy in a rapid and uncontrolled manner. The resulting instability can cause the casualty to assume extreme trim or list; other components of the lifting system can be overloaded or fail. Buoyancy should be recovered in small increments in deeply sunken ships to avoid instability.

42.A.5 Expedient Blowing Fittings

It is occasionally necessary to restore buoyancy to flooded spaces rapidly to prevent sinking or capsize. Tanks and similar spaces that can be made airtight easily can be dewatered quickly with compressed air. Tanks are made airtight by removing the gooseneck standpipes from the tank vents and hammering damage control (DC) plugs into the vent pipes. A positive seal is ensured by coating the plugs with an epoxy compound; air pressure will force the epoxy into small gaps between the plug and pipe. Compressed air can be introduced into the space through the most appropriate expedient fittings available. After the situation has stabilized, the fittings can be systematically replaced by steel blowing plates with valves, gages, and hose couplings. In many cases, introducing compressed air through vent pipes is the fastest way to regain buoyancy in damaged double-bottom tanks. Ships damaged by grounding can be kept from sinking in this manner.

42.A.6 Removing Liquids Other than Floodwater

Compressed air can be used to remove fuel or other liquids from submerged tanks by leading a discharge hose from the standpipe to a receiving ship or barge. It is preferable to use pressurized inert gas rather than compressed air to move a flammable liquid. Blowing with compressed air increases partial pressure of oxygen in the space, possibly creating an explosive atmosphere, even if the tank was previously inerted. Inert gas systems often use the main engine flue gas that is inherently oxygen deficient. These gases are themselves flammable when mixed with sufficient oxygen. Inert gas-generating systems can create high enough pressures to be used for shallow depths; for higher pressures, it may be possible to discharge an inert gas generator into a volume tank, which in turn is led to the inlet of an air compressor. Carbon dioxide or nitrogen from high pressure flasks can be used, but large quantities are required.

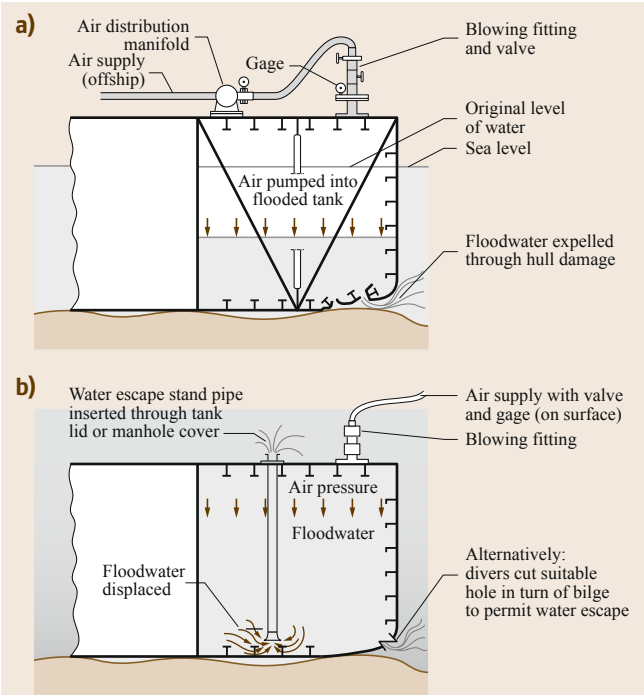


Fig. 42.2a,b Compressed air dewatering: **(a)** dewatering through hull damage; **(b)** dewatering a closed compartment

42.A.7 Combined Use of Air and Pumping

The most common use of combined pumping and blowing is to keep the pressure differential across bulkheads or shell plating within acceptable limits. Excessive pressure differentials can result from either of two conditions:

- When a compartment is dewatered by pumping, the inner sides of the bulkheads, shell, and decks are subjected to atmospheric pressure. The outer sides are subjected to an opposing hydrostatic pressure that varies with depth. If the water depth is great enough, hydrostatic pressure will be excessive.
- To completely dewater a compartment with compressed air, air pressure in the compartment must be greater than the hydrostatic pressure at the bottom of the space. When blowing deep compartments, such as tanks or holds, the high blowing pressure can cause an excessive outward pressure near the top of the compartment.

Combined pumping and blowing is advantageous in those situations where the work and time required to seal the compartment are less than that to shore or strengthen the casualty's structure against excessive pressures.

Combined pumping and blowing is a difficult technique because all the problems of both methods are present:

- Compartments must be sealed for blowing with special attention paid to boundary penetrations for pump suctions, discharges, power lines, hydraulic hoses, etc.
- Pressure on each compartment must be monitored carefully to ensure that the maximum blowing pressure is not exceeded.
- Air flow rate must be matched to water outflow rate (total pumping capacity) to avoid overpressurization.
- Holes must be double-patched.

Air flow must be controlled to ensure that internal pressure does not rise so high that the compartment cannot contain it or fall so low that hydrostatic pressure collapses the structure. Air flow rate is controlled by throttling air inlet lines or varying compressor speed. Correct matching of air and water flow will be verified by relatively constant compartment pressure. If it is not possible to vary air flow, compartment pressure can be maintained between upper and lower limits by intermittently admitting compressed air.

Under some conditions, it may be desirable to use air pressure to increase net positive suction head (NPSH) to improve pump performance. When pumping both flammable liquids and pressurizing the tank with an inert gas improve pump performance and reduces the danger of ignition. Many tanker cargo pumps are designed for optimum operation with a 5–10 psi inert gas pressure.

When some compartments are best dewatered by pumping and others by compressed air, each may be handled independently by the most suitable method. Care must be taken to ensure that pressure differential across bulkheads separating compartments to be blown from those to be pumped does not exceed the design pressure of the bulkhead.

42.A.8 Induced Buoyancy

Although pumping and compressed air dewatering induce buoyancy, the term *induced buoyancy* is specifically applied to methods where water is removed from a space and replaced with a buoyant medium other than air. Two general methods are used:

- Buoyant objects – Buoyant objects are introduced into a flooded space. By displacing water and reducing the space's permeability, a portion of the lost buoyancy is regained. Both purpose-built and improvised systems have been used.
- Cast-in-place foam – A buoyant foam is used to displace water and fill all or part of the space with a homogeneous, continuous mass.

The principal advantage of using induced buoyancy methods instead of compressed air or pumping is that it is not necessary to make the spaces to be dewatered completely air or watertight. However, these methods recover less buoyancy than compressed air, and cast-in-place foam and purpose designed buoyant object systems are generally more expensive to use than compressed air.

42.A.9 Expedient Buoyancy Recovery

Virtually, any buoyant object that is available in sufficient quantity can regain buoyancy by being injected, inserted, or placed in the spaces to be dewatered. Some examples include:

- Lifejackets, empty oil drums, small spheres, or other buoyant objects placed by divers or special-purpose equipment.
- Lift bags or collapsible pontoons placed inside flooded spaces, then inflated.
- Rigid pontoons placed in cargo holds so that they lodge under overhanging deck edges, as shown in Fig. 42.3.
- Wood chips, cork floats, etc., pumped into spaces as a slurry.

A buoyant cargo that has not lost its buoyancy through leakage, rupture, or waterlogging may carry a significant portion of the casualty's weight. The buoy-

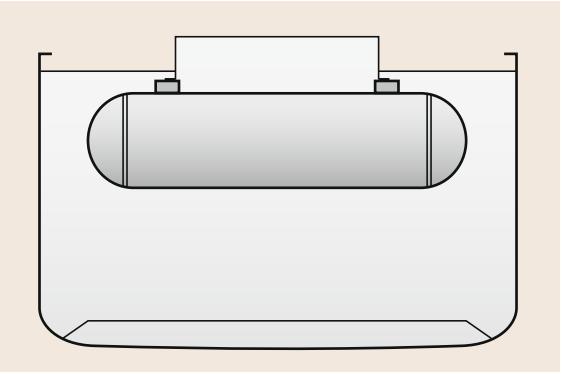


Fig. 42.3 Pontoons placed inside holds

ancy of the cargo reduces the lift required and should not be neglected in the salvage calculations.

42.A.10 Cast-in-Place Foam

In this technique, liquid chemicals are mixed on the surface or in a submerged mixing gun and pumped or blown into interior spaces of the casualty. There, the mixture undergoes a chemical transformation into polyurethane foam. The foam consists of a matrix of tiny gas-filled cells. The gas, usually a fluorocarbon, is blown into the liquid urethane as it is injected into the space and causes the expansion necessary to displace water; the cell walls, after hardening and curing, are strong enough to resist further expansion of the gas as water pressure changes as the vessel refloats. The foam hardens very quickly – within a minute or so of being injected – and forms a rigid, buoyant mass. The foam collects on the top of the compartment, displacing the water. Foam density, strength, and curing time depend on a number of factors, including water depth and temperature, and should be determined by on-site experimentation. Theoretically, cast-in-place foam could be used at very great depths, but foam density increases with water depth, so less buoyancy is recovered. It is unlikely that cast-in-place foam would be practical or economical for water depths greater than 200 ft.

Cast-in-place foam has the following advantages:

- The required equipment and chemicals are transportable by air and small craft.
- The expanding foam blocks and seals small holes, cracks, and splits.
- Buoyancy is recovered without creating free surface.
- The rigid foam possesses shear strength and adheres to internal structural members. The buoyant force is distributed through the casualty; less deck stiffening is required than when using compressed air or other water displacement methods. The hardened foam has some compressive strength and contributes to the overall hull girder compressive strength. The foam also inhibits buckling in shell and bulkhead plating. The contribution of foam to overall strength varies with foam quality and thoroughness of application, and is difficult to quantify. Foam installed in the hull may justify the use of a reduced factor of safety for compressive hull loading.
- The COG of the system, once cast in place, is fixed and predictable. The same is true for the COB so long as the entire foam block is submerged.
- The volume and shape of the foam mass are not subject to change as the vessel rises.
- Foam can be selectively distributed throughout large, undivided spaces, optimizing effects on trim and stability.

The use of cast-in-place foam has several disadvantages:

- The chemicals used to create the foam are toxic, flammable, and vapor-producing. Blending foam components gives off highly toxic vapors, and the foam may create irritating and toxic vapors while curing after the space is dewatered. When the foam is cut or crushed for removal, fluorocarbon gas (freon) is liberated from the foam cells that may displace oxygen in confined spaces, and is considered an environmental pollutant (damaging to upper atmosphere ozone layer).
- Urethane foam is very flammable.
- Foam is relatively expensive.
- Especially trained personnel and sensitive pumping and proportioning controls are required to reliably produce foam of consistent quality.
- The chemical reaction that creates the urethane will produce enough heat to self-ignite if laid down in air in thick layers that cannot dissipate the heat rapidly. The heat can ignite other flammables and induce undesirable chemical reactions.
- The hardened foam is very difficult to remove, especially from cluttered spaces such as machinery rooms.
- Foam components must be batched exactly to suit water and air temperatures, especially in very cold conditions.
- A foam mass in the top of a compartment, as high weight, may cause instability when the vessel is refloated.

The apparent simplicity of foam in salvage is deceiving. It is difficult to produce good quality foam consistently in the field. Foaming is a major operation involving serious safety and fire hazards. Foam should be used only when it is clearly the best alternative.

42.B Appendix: Common Formulas, Calculations, and References

42.B.1 Trim and Stability

Free Communication Effect

A partially flooded, noncenterline space open to the sea introduces the effects of both off center weight and free surface. In addition, floodwater is free to enter or leave the space as the ship inclines. The distribution and weight of floodwater vary with time as the ship inclines. This creates virtual rise in the COG, in addition to that caused by free surface

$$\text{Virtual rise of } G = \frac{A(y2)}{\nabla 1}\,, \tag{42.1}$$

where A is plan area of the flooded compartment, y is transverse distance from the center of the flooded compartment to the ship's centerline, and $\nabla 1$ is volume of displacement after flooding to the waterline.

Free communication exists only when the water level inside the damaged compartment remains the same as the sea level outside the hull. This occurs only when the hull opening is relatively large compared to the volume of the space, and the compartment is vented.

Added Weight Versus Lost Buoyancy

The added weight method assumes that flooding, with or without free communication, increases the weight of the ship by the weight of the floodwater and that none of the hull surface exposed to the buoyant force of the water is lost.

An alternative method, the lost buoyancy method, can be used where floodwater in free communication with the sea is assumed to remain part of the sea, and the flooded portion of the ship no longer contributes buoyancy. The vertical pressure forces about the flooded compartment are assumed to act on the sea rather than on the ship.

Flooding in free communication with the sea can be assessed by either method, but the two methods cannot be mixed during calculations. Table 42.1 itemizes the important points of the two methods. The method used is a matter of personal preference, although the added weight method is more commonly used.

Table 42.1 Added weight versus lost buoyancy

Item	Added weight	Lost buoyancy
Change in displacement	Yes	No
Change in volume of displacement	Yes	No
Change in draft, trim, and list	Yes	Yes
Shift of center of gravity	Yes	No
Shift of COB	Yes	Yes
Shift of metacenter	Yes	Yes
Free surface correction required	Yes	No
Free communication correction required	Yes	No

Ground Reaction Calculations

Residual Buoyancy Distribution Method. Analysis of as-grounded weight and buoyancy distribution can determine the ground reaction and help to determine its distribution and center of pressure. The area between the weight curve and the buoyancy curve for the stranded waterline is the total ground reaction. While the area between the weight and buoyancy curves is equal to the ground reaction, the shape of this area does not define the ground reaction distribution precisely. The load curve formed by the sum of weight, buoyancy, and ground reaction curves is not zero at every point, even though overall equilibrium exists. For equilibrium to exist, ground reaction must be distributed so that the combined COB and ground reaction is in vertical line with the COG. By distributing ground reaction increments over the grounded length so that the combined COB and ground reaction is in line with the COG, the ground reaction distribution may be described with reasonable accuracy. This method is well suited for use with computers and automated hull form databases.

The buoyancy curve is developed from section areas taken from Bonjean's curves or calculated from offsets. For most strandings, entering drafts can be determined by linear interpolation between forward and after drafts by either calculative or graphical means. If the ship is noticeably hogged or sagged, the buoyancy curve should be developed from drafts taken at several stations.

Change of Displacement Method. Ground reaction can be estimated by entering the curves of form or hydrostatic table with the drafts before and after grounding and reading the displacements for the two conditions

$$R = \Delta b - \Delta g\,, \tag{42.2}$$

where R is the ground reaction, Δb is the displacement immediately before stranding, Δg is the displacement after stranding.

Change of Draft Forward Method. The change of draft forward method considers the ground reaction as equivalent to a weight removal that causes both parallel rise and the change of trim. The center of ground reaction must be known or estimated with reasonable accuracy to determine the trimming lever. The distances

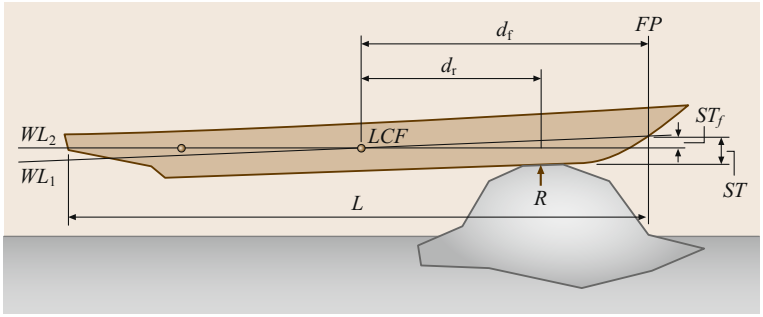


Fig. 42.4 Change of draft forward method

used in the following derivation are shown in Fig. 42.4.

$$R = \frac{\Delta T_f(\text{TPI})(\text{MT1})(L)}{(L)(\text{MT1}) + (d_r)(d_f)(\text{TPI})} , \tag{42.3}$$

where Δt is the total change of trim [in], ΔT_f is the change of draft forward $= T_{fb} - T_{fa}$, d_f is the distance from the center of flotation to the forward perpendicular, d_r is the distance from the center of flotation to the center of ground reaction, R is the ground reaction [tons], MT1 is the moment to trim one inch, TPI is the tons per inch immersion [TPI].

The basic relationship can also be solved for d_r

$$d_r = \frac{1}{\text{TPI}(d_f)} \left[\frac{\Delta T_f(\text{MT1})(\text{TPI})L}{R} - L(\text{MT1}) \right] . \tag{42.4}$$

Tons per Inch Immersion Method. A simple, but often satisfactory estimate of the ground reaction can be made by multiplying the change in the mean draft on stranding by the TPI

$$R = (T_{mbs} - T_{mas})\text{TPI} , \tag{42.5}$$

where T_{mbs} is the mean draft before stranding, T_{mas} is the mean draft after stranding.

The TPI method is often used because mean drafts and TPI can be estimated in the absence of detailed information.

This method considers only the bodily rise of the ship and is suitable for a first estimate of ground reaction when trim has not been changed greatly by the stranding. If the stranding causes a significant change of trim, the accuracy of the method can be improved by correcting the mean draft for trim by the displacement correction for trim.

Change of Trim Method. The change of trim method is most useful when the total trim exceeds 1% of the ship's length, the center of pressure of the ground reaction is known or can be estimated with reasonable accuracy, and the change of trim is the dominant effect of the stranding. Ground reaction is treated as a force that causes only a change of trim.

$$R = \frac{\text{MT1}(\Delta t)}{d_r} , \tag{42.6}$$

where Δt is the total change of trim [in].

Neutral Loading Point. When a weight is added to a floating ship at a point other than the LCF, there is a point on the opposite side of the LCF where the opposing changes in draft caused by trim and parallel sinkage are exactly equal and draft remains constant. By applying this principal to a stranded ship, a neutral loading point (NP), where weight can be added or removed without changing the ground reaction, can be defined. There is no tendency for the hull to move at the point of ground reaction, so the ground reaction remains unchanged. Using the referenced dimensions as shown in Fig. 42.5, location of the neutral loading point is

$$d_n = \frac{(\text{MT1})(L)}{(\text{TPI})(d_r)} , \tag{42.7}$$

where: d_n is the distance from LCF to NP, MT1 is the moment to change trim one inch, L is the length between perpendiculars, TPI is tons per inch immersion, d_r is the distance from the center of ground reaction to LCF (location of LCF should be based on stranded drafts).

The neutral loading point is a datum which helps predict the effect of weight changes at other locations. Weights added aft of the neutral loading point will decrease the ground reaction; weights added forward will increase the ground reaction. The opposite is true for weight removed. As weight is added or removed, the ship will trim; if trim is sufficient to alter MT1, TPI, or the position of LCF, the location of the neutral loading point must be recalculated.

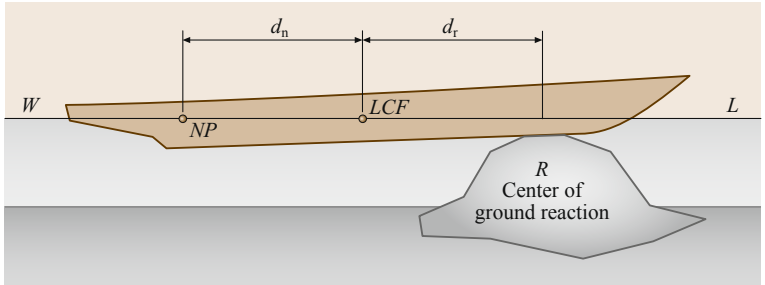


Fig. 42.5 Neutral loading point

The change in ground reaction (ΔR) resulting from a weight change at any point along the length of the ship can thus be predicted by the following relationship

$$\Delta R = w\left(\frac{d}{d_{nr}}\right) , \tag{42.8}$$

where w is the weight added or removed, d is the distance from the added or removed weight to the neutral loading point, d_{nr} is the distance from the neutral loading point to the center of ground reaction $= d_n + d_r$

Tide and Ground Reaction. The waterline of a stranded ship rises and falls with the tide. For a ship that cannot trim, the change in ground reaction caused by the tide is nearly equal to the change in height of the tide multiplied by TPI. For a ship that can trim with tide changes, the change in ground reaction can be estimated by relating the change of ground reaction to change in draft at LCF. For a change of trim, draft is constant at the center of ground reaction. The change of draft at LCF from trim is then given by

$$\Delta T_{\text{LCF, trim}} = \Delta t\left(\frac{d_r}{L}\right) , \tag{42.9}$$

where Δt is the change of trim [in], d_r is the distance from the center of ground reaction, or assumed pivot point, to the center of flotation, and L is the length between perpendiculars.

The total change in draft at LCF is the sum of the changes caused by trim and rise or fall of tide. The change in draft because of tide is simply the change in tide height. The two changes are opposed; a falling tide tends to decrease draft, but the rotation of the ship about the pivot point tends to increase draft at LCF. A rising tide has the opposite effect. Total change in draft at LCF is then

$$\Delta T_{\text{LCF}} = \Delta h - \left(\frac{d_r}{L}\right) , \tag{42.10}$$

where Δh is the tide change [in].

Change in ground reaction can be estimated by multiplying change in draft at LCF by TPI.

$$\begin{aligned}\Delta R &= \left[\Delta h - \Delta t\left(\frac{d_r}{L}\right)\right]\text{TPI} \\ &= \Delta h(\text{TPI}) - \Delta t\left(\frac{d_r}{L}\right)\text{TPI} .\end{aligned} \tag{42.11}$$

Expressing Δt as $\Delta R d_r/\text{MT1}$ implies an assumption that the ship is trimming about its center of flotation; it is not. This assumption introduces errors into ground reaction predictions for different heights of tide. As salvage operations progress, drafts should be taken regularly and ground reaction estimated for the height of tide at that time. Plots of ground reaction as a function of height of tide can be used to predict ground reaction. The relationship between ground reaction and state of tide will not remain constant. The working of the casualty on the bottom will deform both the hull and seafloor, shifting the pivot point and altering the response to tidal fluctuations. The degree to which the vessel response is altered depends on the degree of motion of the casualty, range of tide, effects of currents and waves, and strength of the supporting seafloor. Analysis of ground reaction versus tide plots will show trends that can be used to refine predictions of ground reaction.

Effect of Grounding on Center of Gravity. Ground reaction is equivalent to removing an equal weight from the keel, and causes a virtual rise in the COG similar to that caused by the block reaction on a ship in drydock

$$\text{GG}_1 = \frac{R(\text{KG})}{(W-R)} . \tag{42.12}$$

The effective height of the COG can be calculated directly

$$\text{KG}_1 = \frac{(\text{KG})(W)}{(W-R)} , \tag{42.13}$$

where GG1 is the virtual rise of the COG, KG1 is the effective height of the COG when the ship is aground, KG is the original height of the COG above the keel, W is the weight of the ship, and R is the ground reaction.

Grounding Off Centerline. A stranding with the center of ground reaction off the centerline experiences both a loss of displacement and an upsetting moment. If free to incline, the ship will assume a list. The upsetting moment can be calculated by multiplying the distance off center of the effective point of ground reaction by the magnitude of the ground reaction as shown in Fig. 42.6. Summing moments about the effective point of grounding

$$Rs = Bt = W\,\mathrm{GZ} = W\,\mathrm{GM_{eff}}\sin\theta \text{ (at equilibrium)}\,, \qquad (42.14)$$

where R is the ground reaction, s is the distance off centerline of effective point of ground reaction, B is the buoyant force (displacement aground, $W-R$), t is the lateral distance from effective point of ground reaction to COB, W is the gross weight of the ship (displacement before stranding $\pm$ weight changes after stranding), GZ is the righting arm, taken from the stability curve, $\mathrm{GM_{eff}}$ is the metacentric height, as stranded (corrected for free surface and virtual rise of G), and θ is the angle of list.

The ship will incline until the moments are balanced or until it slides off the bank or rock. Unless restrained from sliding off the pinnacle, there is little probability of capsize in still water. In order to capsize, the ship must lift off of the pinnacle; ground reaction and upsetting moment will fall to zero and the ship will settle back onto the pinnacle. Off center ground reaction may reduce righting energy to the point where external forces (wind, waves, towline pull) or internal weight shifts (free surface, cross-flooding, intentional weight movements, sliding objects) can capsize the ship. Fine-lined ships aground on rocky shores with large tidal ranges are most at risk.

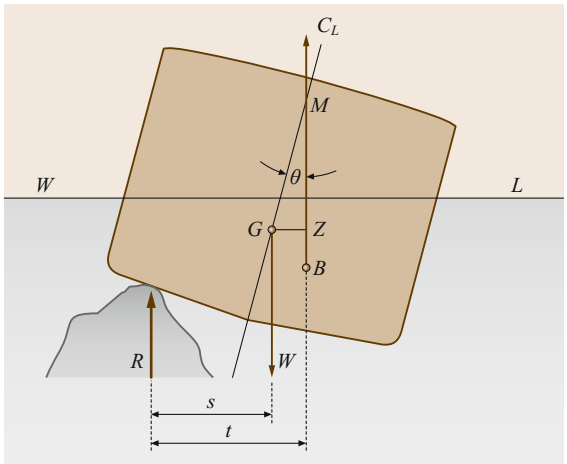


Fig. 42.6 Grounding off centerline

Friction and Freeing Force. Friction between the seafloor and the casualty's hull must be overcome to pull the ship free. On rock, coral, and cohesionless soils, the frictional force resisting the movement of a stranded ship is a function of the ground reaction component normal to the bottom and a coefficient of friction (μ). For the shallow bottom slopes encountered in most strandings, the normal component of ground reaction is essentially equal to the ground reaction (R) and

$$F_\mathrm{f} = \mu R\,, \qquad (42.15)$$

where F_f is the friction force.

Table 42.2 provides coefficients of friction for seafloor and shore soils based on the operational experience.

For cohesive soils, friction is a function of the contact area and soil shear strength

$$F_\mathrm{f} = sA\,, \qquad (42.16)$$

where s is the soil shear strength and A is the area of hull in contact with the seafloor.

42.B.2 Hull Strength in Salvage

Ships are designed and constructed to withstand the shear forces and bending moments expected in their normal service. In an intact floating ship, maximum bending moment occurs at or near the midships section. Maximum shear force occurs near the quarter-length points. Ships have heavier scantlings in high-load areas to keep stresses within acceptable levels. Two conditions common in salvage operations cause unusual hull stress distribution:

- The ship is loaded in ways the designer has not foreseen. Due to flooding, grounding, or other unusual conditions of loading, maximum bending moment occurs at a section other than midships, and maximum shear is at points other than the quarters.

Table 42.2 Coefficients of friction for strandings

Seafloor type	Coefficient of static friction [μ]	Remarks
Silt or mud	0.2–0.3	Suction effects not included
Sand	0.25–0.4	μ increases with grain size
Gravel or pebble	0.4–0.5	
Hardpan	0.5–0.8	Densely consolidated sand and clay, may include gravel
Coral	0.5–0.8	Lower values for dead or pulverized coral
Rock	0.8–1.5	See accompanying text

- Damage alters the geometry of the ship sections, so that maximum stress is in a section other than that of maximum bending moment or shear. Damage, even over a short distance, disrupts the continuity of longitudinal members and reduces the load-carrying ability of the structure for some distance on either side of the damaged section.

Casualty hull girder stresses should be monitored carefully throughout the salvage operation. The salvage engineer takes three actions with regard to hull girder strength:

- Analyzes the casualty's initial condition.
- Determines the casualty's ability to carry loads imposed by environmental conditions and planned salvage actions.
- Determines methods to reduce excessive hull stresses by altering load distribution or reinforcing critical sections.

Initial determination of strength deck and keel bending stresses, including the effects of damage and altered loading, sets a baseline for analysis. As a minimum, stress levels should be determined at damaged sections, where shear force or bending moment is maximum, and where the salvage engineer's judgment indicates there may be problems. The potential effects of major weight changes on load, shear, and bending moment should be examined before making the changes. Curves of maximum acceptable (limiting) shear forces and bending moments, based on actual section geometry, should be developed. Plotting bending shear force and bending moment curves and comparing them with allowable limits determines if planned actions will result in acceptable conditions.

Hull girder failure in bending usually starts with the compressive failure of the top or bottom of the girder. Compressive failure of longitudinally stiffened deck and single-bottom structures almost always takes the form of local inelastic buckling of longitudinals and attached plating. Hull girder failure may not be visible; the failure area may be obscured by other structures or the ship contents. Tensile failure is characterized by athwartship fractures in the outer flange of the structure. The fracture is accompanied by a loud report.

Damaged Strength

The hull's structural design keeps longitudinal bending stresses at acceptable levels under expected loads. Longitudinal bending stresses are highest in the sheer strake, strength deck, keel, bottom longitudinals, and bottom plating. In salvage operations, strength members should be inspected carefully for damage. Salvors should avoid intentional damage to members subject to high stress. Damage to members close to the neutral axis in the midships region, such as holing near the design or service waterline, has much less impact on longitudinal strength than damage to the strength deck and upper side strakes or bottom. Damage in the vicinity of the neutral axis near the quarter-length points can result in very high shear stresses.

A serious consequence of damage is the loss of structural continuity in critical strength members. The application of simple beam theory to ship girders is predicated on the assumption that the ship responds to shear forces and bending moments essentially as a single continuous beam. Loss of continuity in longitudinal members reduces the ability of the structure to spread shear loads across adjacent sections – both shear and bending stresses are increased. Loss of continuity in transverse members reduces the rigidity of the structure and allows longitudinal members to act independently, reducing longitudinal strength. Loss of transverse continuity in regions subject to shear or torsional loading increases shear stress and deformation. Increased shear stress and/or deformation in a section alters the bending stress distribution in that section. In addition, discontinuities in load-bearing members always act as stress raisers or concentration points. These effects, singly or in combination, can raise stresses in the hull girder to unacceptable levels.

When subjected to high enough stress, the hull will fail catastrophically. Usually, total hull failure begins with the compressive failure of the extreme fibers:

- Buckled plating, flange stiffeners, and tripped web stiffeners are evidence of compressive failure. Load shirking by the buckled members and hull deformation cause increased stress in other structural members that may then fail sequentially.
- Shear failure is characterized by wrinkles or cracks in side plating at a 45° angle to the line of stress.
- While not causing immediate failure, cyclic stressing of the hull to levels near the yield limit will reduce the endurance limit greatly, causing fatigue failure at tensile stresses below the ultimate stress. Casualties exposed to severe cyclic loading from wave or tidal action may fail in days or even in hours.

Structural members cut or torn away by battle damage, collision, or grounding cannot contribute to a section's ability to carry loads. Other types of damage reduce load-carrying capacity:

- Buckled members that have lost essentially all of their abilities to carry compressive loads, but keep a large portion of their original tensile strength.

- Large areas of heavily indented or dished plate that cannot carry compressive loads as high as undamaged plating. The tensile strength remains about the same.
- Cracks may appear around the edges of other damage or in otherwise undamaged structure. The great danger of cracks is that they will propagate under tensile loads perpendicular to the crack axis.
- Structural members exposed to high heat or fire suffer a loss of material from wastage due to accelerated corrosion, melting, or burning. Metal strength properties are altered unpredictably by the heating and cooling process, but strength and hardness are always taken as lowered. Fire-damaged members may have very little residual strength in either tension or compression.
- Holes, cuts, or tears act as stress raisers in tensioned members and induce buckling or tripping in members under compression.
- Accesses should not be cut where the loss of material will reduce the section modulus unacceptably or interrupt strength member continuity. If holes are cut in decks or shell plating, the effects of reduced plating area and stress concentration on hull girder strength must be determined, or the openings reinforced to compensate for the lost material.

Table 42.3 Standards for steel double barrel bitts

Bitt size (barrel diameter) [in]	Working load at 1/2 height, pounds	
	Welded[a]	Cast[b]
6	30 130	29 450
8	47 530	47 210
10	92 590	86 650
12	121 070	114 020
14	180 350	159 120
16	236 300	210 970
18	299 360	271 380
20	320 840	313 390
22	390 820	355 560
24	476 610	430 020

[a] based on safety factor of 2 and material yield of 36 ksi (AISI grade 1030)
[b] based on safety factor of 2 and material yield of 30 ksi (AISI grade 6030)

When stresses are within the elastic range, deleting damaged and missing structural members from the moment of inertia and shear area calculations gives a conservative estimate of a damaged hull's ability to carry shear and bending loads.

42.B.3 Lifting and Heavy Rigging

Identify suitable rigging points on a ships hull. Tables 42.3 and 42.4 provide reference load allowances for standard bitts and cleats, respectively.

Determination of Minimum Ground Leg Scope

The minimum ground leg scope is determined in the following manner:

1. Determine the *anchor depth* by summing:
   a) The height of the deck upon which the heaving gear is rigged above the waterline. (When spring buoys are used, the height of the deck above the water is not used.)
   b) The depth of water at the anchor.
   c) The embedded depth which equals:
      - 0 ft for firm sand or clay, coral, or rock
      - 5 ft for medium density sand or clay
      - 10 ft for soft mud.

   These depths are illustrated in Fig. 42.7.
2. Enter Table 42.5 with the anchor depth and read the basic ground leg scope. The basic ground leg scope includes the drag required to set the anchor properly.
3. To obtain the minimum ground leg scope, add the distance the ship must travel to refloat and the length of wire rope on deck to the basic ground leg scope. These distances are illustrated in Fig. 42.7. When the beach gear is laid to a salvage ship or barge, the distance the ship must travel is omitted from the calculation.
4. The total length of components that make up the ground leg should equal or exceed the minimum ground leg scope. Shorter scopes will cause anchor drag.
5. As chain and wire rope come in standard lengths, the next longer scope that can be made up with the components on hand is used.

Table 42.4 Steel welded horn-type cleat (after [42.4])

Cleat size (distance across horns) [in]	Rope Circumference [in]				Rope diameter [in]	Test load [lbs]
	Manila	Nylon	Dacron	Polypropelene	6 × 37 fiber core wire rope	
10	1 3/4	1	1	1 1/2		4100
16	3	2	2	2 1/2	5/16	9000
24	5	3	3	4	1/2	23 000
30	6	3 1/2	4	5	5/8	36 000

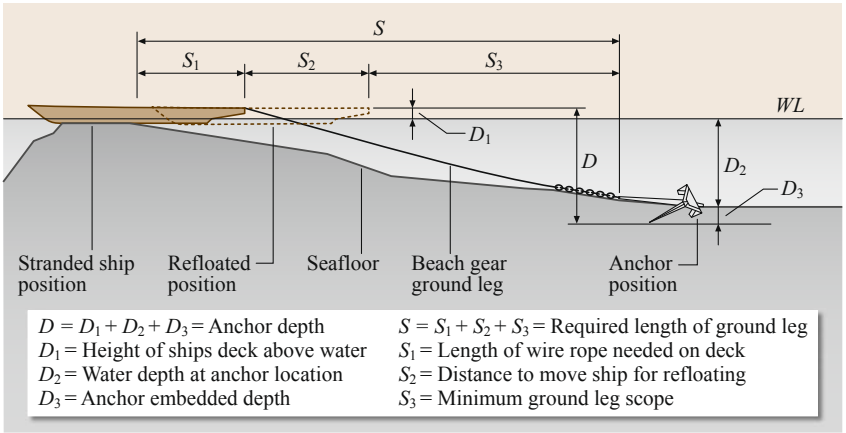


Fig. 42.7 Distances for determining ground leg scope

Table 42.5 Basic ground leg scope

Anchor (D)	1 – 5/8″ wire rope and one shot of 2 – 1/4″ chain	1 – 5/8″ wire rope and two shots of 2 – 1/4″ chain	1 – 5/8″ wire rope and three shots of 2 – 1/4″ chain
60	1120 ft	787 ft	NA
72	1275	907	732 ft
84	1420	1022	823
96	1557	1133	912
108	1637	1240	998
120	1810	1345	1083
132	1929	1445	1166
144	2043	1543	1248
156	2153	1639	1323
168	2260	1732	1407
180		1822	1485
192		1904	1561
204		1998	1636
216		2082	1709
228		2166	1782
240		2247	1854
252			1924
264			1993
276			2062
288			2130
300			2197

Table 42.6 Stress factor K in^{-1} × 1000

r Bilge radius [in]	*h* Hull plating thickness [in]						
	1/4	3/8	1/2	5/8	3/4	7/8	1
6	625	483	405	355	320	293	272
12	252	189	156	135	121	110	101
15	190	142	116	100	89	81	74
18	152	112	91	78	69	63	58
21	126	92	75	64	56	51	47
24	107	78	63	54	47	43	39

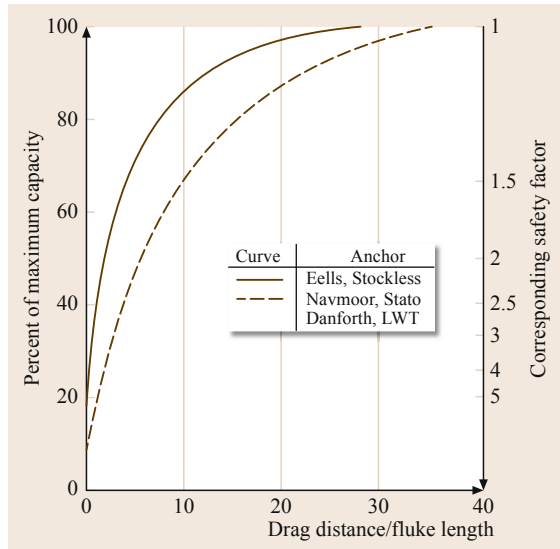


Fig. 42.8 Percent holding capacity versus drag distance in mud

In sand, maximum holding power is reached in less than 10 fluke lengths of drag. Figures 42.8 and 42.9 will help predict anchor drag distances in mud and select an anchor that will develop the required holding power in an acceptable drag distance.

Figure 42.10 shows a typical lift configuration. Each wire is rigged from one lift craft under the wreck to the other lift craft. There is a pair of wires at each transverse location. Each pair of wires has four lifting parts. The total of the vertical components equals the wreck's weight in water. The wire tension may be calculated by

$$T = \frac{W}{2N(1 + \sin \Phi)} , \qquad (42.17)$$

where N is the number of lift wire sets, W is the weight of wreck in water, T is the lift wire tension, and Φ is

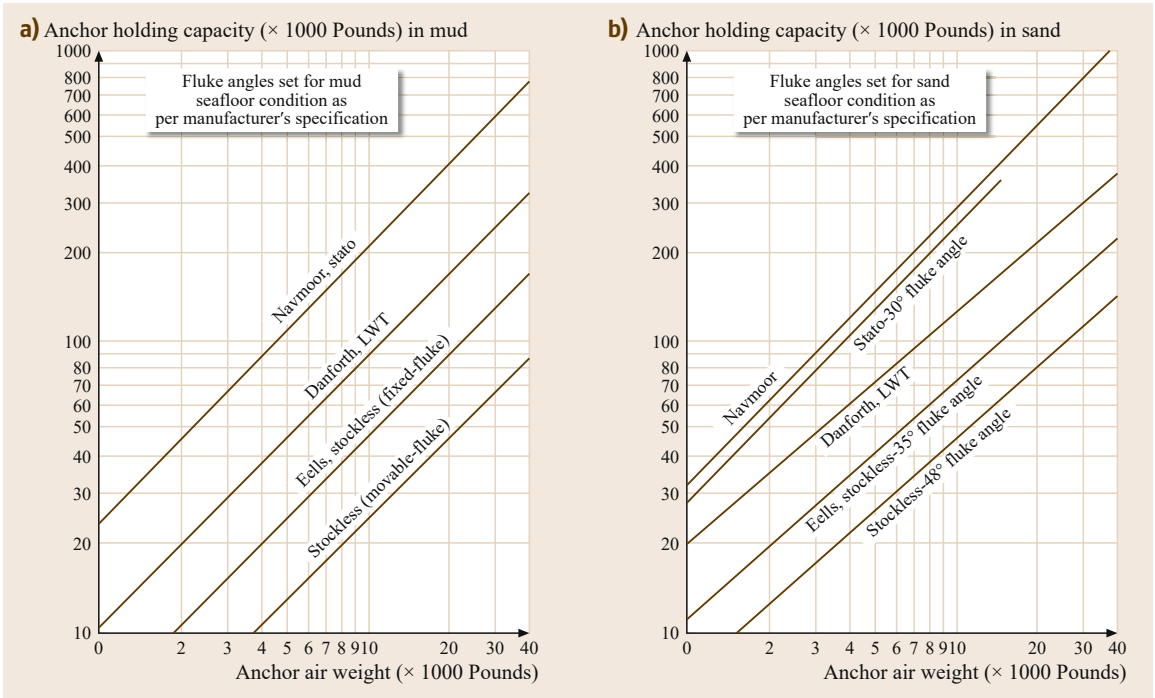


Fig. 42.9a,b Anchor holding capacity

the angle between bottom of lift craft and the outboard wire.

The possibility of lifting wire slicing, particularly at bilges, can be estimated with the empirical data presented in Table 42.6 and Fig. 42.11, a nomogram for the values of lift wire tension between 5 tons and 200 tons.

To estimate the possibility of bilge slicing occurring, with Table 42.6:

1. Calculate the lift wire tension T by taking direct readings from the hook, puller, or purchase readout gages.
2. Enter the K stress factor table with bilge radius and hull plating thickness to get K.
3. Draw a line connecting K and T; the estimated hull stress in way of each lifting wire is shown where the line crosses the principal stress line Z.

For a factor of safety of 1.5, the allowable stress for steel should be taken at 22 000 psi.

Wire Rope Selection

Breaking strength, fatigue resistance, crushing resistance, etc., and availability establish the base for conventional wire rope selection. An approximate selection formula is written as

$$\mathrm{DSL} = \frac{(\mathrm{NS})K_\mathrm{b}}{K_\mathrm{s}}\,, \qquad (42.18)$$

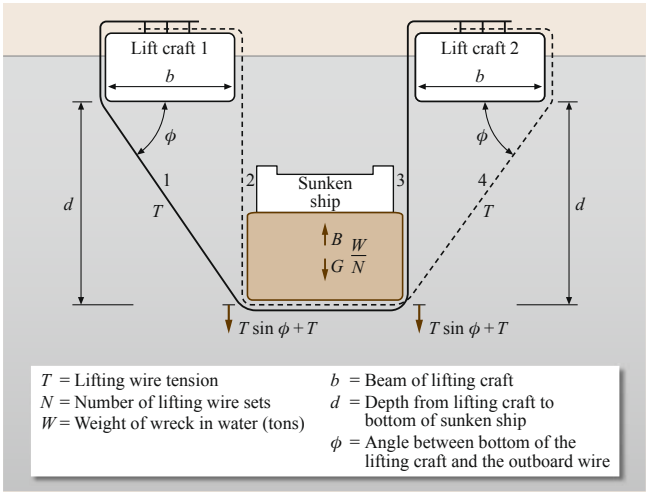


Fig. 42.10 Lifting wire tension sketch

where DSL is the demanded static load = known or dead load plus additional loads caused by sudden starts and stops, shocks, bearing friction, etc., NS is the nominal strength = published test strength, K_b is the bending factor from the curves in Fig. 42.12, and K_s is the safety factor, chosen as discussed below. In that sharp bends decrease both the strength and service life of wire rope, efforts should be taken to maintain a large ratio of sheave or drum diameter (D) to rope diameter (d).

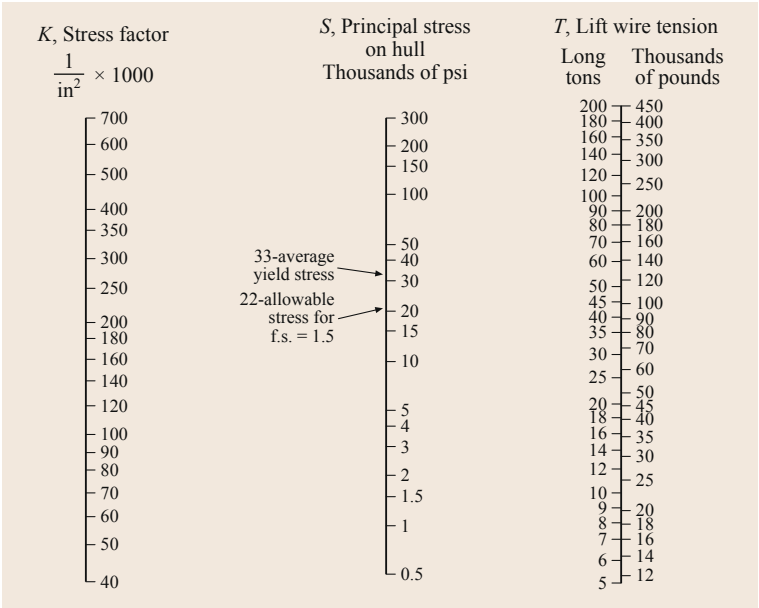


Fig. 42.11 Hull stress nomogram

Figure 42.12 provides bending factors used to calculate wire rope strength reductions as well as service life reductions.

A safety factor is selected after careful consideration of loads, accelerations, shocks, rope speeds, rope attachments, sheave arrangements and sizes, environment, etc. Standard salvage practice is to use safety factors of $3-1/2$ for standing rigging, 5 for running rigging, and 8–12 for situations with danger to life.

42.B.4 Compressed Air

Compressed Air Basics

Air compressors are rated in terms of their discharge pressure and flow rate, for example, a 500 psi/900 cfm compressor. Flow rate is given in terms of air at a standard condition. The following are the standard conditions:

- Pressure – one atmosphere (14.7 psi)
- Temperature – 68 °F
- Relative humidity – 36%
- Density – 0.075 lbm/ft^3.

In English units, flow rates are expressed in standard cubic feet per minute (SCFM). The actual volume occupied by the air delivered by a compressor depends on the pressure of the system or receiving vessel (volume tank, ships compartment, etc.). Volumes of air at standard conditions are expressed in standard units, for example, standard cubic feet (SCF). The quantity of standard air required to fill an actual volume, under nonstandard conditions, measured in actual units (actual cubic feet, ACF) can be found by a modification of

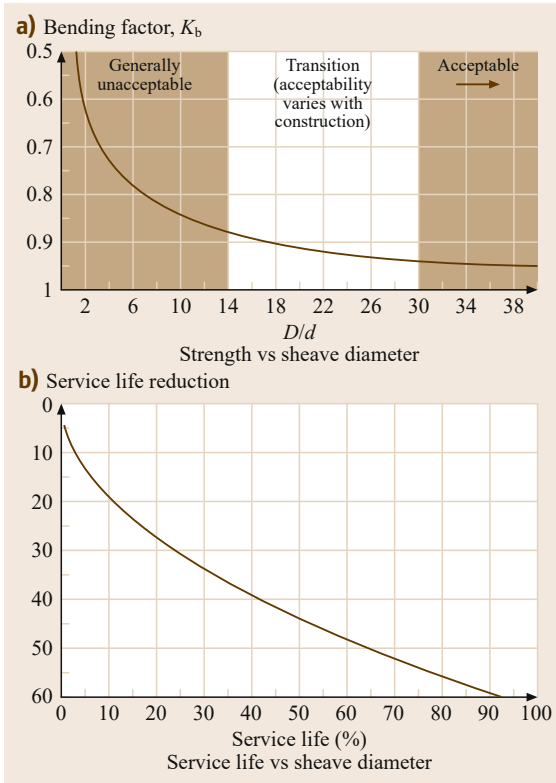


Fig. 42.12a,b Effect of sheave diameter on 6×9 and 6×37 class ropes

the gas law

$$V_s = \left(\frac{P}{P_{atm}}\right)\left(\frac{T_a}{T_w}\right)(V_a)\,, \tag{42.19}$$

where V_s is the standard air volume (length^3), P is the pressure in the space where air is required, absolute [force/length^2], P_{atm} is the atmospheric pressure, absolute [force/length^2], T_a is the water temperature at depth, absolute $= T_{Fahrenheit} + 460 = T_{Centigrade} + 273$, T_w is the water temperature at depth, absolute $= T_{Fahrenheit} + 460 = T_a$ the air temperature, absolute $= T_{Fahrenheit} + 460 = T_{Centigrade} + 273$.

Pressure can be expressed in feet of seawater (FSW):

- 1 FSW = 0.445 psi
- 33 FSW = 14.7 psi = 1 atmosphere.

When pressures are expressed as FSW, standard volume is given by

$$V_s = \frac{D+33}{33}(V_a)\left(\frac{T_a}{T_w}\right)\,. \tag{42.20}$$

Compressors

Compressors and compressed air systems are classified by their pressure rating:

- Low pressure (LP) – 150 psig or less
- Medium pressure (MP) – 150–1000 psig
- High pressure (HP) – 1000 psig or higher.

Compressors, like pumps, are classified by their design and operating features. There are two principal compressor types:

- Positive-displacement (reciprocating and rotary) compressors
- Dynamic (centrifugal and axial) compressors.

Positive-displacement compressors collect a fixed volume of air within a chamber and compress it by reducing the chamber volume. Dynamic compressors operate by transferring momentum to the air via a high-speed rotor. Blowers operate on the same principles as compressors, but at lower pressures. Thirty-five psig is the commonly used arbitrary dividing line between compressors and blowers.

Compressor displacement is the actual swept volume of the air chamber. Capacity is the quantity of air delivered by the compressor, in SCFM. Efficiency (η) is the ratio of capacity to displacement swept per minute

$$\eta = \frac{\text{SCFM}}{(\text{Displ})(\text{RPM})}\,, \tag{42.21}$$

where SCFM is the compressor capacity, standard cubic feet per minute, Displ is the compressor displacement [ft^3], and RPM is the compressor speed, rotations per minute.

If compressor displacement is given without capacity or efficiency, capacity can be estimated by assuming 80% efficiency.

Positive-displacement compressors can reach very high pressures. Unless a relief valve is installed, damage to the compressor, prime mover, or system is very likely. Dynamic compressors will reach a maximum pressure where the momentum imparted to intake air is insufficient to overcome the back pressure on the outlet; flow is zero at this pressure. All dynamic compressors have a minimum flow point, called the surge limit, below which the operation of the machine is unstable. Operation at or below the surge limit must be avoided.

Blowing and Differential Pressures

The pressure required to dewater a space is a function of water depth at the opening or bottom of the standpipe through which the water will be forced

$$P_b = 0.445D + P_L\,, \tag{42.22}$$

where P_b is the blowing pressure (psig), D is the water depth (feet), 0.445 the pressure increase per foot of seawater, and P_L the pressure required to overcome air line losses and standpipe friction, normally taken to be 2 psi.

In deep compartments, the differential pressure required to completely dewater a space may be greater than the bulkheads and top of the compartment can withstand. If so, it will be necessary to reinforce the ship's structure or limit the differential pressure by varying the length of the standpipe or the height of the vents

$$l = \left(\frac{P_d - P_L}{0.445}\right)\,, \tag{42.23}$$

where l is the length of the standpipe or distance of vent from top of compartment, P_d isthe maximum acceptable differential pressure (psig), and P_L is the pressure required to overcome air line losses and standpipe friction, normally taken to be 2 psi.

Air Flow Requirements

The standard volume of air (V_s) required to completely dewater a space is based on the pressure at the opening or bottom of the standpipe

$$V_s = \frac{D+33}{33}(V_a)\left(\frac{T_a}{T_w}\right)\,, \tag{42.24}$$

where D is depth to the vent or bottom of the standpipe (feet), V_a is the water volume, actual cubic feet = space

volume × permeability, T_w is the water temperature at depth, absolute, and T_a is the air temperature, absolute.

If the standpipe or hull openings are too small, water will flow out of the compartment at a slower rate than air flows in, creating a hydraulic block that can cause pressure to build up. To avoid damage to the space or air system, it will be necessary to secure blowing periodically until sufficient water has been forced out of the space to lower the pressure. If compartment pressure is not carefully monitored, the space can be easily over pressurized, causing damage or injury. To avoid over pressurization, the water flow rate out of the compartment must equal the actual air flow rate (Q_a) into the compartment. To find Q_a

$$Q_a = \frac{Q_s}{\text{ATA}}\,, \tag{42.25}$$

where: ATA is the absolute pressure in atmospheres $= \frac{D+33}{33}$ and Q_s is the air delivery rate.

By transposing the equation for flow through a hole or orifice, required outlet area can be found

$$Q = C_d A\sqrt{2gh_{eq}}\,, \quad A = \frac{Q}{C_d\sqrt{2gh_{eq}}}\,, \tag{42.26}$$

where: Q is the water flow rate, $\text{ft}^3/\text{sec} = Q_a$, C_d is the discharge coefficient, from Fig. 42.13, $A =$ area of outlet [ft^2], g is the acceleration due to gravity $\approx 32.2\,\text{ft/sec}^2$, and h_{eq} is the blowing pressure, expressed as an equivalent head of seawater, ft $= P_b/0.445$.

The air pressure in the compartment will equal the hydrostatic pressure at the depth of the water level in the compartment; as blowing progresses, the water level falls, air pressure increases, and flow rates (water and actual air) decrease. The standpipe or opening should be sized to accommodate the maximum flow rate – the flow corresponding to the depth at the top of the compartment.

The variation in flow rate and pressure through the dewatering may give the impression that higher driving pressure shortens dewatering time. While raising the driving pressure increases initial water flow rate, total dewatering time depends on the volume flow of the compressed air system; the compartment will not be dewatered until the air system has delivered sufficient actual cubic feet of air to fill the required volume. Flow rates of centrifugal compressors are related to delivery pressure and the higher flow rate accompanying the increased pressure will in fact shorten dewatering time.

42.B.5 Pumping

Pump Theory and Terminology

Pump terminology is based on the concept of head. Head is a measure of the energy possessed by the liquid because of its pressure (pressure head), velocity (velocity head), or elevation above a datum (static head). Head is most commonly expressed in units of distance (height) or pressure.

In the English system, pressure is usually measured in pounds per square inch (psi) and density in pounds per cubic foot. A conversion factor of $144\,\text{in}^2/\text{ft}^2$ must be used to obtain pressure head in feet

$$H = \frac{144P}{\gamma}\,, \tag{42.27}$$

where H is head, feet of liquid, P the pressure (psi), and γ the fluid density [lb/ft^3].

Densities of various liquids are tabulated in Table 42.13. Additional head terms are used to describe work done by the pump or energy available to the pump

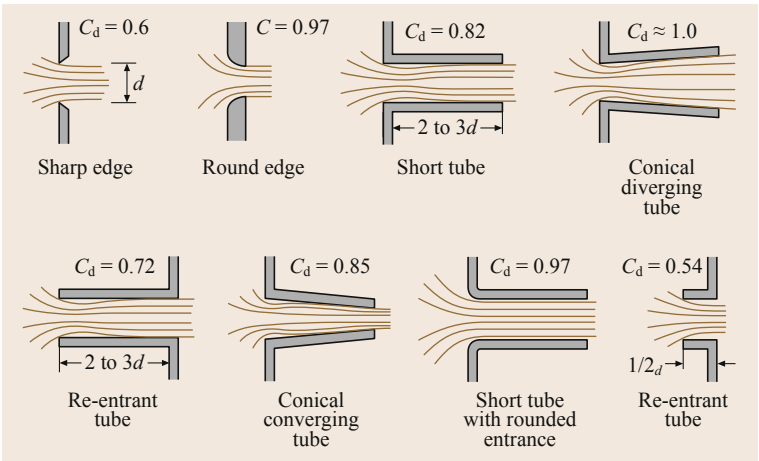


Fig. 42.13 Discharge coefficients

(because of the system configuration), and pump operating characteristics and requirements. These terms are described in the following paragraphs and illustrated in Fig. 42.14.

Atmospheric Head (H_a)

Atmospheric head is the atmospheric pressure expressed as feet of the pumped liquid. For seawater at normal atmospheric pressure

$$H_a = \frac{(144\,\text{in}^2/\text{ft}^2)(14.7\,\text{lb/in}^2)}{(64\,\text{lb/ft}^3)} = 33\,\text{ft}\,. \qquad (42.28)$$

Vapor Pressure Head (H_{vp})

Vapor pressure head is the fluid vapor pressure expressed as feet of the pumped liquid. For seawater at 68 °F, $P_{vp} = 0.34$ psi, and

$$H_{vp} = \frac{144(0.34)}{64} = 0.765\,\text{ft}\,. \qquad (42.29)$$

Vapor pressures for some common liquids at 68 °F are given in Table 42.7.

Static Suction Head (H_s)

Static suction head is the vertical distance between the liquid surface and the pump inlet. If the pump is above the liquid surface, suction head is negative. When working against negative suction head, or suction lift, a pump must create a vacuum so that the pressure differential between the liquid surface and the suction side of the pump is sufficient to lift the liquid, overcome vapor pressure in the pump casing and friction in the suction line, and accelerate the liquid. Maximum suction lift is equal to the atmospheric head (H_a) less vapor pressure head (H_{vp}), and friction head in the suction line (H_{fs})

$$\text{Max Suction Lift} = H_a - H_{vp} - H_{fs}\,. \qquad (42.30)$$

Actual suction lifts will be slightly less because of work required to accelerate the liquid and losses due to pump inefficiencies. The maximum practical suction lift for seawater at atmospheric pressure is about 25 ft for most centrifugal pumps; positive-displacement pumps can achieve slightly higher lifts. For efficient pump performance, suction lift should be minimized, below 15 ft if possible. Most centrifugal pumps must be primed to achieve any suction lift.

Table 42.7 Vapor pressures a 68 °F

Liquid	Vapor pressure [psi]
Water	0.34
Ethyl alcohol	0.85
Turpentine	0.0077
Diesel fuel (typical)	0.041
Heavy oils	0.042
Vegetable oil	0.030

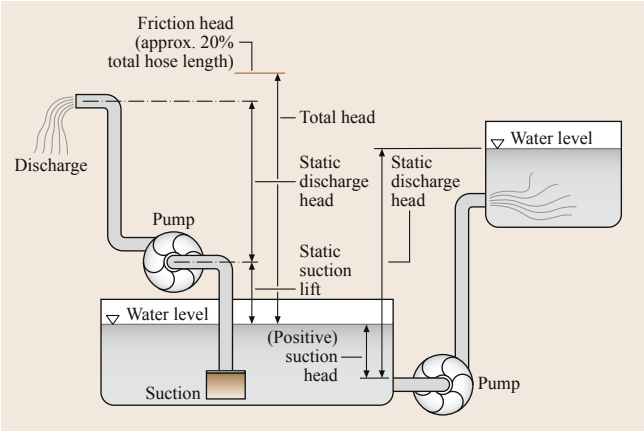


Fig. 42.14 Pumping heads

Static Discharge Head (H_d)

Static discharge head is the vertical distance from the pump to the point of free discharge, and represents the energy the pump must impart to the liquid to raise it to that height in an ideal system. If the discharge is submerged, the static discharge head is measured to the liquid surface.

Friction Head (H_f)

Friction loss in a real system is equivalent to lifting the liquid an additional height in a frictionless system. Head required to overcome friction in pipes, hoses, valves, fittings, etc., is sometimes given as pressure drop, or head loss, as a function of flow rate and length for specific types of hose, pipe, or fittings. Friction head can also be calculated as a function of velocity head, as in the Darcy–Weisbach formula

$$H_f = \frac{fLV^2}{2Dg}\,, \qquad (42.31)$$

where f is the Darcy–Weisbach friction factor, dimensionless, L is the length [ft], D is the diameter [ft], V is the flow velocity [ft/s], g = acceleration due to gravity [ft/s^2] ≈ 32.2 ft/s^2.

For laminar flow ($R_e \leq 2000$), $f = 64/R_e$, where R_e is the dimensionless Reynolds ($\Re$) number

$$R_e = \frac{DV}{\nu} = \frac{\rho DV}{\mu}\,. \qquad (42.32)$$

With kinematic viscosity (ν), fluid density (ρ), and absolute viscosity (μ), expressed in units consistent with the units of velocity and diameter. Friction head

is a function of fluid velocity, flow area, and surface roughness; for a given flow diameter, friction head increases with flow rate. Friction factors can be seen in Fig. 42.15. Friction loss in valves, fittings, and other obstructions is normally expressed as a function of velocity head

$$H_f = K\frac{V^2}{2g}\,, \tag{42.33}$$

where K is an obstruction loss factor taken from Table 42.8.

More extensive lists of friction factors or friction loss tables can be found in pipefitter or civil engineering handbooks. If unknown, friction head can be taken as not more than 20% of the total length of suction and discharge piping or hose for most salvage applications.

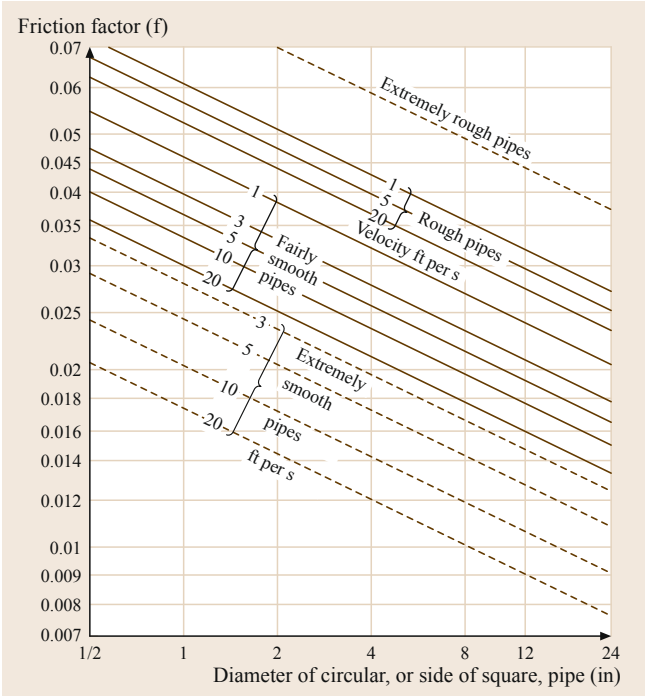


Fig. 42.15 Friction factors for pipe flow

Table 42.8 Obstruction factors (after [42.5])

Fitting	Obstruction factor [K]
Globe valve, wide open	10
Angle valve, wide open	5
Gate valve,	
wide open	0.19
half open	5.6
Strainer	2
Smooth bend	0.3
Short radius elbow	0.9
Long radius elbow	0.6

Total Dynamic Head Required (TDHR)

Total dynamic head required is a measure of the energy necessary to move the liquid through the system to the desired discharge. This measure is equal to discharge head, plus friction head, less suction head

$$\text{TDHR} = H_d + H_f - H_s\,. \tag{42.34}$$

As negative suction head (suction lift) increases, TDHR is also increase. Power used to overcome suction lift is not available to impart energy to the pumped liquid; discharge capacity and/or head are reduced.

If the pump is below the surface of the liquid to be pumped, there is a positive hydrostatic pressure on the inlet side of the pump; no work is required to raise the liquid to the pump. The positive suction head represents increased pump capacity as reflected by the reduced TDHR. Figure 42.16 shows the effect of positive suction head on a typical centrifugal pump.

Total Dynamic Head Available (TDHA)

TDHA is a measure of the total energy imparted to the fluid by the pump, and is equal to the sum of pressure and velocity head at the pump outlet

$$\text{TDHA} = H_{p2} + H_v\,. \tag{42.35}$$

Net Positive Suction Head (NPSH)

Net positive suction head is the absolute suction pressure, less vapor pressure at the pump inlet, under actual pumping conditions. It is the sum of atmospheric head and static suction head, less vapor pressure head and friction head in the suction lines

$$\text{NPSHA} = H_a + H_s + H_{vp} + H_{fs}\,. \tag{42.36}$$

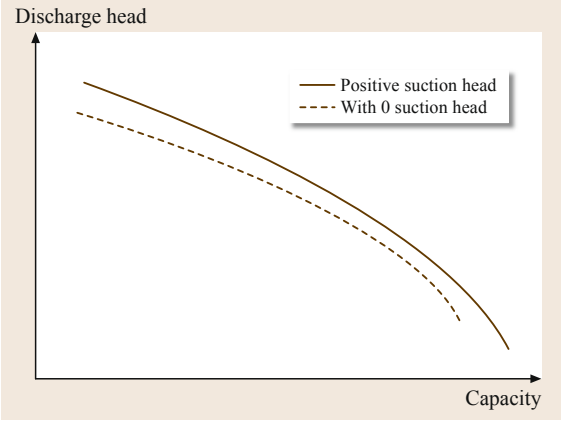


Fig. 42.16 Effects of positive suction

Net positive suction head required (*NPSHR*) is a measure of the minimum fluid energy necessary for satisfactory pump operation and is usually specified by the pump manufacturer. NPSHR is a means of quantifying the work required to accelerate the liquid and other losses. Some pump manufacturers prefer to define net inlet pressure required (NIPR); NPSH multiplied by fluid density gives net inlet pressure (NIP).

Pump Classification

Pumps are classified by their design and operating features:

- Movement that causes the pumping action
- Submersible or nonsubmersible
- Characteristics of the fluid pumped
- Flow characteristics
- Self-priming or non-self-priming
- Constant or variable speed
- Variable or constant capacity.

Major pump types are:

- Dynamic pumps, including centrifugal, axial flow, and mixed flow pumps
- Positive-displacement pumps, including diaphragm, rotary, and reciprocating pumps
- Jet pumps and air lifts.

Table 42.9 gives general characteristics of each type of pump.

Positive-Displacement Pumps

Positive-displacement pumps discharge a specific volume for each stroke or revolution; at a constant speed, capacity is essentially the same for any pressure within the capability of the driver and strength of the pump. At some discharge heads (pressure), the power required to maintain pump speed will exceed the capacity of the prime mover; as the pump slows and eventually stalls, capacity falls off rapidly. In practice, some component of the pump or piping system will fail before the prime mover is overloaded. To avoid failure, a relief valve may be installed in the discharge line.

Table 42.9 Pump characteristics

Characteristic	Displacement	Centrifugal and axial	Jet
Flow rate	Low	High	High
Pressure rise per stage	High	Low	Low
Constant variable over operating range	Flow rate	Pressure rise	–
Self-priming	Yes	No	Yes
Outlet stream	Pulsing	Steady	Steady
Works with high-viscosity fluids	Yes	No	Yes

Dynamic Pumps

Dynamic and jet pumps operate by imparting the kinetic energy of a rotating impeller (centrifugal, axial flow, and mixed flow pumps) or fluid stream (jet pumps) to the pumped liquid; the energy imparted to a liquid is converted to velocity and pressure. To some extent, velocity of the moving fluid can be converted to pressure and vice versa. Velocity head is related to capacity

$$Q = A\sqrt{2gH_v} \Rightarrow H_v = \frac{Q^2}{2gA^2} \,. \tag{42.37}$$

Jet Pumps

Jet pumps use the low pressure created by the passage of a working fluid through a restricting nozzle to draw the pumped fluid into a mixing chamber, where it is entrained with the working fluid and carried out of the eductor. Jet pumps using a liquid as a working fluid are called eductors; those using gas (such as steam or compressed air) are called ejectors. Eductors used in salvage are almost always water driven. Head and volume capacity depend on nozzle and mixing chamber geometry, pressure and flow rate of the working fluid (supply flow), suction lift, and discharge head. For most eductors designed for damage control or salvage dewatering, discharge flow (including the supply flow) is about $1-1/2$ to 2 times the supply flow, with a maximum working head of about 70 ft of water.

All eductors have the minimum supply pressure and flow requirements below which they will not operate. If the supply pressure and flow are inadequate, the supply water may dump through the eductor suction, flooding the space. Because they have no moving parts, eductors are extremely rugged and reliable. They are particularly suited for pumping slurries and contaminated water. The annular ring-type (*Peri-jet*) eductor uses several nozzles arranged around the lower periphery of the mixing chamber rather, than a single nozzle directly below the mixing chamber. This arrangement results in a straightline fluid path through the eductor that is clear of obstructions. A 4-in Peri-jet eductor can pass 2 in contaminants, although large quantities of solid contaminants may clog long discharge hoses.

Large eductors are often installed as bilge and ballast pumps. Some tankers use eductors as cargo pumps. The working fluid is cargo, delivered under pressure by other types of cargo pumps.

Air Lifts

An air lift operates by introducing compressed air into the lower end of a submerged or partially submerged

pipe. The air–liquid mixture inside the pipe is less dense than the liquid outside the pipe. The mixture rises and creates a low pressure at the bottom of the pipe that draws liquid and loose solids up the pipe. Air lifts can handle all kinds of slurries, including those consisting of relatively large particles, such as coal, gravel, ore, canned goods, or other odd-shaped objects. Air lifts are frequently used to clear mud and loose sediments from diver-working areas and to remove sediments from inside wrecks.

The efficiency of an air lift depends on air pressure and flow rate, ratio of immersed to emerged pipe length, water depth, position of the air inlet on the pipe, and the nature of the material being lifted. Air lifts will normally lift loose material only in the immediate vicinity of the lower end. Disturbing heavy or hard-packed material with water jets or other means will improve air lift efficiency. Clay, paper pulp, and similar materials tend to choke air lifts. Although air lifts are not particularly efficient as pumps, they are easy to build in the field and can move slurries, semisolid materials, and contaminants that would clog or damage other pumps.

Air lifts differ from other pumps in that the energy (air pressure) required to lift the liquid is governed by submergence, h, rather than by total head, H. For most efficient operation, air lifts should be submerged for at least two-thirds of their length, although they can operate with as little as 35% submerged. Air pressure at the air line outlet at the bottom of the lift only slightly greater than hydrostatic pressure is sufficient for operation. An empirical formula, based on experience with air lifts in mining applications, estimates required air flow

$$Q_{\text{air}} = \frac{H}{C \log(ATM)} , \tag{42.38}$$

where: Q_{air} is the air flow, standard cubic feet per minute per gallon of liquid per minute lifted, H is the total vertical lift [ft], C is the empirical factor, based on ratio of submergence to lift, from Table 42.10, ATM is submergence, expressed as equivalent atmospheres $= (h + 33)/33$ for seawater, $(h + 34)/34$ for fresh water, $(h + 34\gamma_{\text{g}})/34\gamma_{\text{g}}$ for other liquids, h is submergence (depth to foot of air lift) [ft], and γ_{g} is the liquid specific gravity.

Table 42.10 Air lift constants, C (after [42.6])

Submergence percent	C	
	Outside air line	Inside air line
75	366	330
65	348	306
55	318	262
45	272	214
35	216	162

Pipe size is determined from the initial estimates for air and liquid flow

$$d = 13.54 \sqrt{\frac{Q}{A}} , \tag{42.39}$$

where d is the air lift (discharge) pipe diameter [in], Q is the air–liquid mixture flow rate [ft^3/min], and V is the velocity of the mixture in the pipe [ft/min].

For pipes of uniform diameter and lifts of 40–200 ft, best discharge velocity of the air-liquid mixture ranges from 2000 ft per minute at 70% submergence to 700 ft per minute at 35% submergence. Maximum velocity at the bottom of the pipe ranges from 450 ft per minute at 70% submergence to 800 ft per minute at 35% submergence.

Air lifts cannot completely dewater a space because they cannot operate with more than about two-thirds of the pipe out of the water. When depth allows 25% submergence, a compound lift can be employed. Water and solids are lifted half way in one lift and allowed to run to the closed bottom of a large-diameter pipe at the same depth as the first lift. A second air lift can operate in the large pipe with 50% submergence. In deep mine shafts, water has been raised as much as 1385 ft by a series of air lifts in series using air at 60–80 psi. Air lifts are useful for removing sediments, bulk cargo, and other materials from the bottoms of spaces before dewatering with other pumps.

Salvage Pumps

Special-purpose pumps are used in salvage to meet particular requirements; salvage pumps are usually portable, general purpose dewatering pumps, adapted for marine use. A good salvage pump must have the following features:

- Rugged construction and protective framework or packaging to reduce the risk of accidental damage
- High pumping-capacity-to-pump-weight ratio
- Discharge head greater than 60 ft
- Ability to pump a variety of contaminated liquids
- Comparatively simple construction for rapid routine maintenance and repair.

Submersibility or a low NPSHR, and the ability to self-prime as well as to handle a wide range of fluid viscosities and specific gravities are also desirable. Experience has shown the following pumps to have the widest application to salvage:

- Electric- or hydraulic-motor-driven submersible pumps (centrifugal or axial flow)

- Self-contained, heavy-duty, diesel- or gasoline-engine-driven centrifugal pumps
- Pneumatic diaphragm and centrifugal pumps
- Eductors and air lifts.

Table 42.11 compares the advantages and disadvantages of commonly used salvage pumps.

Pumping Calculations

Salvage pumping calculations are usually undertaken to answer one of the three questions:

- What pumping capacity (and number of pumps) will be required to dewater a flooded space in a given time?
- How much time will be required to dewater a space with a specified pump capacity?
- At what rate will the water surface be lowered?

Suction and discharge heads do not remain constant, but change as pumping proceeds. For a fixed position pump, suction lift increases as the water level drops. If the pump is lowered to limit suction lift, discharge head will increase. Discharge head of a submersible pump remains constant, but the positive suction head decreases as the water level drops.

The means of accounting for friction and vapor pressure heads depends on the format of the pump performance data. Friction head for assumed lengths of discharge and suction piping is sometimes accounted for in performance curves or tables. The vapor pressure head is not factored into the performance curves of most pumps. Increasing fluid vapor pressure decreases a net positive suction head, so vapor pressure head can be taken into account for by adding it to the actual suction lift to determine an effective suction lift (negative suction head). The entering arguments to the performance curves or tables are thus

$$H_{s\,\mathrm{eff}} = H_{s\,\mathrm{actual}} + H_{\mathrm{vp}}\,,$$
$$\mathrm{TDHR} = H_{s\,\mathrm{eff}} + H_{\mathrm{d}} + H_{\mathrm{f}}\,, \qquad (42.40)$$

where $H_{s\,\mathrm{eff}}$ is the effective suction lift, $H_{s\,\mathrm{actual}}$ is the actual suction lift, measured from liquid surface, H_{vp} is the vapor pressure head, TDHR is the total dynamic head required, H_{d} is the static discharge head, and H_{f} is the friction head of hose or piping not accounted for in performance data.

Dewatering time for a given pumping capacity is the volume of water to be removed, divided by the pumping capacity. Since pumping capacity will vary as suction and discharge heads change through the pumping op-

Table 42.11 Pump characteristics

Pump type	Discharge head	Self-priming	Suction lift	Suitability to pump			Typical installations
				Entrained solids	Slurries	Viscous liquids	
Reciprocating	High	Yes	High	Fair	Good	Good	General marine use with steam propulsion plans. Slurry pumps. Oilfield mud pumps
Rotary	High	Yes	High	Poor	Poor	Good	Marine and industrial fuel, cargo, lube oil systems
Diaphragm	High	Yes	High	Fair	Good	Good	General-purpose. Metering
Centrifugal	Medium[a]	No[b]	Low	[c]	[c]	Poor	General use for nonviscous liquids, F/F systems. Water pumps for agricultural, construction, mine, industrial, salvage use
Mixed flow	Low[a]	No	Very low	[c]	[c]	[d]	General use for nonviscous liquids, F/F systems. Water pumps for agricultural, construction, mine, industrial, salvage use
Axial flow	Very low[a]	No	Very low	[c]	[c]	[d]	General use where high discharge head not required. Submersible pumps
Jet pumps	Low	Yes	Medium	Good	Good	Fair	Distilling plant air ejectors. Bilge dewatering/drainage. Portabtle dewatering. F/F foam proportioning systems

[a] Head can be increased by series multistaging or operating pumps in series, [b] Some centrifugal pumps are self-priming at low (< 12 ft) suction lift, [c] Pimps designed as trash or salvage pumps, [d] Impeller must be designed for specific viscosity range

eration, an incremental solution is required. The time required to pump out layers of convenient thickness (10 ft, for example) of the flooded space is calculated based on the pump capacities corresponding to the average suction and discharge heads for each layer. The total dewatering time is the sum of the dewatering times for all the layers.

To determine the rate of fall of the water surface, the volume of one unit (inch, foot, centimeter, etc.) is first calculated or estimated. The volume of the unit layer divided by the pumping capacity is the rate of fall of the water surface.

Pumps in Unusual Service

Pumps are designed to operate with a specific set or range of fluid properties. If fluid properties of the pumped liquid vary significantly from those for which the pump was designed, pump performance and service life will be affected. There are five fluid properties to be considered:

- Specific gravity
- Viscosity
- Vapor pressure
- Abrasiveness
- Corrosivity.

Viscosity

Fluid viscosity affects the pump performance in a complex manner. Figure 42.17 shows the effect of viscosity on pump capacity and discharge head. The chart is entered on the bottom horizontal scale with the pump's rated capacity (water). A vertical line is followed from the capacity until it intersects the corresponding head on the diagonal head scales. A horizontal line is then followed to the left until it intersects the viscosity of the liquid being pumped. A vertical line from this intersection intersects the correction factor curves; correction factors are read from the scale on the left. The example (dotted line) shows that a pump rated for 55 gpm at a 100-foot discharge head has a capacity of 45 gpm (0.82×55) against a 103 ft discharge head (1.03×100) when pumping a liquid with a viscosity of 400 SSU (Saybolt universal seconds).

Approximate viscosity limits of basic types of pumps are:

- Centrifugal: 3000 SSU
- Axial flow: 8000 SSU
- Reciprocating: 10 000 SSU
- Rotary: 2 000 000 SSU.

As a rule, pump speed must be reduced 25–35% for each tenfold increase in viscosity above 1000 SSU to avoid overloading the pump prime mover. This results in a 10% reduction in efficiency. The suction requirements of the pump also vary with viscosity and are referred to by viscous liquid pump manufacturers as net inlet pressure required (NIPR).

Viscosity varies with temperature and extreme pressure. The pressures encountered in dewatering and weight removal operations will have no significant effect on fluid viscosity. Table 42.12 gives viscosities for some common liquids.

42.B.6 Petroleum, Oils, and Lubricants

Definitions

There are descriptive terms for the characteristics of petroleum products that influence behavior during handling. Knowing these terms leads to a better understanding of the mechanics of POL operations allowing the user to anticipate and avoid some of the problems that occur during transfer operations.

Viscosity

A measure of the resistance to flow (internal friction) of the product at a particular temperature. Kinematic viscosity is expressed in centistokes (cSt). The viscosity of water at 68 °F is approximately 1.0 cSt. Highly viscous liquids are thicker and flow more slowly than less viscous liquids. A viscosity of about 875 cSt is the upper limit of pumpability for centrifugal pumps. An-

Table 42.12 Viscosities

Liquid	Viscosity	
	SSU	Centistokes
Water @ 32 °F	33.0	1.79
Water @ 70 °F	30.9	0.98
Water @ 212 °F	29.3	0.29
Ethyl Alcohol @ 68 °F	31.7	1.52
Freon 12 @ 70 °F	0.27	21.1
No. 1 Fuel Oil @ 70 °F	34–40	2.39–4.28
No. 2 Fuel Oil @ 70 °F	36–50	3.0–7.4
No. 3 Fuel Oil @ 70 °F	33–40	2.69–5.84
No. 5 Fuel Oil @ 70 °F	50–125	7.4–26.4
No. 5 Fuel Oil @ 100 °F	42–72	4.9–13.7
No. 6 Fuel Oil @ 122 °F	450–3000	97–660
No. 6 Fuel Oil @ 160 °F	175–780	37–172
Jet Fuel @ −30 °F	52.0	7.9
Lubricating oils		
SAE-5W @ 0 °F	6000 max	1295 max
SAE-10W @ 0 °F	6000–12 000	1295–2590
SAE-20W @ 0 °F	12 000–48 000	2590–10 350
SAE-20 @ 210 °F	45–58	5.7–9.6
SAE-30 @ 210 °F	58–70	9.6–12.9
SAE-40 @ 210 °F	70–85	12.9–16.8
SAE-50 @ 210 °F	85–110	16.8–22.7

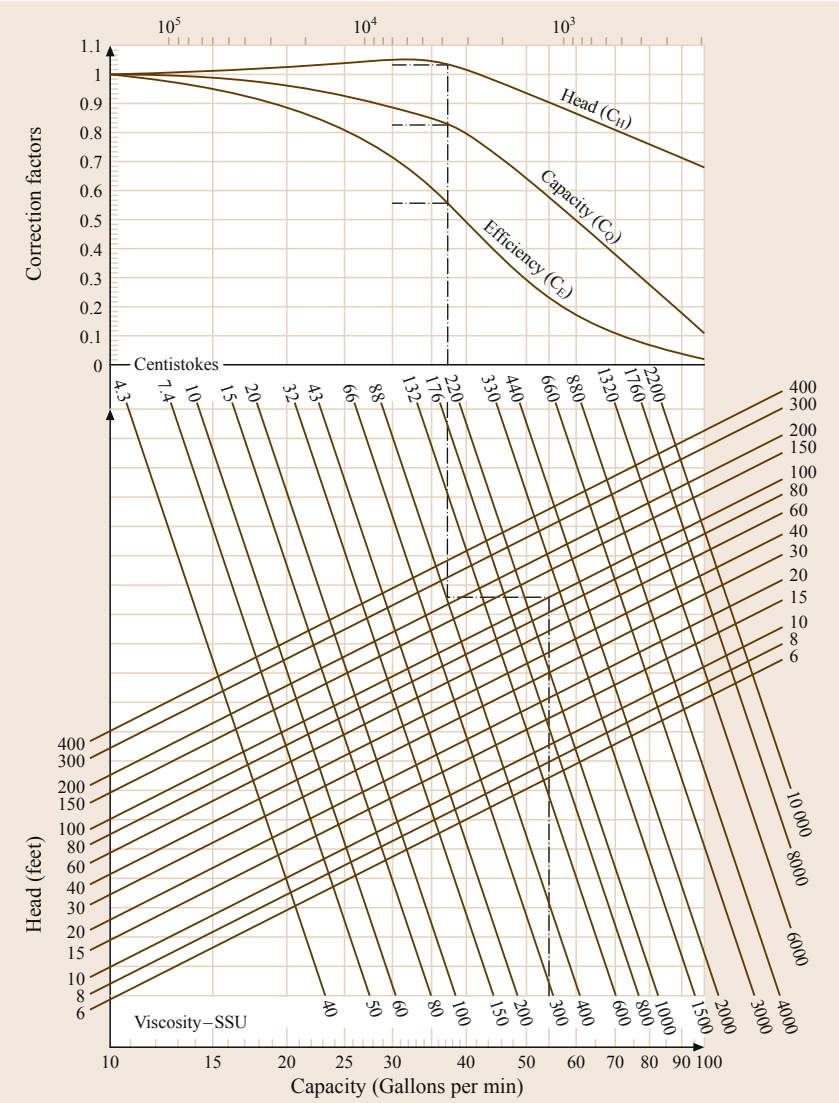


Fig. 42.17 Pump performance correction factor chart

other type of viscosity, absolute viscosity, is measured in poise. The absolute viscosity of water at 32 °F is 1.8086 centipoise. The two viscosities are related by

$$\text{Kinematic viscosity} = \text{Absolute viscosity/density}\,.$$

Both kinematic and absolute viscosities are used in petroleum operations. Saybolt Universal Second (SSU) and Saybolt Furol Second (SSF) are units of kinematic viscosity given by readings on a Saybolt viscometer and are used primarily in marine petroleum operations. The Saybolt Universal viscometer is used for liquids having viscosities below 1000 cSt. The Saybolt Furol viscometer is used for more viscous fuel oils (*furol* is an acronym for fuel and road oils). The Furol viscosity readings are roughly 1/10 the Universal readings. For liquids whose viscosity exceeds 50 cSt at 37.8 °C (100 °F), one SSU is approximately 0.2158 cSt or 0.2158 mm^2/s. For very viscous liquids (viscosity exceeding 500 cSt) at 50 °C (122 °F), one SSF is approximately 2.120 cSt or 2.120 mm^2/s. The Saybolt seconds are considered obsolete, but they have been used traditionally in the petroleum industry and are common in technical articles. Figure 42.18 provides a means for converting between Saybolt Seconds and centistokes with sufficient accuracy for salvage work.

Specific Gravity

The specific gravity of seawater is 1.025 and must be taken into account in some oil–water calculations. Spe-

Table 42.13 Liquid densities

	Density [lb/ft]	Density [lb/gal]	Volume [ft³/ton]	Volume [gal/ton]	Specific gravity
Alcohol, ethyl (100%)	49	6.6	45.7	342	0.789
Alcohol, methyl (100%)	50	6.7	44.8	335	0.795
Acid, muriatic (40%)	75	10.0	29.9	223	1.20
Acid, nitric (91%)	94	12.6	23.8	178	1.50
Acid, sulphuric (67%)	112	15.0	20.0	150	1.30
Acid, hydrochloric (37%)	76	10.0	29.9	224	1.20
Battery electrolyte					
fully charged	81	10.8	27.6	207	1.30
discharged	69	9.2	32.6	244	1.11
Beer	63	8.4	35.5	265	1.01
Ammonia @ 32 °F	39	5.2	57.6	431	0.62
Chlorolorm	95	12.7	23.6	176	1.52
Diesel fuel (DFM, Nato F-76)	52	7.0	42.7	320	0.83
Ether	46	6.2	48.7	364	0.74
Ethylene Glycol (anti-freeze)	70	9.4	31.9	239	1.12
Fuct oil, No. 6	60	8.1	37.1	279	0.96
Fuel oil, No. 5	58	7.8	38.4	287	0.93
Fuel oil, No. 2	55	7.3	40.9	306	0.88
Fuel oil, No. 1	51	6.8	44.3	332	0.82
Gasoline	44	5.9	50.6	379	0.70
Jet fuel (JP-5)	51	6.9	43.5	326	0.82
Kerosene	50	6.7	44.9	336	0.81
Milk	64	8.6	34.8	260	1.025
Linseed oil	59	7.8	38.3	286	0.95
Lye, soda (66%)	106	14.2	21.1	158	1.70
Oil, vegetable	58	7.8	38.6	289	0.91 to 0.94
Oil, lubricating	56	7.5	39.9	293	0.88 to 0.94
Olive oil	57	7.6	39.2	293	0.91
Petroleum, crude	44	5.8	51.3	383	0.70
Sugar-in-water solution					
20% @ 68 °F	67	9.0	33.2	248	1.07
40% @ 68 °F	73	9.8	30.5	228	1.12
50% @ 68 °F	80	10.7	27.9	209	1.28
Turpentine	54	7.2	41.5	310	0.86 to 0.87
Vinegar	67	9.0	33.2	249	1.07
Water, pure, @ 39 °F	62.426	8.3	35.9	269	1.000
Water, seawater, standard	63.987	8.6	35.0	262	1.025
Water, ice	56	7.5	40.0	299	−88 to 0.92

Note: Liquids consisting of a mixture of compounds, such as petroleum products and vegetable derivatives, may vary in density from sample to sample. The densities given in this table are average or typical values. Liquid densities, especially those of petroleum products, can also vary significantly with temperature.

cific gravity is also expressed as American Petroleum Institute (API) gravities (degrees API). API gravities are based on the arbitrary assignment of a gravity of 10° to fresh water. Liquids with higher gravities are lighter than water; those with lower API gravities are heavier than water. Conversion between specific gravity and degrees API is described below. Table 42.14 gives specific and API gravities and weight-volume data for water and representative petroleum products.

Conversion from specific gravity to Degrees API is by

$$\text{Degrees API} = \left(\frac{141.5}{\text{specific gravity at } 60\,^{\circ}\text{F}}\right) - 131.5\,. \tag{42.41}$$

Table 42.14 Specific gravity, API gravity and weight of representative petroleum products

Product	Specific gravity	API gravity	lb/gal	bbl/ton
Fresh water	1.000	10.00	8.33	6.30
Asphalt	1.040	4.61	8.67	6.06
Benzol	0.882	28.87	7.36	7.14
Crude oil, heavy	0.983	12.40	8.19	6.41
Crude oil, light	0.797	46.03	6.64	7.90
Diesel oil, distillates	0.845	36.05	7.05	7.46
Gasoline, aviation	0.708	68.40	5.90	8.90
Gasoline, motor	0.741	59.41	6.17	8.50
Gasoline, natural	0.741	59.41	6.17	8.50
Grease	1.000	10.00	8.33	6.30
Kerosene, jet fuel	0.813	42.57	6.76	7.75
Lubricating oils	0.900	25.72	7.50	7.00
Mineral spirits	0.731	62.11	6.10	8.62
Paraffin wax	0.801	45.26	6.67	7.87
Petrolatum	0.801	45.26	6.67	7.87
Petroleum coke	1.145	−7.97	9.55	5.50

Table 42.15 Flammability classification of petroleum products

Class	Flash point	Reid vapor pressure
A (flammable)	80 °F or below	14 psi or above
B (flammable)	80 °F or below	Between 8.5 and 14 psi
C (flammable)	80 °F or below	Below 8.5 psi
D (combustible)	80–150 °F	NA
E (combustible)	Above 150 °F	NA

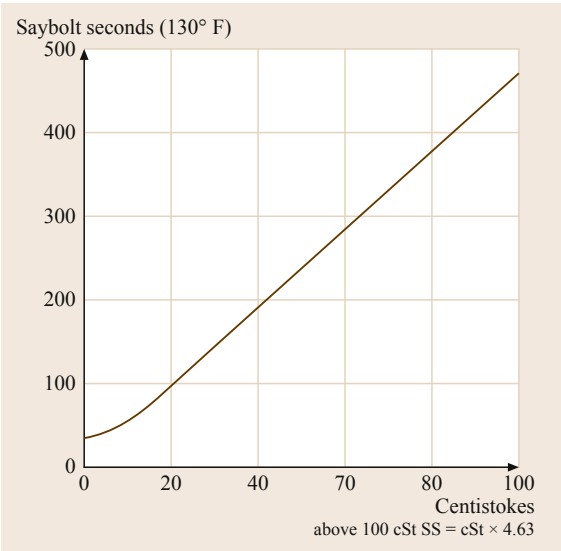


Fig. 42.18 Approximate conversion between saybolt seconds and centistokes

Flammability

A relative measure of how easily a product burns. The U.S. Coast Guard has divided POL products into two broad categories: flammable – products with flash points below 80 °F, and combustible – products with flash points above 80 °F. Products are subdivided into five grades: A, B, and C are flammable, whereas D and E are combustible. Grade A, B, and C products are defined by their Reid vapor pressure – the pressure generated by vapor when the product is placed in a sealed container and heated to 100 °F. Grade D and E products are defined by their flash points. Both Grade D and E products are relatively safe to handle. Table 42.15 gives the specifics of flammability classification.

Explosive Range

The range of hydrocarbon gas concentrations in air that is capable of being ignited and of burning. The range is bounded by upper and lower explosive limits. Concentrations of vapor lower than the lower explosive limit (LEL) are too low in hydrocarbon vapor (lean) to burn and concentrations above the upper explosive limit (UEL) are too high in hydrocarbon vapor (rich) to burn. The explosive range is usually between 1% and 6% of hydrocarbon vapor in air but varies with the petroleum product. Small quantities of volatile POL products can produce vapor concentrations within the explosive range. The explosive range is affected by the oxygen content of the atmosphere; explosive atmospheres cannot form if the oxygen content is 11% or less by volume. Figure 42.19 illustrates the explosive range as a function of oxygen content in a hydrocarbon vapor atmosphere.

Toxicity

All POL products are toxic to humans to some extent. Toxic effects range from contact dermatitis to almost immediate unconsciousness and death. Toxicity is quantified by allowable exposure described by

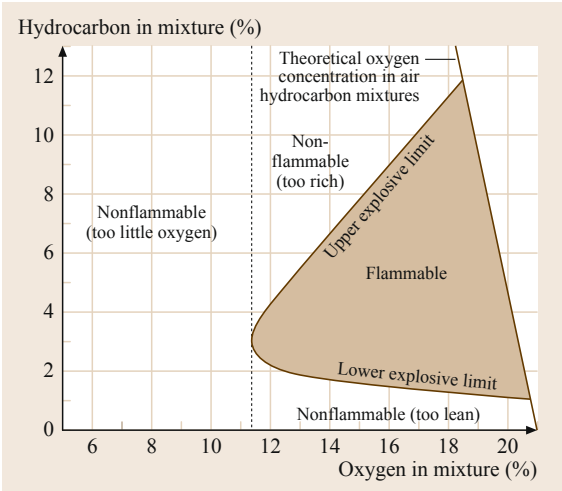


Fig. 42.19 Explosive range

concentration in parts per million and time. National Institute of Occupational Health Guide to Chemical Hazards and NAVOSH instructions contain information on toxicity of POL products. The characteristics of the products to be handled should be known before undertaking emergency transfer operations.

Oily-Water Hydrostatics

An understanding of how oil in ruptured tanks behaves is basic to correct handling of oil in casualties. The weight of the petroleum product, as indicated by specific gravity, determines the mechanism by which spills occur. POL with a specific gravity less than 1.0 is lighter than pure water and floats on the surface of water. If a tank containing oil is holed, either the oil spills from the tank or water enters the tank until the hydrostatic pressures of the water outside the hole and of the oil inside balance. When the balance is achieved there is no flow.

The effectiveness of water bottoms is limited for water-soluble liquids or liquids with a specific gravity very near one. Water bottoms cannot be created at all under liquids with specific gravities greater than 1. Many bulk chemicals fall into this category, as well as some crude oils and bunker fuels.

$$P_w = P_o \,,$$

where P_w is the hydrostatic pressure of the water outside the hull in pounds per square inch (psi) and P_o is the hydrostatic pressure of the oil inside the hull in psi.

The hydrostatic pressure of a fluid is proportional to the specific gravity of the fluid and its depth so that at hydrostatic equilibrium

$$(SG_w)(D_w) = (SG_o)(D_o) \,, \tag{42.42}$$

where SG is the specific gravity of the water in which the ship is floating (SG_w) or of the petroleum product in the tank (SG_o) and D is the depth of the water (D_w), or of the petroleum product (D_o).

For seawater

$$D_o = (1.025)\left(\frac{D_w}{SG_o}\right) \,. \tag{42.43}$$

Hot Taps

Hot taps cut through the oil container with an extendable rotary cutter enclosed in a watertight and oil tight cylinder and operated through an open gate valve. When the cut is completed, the cutter is withdrawn through the valve, the valve closed, and the cutter replaced with a hose, pipe, or other means of carrying the oil away.

High-Viscosity Oil

For vessels carrying very viscous oils, the addition of annular water injection to the pump suction and discharge may be necessary to pump the oil to the surface. For some oils, such as Bunker C oil (#6 or IFO380) in cold water, the tank may also need to be heated to get the oil to flow. Even if the oil will flow to the pump, it may require an annular water injection system, as well as heat, to bring pumping rates up to an acceptable flow rate without over pressurizing the offloading hose.

42.C Appendix: De-Beaching

42.C.1 Refloating Stranded Ships

Stranded ships are refloated by the following actions:

- Moving the ship to water deep enough to float it at the draft corresponding to its weight (displacement).
- Deepening the water around the casualty.
- Reducing the required draft at the grounded portion by removing weight lifting or altering trim.

In practice, a combination of methods is normally used. In most cases, the stranded ship is lightened until the required freeing force is less than

the available tractive forces, then pulled into deeper water.

42.C.2 Moving Stranded Ships

The force required to move a casualty over its strand is the sum of the forces required to:

- Overcome friction between the ship and seafloor.
- Move loose seafloor material that may be pushed ahead of the ship.
- Break or crush obstructions or impalements, such as rock outcroppings, coral heads, etc.
- Overcome suction on soft bottoms.

Friction is a function of ground reaction as modified by other factors, such as the coefficient of friction of the bottom, the area of the hull in contact with the bottom, and the casualty's list and trim. Freeing force is reduced by decreasing the effects of these factors, as well as by decreasing ground reaction.

42.C.3 Reducing Ground Reaction

Ground reaction is reduced chiefly by removing weight, although external lifting may be employed. In situations where the casualty is aground over only a short length at one end, trimming moments induced by weight shifts or additions can reduce ground reaction.

- *Weight Removal*: Removal of floodwater and other weights is one of the principal means of reducing ground reaction. The ship must be controlled by ground tackle, tugs, or her own propulsion to prevent being driven further ashore as she refloats. Weight should be removed so that the ship will lift off her strand with her keel approximately parallel to the beach slope, or so that the bow remains in contact after the stern lifts. The affects of weight removal on casualty strength and stability should be considered.
- *Lifting*: Lifting is a means of spreading the casualty's weight over additional vessels or buoyancy devices in order to reduce its draft enough to float free, or to reduce ground reaction. Pontoons of any description may be placed alongside the stranded ship and rigged to the hull or to slings under the hull to provide lift and reduce ground reaction. Where space and water depths permit, cranes and sheer legs may be brought alongside and rigged to lift the stranded ship to reduce the ground reaction. When sheer legs and cranes are used, the refloating should be slow and controlled to prevent sudden high loading of lifting gear or side loading of cranes as the ship comes afloat and the center of buoyancy moves forward.

42.C.4 Reducing Friction

Methods used to reduce friction depend on the type of seafloor. Cohesive and cohesionless soils develop friction forces in different ways; actions that effectively reduce friction on one type of soil may have little or no effect on another type. The following general guidelines apply:

- *Hard seafloors* (*rock, coral, hardpan, very stiff clay*): Friction is essentially a function of ground reaction and is independent of contact area. Hull contact is generally not continuous. Reducing the ground reaction is the only way to reduce friction – reducing contact area and attempts to induce vibration are generally ineffective. Friction-reducing water films cannot be maintained between the small areas of hull-to-seafloor contact because of the high unit pressure.
- *Cohesionless soils* (*sand, gravel, crushed coral, nonplastic silts*): Friction is essentially a function of ground reaction and is independent of contact area. Reducing the ground reaction reduces friction. Sands and gravels have good resistance to lateral flow under distributed vertical loads, so increasing pressure alone is not effective. Heavy vertical loads on wedge shapes, such as a narrow forefoot, may cause lateral flow in loose soils. Low-frequency vibrations can cause sand to assume fluid properties. High-volume water flow through a sand bed can create quicksand, with fluid behavior. It is possible to maintain short-duration water films between the soil and the hull with high-volume water flow.
- *Cohesive soils* (*clay, plastic silts*): Friction is a function of soil shear strength and contact area. Soil strength and adhesion depend on its ability to inhibit water flow. These soils have low resistance to lateral flow under vertical pressure. Decreasing contact area and increasing unit pressure are effective means of reducing friction. Vibration has little effect, but active disturbance and other measures to open water flow paths to the hull effectively decrease friction. Disturbing adjacent soil masses is also helpful; it reduces the distance water must flow through the essentially impermeable, undisturbed soil.

If some motion, however small, between the casualty and the bottom can be induced, the freeing force will be based on the smaller dynamic coefficient of friction. Initial motion may also allow a film of water to

develop between the hull and bottom, further reducing friction. Specific actions to reduce friction include the following:

- *Wrenching*: Wrenching uses the length of the casualty as a lever arm to multiply pulling force and rotate the casualty about the center of ground reaction. The induced motion reduces friction to dynamic levels and opens water flow paths in cohesive soils.
- *Sallying*: Ships with significant deadrise can sometimes be sallied by shifting weights back and forth athwartships. Contact area is decreased and unit pressure increased momentarily on each cycle as the ship rolls across its keel.
- *Vibration*: Inducing vibration in the hull causes very small movements between the casualty and the bottom, and may reduce the coefficient of friction. Vibration transmitted into the seafloor can fluidize cohesionless soils and may improve water flow through cohesive soils very slightly. Vibration can be induced by operating the casualty's machinery, moving heavy vehicles within the ship or on deck, veering anchor chain, etc. Casualty machinery should be operated only if all support systems (cooling water, lubrication, etc.) are operational.
- *Reducing Contact Area*: Measures taken to reduce ground reaction often reduce contact area by altering trim or removing ground from under the casualty. Weight shifted or added forward may reduce contact area and increase pressure under a narrow forefoot. Generating a strong trimming moment uses the ship's length as a lever arm to break suction; the increased pressure under the forefoot can reduce friction and cause lateral flow in soft soils. If the casualty is to plow through the bottom, it will take less force to plow a narrower furrow.
- *Disrupting Soil Continuity*: Water or air jets can be directed under the casualty to disturb the soil. On mud bottoms, hogging lines can be dropped over the bow and drawn along the hull. Disturbing cohesive soils reduces their shear strength and opens water flow paths. Similar disturbances may fluidize cohesionless soils and induce quicksand-like behavior.
- *Increasing Pressure or Ground Reaction Forward*: Experience has shown that trimming ships with a narrow forefoot or very fine lines forward stranded on sand or gravel bottoms hard by the bow often reduces resistance to sliding. The reasons for this reduction in sliding friction are not clear. The extreme pressure may cause a reduction in coefficient of friction, or be sufficient to cause lateral flow in the soil. The high bearing pressure may actually increase friction to the point that the underlying soil fails in shear before sliding is initiated between the hull and soil. If this occurs over a reduced area, the total force required may be less than that required to overcome friction between the hull and seafloor over a larger area. If the ships form is narrow enough, the ship may penetrate into the seafloor, gaining buoyancy as it settles.
- *Drainage Tubes*: Perforated pipes inserted into the soil in contact with the casualty's hull will allow water to flow between the hull and the soil.
- *Slipways*: If a casualty or portions of it can be lifted by jacks or other means, slipways can be built underneath, allowing the ship to be moved across a lower friction surface. Effective slipways are arranged so that the ground reaction is shared equally by the ways. The slipways and underlying soil or rock must be strong enough to support the ship without deforming, so the ship can slide smoothly. Jacks consisting of modified aircraft landing gear, complete with tires, arranged so that the tires could roll as the casualty was pulled across them, have been used successfully on flat-bottomed hulls.

42.C.5 Increasing Water Depth

Water depth under and around a casualty can be increased to:

- Obtain sufficient water depth to refloat the vessel.
- Reduce ground reaction by increasing buoyancy.
- Free one end of a vessel to allow it to be pivoted by other methods.

Water depth inside small coves or estuaries can be increased by closing the entrance with sheet-steel piles or cofferdams. In some navigable rivers and canals, water level can be controlled to some extent by lock gates and dams – it may be possible to raise water level by increasing flow past upstream dams. On very soft soils, it may be possible to increase the ship's weight, by flooding or other means, so that she settles more deeply into the seafloor. After she has settled, the excess weight can be removed to allow her to float free. More commonly, water depth around a casualty is increased by removing ground from under it. Ground removal is accomplished by scouring or dredging soft bottoms, or by blasting hard bottoms. These methods can also be used to cut channels to deep water through bars or behind vessels stranded high and dry.

42.C.6 Temporary Reduction

Dynamic friction is almost always less than the static friction between two objects. If freeing force can be re-

duced long enough for the pulling system to start the ship moving, it can usually be kept moving:

- Swells increase the buoyancy of the stranded ship and decrease the ground reaction as they pass. High seas or heavy swells running during a retraction decrease the pulling force required to refloat the ship. If the pulling force is enough to start the ship moving at the top of a swell when ground reaction is lowest, the coefficient of friction is lowered instantly to the dynamic level.
- Jacking reduces freeing force by changing the nature of the ground reaction, rather than reducing it. Hydraulic jacks of 60 tons capacity or greater are employed to temporarily lift the ship. By taking up part of the ship's weight on the jacks, the amount of weight bearing on the high friction interface between the ship's bottom and seabed is decreased. The jacks are rigged on long spuds that can pivot at their bases, allowing the casualty to be moved when the friction force is sufficiently reduced. Jacks are placed symmetrically about the estimated position of the center of ground reaction and are secured with a retrieving line led to the deck.

For jacking to be successful, the seafloor must be hard enough, or must be reinforced, to support the jacking forces. On rock seafloors, concrete rubble-filled beds or heavy timbers topped by steel plate are adequate foundations. On sediment seafloors, plate or timber mats are used to spread the load until the unit pressure is less than the bearing capacity of the soil. Crushed coral, stone, shell, or gravel can be laid in to increase soil bearing strength. Similarly, the hull of the ship must be protected from the jacking forces. If these forces are not spread out along the hull, they will cause local damage at the point of application and may even rupture the hull. Steel weldments or heavy steel angles welded to the hull and padded with timbers are suitable jacking pads. The load is transmitted to the ship structure by shear stress in the welds and side plating.

42.C.7 Applying Forces

Forces are applied to a stranded casualty to achieve the following effects:

- Move the casualty directly off the beach.
- Swing the casualty into a more favorable position.
- Induce wrenching motion to break suction or reduce friction.
- Hold the casualty in position.
- Control the casualty's motion during refloating/retraction.

Pulling systems are combinations of mechanical components that work together to apply a controlled, essentially horizontal force to a stranded ship. Pulling systems include, but are not limited to, the following:

- *Tugs*: Tugs can pull directly on the casualty, or power purchases for beach gear systems. Tugs are particularly useful for applying wrenching forces, and can be placed in action quickly to restrain a casualty.
- *Ground tackle*: Heaving systems located on the casualty, on salvage vessels, barges, or on shore can tension ground tackle to move or restrain the casualty.
- *Winch pulling systems*: Heavy winches mounted on barges, salvage vessels, or ashore can pull directly on the casualty.

Tugs and ground tackle are the pulling systems used most frequently in salvage. Tugs attached to the stranded ship with a towline develop pulling forces with their engines. Salvage ground tackle is a system of anchors, ground leg, and hauling gear rigged to pullers, purchases, or winches on a platform. The platform may be the stranded ship, an assisting ship, a barge, or the shore. In many salvage operations, the total pulling force is developed by a combination of systems. Pulling systems are tailored to the particular stranding to gain maximum effect and minimize interference. In rare circumstances, shallow draft tugs, jacks, heavy vehicles, or barges rigged with ground tackle can be used to push a casualty off the beach. Figure 42.20 shows typical pulling arrangements. Figure 42.21 shows two examples of innovative pulling systems tailored to particular situations.

42.C.8 Tug Bollard Pull

Bollard pull is the pulling force or tow wire tension generated by a tug. Bollard pull is essentially the propeller thrust at zero speed and is related to engine power and propeller characteristics. Propellers in shrouds (Kort nozzles) and controllable-pitch propellers produce greater thrust than fixed-pitch propellers for the same amount of horsepower. As refloating efforts generally require a tug to maintain steady bollard pull for much longer than 5 min, bollard pull should be based on the continuous duty or intermittent duty power rating.

Bollard pull is measured by a standardized trial conducted when the tug is new and after major modifications. If there is no bollard pull certificate, static bollard (BP) can be expressed as a function of brake

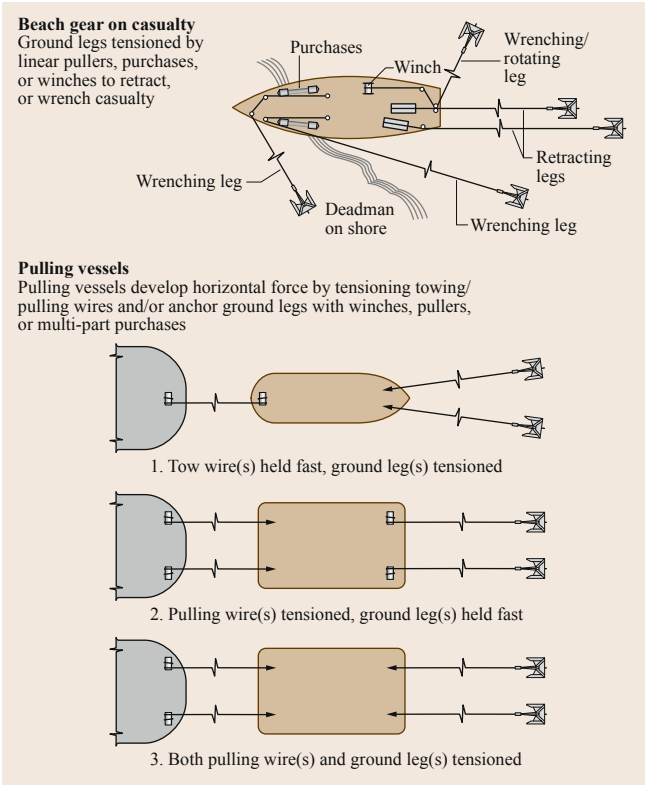


Fig. 42.20 Typical pulling arrangements

horsepower (BHP) and propeller design

$$\mathrm{BP} = 0.011 \times \mathrm{BHP} \text{ for open fixed-pitch propeller,}$$
$$\mathrm{BP} = 0.012 \times \mathrm{BHP} \text{ for open controllable-pitch propeller,}$$
$$\mathrm{BP} = 0.013 \times \mathrm{BHP} \text{ for shrouded fixed-pitch propeller,}$$
$$\mathrm{BP} = 0.016 \times \mathrm{BHP} \text{ for shrouded controllable-pitch propeller,}$$

where BP is the bollard pull in short tons and BHP is the brake horsepower of the tug's main engines.

BHP can be estimated from shaft horsepower SHP (power at the propeller)

$$\mathrm{BHP} = \mathrm{SHP} \times 1.05 \, .$$

42.C.9 High-Capacity Winches

Heavy winches can heave directly on the casualty from a firmly anchored barge or vessel, stationary platform, or location ashore. Large winches can be mounted on barges or ashore to power the beach gear systems. Mounting large winches on the casualty is not usually practical, but installed heavy winches can be used to tension the beach gear ground legs.

42.C.10 Ground Tackle

A disadvantage of linear pullers is that the ground leg cannot be released quickly. If quick release is desired, a 1 5/8-inch wire pendant with a carpenter stopper can be used to pull on the ground leg. The principal advantage of purchase systems is the light weight and portability of individual components, compared to other heaving systems. Purchase systems enable relatively low-powered line-handling equipment to exert high forces against the ground leg anchors. The standard set purchase requires a five- to eight-ton line pull, for example. By using luff-on-luff purchases, chain hoists, or grip hoists to increase mechanical advantage, manual labor can power the purchase, albeit at a very slow speed.

42.C.11 Heaving System Location

Heaving systems can be sited on the casualty, on salvage vessels or tugs, on barges, and occasionally ashore.

- *On the casualty*: The preferred placement for beach gear is on board the stranded ship. Beach gear is usually more effective when tensioned from the casualty where the salvor can observe heaving operations and their effect on the stranded ship directly, and has maximum control over the operation.

It is often advantageous to heave on the beach gear when a heavy swell or surf is running, because the lift of the passing waves momentarily reduces ground reaction. A salvage ship in harness with taut ground legs over the bow is effectively in a multipoint moor. In rough weather, the restrained ship motion can create excessive loads in the towline or ground legs.

On steeply shelving beaches, ground legs must be laid close to shore in shallow water to keep the pull roughly horizontal; placing the salvage ship between the casualty and the anchors may not permit enough scope for the ground legs to hold.

- *On a salvage vessel*: When heaving systems are located on the salvage vessel, a towline from the salvage vessel transmits the ground leg tension to the casualty. The salvage vessel pulls itself toward the ground leg anchors, pulling the casualty along with it. If deck fittings are of sufficient strength, the pulling vessel can be made up alongside the casualty.

Beach gear heaving systems are rigged routinely on salvage vessels when it is impossible or inconvenient to rig them on the casualty, or additional beach gear legs are required. Conditions that may preclude heaving from the casualty are:

1. Weather, surf, location, or other conditions prevent the transfer of necessary equipment to the casualty.
2. The casualty does not have sufficient deck space, operating winches, or attachment points for purchase systems.
3. An immediate heavy pull is warranted; heaving systems can be rigged on the salvage tug while enroute to the casualty, ready for use as soon as the ground legs are laid.

Heaving systems can be rigged on the salvage vessel in sheltered waters before moving to the salvage site.

- *On a barge*: Barges are used as heaving system platforms in the same manner as salvage vessels. Typical situations where heaving systems are rigged on barges include the following conditions:
  1. Operations require more legs of the beach gear than can be operated from the casualty and salvage ships.
  2. Neither the casualty nor the assisting vessels have adequate deck space or fittings to operate beach gear.
  3. It is undesirable to foul the deck of either the casualty or the assisting vessels.
  4. A beach gear leg leading across shallow water cannot be tensioned from the casualty.

If the towline and heaving system are connected directly to the barge, the ground leg tension is transmitted through the barge structure. Ordinary flat-topped barges are not built for this kind of loading and may be severely damaged if not properly reinforced. Padeyes should be installed. Alternatively, a length of plate strong enough to carry the pulling loads can be tack welded to the deck. The winch and padeyes for the holding stopper and block are then welded to the plate.

- *Ashore*: Heaving systems and ground legs located ashore can be used to wrench or turn a casualty. Winches or heavy vehicles located ashore can power purchases on the casualty's deck. Winches, standing blocks, and stoppers can be secured to concrete footings, deadmen, or natural rock outcroppings. Advantages of locating heaving systems ashore normally include accessibility and large working area. The most obvious disadvantage is that the heaving system cannot follow the casualty as it is retracted. Lines led to vehicles must be attached so that they can be quickly released to avoid pulling the vehicle into the water; winches must be able to spool their wire off as the ship retracts.

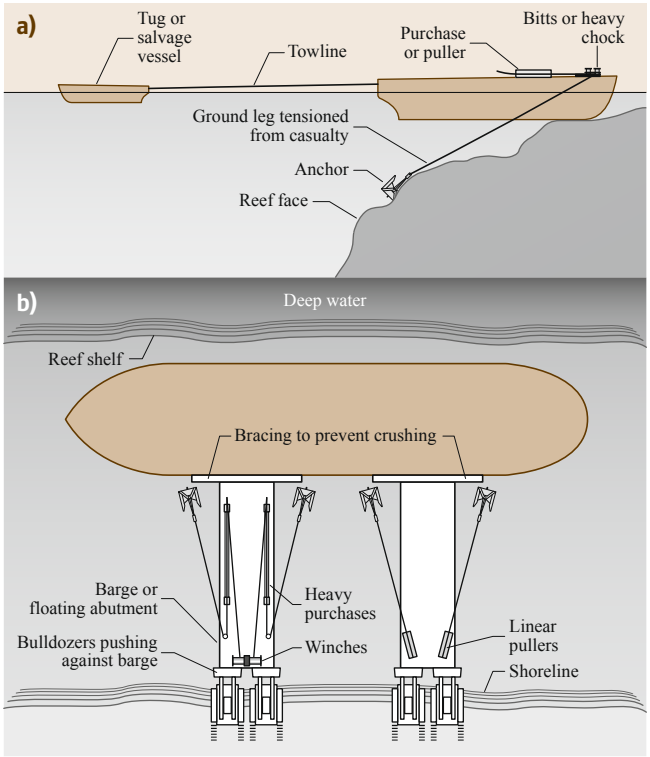


Fig. 42.21a,b Pulling arrangements for unusual situation circumstances: **(a)** water depth astern of casualty too deep for the beach gear anchores to be effective, **(b)** shallows in shore and steep shelf near shore favor pushing

42.C.12 Underwater Excavation

Underwater and semisubmerged excavation and earthmoving are conducted in support of several types of salvage operations:

- Removal of seafloor material from under stranded ships to reduce friction and/or ground reaction.
- Removal of accumulated sediments from sunken ships to reduce weight.
- Dredging channels for the retraction of stranded ships.
- Digging trenches into which stranded ships can slide or be pulled.
- Tunneling under sunken ships to pass messenger wires.
- Disturbing cohesive soils under and around sunken or stranded ships to reduce suction effects.

- Cutting new channels or widening/deepening existing channels in harbor clearance operations.
- Digging holes to topple wrecks into, or drawing seafloor material from under a wreck, during wreck burial operations.
- Removing sediments as they build up around a stranded casualty to prevent severe hull stresses and keep machinery sea suctions clear.
- Removing accumulated sediments to gain access to objects or portions of a casualty.

There are five general methods of excavating and/or moving seafloor material available to the salvor:

- Scouring
- Air lifting
- Jetting
- Dredging
- Blasting.

Table 42.16 compares the relative advantages of various excavation methods for different circumstances.

42.C.13 Scouring

Water currents can scour loose soils from around a ship or out of a channel. Scouring currents can be produced by the propeller wash of tugs or workboats, a stranded ship's propeller wash, by pumps, or by channeling natural currents. Breakwaters or groynes built perpendicular to the beach can set up currents that will prevent sediment buildup around the ship or will scour away the ground. Scouring is most effective on loose, fine-grained sediments, such as sand or silt. Scouring is usually less efficient than dredging. The method of scouring chosen depends upon the assets available, the conditions at the site, and the amount of ground to be moved:

- Tugs or workboats, trimmed by the stern to direct the propeller wash downward, can be moored alongside a stranded casualty with their sterns directed toward the area from which ground is to be removed. The tug lies alongside the ship at an angle of 40–50° to her heading, then builds up to full power and gradually works her way aft. Lines from the stranded ship and the tug's towline may be slacked or hove taut to change the direction of the wash. The wash from the tug's propeller scours against the stranded ship's bilge, carrying seafloor material down the side and clear of the casualty. Tugs can also work from amidships forward and scour both sides simultaneously. Twin-screw vessels with outward turning propellers produce more effective scouring streams than single-screw vessels. The swirling propeller streams are angled toward each other at the bottom of the propeller cir-

Table 42.16 Comparison of underwater excavation methods

Excavation factor	Excavation method				
	Scouring	Air lifting	Jetting	Dredging	Blasting
Type of seafloor	Loose sediments	Loose sediment, Cobbles	Loose sediments	All except rock and coral	Rock, coral, hardpan, stiff or hard clay
Water depth	Less than 30 ft below scouring vessels propeller(s)	25–75 ft	Unlimited	depends on dredge type; unlimited for depths encountered in salvage	Unlimited
Horizontal spoil transport	Short	Short	Short	Short to long, depending on dredge type	Short
Vertical spoil transport	Negligible	Up to 70 ft	Short	Depends on dredge type	Short
Current dependence	Not required but can be beneficial (or detrimental)	Not required, may be detrimental	Required for most efficient operator	Not required, may affect dredge maneuvering	Not required but beneficial so long as charges are not disturbed
Ancillary equipment[a]	Tug or powerful workboat	LP air compressor	HP pump	HP pump	Rock drill for most efficient blasting
Relative shipping space/weight	Large	Large (can be reduced if lift pipe shipped in sections)	Small to medium	Small to medium	Large (if drilling equipment shipped)

[a] In addition to work platform

cle, where they merge into a nearly straight stream directed aft at the level of the vessels keel. Twin-screw vessels of less than 500 shaft horsepower and single-screw vessels of less than 1000 horsepower are generally ineffective scouring platforms. Scouring by tugs can move moderate amounts of material from under specific areas of the ship. Tugs with controllable-pitch propellers should not be used for scouring because sand and other abrasive material stirred up may damage the pitch-control mechanism. Figure 42.22 shows scouring by a tug. Some salvage vessels have been fitted with propeller nozzles or propwash deflectors to better direct the propwash against the seafloor. If there is a possibility of the casualty working free from her strand while a scouring vessel is working alongside, the casualty should be restrained by ground tackle. This method of scouring can also be used to cut channels in sand, gravel, or mud. If an appropriate vessel can be trimmed by the stern enough that her propwash strikes the bottom, she can scour material aay from a series of areas while held in place by anchors.

- A stranded ship's propeller may be run *ahead* to wash ground away from the after section of the ship. The effect will be limited to the area immediately forward and for some distance aft of the propeller. When the ship's propeller is used for scouring, machinery cooling water intake should be shifted to high sea suctions to minimize the infusion of seafloor material into the ship's machinery, or pumps rigged to provide machinery cooling water. As an alternative, the sea suction(s) can be blocked with a plate fitted with a large-diameter pipe or noncollapsible hose fitting, with the pipe led to an area of clear water. The ship must be restrained by ground tackle to prevent her from driving herself further aground. Scouring should not be attempted when the ship has controllable-pitch propellers or other underwater installations that may be damaged by grit. This method of scouring can be used to open a channel in an obstructing bar behind the ship as she is drawn off her strand by tugs and/or beach gear. Stranded ships have worked themselves across miles of mud flats in this manner.
- Jetting pumps or other high-pressure pumps may be used to scour limited areas. Pumps may be operated from the stranded ship, but it is usually better to locate them on tugs or barges that are closer to the water and more mobile.
- Many small water jets are often more effective than one large jet. High-volume flow through perforated pipes or hoses rigged alongside the ship can scour seafloor material away and/or effectively prevent sediment buildup.

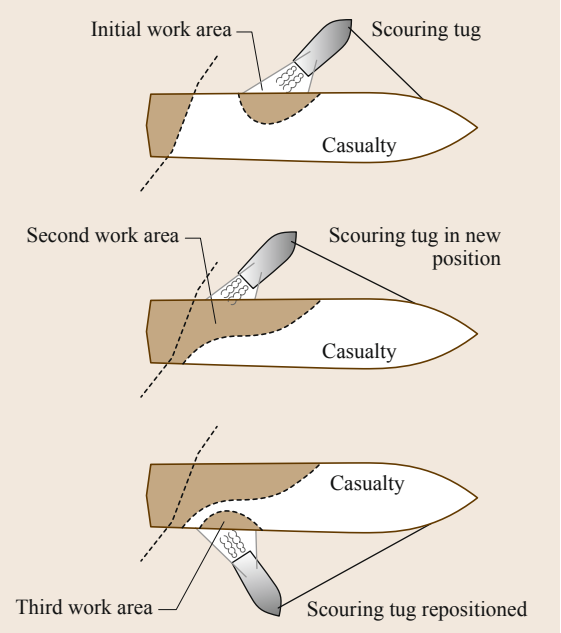


Fig. 42.22 Salvage tug scouring away soft material

42.C.14 Air Lifting

Air lifts can effectively remove loose materials at moderate depths (25–75 ft). The amount of material lifted depends on the size of the air lift, submerged depth of the pipe, air pressure and volume, and discharge head or lift.

42.C.15 Jetting

In its simplest form, jetting moves quantities of mud, silt, or sand by directing a high-velocity water stream at the material to be moved. This process is normally performed by a diver with a nozzle and a high-pressure water hose. It can be performed from the surface in shallow water with lances improvised from 10 and 12 ft firefighting applicators or lengths of pipe. A flow rate of about 100 gpm and a discharge pressure between 50 and 150 psi over bottom pressure is adequate for most jetting. Jetting is most efficient when there is a strong current to carry the disturbed material away from the work area.

42.C.16 Dredging

Dredges can remove large quantities of seafloor material from around and under a casualty and dig channels to deep water. Dredges are most effective in soft soils,

but some can dig coral, hardpan, stiff clay, and some kinds of soft rock (limestone, shale, etc.). The equipment used for dredging depends on the casualty location and attitude, seafloor type, and water depth.

A dredge's *production rate* is its rate of removing seafloor material, or *spoil*. Production rate depends on the dredge type, size, digging depth – depth to the bottom of the cut – and seafloor type.

If dredging appears to be a viable technique for a salvage operation, the nearest Army Corps of Engineers district office should be consulted for advice concerning the best type and size of dredge to use, and the probable production rate.

The two basic types of dredges are mechanical (bucket dredge) and hydraulic.

42.C.17 Mechanical Dredges

Mechanical dredges are classified as *grab*, *dipper*, and *ladder* dredges. Mechanical dredges move spoil by physically picking it up and depositing it elsewhere.

A *grab dredge* is a grab bucket operated from a derrick mounted on a barge. Either clamshell or orange peel buckets are used, depending on the consistency of the spoil. Digging action depends on the weight of the bucket, so grab dredges work best in soft soils. The production rate for a dredge with a one cubic yard bucket is 45–55 cubic yards of mud per hour in 15–20 ft of water. The production rate in clay is about half as much. Digging depth is limited only by the length of the lift wires, although the production rate decreases rapidly with increasing depth.

A *dipper dredge* is simply a power shovel operating from a barge. It is most effective in hard seafloors, such as gravel, broken rock, or shale. Bucket capacities vary from 1 to 5 cubic yards. A dredge with a 1 cubic yard bucket may move 50–250 cubic yards of mud per hour; about half as much clay per hour. Digging depth is limited by the length of the boom; 65 ft is about maximum. Dipper dredges are uncommon outside the United States.

A *ladder dredge* moves spoil with an endless chain of buckets running on an inclined ladder, like a department store escalator, that is lowered to the bottom. Bucket sizes vary from 5–55 cubic feet. The bucket cycle will average 20–30 buckets per minute in silts and muds, 18–24 in medium soils, and 9–12 in stiff clay, with an average of 85% bucket fill. Maximum digging depth is usually about 40 ft, but may be as deep as 75 ft.

42.C.18 Hydraulic Dredges

Hydraulic dredges move spoil with high-capacity centrifugal or jet pumps. *Plain suction dredges* draw seafloor material through a suction pipe into a pump in the hull. The pump discharges into hoppers within the hull or into barges alongside. Some suction dredges have water jets at the lower end of the suction pipe to break up the seafloor material.

42.C.19 Salvage Dredging

The principal dredging operations in salvage are:

- Digging retraction channels for stranded ships
- Trenching
- Ponding.

When dredges cut trenches close alongside casualties in soft or fluid soils, soil from under the ship will flow into the hole, and the ship will settle, increasing buoyancy and decreasing ground reaction. Because of the currents set up at the suction head, hydraulic dredges can draw material from under the ship. It may be possible to dig deep enough to refloat the ship. Ships can be pulled or heeled into trenches dug alongside when the soil is too firm to flow. In a similar manner, dredges can trench alongside wrecks to be buried, allowing them to settle as sediments are drawn from under them.

If a ship is high and dry, a basin can be dug by earth-moving equipment, leaving columns or ridges of seafloor material under the ship to support her as blocks do on a drydock. A channel from the sea is opened into the basin with a dredge. The inrushing water, assisted as necessary by high-pressure water jets, washes the supports from under the ship, allowing it to refloat.

Salvage dredging is a complicated operation requiring time, work, planning, and careful coordination with other work.

42.D Appendix: Refloating

42.D.1 Stability of Sunken Ships

The initial stability of a ship, sunken and resting upright on the seafloor, depends primarily upon whether the main deck is above water, partially submerged, or completely submerged.

42.D.2 Main Deck Above the Surface

If the ship is sunk with the main deck above the surface of the water, there is a waterplane and the metacentric radius (BM) and metacentric height (GM) can be calculated or estimated. In such cases, the COB (B) may lie above the center of gravity (G). As can be seen from Fig. 42.23, the hull will be stable in this condition. When buoyancy is restored, the COB moves down in the hull, crossing the position of the COG, eventually lying below it. As this happens, the metacentric radius increases because the moment of inertia of the waterplane (IT) remains the same and the displacement volume (V) decreases. Ideally, the ship will remain positively stable throughout the process. If, however, the ship was unstable in her afloat condition, she will be unstable as that condition is restored. The addition of high weight or the removal of low weight during the salvage operation can cause an unstable afloat condition. Far more common is a loss of positive stability caused by free surface in an otherwise stable ship being refloated. It is possible that the free-surface effect will be so great that the ship cannot be made positively stable under some conditions. When this occurs, precautions must be taken to prevent capsizing. Methods to prevent capsizing of sunken ships during refloating are discussed in Appendix 42.D.13.

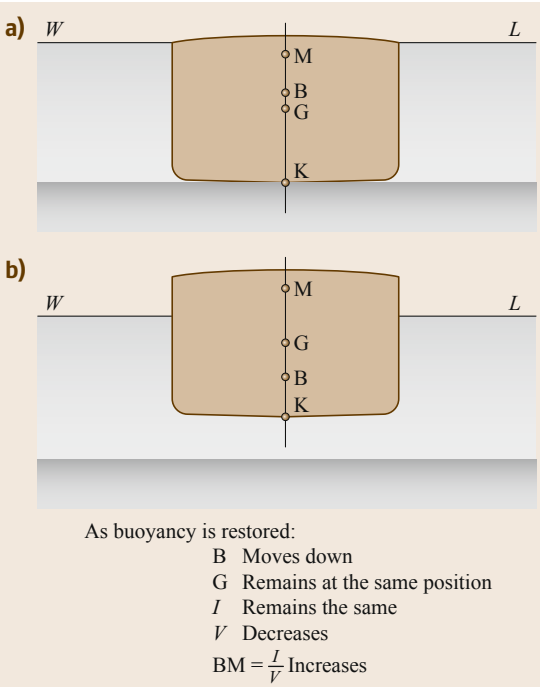


Fig. 42.23a,b Sunken ship stability, main deck above water **(a)** before refloating **(b)** after refloating

42.D.3 Main Deck Partially Above the Surface

If the ship is sunk so that the main deck is partially above the surface, the metacentric radius may be calculated based on the moment of inertia of the existing partial waterplane. Because of the relatively small value of the moment of inertia of the waterplane and the large underwater volume, the metacentric radius may be quite small. Depending on the location of the COG, the metacentric height may be positive, negative, or in rare cases, zero. As buoyancy is restored, the length of the waterplane increases causing the moment of inertia of the waterplane to increase and the underwater volume to simultaneously decrease. The overall result is that the metacentric radius increases and the ship becomes potentially more stable. As when the main deck is completely above water, the overall stability of the ship depends upon the position of the COG and the free surface in partially flooded spaces. Figure 42.24 illustrates the stability situation in a ship being raised in which the main deck was initially partially above water.

42.D.4 Main Deck Submerged

By far, the most complex and difficult stability situation occurs when a ship is to be raised from a position where the main deck is completely submerged. Under this condition, there is no metacentric radius because there is no waterplane. The positions of the metacenter and the COB are coincident, and the distance between the centers of gravity and buoyancy (BG) becomes the measure of stability. It should be emphasized that B and G must be in line vertically, both transversely and longitudinally, at equilibrium. If B is above G, as shown in Fig. 42.25, the ship is stable. If B and G are colocated, stability is neutral. If B is below G, the ship is unstable. When a stable ship is displaced from an upright position, a righting couple will be formed by the weight and buoyancy. In an unstable ship, the couple formed by weight and buoyancy acts to upset the ship.

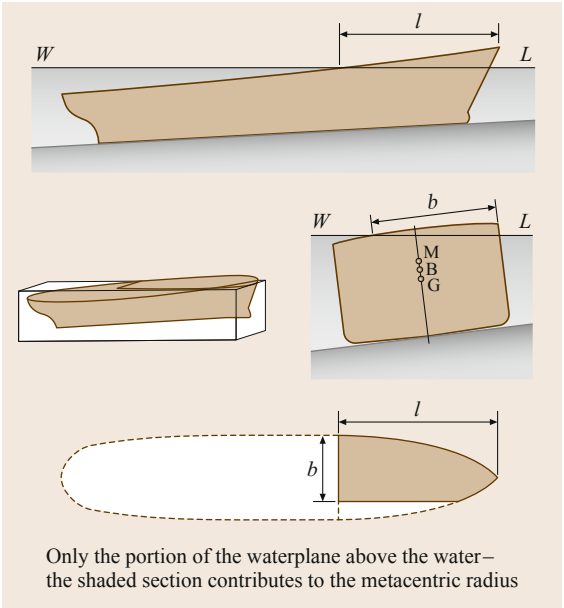


Fig. 42.24 Sunken ship stability, main deck partially above water

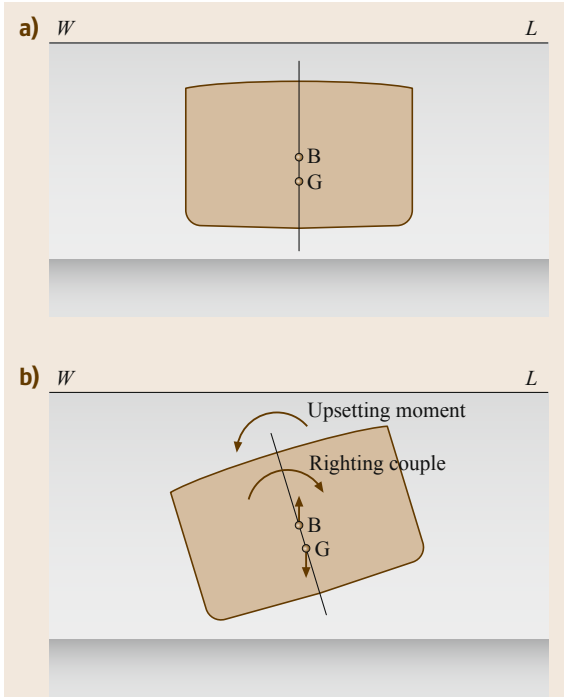


Fig. 42.25a,b Sunken ship stability, completely sunken ship: **(a)** Ship upright, clear of bottom no waterplane, **(b)** if G is above B, ship is unstable

Stability of a completely submerged ship is generally not a concern if the ship is resting and restrained from capsizing by the seafloor; however, it will be of concern if the ship has floated free, but is still completely submerged.

As the ship begins to surface and a waterplane is developed, a metacentric radius is formed and normal stability considerations apply. The metacentric radius is quite small at first; the metacentric height probably will be negative and the ship unstable, especially when there is an appreciable free surface. As additional waterplane is gained and the underwater volume decreased, the ship becomes more stable. The period between the time the ship being raised begins to develop a waterplane and when it becomes positively stable is critical. During this period, the ship must be stabilized to prevent capsizing. Figure 42.26 illustrates the stability of a ship being raised from a completely submerged condition.

42.D.5 Longitudinal Stability

Longitudinal stability is the measure of a ship's ability to return to its original position after being disturbed by a force that rotates it around a transverse axis. Longitudinal stability is important to refloating operations because changes in the longitudinal stability of a stranded or sunken ship will not be apparent since the ship does not respond in the same manner as a ship afloat. The changes must be calculated to ensure that salvors have an accurate assessment of the actual longitudinal stability situation. Because of the contribution of length to the longitudinal moment of inertia of the waterplane, ships with any significant length of waterplane are inherently longitudinally stable. In ships with little or no internal transverse subdivision, free surface may present a major problem while the ship is being raised.

The greatest danger to longitudinal stability from free surface lies not in the reduction of longitudinal metacentric height but in the trimming moment caused by the mass of water rushing to the low end of compartments as the ship trims from redistribution of weight. If transverse subdivision is nonexistent or inadequate, the trimming moment may cause the ship to trim sufficiently to allow downflooding with subsequent loss of buoyancy and plunging. In a submerged hull, trim is affected by the longitudinal separation of B and G. Longitudinal and transverse stability are essentially the same, in this situation, because there is no waterplane. A ship with no waterplane and no longitudinal metacentric radius has only the longitudinal righting moment provided by the relative positions of the centers of gravity and buoyancy. Care must be taken to raise the ship so that trim does not develop, or to keep it in contact with the bottom.

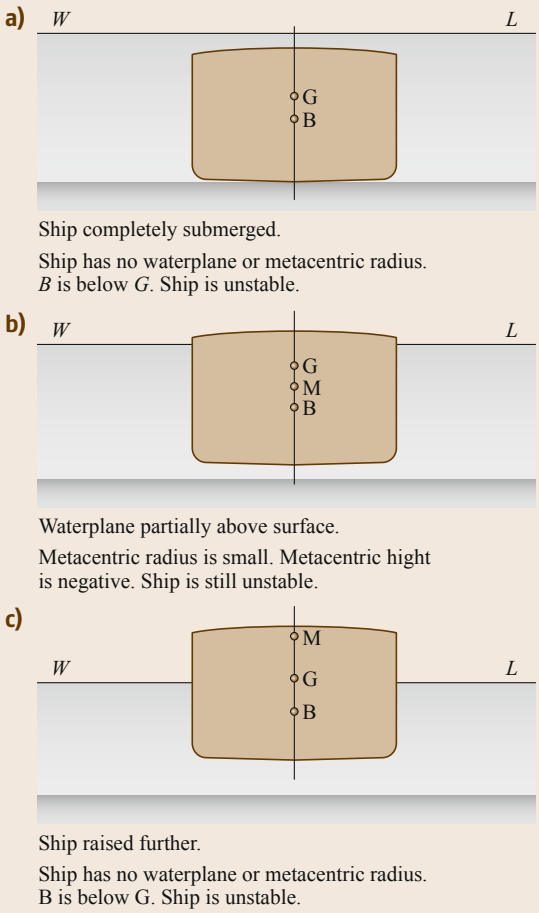


Fig. 42.26a–c Development of stability in raising a completely sunken ship

42.D.6 Longitudinal Position of Center of Gravity (LCG)

The LCG is as important to longitudinal stability as the height of the COG is to transverse stability. Its position is determined solely by the distribution of weight along the length of the ship. The longitudinal position of the COG is measured in feet from the midships section or the forward perpendicular. It is determined in a manner similar to determining the height of the COG above the keel in that the sum of the moments of the weights about either the forward perpendicular or the midships section is divided by the total weight to obtain the desired position. The following steps are necessary:

1. Classify all the weights in the ship.
2. Determine the longitudinal distance of each weight from the reference.
3. Multiply each weight by the longitudinal distance from the reference to determine the moment of the weight.
4. Total the weights and the moments of weight.
5. Divide the total of the moments of weight by the total weight to determine how far the longitudinal position of the COG lies from the reference.

42.D.7 Longitudinal Position of the Center of Buoyancy (LCB)

The LCB is measured in feet from either the forward perpendicular or the midships section. For a ship in equilibrium, the longitudinal position of the COB is in the same vertical line as the longitudinal position of the COG. At any given time, there is only one point that is the COB; the height and longitudinal position are two coordinates of that single point. Determination of the longitudinal position of the COB is lengthy and tedious, requiring calculation of the underwater volume of the ship and its distribution. The longitudinal position of the COB may be obtained from the curves of form. In salvage, the longitudinal position of the COB is important primarily in making strength calculations when buoyancy must be distributed to place the longitudinal positions of the centers of gravity and buoyancy in the same vertical line.

42.D.8 Longitudinal Center of Flotation (LCF)

The LCF is the point about which the ship trims. It is the geometric center of the waterline plane. The longitudinal position of the center of flotation is measured in feet from either the midships section or the forward perpendicular. In ships of normal form, it may lie either forward or aft of the midships section. In fine-lined ships, the LCF is usually slightly abaft the midships section. The longitudinal position of the center of flotation is required to calculate final drafts when trim changes. It can be calculated if the exact shape of the waterplane is known, or it may be obtained from the curves of form. If the position cannot be obtained, it can be assumed to be amidships.

42.D.9 Longitudinal Metacenter (ML)

The ML longitudinal metacenter is an imaginary point of importance in longitudinal stability.

Like the transverse metacenter, it is located where the force of buoyancy's lines of action running through the LCB intersects the vertical line through the COB of the untrimmed ship. While the COG and the COB are points whose heights above the keel and longitudinal position are two coordinates of the same point, the

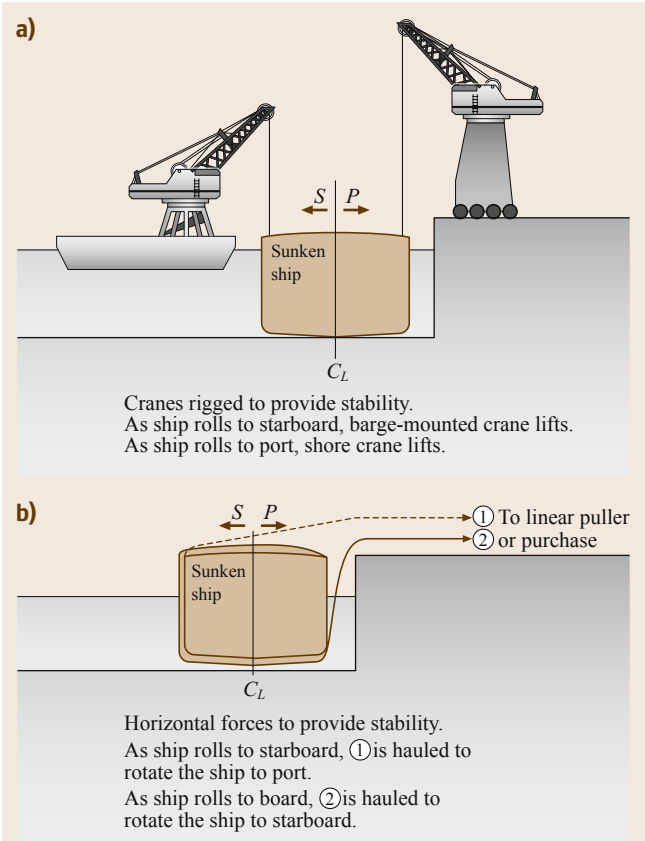


Fig. 42.27a,b Stabilizing sunken ships with cranes or tackles

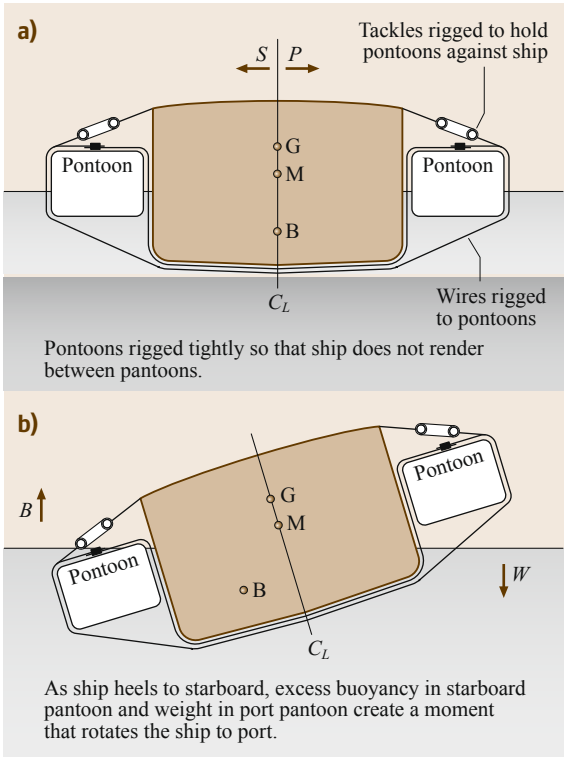


Fig. 42.28a,b Pontoons rigged to stabilize sunken ship during refloating

transverse metacenter and the longitudinal metacenter are separate points, each with its own set of coordinates.

42.D.10 Longitudinal Metacentric Radius (BML)

The BML is the distance between the COB and the ML. The BML is measured in feet. It is defined as the moment of inertia of the waterline waterplane about a transverse axis divided by the volume of displacement.

42.D.11 Height of the Longitudinal Metacenter (KML)

The KML is the distance between the metacenter and the keel, and is measured in feet. The height of the ML is the sum of the height of the COB and the BML.

42.D.12 Longitudinal Metacentric Height (GML)

The GML is the distance between the height of the COG and the longitudinal metacenter measured in feet. The GML is the difference between the longitudinal height of metacenter and the height of the COG.

42.D.13 Keeping the Ship Upright

As the ship is raised, various methods are used to prevent it from capsizing while it passes through ranges where it is inherently unstable or develops instability from free surface. The most common method is to refloat one end of the vessel while keeping the other end firmly in contact with the bottom. The contact with the ground prevents the ship from taking on a dangerous list or capsizing. In deep water, keeping one end of the ship in contact with the ground and limiting the rise of the other end prevent extreme trim.

Before the grounded end is raised, free surface in the floating end is reduced by dewatering. Spaces low in the ship, such as doublebottom tanks, may be flooded and pressed up to both eliminate free surface and to lower the COG and increase the metacentric height.

At the same time, topside weights are removed and other portable weights moved as low in the ship as possible to lower the COG. With free surface minimized from the floating part of the ship and the COG as low

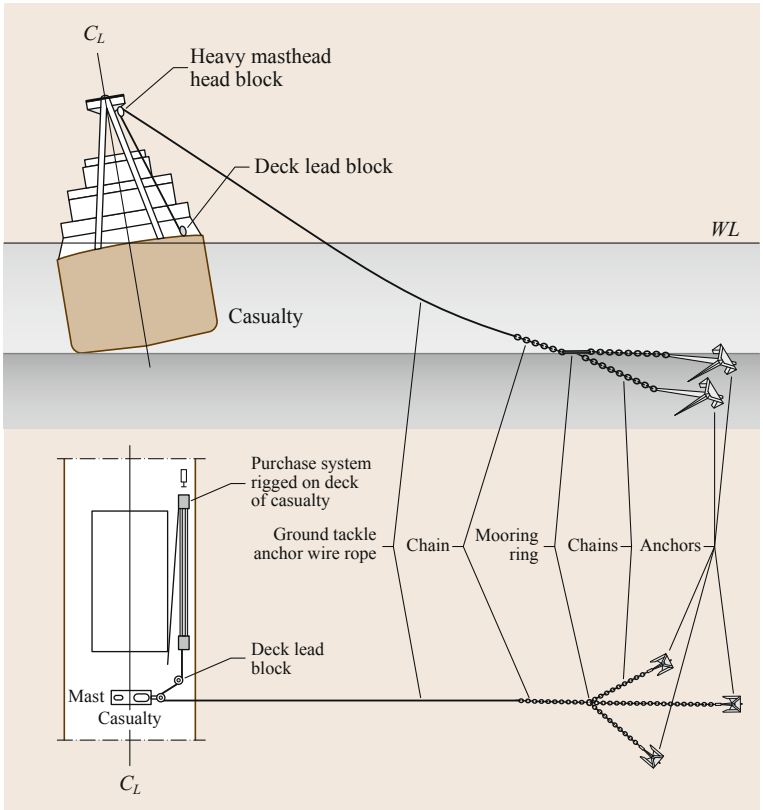


Fig. 42.29 Restraining a casulty with masthead tackle

as practicable, the grounded end may be raised. These methods are not always adequate; careful stability calculations with detailed consideration of free surface should be made before attempting to raise the grounded end.

Large free surfaces can be broken up by repairing damaged bulkheads and by building temporary bulkheads within flooded spaces. The work of building temporary bulkheads is considerably reduced if the bulkhead is built with high-pressure concrete pumped into simple forms. Both bulkhead reinforcements and temporary bulkheads should be built wider at the base than at the top to assist in lowering the COG.

A force may be applied to the ship being raised to produce a moment that counters an upsetting moment. This is done by attaching cranes or tackles rigged to apply a vertical force near the side of the ship. When the ship begins to list, a force is applied to counter the list and bring the ship back to the upright position. Figure 42.27 shows methods of accomplishing this technique.

As illustrated in Fig. 42.28, pontoons, barges, or lift craft may be rigged to provide a force that counters a heeling moment and keeps the ship upright. The pontoons must be rigged tightly to the ship so that when the ship begins to heel, she will also attempt to submerge the pontoon on the low side. The additional buoyancy of the submerging pontoon, coupled with the loss of buoyancy from the pontoon on the high side, creates a moment that rotates the ship back to an upright position. Tightly rigged pontoons and lift craft not only provide an uprighting force from their buoyancy but also act as an increased waterplane of a system composed of the ship and pontoons. The waterplane of the pontoons increases the metacentric height and overall stability of the ship–pontoon system. This advantage is gained only when the ship and pontoons are so tightly rigged that they function as a unit. If the ship is free to render in a cradle formed by the rigging, the pontoons provide only buoyancy.

Ships sunk in harbors have been kept upright by rigging purchases from their mastheads to anchoring points ashore. If this method is attempted, the ship should be held securely by mooring lines, as there will be a tendency for the ship to kick out from under the strain of the purchases on the masts. Figure 42.29 illustrates this technique.

One of the most secure methods of controlling the ship is to rig the beach gear and haul the ship into shallower water as it is lightened or lifted.

The beach gear is kept under constant heavy tension so that the ship moves into shallower water in constant contact with the bottom until it reaches a location where it may safely be dried out and refloated. Beach gear rigged ashore may be hauled with winches, linear pullers, or heavy vehicles. Tugs may be used to help move and direct the ship. Ashore, heavy tracked vehicles may also be used to haul lines for positioning the ship.

Keeping a ship in contact with the bottom, either at one end or all along her length, also assists in controlling trim and preventing the loss of longitudinal stability caused by water rushing to the low end.

42.D.14 Strength of Sunken Ships

The local, longitudinal bending, and shear strength of sunken ships are often impaired by the damage that led to their sinking. The strength of the sunken ship can have a major effect on the methods used to salvage it.

Weakened local areas can usually be reinforced adequately by simple double patches, or may sometimes be ignored where patches are placed to restore the watertight envelope. The decision whether or not to repair locally weak areas depends upon the loads that will be placed in those area during the salvage operation, the nature of the potential failure, and the consequences of a failure.

Local cracks can be dangerous, particularly if they are in high-stress areas. These cracks may grow as the operation proceeds and may eventually lead to serious failure.

Longitudinal and shear strength in sunken ships are evaluated in the same way they are for intact ships.

The most common type of hull failure is compressive failure of the deck or bottom structure. Compressive failure can be recognized by athwartships buckling – up-and-down wrinkles in the plate – in way of the failed section. Buckled structure has essentially no ability to carry compressive load but may carry almost its entire design tensile load. Ships that have failed in compression may be raised by distributing weight and buoyancy in the ship so that the failed plating is in tension throughout the salvage operation.

Shear failure may be dealt with most simply by adjusting weight and buoyancy to keep shear stresses in the failed area to a maximum of 25% of design shear stress. Attempts to double plate in the affected areas generally are unsuccessful because the deformation of the plating and internals usually prevents making the structure sufficiently straight and continuous for proper load carrying. Massive hull damage that leaves the ship in one piece but with a hinge likely or already developed requires a decision about the basic techniques to be used. The ship may be:

- Cut into pieces with each piece refloated separately
- Refloated by zero-stress/zero-shear techniques
- Wrecked in place.

When portions of the hull can be made watertight and stable, the wreck may be cut into sections and each section refloated individually. Each section may be handled by methods that are most appropriate for it. For instance, one may be refloated while another is wrecked in place. However, cutting the wreck into sections may be time-consuming and expensive.

Often the most efficient and sophisticated means of floating a badly damaged sunken ship is by using zero-shear/zero-stress methods. With these methods, the ship is loaded so that throughout the operation, shear force and bending moment are zero at the hinge or section where a hinge is likely to develop. Zero-stress/zero-shear techniques require detailed engineering analysis and planning, as well as careful attention to the hull loading throughout the operation.

Wrecking in place is always an option but is dependent upon the availability of equipment to cut the wreck effectively and to lift and handle the pieces. Wrecking in place may be the most laborintensive, time-consuming, and expensive option; on the other hand, it is usually a zero-risk-of-failure option.

42.D.15 On-the-Side Refloating

Methods of restoring enough buoyancy to refloat a ship on its side include:

- Sealing off enough major spaces to allow dewatering by compressed air, pumping out, inducing buoyancy, or a combination of these methods
- Deploying enough lifting power to lift the ship bodily on its side
- A combination of lifting and restoration of buoyancy.

The transverse and longitudinal stability of a ship to be refloated on its side must be examined thoroughly by the salvage engineer. Figure 42.30 shows the relative positions of both the transverse and longitudinal metacenters in on-the-side refloating. Ships that capsize and sink in the middle of wide rivers or navigable waterways are traffic hazards. One-way traffic on one side of the sunken ship probably can continue, but two-way traffic may be either restricted or impossible. In such cases, refloating the sunken ship on its side before dragging it to the most suitable channel edge

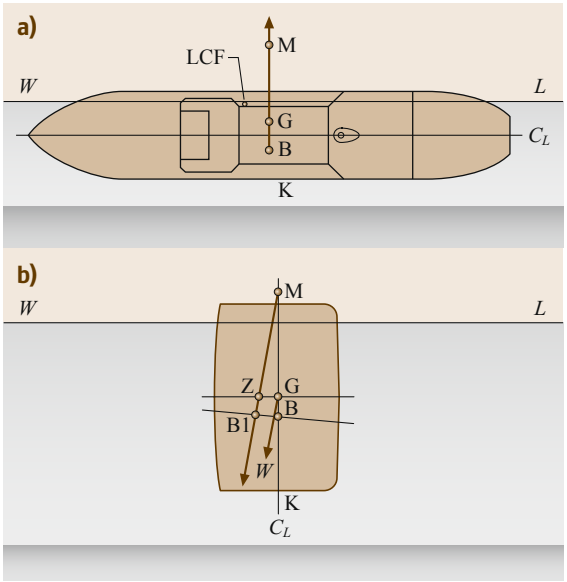


Fig. 42.30a,b Stability of a ship refloated on its side **(a)** longitudinal metacenter **(b)** transverse metacenter

for righting is convenient. An operation of this type involves:

- Setting up hauling equipment on the selected channel bank or shoreline
- Preparing the ship for refloating on its side or at an acute angle to reduce ground reaction
- Rigging attachments and lifting points for the floating cranes, lift barges, or other lifting devices that will stabilize the ship
- Removing structure, top hamper or masts, stacks, or equipment that increases the capsizing moment.

It is important to complete all preparations and system tests before rigging hauling equipment that blocks a working channel or fairway. When there is no operational necessity for salvors to obstruct channel traffic, it is better to allow normal traffic to continue for as long as possible. Safe navigational practice dictates that channel traffic is either restricted or stopped while a sunken ship is hauled to shore.

42.D.16 Upside-Down Refloating

Ships may capsize and sink in water that is deeper than the beam of the ship. The sunken ship becomes an underwater obstruction and hazard to maritime traffic. Water depth around and over the ship at high water determines the extent and seriousness of the hazard and the urgency of removal. Refloating the ship upside down is sometimes suitable in such cases.

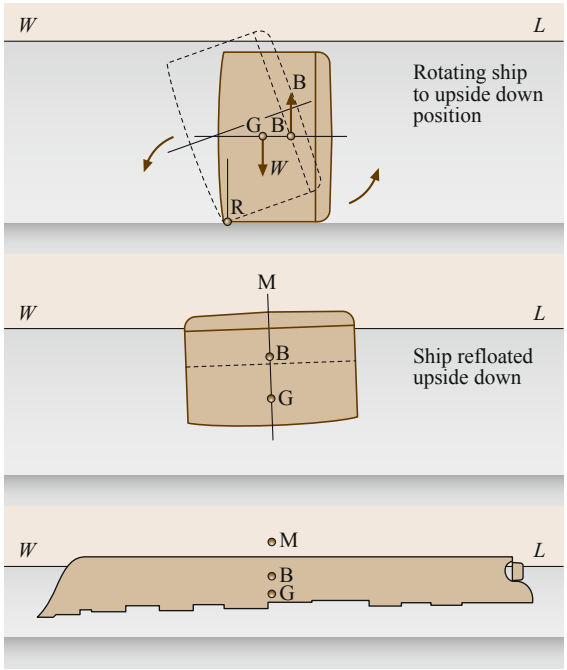


Fig. 42.31 Refloating ships upside down

Floating a ship upside down is particularly suitable when:

- The ship is capsized to more than 90°.
- The ship's bottom is relatively intact or can be made airtight.
- The top hamper, superstructure, and other items that will increase the navigational draft of the inverted ship can be removed easily.
- The channel to the ultimate destination is deep enough to allow the inverted ship to pass.
- The refloated ship is to be scuttled in deep water, scrapped in a drydock, or taken to a location that can accept the upsidedown ship.

Ships are refloated upside down by restoring their buoyancy with compressed air. Therefore, it is important that the ship's bottom plating is intact or can be made airtight with minimal work. Rotation to the completely inverted position usually is accomplished by a combination of inducing buoyancy and applying a comparatively small amount of external buoyancy or rotational force to help the vessel to capsize completely. Figure 42.31 shows a ship capsized and sunk and the sequence of restoration of buoyancy, rotation, and upside-down refloating.

Transverse and longitudinal stability are calculated the same way for an upside-down ship as for the same ship floating normally. Normally, the stability charac-

teristics of an upside-down ship differ from those of the same ship floating upright by having:

- Greater transverse stability
- Somewhat less longitudinal stability
- Greater initial stability
- Considerably greater resistance to external inclining forces.

Upside-down ships usually are very stable and handle easily when the waterline is about at the tank top level. Ships without double bottoms should have a freeboard of about three feet.

Air leaks from ships under tow or sitting for long periods while they are upside down. If the air is not replenished, the ship will sink when sufficient buoyancy is lost. Compressors are provided on board or connected to the casualty for replacing lost air. This is particularly important when towing upside-down ships in harbors or on long coastal passages.

42.D.17 Righting Capsized Ships

Righting a capsized ship is almost always an expensive and complex operation. It is usually done to remove a ship that is obstructing a berth, harbor area, or access channel, although increasingly wrecks are being salvaged for environmental or aesthetic reasons. There is no guarantee that a righted and refloated ship can be returned to service. More often than not, the combined costs of righting, refloating, repairing, and refurbishing make returning the ship to service financially impractical. Almost every righting operation involves the removal of considerable superstructure under less than optimum conditions. These removals increase repair costs considerably.

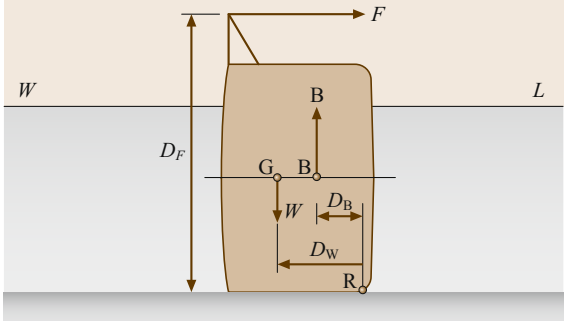
Before deciding what method, or combination of methods, is to be used to right a capsized ship, there are several important engineering and technical questions to be investigated and answered. These include:

- Calculations of righting moments to be developed to overcome the capsizing moment
- Investigations and, where appropriate, calculations to determine the physical point about which the ship will rotate, such as seafloor/soil load-bearing and shear calculations
- Investigation of local hull stresses in the ship during righting operations
- Determination of load-bearing abilities of hull areas critical to righting
- A detailed transverse and longitudinal stability analysis at selected stages of the parbuckling process (a parallel series of hull shear and bending moment analyses may be necessary)
- Weight reductions, additions of buoyancy, and other methods to reduce righting forces or lower the capsizing moment must be investigated and calculated.

42.D.18 Initial Calculations

Ships are righted by creating a moment with buoyancy, externally applied forces or both, that act around a pivot point to overcome the moment of weight, acting through the COG, that is holding the ship capsized. The location of the pivot point lies at or near the turn of the bilge. Normally, a point near the turn of the bilge is selected for initial calculations. The pivot point may be moved farther up the hull by altering or dredging away the sea bottom supporting the ship. Moving the pivot point up the hull shortens the weight's moment arm. The righting force should be applied as far as possible from the pivot point.

Initial calculations on capsized ships establish the moment of weight to be overcome to right the ship. An order of magnitude of capsizing moment is obtained by arbitrarily basing calculations on a point of rotation at the bilge in contact with the seafloor. A lever arm is measured by projecting vertical lines from the free-floating center of gravity, G, to the seafloor.



Ship assumed to rotate abaut the point R.

Righting moment (clockwise):

Any remaining buoyancy acting through the center of buoyancy.

Force F applied at 90 degrees to headframes to rotate ship.

Moment = $B \times D_B$
Moment = $F \times D_F$

Resisting moment (counterclockwise):

Weight W acting through the center of gravity.

Moment = $W \times D_W$

Fig. 42.32 Measurement points for calculation of rotation forces on righting ships

The distance between the pivot point, R, and the intersection of the vertical through G with the seafloor, W, as shown in Fig. 42.32, is the arm for the moment of weight that resists righting.

Figure 42.33 shows the change in measurement points as force is applied.

Note

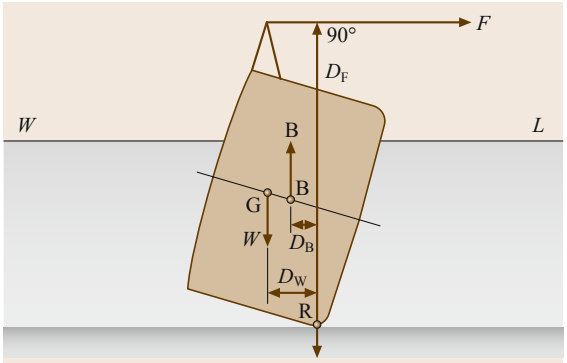
Salvors making righting calculations must be careful to keep their units straight. Ships' displacements and weights are given in long tons (2240 pounds), while lifting and pulling systems are rated in short tons (2000 pounds). Salvors in the field may find it easier to convert all units to pounds or kips (1000 pounds).

The initial moment is the greatest that is required during the righting. As the ship rotates her COG gets closer to the pivot point, reducing the moment arm and the moment. Once the COG final passes the pivot point, the weight provides a moment that assists in the righting.

42.D.19 Righting Methods

Figures 42.34–42.42 illustrate basic methods for righting capsized ships. Most of these methods involve righting the ship with the bilge firmly in contact with the seafloor, a static righting. There are circumstances when a static righting is not practical. If the ship is refloated in its capsized condition and then righted while in a floating condition, different criteria apply:

- Method 1 (Fig. 42.34): Selective sealing of major spaces in the ship allows controlled dewatering to restore buoyancy. Buoyancy, along with the addition of ballast to the high side, produces a couple to right the ship.
- Method 2 (Fig. 42.35): Inducing buoyancy into selected spaces with compressed air, and adding water ballast to provide a couple. A small external force to provide the initial rotating moment is usually necessary. This system is a variation of Method 1, but usually involves compressed air dewatering.
- Method 3 (Fig. 42.36): Applying external static forces to lever arms mounted on the hull. This method is most often combined with Methods 1 and 2.
- Method 4 (Fig. 42.37): Applying external counterweights to the high side of the hull with buoyant lifting systems attached to the low side.
- Method 5 (Fig. 42.38): Applying a direct, external, rotational force or pull to the low side of the hull with external pulling or heaving systems. This is not a particularly common method because of the difficulty of generating enough righting moment unless the ship is made buoyant first.
- Method 6 (Fig. 42.39): Extending lever arms, or headframes, from the hull and applying righting forces at the head of this system. This is one of the



Ship rotates abaut the point R.
Righting moment (clockwise):
Any remaining buoyancy acting through the center of buoyancy.
Moment = $B \times D_B$
Force F applied to headframes.
Moment = $F \times D_F$
Capsizing moment (counterclockwise):
Weight W acting through the center of gravity.
Moment = $W \times D_W$

Fig. 42.33 Alteration in moments as ship rotates

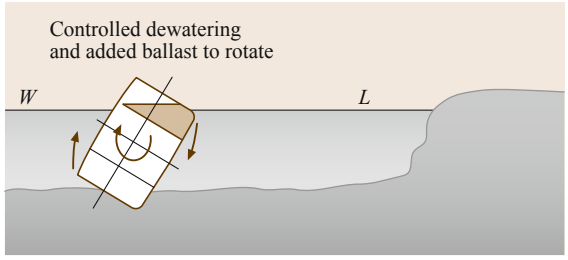


Fig. 42.34 Righting method 1

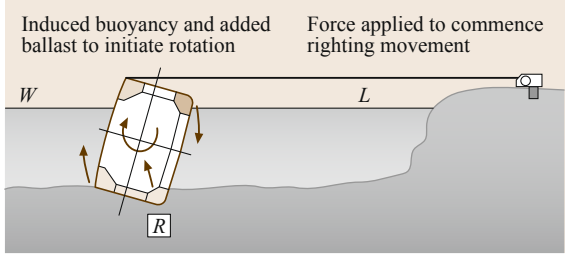


Fig. 42.35 Righting method 2

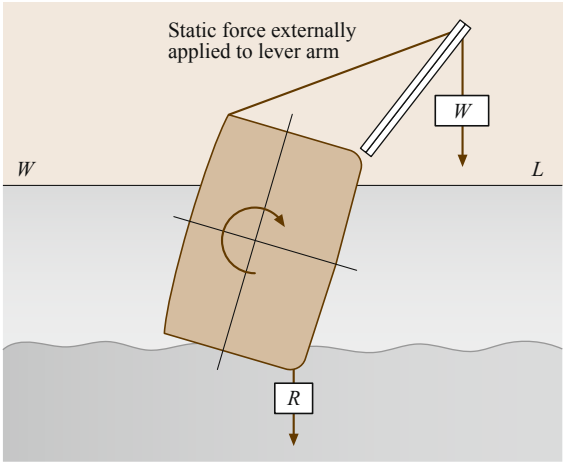


Fig. 42.36 Righting method 3

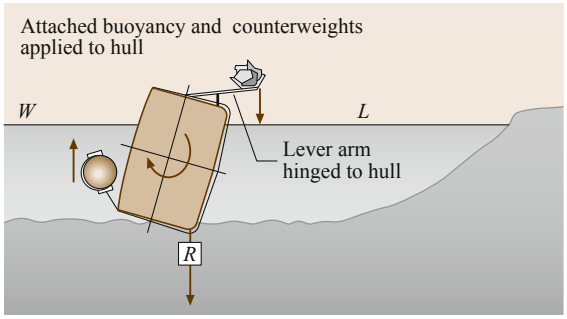


Fig. 42.37 Righting method 4

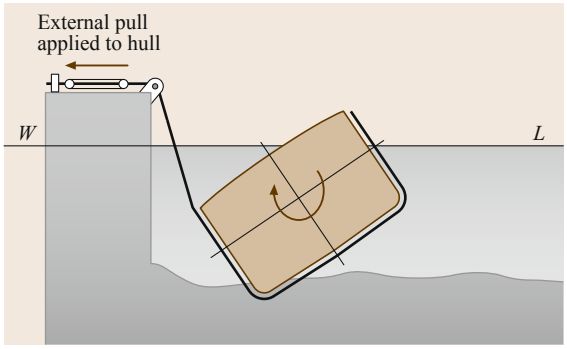


Fig. 42.38 Righting method 5

most common righting or parbuckling methods involving external haulage.

- Method 7 (Fig. 42.40): Applying a combination of direct, dynamic lifts to the low side of the hull, and an external pull to the high side of the hull. This largely mechanical system is used when sufficient hauling/lifting power is readily available and sealing off the hull for induced buoyancy is not practical.

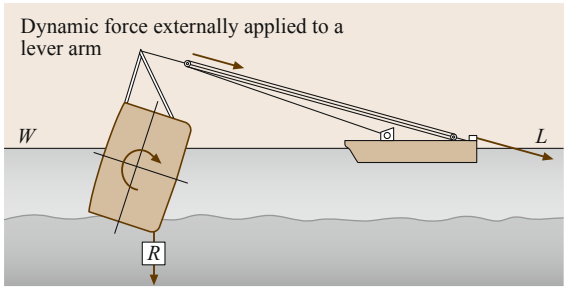


Fig. 42.39 Righting method 6

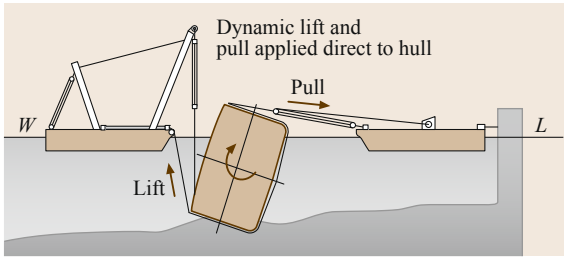


Fig. 42.40 Righting method 7

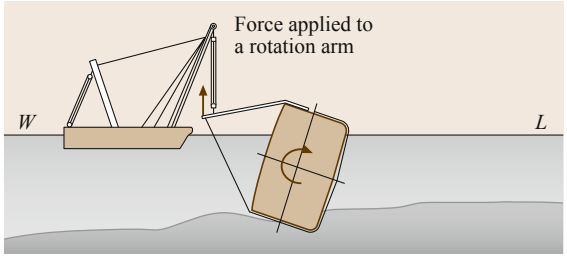


Fig. 42.41 Righting method 8

- Method 8 (Fig. 42.41): Constructing and fixing righting beams to the high side of the capsized ship, then applying a lifting force to these righting beams. This method is usually satisfactory in conjunction with large floating cranes or salvage sheer legs.
- Method 9 (Fig. 42.42): A combination of methods including:
  1. Restoring buoyancy by dewatering selected spaces
  2. Adding rotational ballast to the high side
  3. Applying a dynamic pull to the high side of the ship, along with a mechanical lift on the low side.

42.D.20 Variable Factors

The method or combination of methods selected will depend upon several variable factors that include, but are not limited to:

- Quantity of buoyancy that can be restored to the sunken ship
- Availability of heavy lifting and heavy hauling equipment in the casualty area
- Working considerations, including time and effort, required to restore buoyancy factors that may prevent the use of one or more obvious methods because the method will seriously disrupt port operations
- Environmental factors.

42.D.21 Offshore Operations

A ship that is capsized and sunk a considerable distance offshore presents several problems. The problems are rarely insurmountable, but can add greatly to the cost and time of the operations. Some of the most common offshore or exposed-location righting difficulties are:

- Hauling and lifting systems usually are operated from barges or floating cranes that are subject to weather limitations. Parbuckling, lifting, or combined parbuckling and lifting can be done only in good weather.
- When righting forces are generated by barge-mounted hauling systems and holding ground is poor, it may be necessary to build piled anchorage systems or use propellant-embedment anchors. Piled anchorages normally must be removed or cut off at the mud line when the job is finished.

Figure 42.43 shows an exposed offshore area where a parbuckling operation is undertaken with a pulling barge.

42.D.22 Righting Plans

With a proposed righting plan sketched out on paper, salvage officers and salvage engineers make a twofold analysis of the plan.

42.D.23 Calculations

Calculations are made to ascertain the following:

- Movement of the COG at successive intervals (usually 5° or 10°) of righting and applied pull required at each stage.
- Stage of inclination at which ship reaches a neutral point and capsizing moment becomes the righting moment.
- Changes, if any, at the assumed point of rotation and an accurate assessment of sinkage that may occur during rotation.

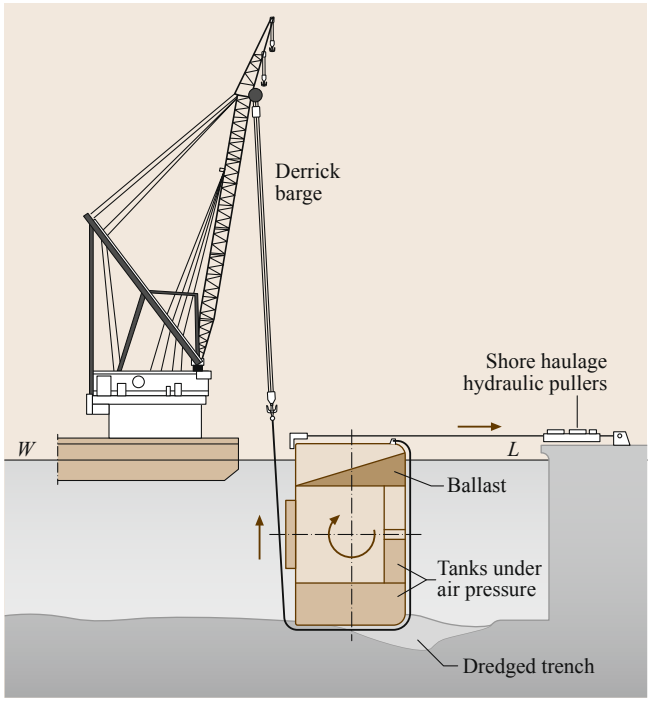


Fig. 42.42 Combination of methods used to upright a capsized ship

- Soil load-bearing characteristics and, if necessary, arrangements for detailed soil analysis.
- Existing quantities of siltation, residual stores, and cargo; in many instances, estimates will have to be made.
- Soil load-bearing characteristics of areas proposed for mounting the shore-based hauling systems (design calculations for hauling anchorages are a special task in themselves).
- Hull strength calculations to determine the amount of local stiffening or reinforcement required where head frames are mounted.
- Checks on pressure differentials created by dewatering major spaces, some of which may be partially submerged.
- That the necessary pulling power can be developed by the proposed methods, and that any proposed mechanical system is suitable for the task.
- Whether righting operations would be simplified by a trench dredged along the bilge.
- Whether an air bubble introduced temporarily into selected compartments or further dewatering of major compartments would help to right the ship.
- Whether a weight reduction program, including cutting down superstructure sections, masts, stacks, and other structural members, is necessary and useful in the overall righting and refloating plan.

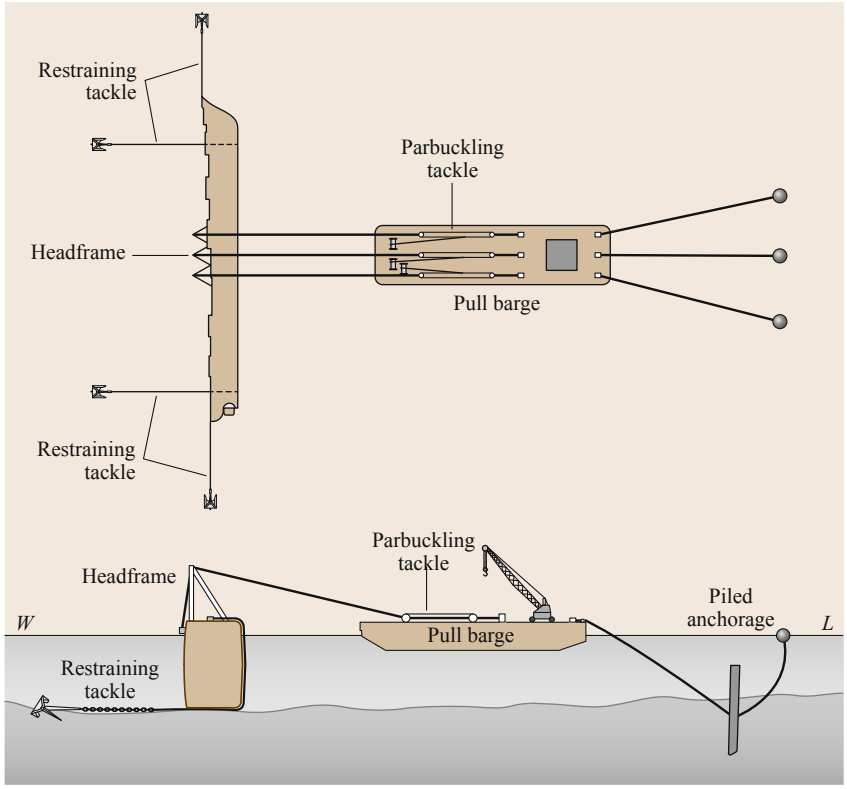


Fig. 42.43 Offshore parbuckling operation

42.D.24 Headframes

The use of headframes offers significant leverage in parbuckling operations. Numbers and individual strengths of headframes are determined by a combination of rigging and engineering factors, including:

- Pull applied to each headframe, a function of the hauling system capacity. Headframes increase in size, strength, and foundation complexity with increasing pull.
- Whether wires pull directly from headframe tops or are connected to fittings on the ship's hull and led over shoes or guides set into headframe tops.
- Whether doubling sheaves are used.
- Availability of suitable structural steel to construct headframes.

42.D.25 Types of Headframes

Headframe construction usually follows one of four principal designs.

Braced Girder Design

In the braced girder design (Fig. 42.44), the bracing leg or legs extend outward from the headframe tops toward the pull direction. The design is similar to a bipod or tripod mast and is connected rigidly to the ship's hull. Connections are made at the main or strength deck level. Braced girders re-

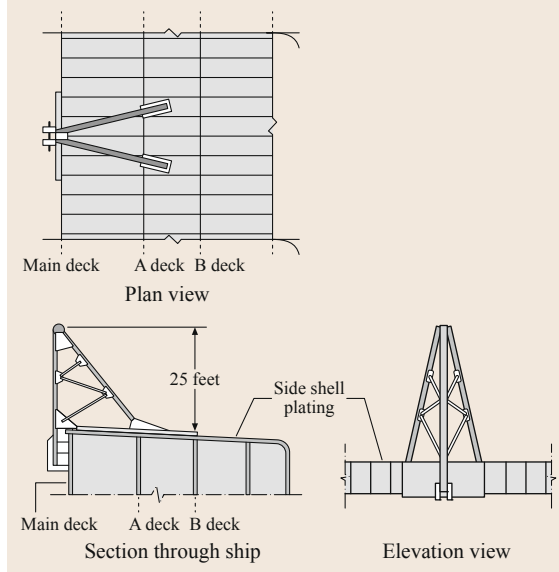


Fig. 42.44 Braced-girder headframe

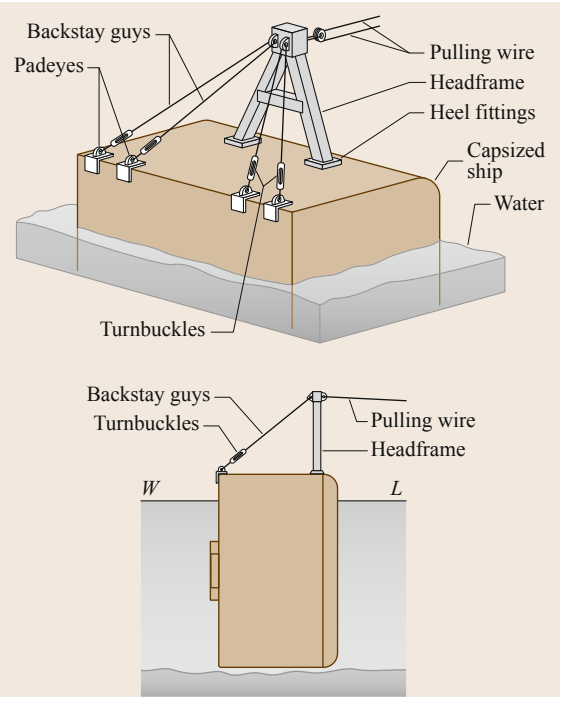


Fig. 42.45 Stayed girder headframe

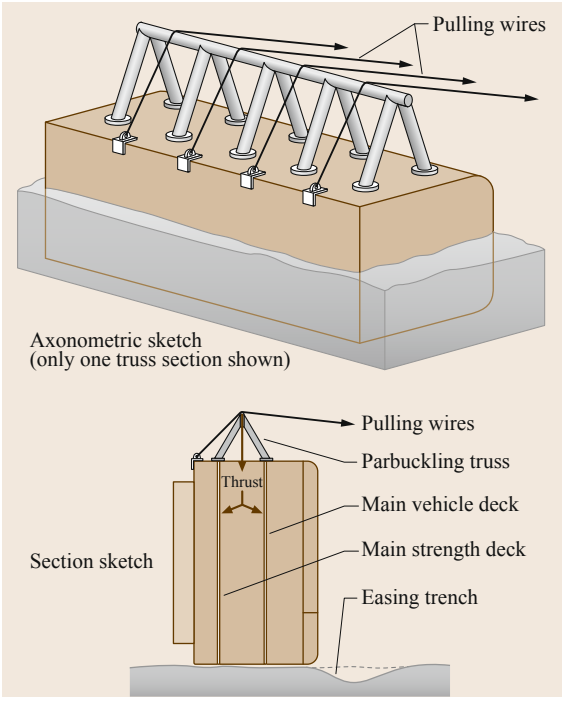


Fig. 42.47 Triangular braced semicontinuous girder headframe

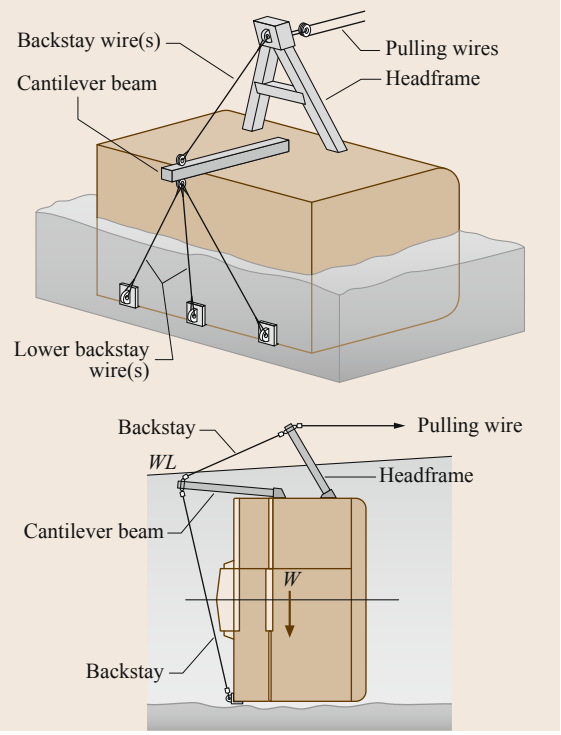


Fig. 42.46 Cantilevered strut headframe

quire more shoreside or shop fabrication than any other design.

Stayed Girder Design

In the stayed girder design (Fig. 42.45), headframes are stayed back to hull connections on the sunken ship. This design is advantageous when there is a shortage of structural steel. Stayed girder designs require numerous wire backstays and are more time-consuming and costly field rigging work than the braced girder design. It is difficult to achieve uniform stayback tension and to control quality because of the number of anchorages, padeyes, and field connections.

Cantilevered Strut Design

In the cantilevered strut design (Fig. 42.46), each headframe is a simple sheer leg braced to a horizontal strut or outrigger. The system is associated most commonly with wooden construction. In addition to difficulty in obtaining uniform strength with wooden spars or logs, this design requires a complex backstay rigging system. Backstay wire ropes are arranged from the headframe to the horizontal strut and from the horizontal strut to the ship. The principal advantage of this system is that when the hauling wire is led over the headframe and connected to the ship's hull, the righting

pull remains at right angles to the hull throughout the operation.

Triangular Braced Semicontinuous Girder Design

In the triangular braced semicontinuous girder design (Fig. 42.47), each headframe resembles triangular leg sections of offshore oil drilling rigs. The headframe structures usually are fabricated from heavy wall tubulars. A continuous length of pipe or girder has six- or eight-leg struts welded so that they are perpendicular to one another at the pipe or girder. The struts are welded to the hull with the girder elevated. It is usual with this system to lead pulling wires over the girder and connect them to special brackets, bollards, or padeyes welded to the hull.

Variations of all four systems have been successful, depending upon material availability, specific problems on the ship, and local conditions. As a rule, it is better practice to pass the hauling wires over the headframe and attach them directly to the hull than to connect them to the top of the headframe. In Fig. 42.47, triangular braced semicontinuous girder headframe, one of the advantages of this system is the ability to share leg thrust between two decks: the main vehicle deck and the main strength deck.

42.D.26 Connection of Pulling Wire

Main hauling wires are connected to either headframes or the hull by several methods. Figure 42.48 shows connection methods.

42.D.27 Connection to Headframes

Main hauling wires are shackled directly, or made up to bolted or pinned connections at the tops of individual head frames. Direct attachment of main hauling wires to headframes is one of the more common connection methods. The headframes incorporate a heavy joining lug, padeye or plate shackle arrangement to which hauling wires are shackled or bolted. This method has the advantage that all connection components can be built to uniform specifications when headframes are fabricated. Headframe connection material and component requirements are analyzed by the salvage engineer as part of the overall analysis of the headframe design.

42.D.28 Connection to the Hull

Main hauling wires are bolted or shackled to especially fabricated anchorage points welded to strong points on the hull. In such cases, hauling wires are led *over* the top of each head frame to the hull anchorage point.

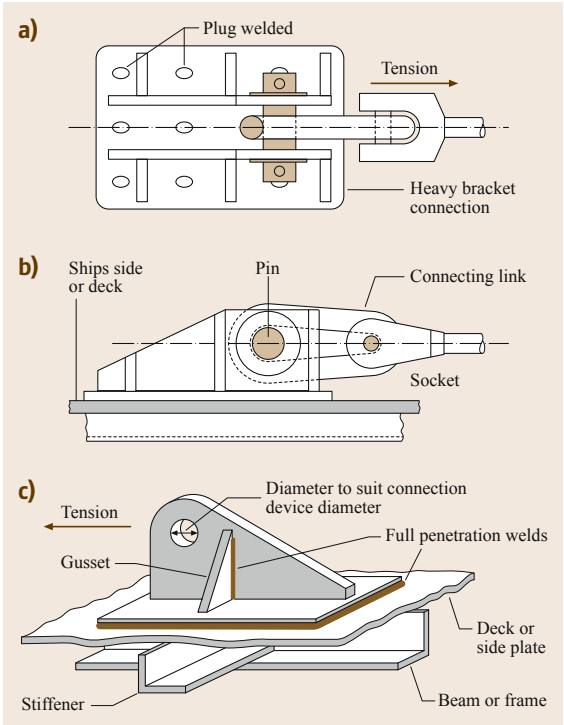


Fig. 42.48a–c Typical connection points for pulling and restraining wires: **(a)** Plan view, **(b)** elevation sketch, **(c)** heavy padeye connection

Special anchorage points welded directly to the capsized ship's hull or deck are usually made for wires of between 2 1/2 and 3 1/2-inch diameter. Typically, such anchorage points consist of two heavy steel plates welded to plate foundations attached to the capsized ship's deck. Each anchorage point is located beneath and aligned with a parbuckling headframe or deck edge bolster serving a hauling wire. This connection has the added advantage that foundation plate designs pick up several frame stations or strong points where analysis indicates load spreading is necessary.

42.D.29 Padeyes

Main hauling wires are attached to padeyes where enough local strength exists at wire connection points. Where analysis shows that not enough strength exists, it is usually not worth the time and effort to stiffen the structure for padeye connections. In such a situation, it is usually preferable to have purpose-built anchorage points fabricated ashore. The eye opening must be large enough to accept the pin of a shackle of strength equal to that of the padeye. There must be enough metal around the eye to prevent failure in bearing or tension. Padeyes should be installed so that loads are in

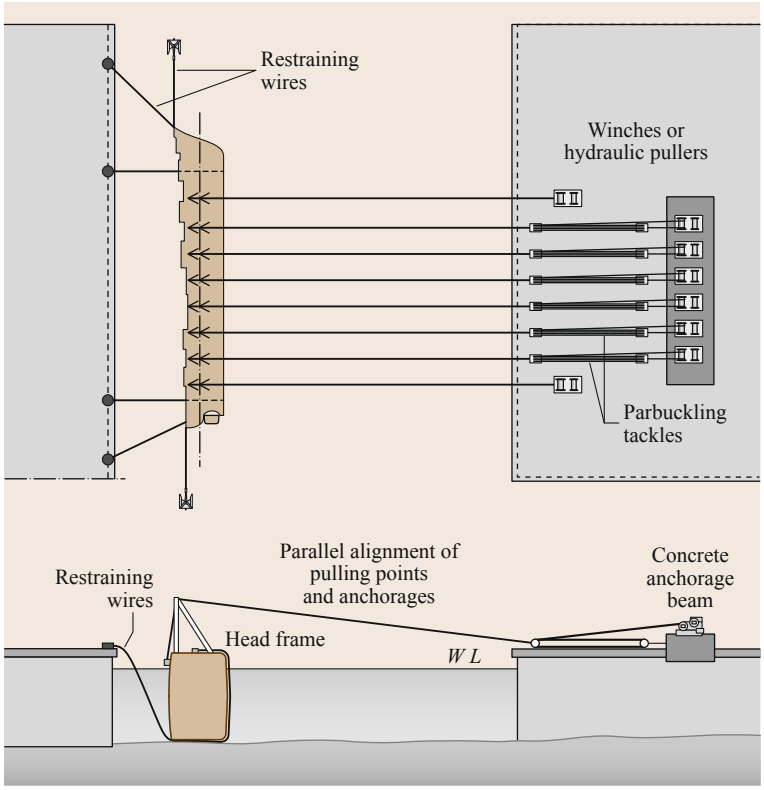


Fig. 42.49 Parallel alignment of pulling points and tackle

their own plane. Doubler plates and/or underdeck reinforcements spread the padeye loads through the ship's structure. Padeyes should be located to take advantage of existing stiffeners.

42.D.30 Special Bollards

Short, heavy-wall-thickness pipe bollards, welded to suitable long girders or structural steel sections, are a relatively efficient method of connecting multiple pulling wires to capsized ships. Each short bollard is fitted with a wide-top flange to prevent pulling wires from slipping off accidentally. Bollard-type connections are not suitable for all righting or hauling operations. This system is most often used with purpose-built pulling barges that deploy several wires to the capsized ship.

42.D.31 Chain Pigtails

Main hauling wires may be shackled to chain pigtails rove through specially cut holes or apertures in the capsized ship's hull. Chain pigtails rove through hull openings are most common when dragging partially submerged capsized ships out of channels. Decisions to use chair pigtail connectors are influenced by diving conditions and time required to prepare and weld special wire connection anchorages to the hull, especially if the attachment would otherwise require wet welding. Chain pigtails are not particularly effective unless the ship has an extremely heavy framing system to withstand the combined pulling and cutting effects of the chains. The method does not allow very detailed analysis of connection strength.

42.D.32 Location of Both Hauling and Lifting Points on Hull

Several factors that act individually and in combination influence the location of hauling and lifting points for mechanical righting. Hull construction and ship strength are considered along with the mechanical method to right the ship.

42.D.33 Headframe Attachment Position

To ensure that each purchase or pulling system develops almost identical righting moments, headframe top and hauling system anchorage point baselines must be in the same plane. To achieve this, it is customary to locate all headframe attachment positions on the capsized ship's parallel midbody.

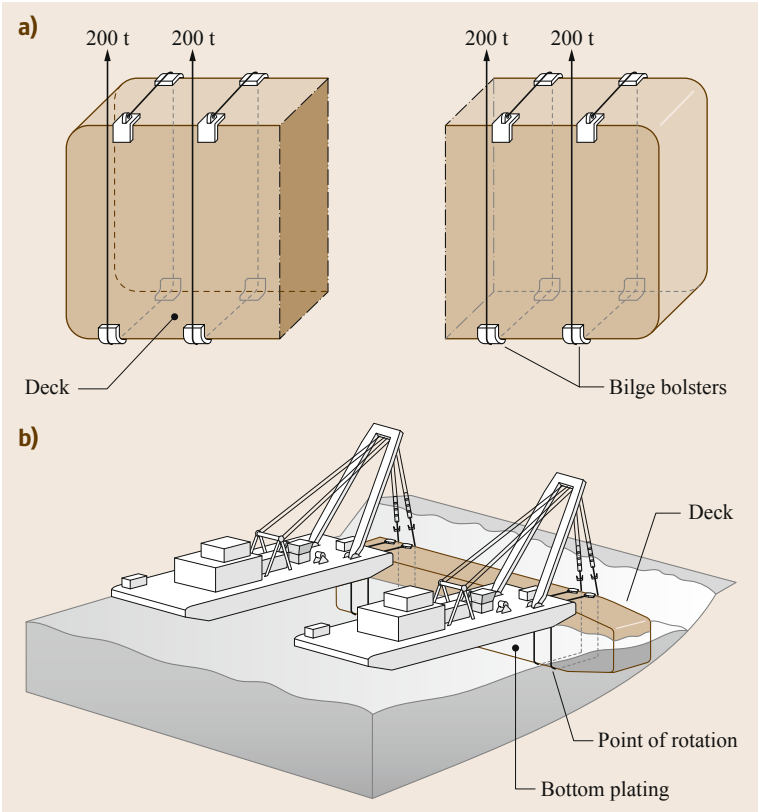


Fig. 42.50a,b Sheer legs uprighting capsized ship: **(a)** Sheer leg lifting wires, **(b)** detail of rigging arrangement

Large areas of parallel midbody seldom exist on modern warships, so salvors may have to construct headframes of different heights to be sure headframe tops are in alignment. Larger fleet auxiliaries and most large merchant ships have long parallel midbodies that simplify headframe mounting arrangements. Figure 42.49 shows this alignment.

Headframes must be set over strength decks to ensure that the strongest hull areas absorb the vertical thrust or pull. Salvors try to set headframes at junctions of sheer strake and weather deck plating, or junctions of main longitudinal strength deck and side shell plating. Where headframes cannot be set on hull junction points, salvors may have to install heavy stiffening.

42.D.34 Floating Cranes and Sheer Legs

Floating cranes and sheer legs in a righting role typically lift in the parallel midbody at the low or seafloor side of the capsized ship, opposite the bottom plating. Lifting slings or straps from cranes normally are passed underneath the capsized ship's hull, taken up along bottom plating, and then alongside the shell plate. This may involve tunneling, sawing, or sweeping messenger wires under the capsized ship. The wires connect to padeyes or special side bolster fittings welded to the high side sheer strake. As large forces or pulls are applied in localized areas, bolsters are fitted to side shell and deck plate junction areas in way of lifting pendants. Figures 42.50 and 42.51 show sheer legs uprighting a capsized ship and a schematic of pontoon attachment for righting.

Salvage personnel and crane barge operators must establish close liaison and mutually agree on connection methods to ensure that crane barge shackles and associated rigging jewelry match bolster fittings or padeyes made by salvors. Some superstructure elements, masts, and other fittings may be removed from the capsized ship to prevent them from damaging the barges during rotation operations.

High-capacity, single main hook offshore derrick barges employed in parbuckling operations may require several under hull pendants and connecting points because of their lifting power. Figure 42.50 shows typical righting arrangements with salvage sheer legs. The sheer legs have deployed both main lift hooks and deck tackles for maximum righting power.

42.D.35 Small-Capacity Buoyant Lift Devices

Salvage pontoons and other small-capacity buoyant lifting devices are connected to capsized ships by fittings and rigging methods that support the buoyant force. Rigging for buoyant lift systems does not involve the degree of engineering or component strength required for large cranes or sheer legs. Figure 42.51 shows a typical salvage pontoon attachment for righting with external buoyancy.

42.D.36 Hauling System Anchorages

Mechanical hauling methods for righting capsized ships require an anchorage system to pull against. The engineering and building or laying of anchorage systems are two of the most important and time-consuming parts of preparing for a mechanical righting operation. Failure or drag of any individual hauling anchor point creates a situation that results in either an embarrassment or a catastrophic stoppage of righting operations. This section describes some of the more important practical matters in selecting and establishing strong points for hauling anchorages.

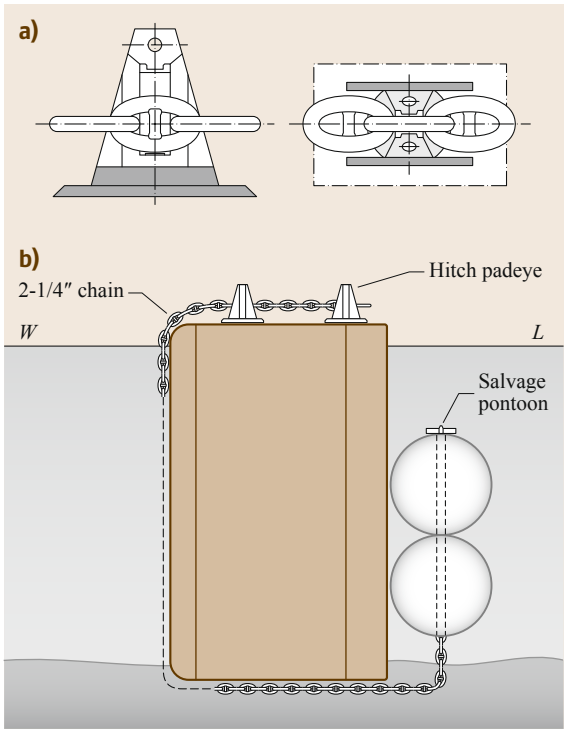


Fig. 42.51a,b Pontoon attachment for righting: **(a)** Hitch padeye, **(b)** typical arrangement of salvage pontoons

Anchorages are divided into two basic classes:

- Shore-based hauling foundations
- Marine-based pulling anchorage systems.

Anchorage or foundation systems for large righting systems require careful engineering and design analysis that consider:

- Total pulling force on the system
- Individual pull or reaction force on each anchorage
- Soil shear and strength characteristics, including the loadbearing capacity of the foundation area
- Methods of installing, removing and, where appropriate, demolishing and/or rehabilitating areas of foundation construction
- The relative merits and disadvantages of proposed anchorage systems and their cost in time and materials
- Effects in cost and time for the construction of specially engineered anchorage systems compared to intelligent improvisations with locally available components.

42.D.37 Shore-Based Anchorages

Anchorages for shore-mounted hauling systems cover a wide range of designs, depending on total pull exerted on each anchorage point and soil characteristics. Some systems designed to absorb 40–60 tons per point may use convenient materials of opportunity and involve simple excavations with backhoes. At the opposite end of the scale, major foundation works and site engineering are required with individual point pulls of 200–300 tons or more.

Basic designs for shore-hauling foundations include:

- Simple deadman systems that consist of an excavated pit with its front, or pressure face, lined with vertical balks or logs of heavy timbers. A chain for the standing purchase block is secured around a horizontal beam laid behind the face timbers. The chain is laid up outside before the pit is filled with soil, crushed rock, or reinforced, ready-mixed concrete.
- Simple piled anchorages, driven or drilled into soil or bedrock, to which tail chains or heavy padeyes are connected. Basic piled anchorages of this type are suitable as attachment points for standard hydraulic pullers or standard beach gear purchase blocks.
- Individually excavated pits in which steel structural beams are placed as foundations for winches. After structural steel members are welded together, the

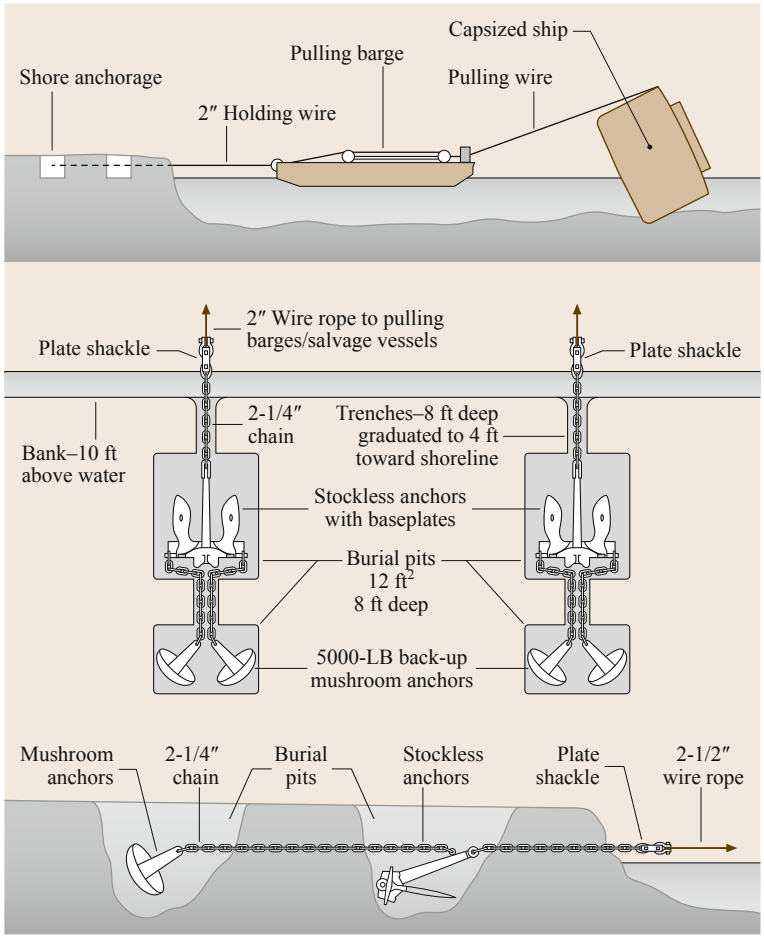


Fig. 42.52 Ancorage for small righting job

pits are filled with concrete to complete each anchorage unit. Variations on this method are quite common; winch foundation blocks.

- Large, composite foundation blocks for mounting several winch or hydraulic puller systems. Typically, such foundations are deeply excavated, prismatic trenches incorporating raker piles and large quantities of reinforced concrete. Excavation, preparation, steel fixing, and concrete pours make such foundations major tasks that may best be subcontracted to military or civilian engineering organizations.
- On occasion, bulldozers, crawler cranes, and other heavy tracked vehicles, including tanks, packed in roughly excavated pits have served as satisfactory shore anchorage points.

Innovative salvors have constructed their own shore-hauling system anchorages for smaller righting tasks. Simple and effective shore-hauling foundations have been constructed by taking beach gear anchors and chain cables ashore, excavating suitable pits, and burying them. Other salvors have successfully used large concrete mooring clumps, discarded structural steel or other materials of opportunity as shore-hauling foundations. Figure 42.52 shows such a system.

42.D.38 Weight Removal

In almost every case where ships capsize and sink, salvors must remove the structure or steel weight before righting the ship. Structure most frequently removed includes:

- Masts, stacks, and cranes
- Superstructure
- Weapons mounts and weather deck fittings
- Large sections of hull structure below the main deck.

No two sunken, capsized ships exhibit identical characteristics to salvors, even when ships are of the same type. In stranding situations, no two casualties are exactly the same. In capsized ships, a weight that must be removed on one ship becomes a useful weight to retain on a sister ship capsized under different circumstances.

42.D.39 Weight Removal

Weight or structural sections of capsized ships are removed for one of four reasons:

- Structures that will prevent righting of the ship, or will cause damage to salvage vessels, are removed to allow operations to proceed.
- Total capacity of available righting forces cannot overcome the calculated capsizing moment unless weights are removed.
- Weight, such as mud, silt, or debris, that enters a capsized ship after it has sunk, usually is removed to reduce righting force as a matter of good salvage practice.
- The weight concerned is of a hazardous or pollutant nature, such as ordnance or fuel oil.

Weight removal plans consider several factors, including:

- Getting maximum reduction of capsizing moment with minimum possible surface and diver work
- Not compromising or making refloating operations more difficult as a result of structural removals
- Availability of lifting equipment, either mechanical or buoyant, that governs the size and weight of sections that can be handled safely.

42.D.40 Weight Removal Methods

Methods of removing weight from capsized ships depend upon the location of weight or structure relative to average high and low water levels. As a general rule, any structural elements that are out of water throughout the tidal cycle are wholly or partially removed by conventional surface burning and cutting. Structural sections or individual weight elements that are wholly underwater are cut away by divers. Subject to equipment availability, some major surface and underwater sections may be removed by mechanical or explosive cutting.

42.E Appendix: Wrecking in Place

42.E.1 Wrecking-in-Place Methods and Techniques

Wrecking-in-place operations employ several specialized methods in addition to normal salvage techniques. These methods can be classified very broadly as:

- Manual cutting by divers and surface workers
- Mechanical demolition using heavy lift cranes
- Explosive sectioning, dispersal, or flattening
- Burial or settling by hydraulic dredging.

Each method has its own particular equipment requirements, operational techniques, and specific advantages and disadvantages. A method that is operationally suitable and acceptable for one wreck may be technically impractical or environmentally unacceptable under different circumstances. Salvors must determine what method, or combination of methods, is most suitable for their wreck and if appropriate equipment is available. Where the opportunity exists, salvors may combine wrecking in place with conventional salvage techniques.

42.E.2 Manual Cutting

Manual cutting by divers and surface crews is the most common method of performing small wrecking-inplace jobs and is used extensively on large wrecking operations. Topside crews cut with conventional oxy-acetylene or oxy-arc burning gear, or semiautomatic cutting machines. Divers make underwater cuts with underwater cutting equipment. Topside manual cutting is the most precise wrecking-in-place method, but it is also labor-intensive. Many environmental and wreck-related conditions influence underwater cutting speed and precision.

42.E.3 Mechanical Demolition

Mechanical demolition is usually performed by heavy lifting and hauling equipment mounted on floating cranes, salvage sheer legs, salvage vessels, or improvised salvage barges. Mechanical demolition methods include:

- Chain and wire cutting or sawing
- Direct ripping or stressing of weakened steel structures

- Tearing wrecks apart with heavyweight dredging grabs or specially designed wreck grabs
- Smashing wreck sections with wrecking chisels.

42.E.4 Explosive Sectioning

Explosives are an important wrecking-in-place tool that can be effective in both wreck removal and wreck dispersal operations. Explosive uses include:

- Cutting or breaking hull and superstructure sections
- Pounding or flattening wrecks into the seafloor
- Breaking wreck sections that cannot be handled by other methods
- Dispersing wreck sections or entire wrecks as part of channel clearing operations.

42.E.5 Burying or Settling Wrecks

Where seafloor conditions are suitable, and sufficient depth of water exists, it may be possible to dredge a trench, or a series of trenches, for the wreck to settle or skid into as the surrounding soil collapses. To reduce the amount of dredging work, divers cut away the wreck's superstructure, masts, and stacks before dredging begins. Before a wreck is buried, a soil analysis should be made to determine if the soil is suitable for this technique. Excavation for wreck settling or burial operations is done in three principal ways, either singly or in combination:

- Mechanical or profile dredging with clamshell or hydraulic cutter dredges
- Explosives
- High-pressure water jetting or sluice pumping.

Wreck burial normally is the least favored method of wreck disposal. While the problem of obstruction of a waterway or berth may be dealt with by burying a wreck, the wreck remains at the site where it may become an obstruction to future dredging or construction. Buried wrecks have been known to work their way back to the surface. Wreck burials should not be attempted without a thorough civil engineering investigation, consultation with long-term planning authorities, and approval from cognizant or governing environmental agencies.

42.E.6 Problems in Wrecking in Place

Wreck removal and wrecking-in-place operations are subject to all the difficulties of other salvage operations plus:

- A large quantity of unknowns about contents of and extent of damage to the ship
- Stringent safety and accident prevention policies and procedures
- Siltation and subsidence into the seafloor
- The physical working environment relative to port activities
- Removal and disposal of remaining cargo, stores, ordnance, and provisions
- Environmental regulations and restraints
- Acceptable and realistic work plans and schedules
- Equipment, manpower, logistics, and funding limits
- Postoperation cleanup and rehabilitation.

42.E.7 Environmental Regulation and Constraint

In all but purely military operations in wartime, wrecking-in-place operations are usually subject to the same environmental protection rules, regulations, and guidelines as other marine or harbor industrial activity. Salvors performing harbor or coastal wrecking operations often believe that environmental protection rules hinder their work. The rules are seldom waived because of the special circumstances of the work. While local regulations concerning matters such as explosives storage and handling and permissible charge rates may be negotiable, wrecking plans should not be based on any exceptions to environmental rules.

Wrecking in place invariably generates pollution in the form of:

- Unintentionally liberated cargo and debris
- Mud, silt, and solids in the water column
- Offensive garbage and trash liberated during the work
- Accidental spills of residual oil.

42.E.8 Manual Cutting

Manual cutting is surface or underwater cutting operations performed with man-portable cutting gear or tools including:

- Oxy-fuel gas cutting torches employed on the surface or underwater
- Electric or hydraulic cutting, grinding, or shearing
- Diver-operated cutting equipment such as oxy-arc, thermic lance, and Kerie cable.

Wrecking in place often requires extensive surface and underwater cutting. Numerous minor operations are

best performed by surface workers or divers cutting steel structural members, piping, and internal fittings with basic flame or oxy-arc techniques. Manual cutting is deployed for large-scale wrecking-in-place operations when:

- Mechanical wrecking systems are unavailable or unsuitable.
- Large portions of the superstructure or hull is above the surface.
- Debris or wreckage obstructing main working areas or cut lines must be removed for access.

Manual cutting supplements other wrecking-in-place techniques. Preliminary phases of mechanical and explosive cutting operations may require preparatory manual cutting to ensure efficiency.

42.E.9 Underwater Manual Cutting

Diver-operated manual cutting systems are used on almost every wrecking-in-place operation. Work assigned to underwater cutting teams ranges from cutting hull-section slinging holes to cutting entire hulls. The success and speed of underwater cutting depends upon several factors including:

- Experience and skill of the divers
- Limitations imposed on divers by diving conditions, current, wreck attitude, and depth
- Location of cut lines relative to hull structural features, machinery, piping, and internal fittings
- Access to the cut line.

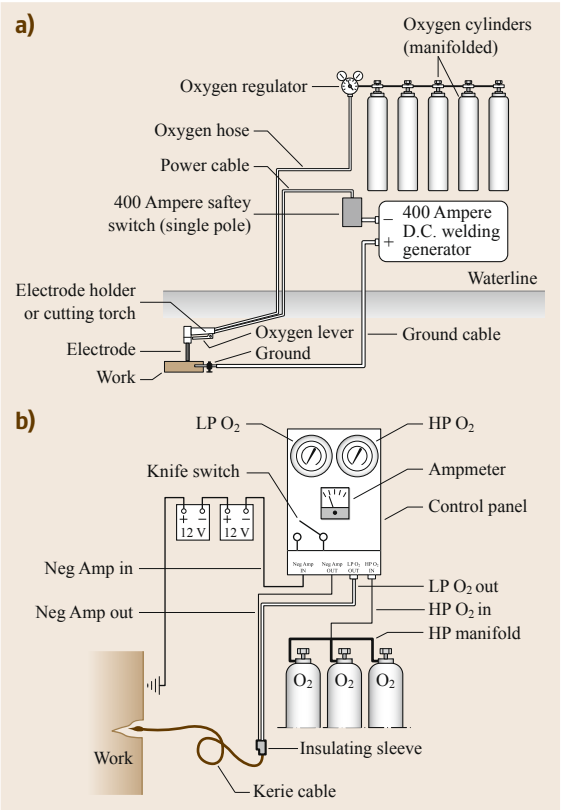


Fig. 42.53a,b Typical underwater cutting systems (**a**) typical oxy-arc arrangement (**b**) typical Kerrie Cable arrangement

42.E.10 Underwater Cutting Processes

Two underwater cutting processes are primarily employed:

- Oxy-arc cutting with exothermic electrodes, steel-tubular electrodes, and exothermic cable
- Shielded metal-arc cutting (cutting with ordinary welding leads and rods).

Oxy-arc is preferred because of its ease of use. There are two types of electrodes (rods) for oxy-arc cutting: exothermic and steel tubular. The exothermic is preferred because it burns independently after an arc is struck and oxygen is flowing.

In shielded metal arc cutting, the metal is cut by the intense heat without the oxygen. Shielded metal arc cutting is particularly suitable for cutting steel 1/4-inch or less thick, and nonferrous metals or corrosion-resistant metal of any thickness.

Figure 42.53 shows typical arrangements for underwater cutting with oxy-arc and Kerie cable systems.

42.E.11 Surface Manual Cutting

Surface cutting is often an important part of wrecking-in-place work. Cutting tasks vary from cutting holes for slings or access to demolition and removal of major hull sections above the waterline. Topside flame cutting teams range from one man with a portable cutting torch to several gangs of burners distributed all over the wreck. Most flame cutting is done with oxy-acetylene or oxy-propane gas mixtures. Cutting gases are compressed flammable and explosive substances that carry with them fire and explosion hazards. Wrecking operations with gas cutting systems require stringent safety programs to prevent accidents.

42.E.12 Lifting

Accidents during lifting operations occur easily when untrained personnel are allowed to rig, sling, or break

out even small lifts. Most wrecking-in-place operations employ bargemounted crawler or revolving cranes for general-purpose lifting. Crane operators are not always able to see lift areas and depend upon signals from the wreck. Many lifting accidents occur because:

- The lift area is not cleared of personnel before lifting.
- Incorrect or misleading signals are given to the crane operators.
- Adequacy of attachment lifting points.
- The sling is inadequate or incorrect for the weight, geometry, or behavior of the lift.
- Movements of the barge-mounted crane in the sea and swell is not taken into account.

Most lifting accidents can be avoided by assigning a small group of qualified salvage personnel as a rigging team. Because lifts take place from different areas, the rigging team moves from site to site. The team's duties include:

- Providing and rigging lifting gear to each section or component structure as it is ready to lift
- Checking that each piece is cut free and ready for lifting
- Planning and directing any final cuts necessary to break out difficult lifts
- Ensuring all hands and their equipment are clear of the working area before lifting begins
- Directing the crane operator by hand signals, whistle, or radio to make lifts
- Controlling lifts until the crane operator slews the lifts away from the wreck.

42.E.13 Mechanical Demolition

Mechanical demolition describes cutting with heavy lifting and hauling equipment. Such systems reduce the amount of diving and surface labor time required for wrecking in place operations. Mechanical cutting can be very effective, either in conjunction with explosive cutting or as a stand-alone technique, provided suitable heavy lifting and hauling equipment is available. Mechanical demolition methods include:

- Chain and wire cutting or sawing wrecks into sections suitable for lifting
- Tearing wrecks apart with specially designed wrecking grabs or heavy dredging grabs
- Direct impact cutting and smashing with wrecking chisels or wreck punches
- Stressing weakened steel structures to breaking point by direct ripping.

Mechanical cutting is usually performed by heavy lifting and hauling equipment mounted on floating cranes, salvage sheer legs, salvage vessels, or improvised salvage barges. Under certain circumstances, some mechanical cutting systems can be operated by shore-based lifting or hauling systems.

42.E.14 Chain Cutting

Chain-cutting sections wrecks to suit lift capacity and local circumstances. Chain cutting does not have precise guidelines to suit every wreck situation. Wrecks may be cut into sections either vertically or horizontally, depending upon wreck attitude and availability of suitable hauling or lifting equipment. Chain cutting is more efficient than wire cutting and is advantageous when:

- Suitable heavy-lift salvage sheer legs or cranes to operate cutting system are available.
- Tidal or river currents severely restrict diving operations.
- Cut lines with oxy-arc or explosives have not been completely successful.
- Damage to the wreck makes precision cutting difficult and dangerous, particularly where wreck sections are partially buoyant and hinged.
- Large quantities of cargo or debris create serious obstructions to clearing away and maintaining access to cut lines.

Figure 42.54 shows heavy salvage sheer legs with a cutting chain rigged into position underneath a wreck.

42.E.15 Advantages and Disadvantages of Chain Cutting

Chain cutting of wrecks has advantages and disadvantages. The advantages include:

- The system is basically independent of divers after the cutting chain has settled into its starting notches.
- Chain sawing usually cuts any given section faster than any diver-operated underwater methods.
- Chains are not hampered by mud, poor underwater visibility, or bottom time limitations.
- Buoyancy exerted by salvage cranes operating chain cutting systems usually has positive effects on cutting speed by increasing tearing and breaking action.

Disadvantages of chain cutting systems include:

- It is difficult and frequently dangerous for divers to examine progress of cutting due to jagged and

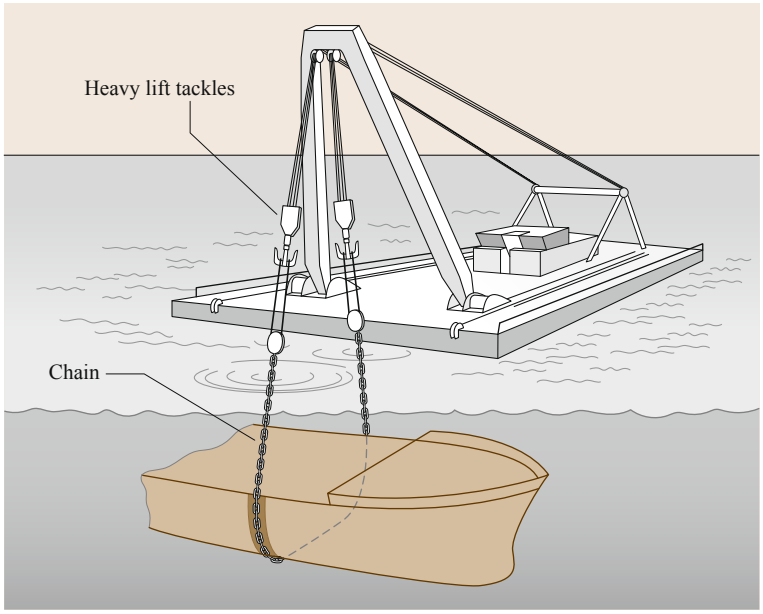


Fig. 42.54 Chain-cutting rig

torn metal edges along cut lines. Visual monitoring of progress can be difficult, if not impossible, even with underwater TV systems.

- Cutting delays may be encountered where chains deflect from planned cut lines and unintentionally cut into heavy beams and girders.
- Chains sometimes break inside the cut line. Extracting broken chain ends and re-rigging a new length of chain into cut lines can be time-consuming and difficult.
- Large salvage sheer legs perform the most efficient chain cutting work. These craft are not always available.

42.E.16 Cutting Chains

Chain cutting requires high-grade chain, such as Di-Lok or near equivalent flash-butt-welded, stud link chain. Cutting chains must be free of flaws, loose studs, and structural distortion. Scrap chain is rarely suitable for cutting. Heavy, good quality, used oil rig quality (ORQ) chain of 2 3/4 to 3 1/2-inch diameter has been successful in chain cutting. As a general rule, chain cutting should not be attempted with chain of less than 2 1/4 -inch diameter. The Di-Lok chain should be reserved for this type of work and lifting operations.

42.E.17 Preparation for Chain Cutting

The effectiveness and speed of chain cutting depend on the lifting capacity and outreach of the cranes available to salvors. Chain cuts are most efficiently performed by salvage sheer legs or heavy lift cranes rigged with two or more lift purchases of equal capacity. Lifting capacity of 150–200 tons per lift purchase appears to be the lowest acceptable level of lifting power for cutting large ship sections. Salvage sheer legs with several 300-ton lift capacity purchases are more suitable, but not always available. Preparatory steps for making chain cuts are:

1. Suitable messenger wires are passed, swept, or sawn under the wreck at each cut station.
2. Cutting chains are passed or dragged underneath the wreck and connected to crane lift purchase hooks with heavy wire slings.
3. Both lift purchases are lightly tensioned to bring the cutting chain into contact with wreck hull. Both contact points are marked and chain slacked off to allow divers to cut *starting notches* with oxy-arc equipment or shaped charges.
4. The size and depth of starting notches depends upon the aspect of the wreck and the diameter of the cutting chain. As a general rule of thumb, at least three or four links of chain should bury themselves in each starting notch.

In some cases, starting notches may have been pre-cut by an advance team so crane barge crews and divers rig cutting chains into pre-cut notches or a previously attempted cut line. Some pre-cutting of heavy structural section, such as propeller shafts and machinery foundations, may be necessary and advisable if hull cuts are to be made in such areas. Chain cutting through machinery spaces should be avoided whenever possible.

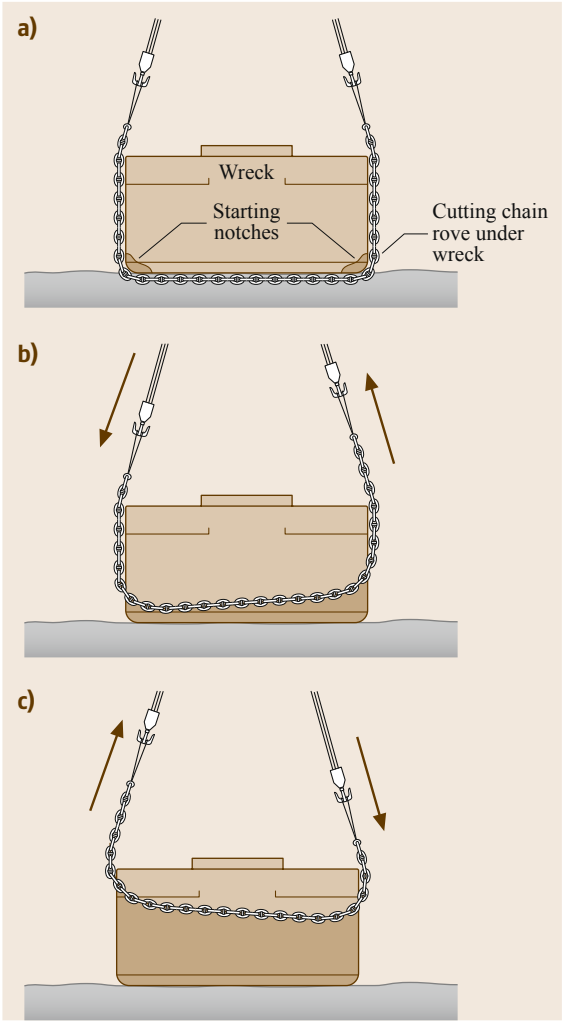


Fig. 42.55a–c Chain-cutting sequence: **(a)** Commencement of cutting sequence, **(b)** cutting sequence in progress, **(c)** cutting sequence almost complete

Figure 42.55 shows two starting notches cut into bilge radius plates of a wreck and the general progress of a chain cut.

42.E.18 Chain-Cutting Operations

Cutting chains are sawn alternately backward and forward through the wreck's hull and superstructure. Cutting is achieved by stressing, shearing, and tearing of steelwork by controlled lifting forces. Each link of the cutting chain acts like a blade on a chain saw. Steel plate and structural sections are crushed or distorted to their failure point as successive chain links wear and rip at the metal.

After a cutting chain is settled in its starting notch, operations usually proceed as follows:

1. A heaving or lifting strain of 150–200 tons is put on one end of chain, while the other purchase system slacks away slowly at about half that tension.
2. Several cycles, alternating each purchase between heaving and slacking, are usually necessary before the chain cuts or breaks into the wreck's hull.
3. Cutting rate is monitored by both hook weight readouts from strain gages and observing travel lengths of purchase tackles. Successively shorter purchase fleets indicate that a cut is proceeding efficiently.
4. When purchase fleet lengths become unworkably short, the long wire slings connecting each lift hook to the cutting chain ends must be replaced by shorter wire slings.
5. Cutting is completed when the cutting chain's bight is torn free of the wreck and recovered to the salvage crane.

42.E.19 Improvising Chain-Cutting Systems

Chain-cutting systems can be improvised from assets of opportunity, with varying degrees of success, depending upon knowledge, experience, and skill of salvors. Extremely powerful winches, such as the oilfield truss winches are suitable for chain cutting in either barge- or horemounted configurations.

Figure 42.56 shows an improvised chain sawing system rigged on a single-hook sheer legs crane. This system operates with a deck purchase leading over a bow fairlead, working in conjunction with the sheer leg purchase tackle. In this system, the main cutting and tearing load is applied by the Aframe purchase, with the deck purchase backhauling cutting chain after each tension cycle.

Successful chain cuts can be made by pairs of single-hook floating cranes or sheer legs. One derrick is located on each side of the wreck or section to be cut, and the chain saw steadily between both cranes. This method is technically relatively efficient, but requires two floating cranes or sheer legs that should be fairly well matched in size and capacity. Because the operation is performed from two separate vessels, coordination and control between crane operators is a critical factor in a safe and successful operation. Startup is characterized by some degree of trial and error as crane operators and salvors adapt themselves and their craft to the method. A strong mooring system for each crane and between the crane barges is essential to operate the system effectively. Figure 42.57 shows

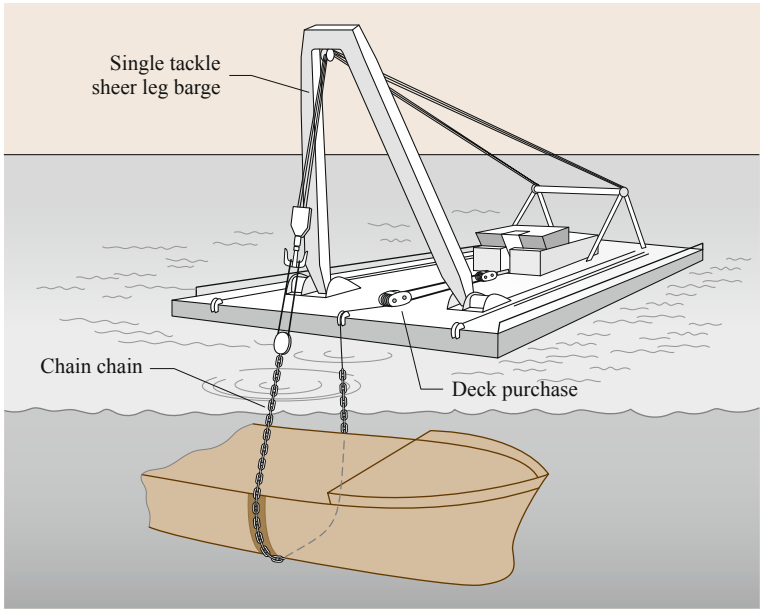


Fig. 42.56 Improvised chain cutting system with a-frame and deck tackle

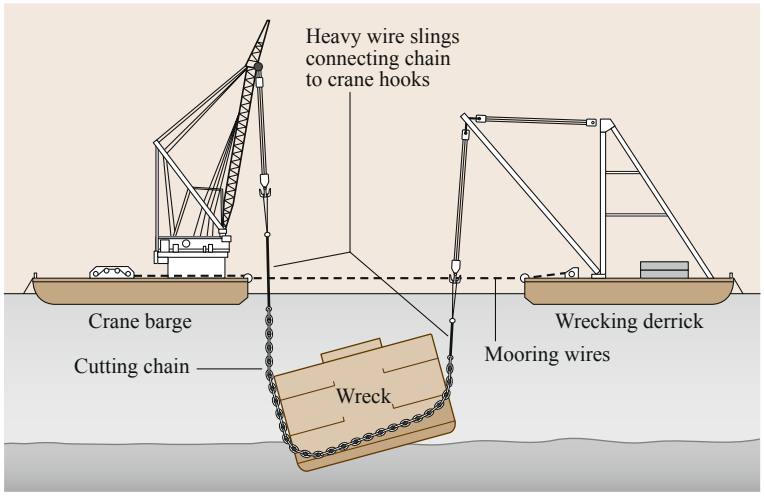


Fig. 42.57 Two crane barges operating a sawing chain

a chain sawing arrangement operated by two floating cranes.

Single kingpin mobile or tracked cranes such as barge-mounted crawler cranes are not designed or constructed to operate chain cutting systems. Improvised chain cutting systems with lattice boom rotating cranes may result in boom or pivot systems being unacceptably overloaded.

42.E.20 Wreck Grabs

Wrecks that lie partially or wholly buried in the seafloor, that have deteriorated with age, or that are seriously damaged, present a difficult wreck removal problem. Diving operations on such wrecks are usually hindered by strong currents, poor visibility, and a high degree of risk. Without mechanical systems, attempts to demolish or remove seriously damaged or partially destroyed sunken wrecks are usually time-consuming and very often extremely costly.

Dredging grabs are usually unable to crush steel effectively, and do not withstand the heavy stresses of wrecking. However, modified *cactus grabs* are extremely useful for gripping and tearing away weakened steel plates. These cactus grabs, also known as *orange peel grabs*, successfully tear away steel structure previously weakened or partially cut by explosives.

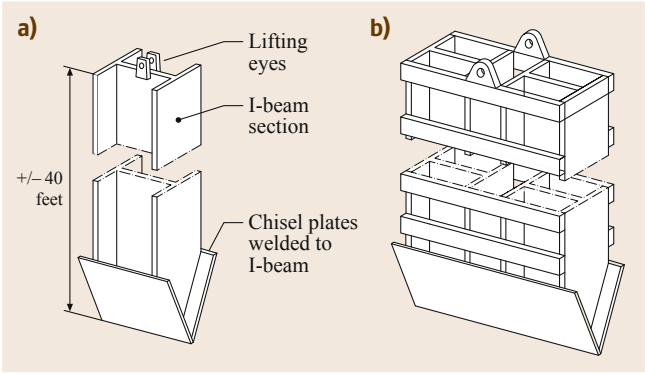


Fig. 42.58a,b Field fabrication wreck chisels: **(a)** Single I-beam chisel, **(b)** double I-beam chisel

Cactus grabs are strongly built, but do not have a wide total jaw opening. The grabs are suitable for demolition and wreckage purposes because they:

- Are strongly built, and relatively simple to operate
- Can be operated independently of divers, reducing risks to personnel
- Have the ability to grip even badly distorted steel structures.

The basic procedure for demolition and removal of sunken wrecks by wreck grabs is:

1. After completion of wreck survey, a decision is made on which end of the wreck will be demolished first. Grabbing operations usually commence at wreck's shallowest end and work steadily toward the more deeply buried section.
2. Heavy moorings are laid out around the wreck, and the sheer legs or salvage crane are aligned facing, or on the same heading as the wreck. Alignment of the crane is determined by the current and the basic demolition sequence.
3. The grab is lowered in the fully open position until it contacts the wreck. Closing tackles are hove up slowly to bite grab jaws into hull wreckage. As the grab jaws close, they penetrate steelwork and begin crushing and tearing.
4. When the grab will not close any more, its lifting tackles are hove up to lift the grab and the wreckage gripped in its jaws. Breaking out a closed grab is a contest of strength between the sheer legs and the wreck.
5. Wreck structure held by grab jaws is torn, crushed, and sheared away from the wreck as the sheer legs heaves up its grab. A combination of brute force and the buoyant upthrust of the sheer legs hull breaks overstressed and damaged steelwork.
6. The wreck grab is brought to surface where recovered wreckage is lowered onto a barge. Bites of 60–80 tons of wreckage can be taken under ideal conditions.

42.E.21 Wreck Punches and Chisels

A wreck punch or wreck chisel is a steel I-beam section cut to a chisel-shaped point at its lower end. Wreck punches smash steel hulls that have not been completely cut by explosive, oxy-arc, or surface cutting techniques. Under some circumstances, wreck punches cut hulls into sections.

Wreck punches are usually made up from heavy I-beams that are stiffened and sometimes boxed in with thick plates welded to beam flanges. Lead billets can be arranged inside beam flanges to add extra weight. Wreck punches are typically made up about 40 ft long, with a weight of between 10 and 15 tons. Wreck punches are operated by cranes. The punch is lifted above the wreck and then dropped repeatedly on the area to be cut. A heavily constructed punch, dropped from sufficient height, obtains enough energy to cut or break plate sections and frames on impact. Figure 42.58 shows typical wreck punches constructed from locally available materials.

Where chain cutting is not practical and wreck grabs are not available, wreck punches are comparatively effective mechanical cutting devices. Most large barge-mounted crawler cranes are capable of operating wreck punches. Heavy oilfield construction cranes can operate very large wreck punches. Improvised heavy wreck punches have been constructed from:

- Large dredge spuds, on which the pointed end was covered with welded steel plate to form a chisel end. Some of these converted spuds weigh 35–40 tons.
- Pairs of large section I-beams welded together and weighted with railroad track and scrap billets.

Not all cranes or derricks are suitable for operating wreck punches. Free-fall capability of the main crane hook is essential for wreck punching. Candidate cranes must be able to release and free-fall a heavy weight from a high boom elevation. Successful wreck punching depends upon a combination of punch weight and drop velocity. Cranes designed or specially adapted for marine clamshell dredging are almost always suitable for operating wreck punches. Auxiliary hooks on large offshore derricks have designed free-fall capability and sufficient lift combined with suit-

ably long booms to handle wreck punches quickly and safely.

42.E.22 Explosive Cutting

Explosives are an important salvage tool in wreck removal, harbor clearance, and wreck dispersal operations. Principal salvage and harbor clearance related explosive uses are:

- Cutting and breaking hull and superstructure sections
- Pounding down, flattening, or burying wrecks into the seafloor
- Dispersing wrecks or wreck sections as part of harbor or channel clearances
- Widening, deepening, and straightening channels
- Demolishing concrete masonry and steel harbor installations that obstruct port or salvage operations.

42.E.23 Burial, Flattening, and Reduction of Wrecks

In some situations, the combination of time, cost, and physical conditions make total wreck removal uneconomical or impractical. Under those conditions, the hazards of sunken wrecks may be sufficiently reduced without physically removing major portions of the wrecks. Three common techniques of wreck reduction are:

- Wreck burial
- Flattening wrecks
- Cutting down wrecks.

Wreck burial may be conducted as a wartime military operation with heavy explosive charges or as a peacetime task involving special dredging techniques. In the past, cutting down wrecks was a common method of increasing navigational depths over sunken wrecks. Military circumstances may still require salvors to cut down wrecks, but civilian port authorities do not encourage the practice. Each method described in this section results in hull structures remaining at or close to wreck sites, potentially creating future navigation, construction, or environmental hazards.

42.E.24 Burial of Wrecks

Burial of a wreck minimizes a navigational hazard, but does not remove a potential obstruction. Buried wrecks have changed their position because of scouring, and in the worst cases, have worked their way back toward the surface. Wreck burials should not be attempted without detailed consultation and approval of port operating, navigational, and environmental authorities. Detailed engineering investigations and dredging expertise are key elements in peacetime wreck burial tasks. Combination of careful profile dredging and explosive or gravity induced skidding undercut and settle wrecks into burial trenches.

Site conditions and lack or cost of wreck removal equipment may combine with suitable soil conditions to permit wreck burial. Seafloor soil characteristics are critical in the decision process. Detailed seafloor investigations and tests are necessary to:

- Establish the level below the seafloor and navigational datums where the bedrock or undredgable material strata is located.
- Confirm that the wreck can be lowered or buried to the clearance depth required by authorities. Clearance depths over the buried wreck are critical to the operation.
- Estimate rates of current-induced soil deposit or back filling that will occur during dredging operations.
- Establish soil characteristics for calculation of trench profiles and skidding angles.
- Decide a suitable method of dredging or combination of dredging and blasting necessary to excavate the burial trench.

Salvage engineering studies of wreck movement are conducted in conjunction with soil engineering and excavation investigations. These studies combine to produce a wreck burial and excavation plan based on either dredging or combined dredging and explosive trenching. Dredging methods include:

- Undercutting one side of the wreck so that it capsizes into a preexcavated trench or burial area.
- Alternately dredging and undercutting on each side of wreck so that it subsides in a rocking motion from port to starboard.
- Dredging a deep trench near the wreck, then profile dredging a sloping skidway from that trench to wreck.

Most work in a conventional wreck burial task is performed by dredging and civil engineering personnel. Salvage aspects of the work in terms of utilizing salvors' skills include:

- Detailed wreck surveys and assistance with moorings and seamanship aspects of soil investigations.

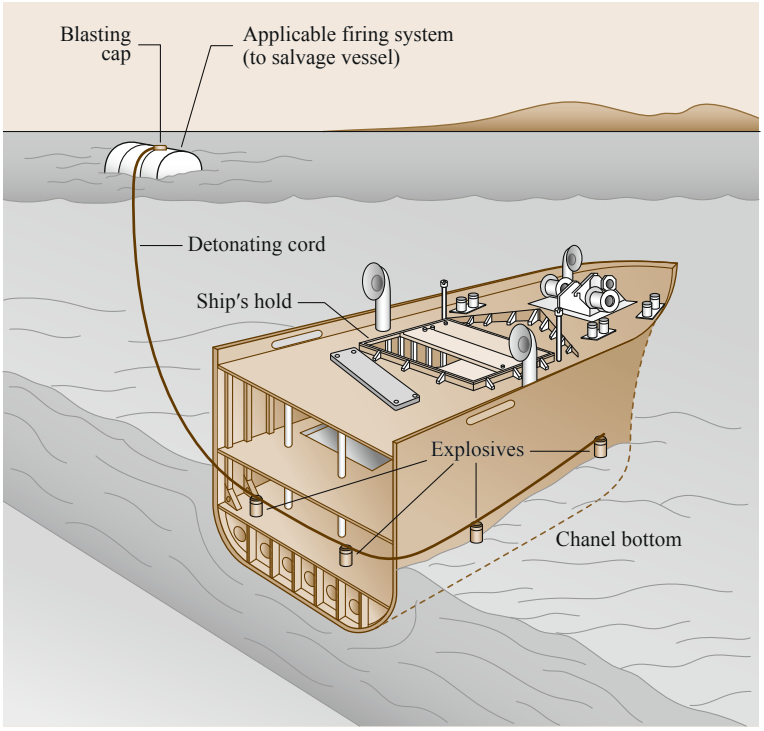


Fig. 42.59 Charge placement for wrecking settling

- Underwater cutting and removal of masts, stacks, superstructure, and other wreckage that may project above cut line.
- Monitoring project and providing technical advice and assistance with explosive charge placement, if required.

42.E.25 Flattening Wrecks

Flattening wrecks encompasses practices that are more specifically known and referred to as:

- Ship settling
- Ship flattening
- Wreck dispersal.

These three practices, grouped together because they have similar end results, employ different techniques, but all have common features including:

- The primary intention of destroying wrecks is to disperse navigational hazards or increase navigable depths.
- Time and circumstances do not permit wrecks to be removed conventionally.
- Wrecks are settled, flattened, or dispersed by explosive demolition.
- Most major wreck components and structures are left in situ as shattered debris.

Ship flattening or settling is usually only performed as an emergency means of channel or harbor clearance. Explosive wreck dispersal may be conducted before follow-up clearance with wreck grabs.

42.E.26 Ship Settling

Ship settling is making a sunken wreck entrench itself deeper into the seafloor. Ship settling may be the sole method of lowering a wreck to increase depth over it, or may be done in conjunction with ship flattening or wreck dispersal. Suitable soil conditions must exist for explosive ship settling to be successful. Soft sand or muddy seafloors permit settling, but hard clay soil may present some difficulties. Wreck settlement is done by:

- Placing heavy charges inside the hull and blowing holes along the bottom or side of the wreck where it is in contact with the seafloor. Damage to hull plating and reduction of bearing surface area causes the hull to sink. Seafloor material oozes into the hull through the holes blasted by the charges.
- Charges placed around the hull and simultaneously detonated to excavate a rough trench around the

wreck. In strong current areas scouring occurs and expedites settling.

These methods may be sequential; first the wreck's side or bottom plating is blasted out, followed by detonation of trenching charges around the wreck.

Figure 42.59 shows internal placement for settlement of an upright wreck.

42.E.27 Ship Flattening

Flattening procedures depend upon how the ship is sunk relative to the seafloor. A wreck lying on its side presents a different problem from one that is substantially upright on the bottom. Upright wrecks are flattened from top downward. Masts, stacks, and superstructure may be cut away with oxy-arc or explosive cutting methods and removed or blasted and allowed to scatter on the seafloor. Some methods of explosive hull flattening follow a combination of pounding down and linear cutting along deck edges to collapse the hull. Combined explosive heavy linear cutting and pounding often results in the wreck collapsing in on itself as a distorted scrapheap.

Another explosive wreck flattening method, more commonly associated with wreck dispersal, places a series of very heavy charges internally. Charges are located inside the wreck's hull with the most powerful charges sited amidships. When the forward, midships, and after end charges are simultaneously detonated, very heavy and opposing pressure surges occur. Internal over-pressure causes side plating and bulkheads to rupture and decks to collapse. Blast and overpressure effects from very heavy explosive charges detonated inside or against wrecks cannot be easily controlled and result in a shattered and distorted wreck.

Diving activities around such wrecks are hazardous and must be conducted with extreme caution. Wrecks that have been flattened or dispersed with heavy explosive charges usually make subsequent wreck removals difficult.

42.E.28 Cutting Down Wrecks

Sometimes navigational obstructions can be reduced by cutting a wreck down to a specified depth below a specified water level. Clearance levels over wrecks are established by port or area operating authorities with regard to present and future traffic. Wrecks can be cut down to specified levels by various wrecking techniques including:

- Oxy-arc cutting
- Explosive cutting
- A combination of oxy-arc and explosives
- Wreck grabbing.

Manual and explosive cutting methods are most common for this type of wreck reduction. Where cost and lack of suitable floating equipment influences wreck removal plans, partial reduction of wrecks is a short- to medium-term solution to a wreck removal problem.

References

42.1 US Navy: *US Navy Salvage Manual, Vol. 1, Strandings and Harbor Clearance and Afloat Salvage*, S0300-A6-MAN-010 (Naval Sea Systems Command, Washington 2013)

42.2 US Navy: *US Navy Salvor's Handbook*, S0300-A7-HBK-010 (Naval Sea Systems Command, Washington 2014)

42.3 C. A. Bartholomew, B. Marsh, R. W. Hooper: *US Navy Salvage Engineer's Handbook, Vol. 1, Salvage Engineering*, S0300-A8-HBK-010 (Naval Sea Systems Command, Washington 1992)

42.4 ASTM STD F1074-87: Standard Specification for Cleats, Welded Horn Type (ASTM International, West Conshohocken 2012)

42.5 Crane Co.: *Flow of Fluids through Valves, Fittings, and Pipe*, Tech. Paper 409 (Crane Co., Stumford 1942)

42.6 R. Peele, J.A. Church: *Mining Engineer's Handbook*, 3rd edn. (Wiley, New York 1948)

43. Oil Spills and Response

Merv Fingas

Oil spills are random phenomenon and occur in many sizes and forms. The largest 50 spill events are listed, noting that the Gulf War spill in 1991 remains the world's largest oil spill. Organized and rapid response to oil spills is very important to minimize environmental damage. Contingency plans detail the planned response. Five types of oils and fuels are frequently spilled and include gasoline, diesel fuel, light and heavy crude oils, and bunker fuels. Two important properties of these oils include the viscosity and density. The viscosity of spilled products can vary over orders-of-magnitude. The behavior of oils when spilled dictates the environmental impact these will have. The most important oil behaviors are evaporation and water uptake. Evaporation is an exponential factor with time and thus most of the evaporation (about 80%) occurs in the first 2 days. Water uptake takes place in any of the five ways:

- Soluble water
- Entrained water
- Meso-stable emulsions
- Stable emulsions
- Those that do not form any of the other types or unstable.

An important facet of oil spill assessment is laboratory analysis. The most common analytical method is by gas chromatography with mass spectrometric detection (GC-MS) or with flame ionization detection (FID). These methods are not only used for quantification but also for identification and measurement of amounts evaporated or biodegraded. Remote-sensing techniques especially that of satellite radar, are used to map oil spills on the sea. Once spilled, oils on the sea are typically contained using booms and recovered using skimmers, the most important of which are weir and oleophilic surface skimmers. Dealing with recovered oil and disposing of it are important steps. Spill-treating agents are occasionally used, especially oil spill dispersants. There are many conditions and considerations for the use of these. In-situ burning of oil is now being used more frequently and under correct conditions, can rapidly remove spilled oil. Finally, spilled oil often contacts shorelines, therefore assessment and careful removal techniques for shorelines are essential.

43.1 **Frequency of Oil Spills** 1068
43.2 **Response to Oil Spills** 1070
43.3 **Typical Oils and Their Properties** 1070
43.3.1 Oil Composition 1070
43.3.2 Properties of Oil 1071
43.4 **Behavior of Oil in the Environment** 1071
43.4.1 Evaporation 1072
43.4.2 Water Uptake 1072
43.4.3 Natural Dispersion 1073
43.4.4 Dissolution 1073
43.4.5 Photooxidation 1073
43.4.6 Sedimentation and Oil–Mineral Particle Interaction 1074
43.4.7 Biodegradation 1074
43.4.8 Tar Ball Formation 1074
43.4.9 Spreading and Movement 1075
43.4.10 Submergence/Sinking 1075
43.5 **Analysis, Detection, and Remote Sensing of Oil Spills** 1075
43.5.1 Laboratory Analysis 1075
43.5.2 Detection and Surveillance 1076
43.5.3 Remote Sensing 1076
43.6 **Containment on Water** 1078
43.6.1 Containment Booms 1078
43.6.2 Boom Failures 1079
43.6.3 Sorbent Booms and Barriers 1080
43.7 **Oil Recovery on Water** 1080
43.7.1 Skimmers 1080
43.7.2 Skimmer Performance 1082
43.7.3 Sorbents 1083
43.8 **Separation, Pumping, Decontamination, and Disposal** 1084
43.8.1 Temporary Storage 1084
43.8.2 Pumps 1084
43.8.3 Separators 1085
43.8.4 Decontamination 1085
43.8.5 Disposal 1085
43.9 **Spill-Treating Agents** 1086
43.9.1 Dispersants 1086
43.9.2 Surface-Washing Agents 1086

43.9.3 Solidifiers 1087
43.9.4 Biodegradation Agents..... 1087
43.10 **In-Situ Burning**..... 1087
43.10.1 Advantages..... 1087
43.10.2 Disadvantages..... 1087
43.10.3 Ignition and What Will Burn... 1087
43.10.4 Burn Efficiency and Rates..... 1088
43.10.5 Use of Containment..... 1088
43.10.6 Emissions from Burning Oil 1089
43.11 **Shoreline Cleanup and Restoration**..... 1089
43.11.1 Fate and Behavior on Shorelines..... 1089
43.11.2 Types of Shorelines and Their Sensitivity to Oil 1090
43.11.3 Cleanup Methods 1092
43.11.4 Recommended Cleanup Methods 1092
References..... 1092

Oil spills receive significant media attention when they are large or in areas of high significance and little attention otherwise. This also reflects the wide diversity of spill sizes and impacts. The movement of petroleum from the oil fields to the consumer involves as many as 10–15 transfers between many different modes of transportation including tankers, pipelines, railcars, and tank trucks. Oil is stored at transfer points and at terminals and refineries along the route. Accidents can happen during any of these transportation steps or storage times.

The rate of spillage has decreased in the past 10 years. This is especially true for tanker accidents at sea. Intensive training programs have been developed to reduce the potential for human error. Despite this, experts estimate that 30–50% of oil spills are either directly or indirectly caused by human error, with 20–40% of all spills caused by equipment failure or malfunction [43.1, 2].

43.1 Frequency of Oil Spills

Oil spills of any size, are a frequent occurrence, because of the heavy use of oil and petroleum products. The United States uses about 4×10^6 t of oil per day and worldwide, about 20×10^6 t are used per day. In the United States, more than half of the approximately 4×10^6 t of oil and petroleum products used per day is imported, primarily from Canada, Saudi Arabia and Africa [43.3]. About 40% of the daily demand in the United States is for automotive gasoline and about 15% is for diesel fuel used in transportation. About 40% of the energy used in the United States comes from petroleum, 25% from natural gas, and 20% from coal.

Spill statistics are collected by a number of agencies in the United States. The coast guard maintains a database of spills into navigable waters, while state agencies keep statistics on spills on land which are sometimes gathered into national statistics. The Bureau of Safety and Environmental Enforcement (BSEE) in the United States maintains records of spills from offshore exploration and production activities.

It can sometimes be misleading to compare oil spill statistics, however, because different methods are used to collect the data. The spill volume or amount is difficult to determine or estimate. For example, in the case of a vessel accident, the exact volume in a given compartment may be known before the accident, but the remaining oil may have been transferred to other ships immediately after the accident. Some spill acci-

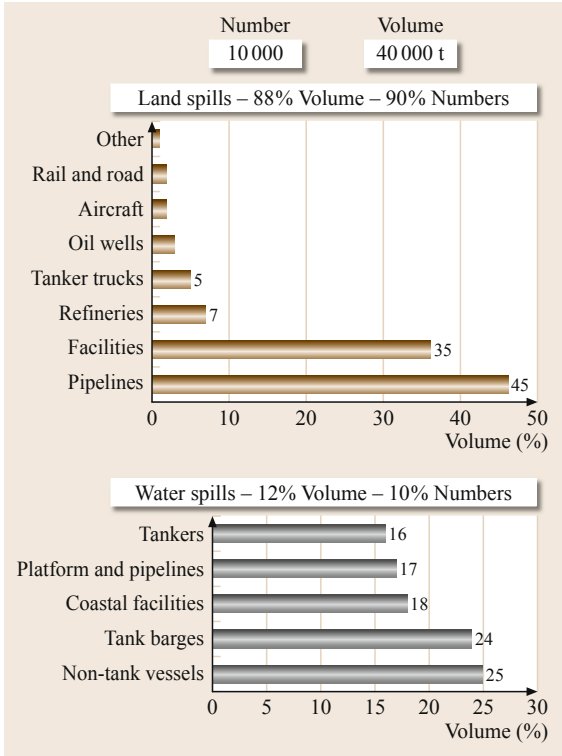


Fig. 43.1 Typical annual spill statistics for the United States (after [43.4]) ►

Table 43.1 List of the world's largest spills (after [43.1, 4])

No	Year	Day	Ship/Incident	Source	Country	Location	Tons ($\times 10^3$)
1	1991	Jan 26	Gulf war	War sabotage	Kuwait	Sea Island	800
2	2010	Apr 20	Deep water horizon	Oil well blowout	United States	Gulf of Mexico	500
3	1979	Jun 5	IXTOC blowout	Oil well blowout	Mexico	Gulf of Mexico	470
4	1979	Jul 19	*Atlantic Empress/ Aegean Captain*	Tanker collision	Off Tobago	Caribean Sea	287
5	1992	Mar 2	Oil well blowout	Oil well blowout	Uzbekistan	Fergana Valley	285
6	1993	Feb 4	Oil platform blowout	Oil well blowout	Iran	Nowruz Field	270
7	1983	Aug 6	*Castillo de Bellver*	Tanker accident	South Africa	Saldanha Bay	260
8	1991	May 28	*ABT Summer*	Tanker accident	Off Angola	Atlantic Ocean	260
9	1978	Mar 16	*Amoco Cadiz*	Tanker accident	France	Brittany	223
10	1991	Apr 11	*Haven*	Tanker accident	Italy	Genoa	144
11	1980	Aug 11	Oil well blowout	Oil well blowout	Libya	Inland	140
12	1988	Nov 10	*Odyssey*	Tanker accident	Off Canada	North Atlantic	132
13	1967	Mar 18	*Torrey Canyon*	Tanker accident	England	Lands End	119
14	1972	Dec 19	*Sea Star*	Tanker accident	Oman	Gulf of Oman	115
15	1981	Aug 20	Storage tanks	Loss from tanks	Kuwait	Shuaybah	110
16	1971	Dec 7	*Texaco Denmark*	Tanker accident	Belgium	North Sea	107
17	1994	Oct 25	Pipeline rupture	Pipeline release	Russia	Usinsk	105
18	1976	May 12	*Urquiola*	Tanker accident	Spain	La Coruna	100
19	1978	May 25	Pipeline rupture	Pipeline release	Iran	Ahvazin	100
20	1980	Feb 23	*Irenes Serenade*	Tanker accident	Greece	Pylos	100
21	1969	Feb 11	*Julius Schindler*	Tanker accident	Portugal	Azores	95
22	1977	Feb 23	*Hawaiian Patriot*	Tanker accident	Off USA	West of Hawaii	95
23	1979	Nov 15	*Independentza*	Tanker accident	Turkey	Bosporus Strait	95
24	1975	Jan 29	*Jakob Maersk*	Tanker accident	Portugal	Oporto	88
25	1979	Jul 6	Storage tanks	Loss from tanks	Nigeria	Forcados	85
26	1993	Jan 5	*Braer*	Tanker accident	United Kingdom	Shetland Islands	85
27	1989	Dec 19	*Kark 5*	Tanker accident	Morocco	Atlantic Ocean	80
28	1992	Dec 3	*Aegean Sea*	Tanker accident	Spain	La Coruna	75
29	1985	Dec 6	*Nova*	Tanker accident	Iran	Persian Gulf	72
30	1992	Apr 17	*Katina P*	Tanker accident	South Africa	Indian Ocean	72
31	1996	Feb 15	*Sea Empress*	Tanker accident	United Kingdom	Milford Haven	72
32	1971	Feb 27	*Wafra*	Tanker accident	South Africa	Atlantic Ocean	70
33	1978	Dec 11	Fuel storage tank	Loss from tanks	Rhodesia	Salisbury	65
34	2002	Nov 13	*Prestige*	Tanker accident	Spain	Off Spain	63
35	1960	Dec 6	*Sinclair Petrolore*	Tanker accident	Brazil	Off Brazil	60
36	1970	Mar 20	*Othello*	Tanker accident	Sweden	Vaxholm	60
37	1975	May 13	*Epic Colocotronis*	Tanker accident	United States	West of Puerto Rico	60
38	1978	Jun 12	Fuel storage tank	Loss from tanks	Japan	Sendai	60
39	1974	Nov 9	*Yuyo Maro 10*	Tanker accident	Japan	Tokyo	54
40	1983	Jan 7	*Assimi*	Tanker accident	Oman	Ras al Had	53
41	1965	May 22	*Heimvard*	Tanker accident	Japan	Hokkaido	50
42	1978	Dec 31	*Andros Patria*	Tanker accident	Spain	Bay of Biscay	50
43	1968	Jun 13	*World Glory*	Tanker accident	South Africa	Indian Ocean	48
44	1983	Dec 9	*Peracles GC*	Tanker accident	Qatar	Persian Gulf	48
45	1974	Aug 9	*Metula*	Tanker accident	Chile	Straits of Magellan	47
46	1970	Jun 1	*Ennerdale*	Tanker accident	Seychelles	Indian Ocean	46
47	1975	Jan 13	*British Ambassador*	Tanker accident	Japan	Iwo Jima	46
48	1994	Oct 21	*Thanassis A*	Tanker accident	Hong Kong	South China Sea	46
49	1978	Dec 7	*Tadotsu*	Tanker accident	Indonesia	Straits of Malacca	44
50	1968	Feb 29	*Mandoil*	Tanker accident	United States	Oregon	43

dent data banks do not include the amounts burned, if and when that occurs, whereas others include all the oil lost by whatever means. Sometimes the exact character or physical properties of the oil lost are not known and this leads to different estimations of the amount lost.

Only a small percentage of oil used in the world is actually spilled. Oil spills in the United States are summarized in Fig. 43.1 in terms of the volume of oil spilled and the actual number of spills [43.4]. In terms of oil spills, many countries have similar trends.

A list of the largest oil spills in the last 50 years is provided in Table 43.1 [43.1, 4]. The spills are listed according to their volume beginning with the largest spill to date – the release of oil during the Arabian Gulf War in 1991. It is important to note that because of the differences in record keeping and spill reporting, the list of the largest spills may vary from source to source.

There is a misconception that oil spills from tankers are the primary source of oil pollution in the marine environment. While it is true that some of the large spills are from tankers, it must be recognized that these spills still make up less than about 5% of all oil pollution entering the sea. The sheer volume of oil spilled from tankers and the high profile given these incidents in the media have contributed to this misconception. Half of the oil spilled in the seas is due to the runoff of oil and fuel from land-based sources rather than from accidental spills.

43.2 Response to Oil Spills

Rapid and effective response to oil spills will minimize damage to the environment. While it is important to focus on ways to prevent oil spills, methods for controlling them and cleaning them up must be rapidly and effectively implemented. An integrated system of contingency plans and response options can significantly reduce the environmental impact and severity of spills.

The purpose of contingency plans is to coordinate all aspects of the response to an oil spill. This includes stopping the flow of oil, containing the oil, and cleaning it up. The area covered by contingency plans could range from a single bulk oil terminal to an entire section of coastline. Oil spills, like forest fires, are not predictable and can occur anytime and during any weather. Therefore, the key to effective response to an oil spill is to be prepared for the unexpected and to plan spill countermeasures that can be applied to a range of possible conditions.

Most contingency plans utilize a tiered response, which means that response steps and plans escalate as the incident becomes more serious. As the seriousness of an incident is often not known in the initial phases, one of the first priorities is to determine the magnitude of the spill and its potential impact.

A contingency plan usually includes background information on the area covered by the plan. This consists of data collected from studies and surveys and often takes the form of a sensitivity map for the area. Sensitivity maps are now computerized in geographic information systems (GIS). These systems allow a composite map or image to be drawn up in layers. This allows personnel to rapidly update and analyze data in the area. Detailed information is usually kept in tables as part of the GIS. Sensitivity maps can also be integrated with computerized oil spill models.

As most oil companies or firms that handle oil do not have staff dedicated to cleaning up oil spills, several companies in the same area often join forces to form cooperatives. By pooling resources and expertise, these oil spill cooperatives can develop effective response programs. The cooperative purchases and maintains containment, cleanup, and disposal equipment and provides the training for its use.

43.3 Typical Oils and Their Properties

43.3.1 Oil Composition

Crude oils are mixtures of hydrocarbon compounds ranging from smaller, volatile compounds to very large compounds. This mixture of compounds varies according to the geological formation of the area in which the oil is found and strongly influences the properties of the oil. Petroleum can be conceived as consisting of groups of compounds called saturates, aromatics, resins, and asphaltenes [43.1].

The saturate group of components in oils consists primarily of alkanes, which are compounds of hydrogen and carbon with the maximum number of hydrogen atoms around each carbon. Thus, the term *saturate* is used because the carbons are saturated with hydrogen. Larger saturate compounds are often referred to as

Table 43.2 Typical oil properties (after [43.5])

Property	Units	Gasoline	Diesel	Light crude	Heavy crude	Intemediate fuel oil	Bunker C
Viscosity	mPa s at 15 °C	0.5	2	5–50	50–50 000	1000–15 000	10 000–50 000
Density	g ml^{-1} at 15 °C	0.72	0.84	0.78–0.88	0.88–1.00	0.94–0.99	0.96–1.04
API gravity		65	35	30–50	10–30	10–20	5–15

waxes. The saturate group also includes cyclo-alkanes, which are compounds made up of the same carbon and hydrogen constituents, but with the carbon atoms bonded to each other in rings.

The aromatic compounds include at least one benzene ring of six carbons. Three double carbon-to-carbon bonds are shared around the ring and add stability. Because of this stability, compounds containing benzene rings are very persistent and can have toxic effects on the environment. The most common smaller and more volatile compounds found in oil are often referred to as (BTEX) or benzene, toluene, ethyl-benzene and xylenes. Polyaromatic hydrocarbons (PAHs) are compounds consisting of at least two benzene rings. PAHs make up between 0 and 30% of the composition of oil.

Polar compounds are those that have a significant molecular charge as a result of bonding with compounds such as sulfur, nitrogen or oxygen. The polarity or charge that the molecule carries results in behavior that is different from that of un-polarized compounds, under some circumstances. In the petroleum industry, the smallest polar compounds are called resins. The larger polar compounds are called asphaltenes and they often make up the largest percentage of the asphalt commonly used for road construction. Asphaltenes are very large molecules and if in abundance in an oil, they have a significant effect on oil behavior.

43.3.2 Properties of Oil

The principal properties of oil are viscosity, density, and specific gravity. These properties for the oils discussed below are listed in Table 43.2 [43.5].

Viscosity is the resistance to flow in a liquid. The lower the viscosity, the more readily the liquid flows. The greater the percentage of light components such as saturates and the lesser the amount of asphaltenes, the lower the viscosity. As with other physical properties, viscosity is affected by temperature, with a lower temperature yielding a higher viscosity. In terms of oil spill cleanup, viscosity can affect the oil's behavior. Viscous oils do not spread rapidly, and do not penetrate soil as readily and hamper the ability of pumps and skimmers to handle the oil.

Density is the mass of a given volume of oil and is typically expressed in grams per cubic centimeter (g cm^{-3}). It is the property used by the petroleum industry to define light or heavy crude oils. Density is also important as it indicates whether an oil will float or sink in water. As the density of seawater is 1.03 g cm^{-3}, even heavier oils will usually float on it. The density of oil increases with time, as the light fractions evaporate. Occasionally, when the density of an oil becomes greater than the density of freshwater or seawater, the oil will sink. Sinking is rare, however and happens only with a few oils, usually residual oils.

Another measure of density is specific gravity, which is an oil's relative density compared to that of water at 15 °C. It is the same value as density at the same temperature. Another gravity scale is that of the American Petroleum Institute (API). The API gravity is based on the density of pure water which has an arbitrarily assigned API gravity value of 10° (10 degrees). Oils with progressively lower specific gravities have higher API gravities. The following is the formula for calculating API gravity

$$\text{API gravity} = \frac{141.5}{(\text{density at } 15.5\,^{\circ}\text{C})} - 131.5\,.$$

Oils with high densities have low API gravities and vice versa [43.1].

43.4 Behavior of Oil in the Environment

The specific oil-weathering behavior processes that occur after an oil spill dictate how the oil should be cleaned up and its effect on the environment. The fate and effects of a particular spill are determined by the behavior processes which are in turn almost entirely determined by the type of oil and the environmental conditions at the time of the spill [43.6].

The processes included in weathering are evaporation, emulsification, natural dispersion, dissolution, photooxidation, sedimentation, interaction with mineral fines, biodegradation, and the formation of tar balls [43.7]. These processes are discussed in order of importance in terms of their effect on the percentage of total mass balance, the greatest loss from the slick

in terms of percentage, and what is known about the process.

43.4.1 Evaporation

Evaporation is usually the most important weathering process for many oils. It has the greatest effect on the amount of oil remaining on water or land after a spill. Over a period of several days, a light fuel such as gasoline evaporates completely at temperatures above freezing, whereas only a small percentage of a heavier Bunker C oil evaporates. The evaporation rates of the oils discussed in this chapter are shown in Fig. 43.2 and [43.1]. The rate at which an oil evaporates depends primarily on the oil's composition. The more volatile components an oil or fuel contains, the greater the extent and rate of its evaporation.

The rate of evaporation is very rapid immediately after a spill and then slows considerably. About 80% of evaporation occurs in the first 2 days after a spill, which can be seen in Fig. 43.2. The evaporation of most oils follows a logarithmic curve with time. Some oils such as diesel fuel, however, evaporate as the square root of time, at least for the first few days. The properties of an oil can change significantly with the extent of evaporation. If about 40% of an oil evaporates, its viscosity could increase by as much as a thousand-fold. Its density could rise by as much as 10%.

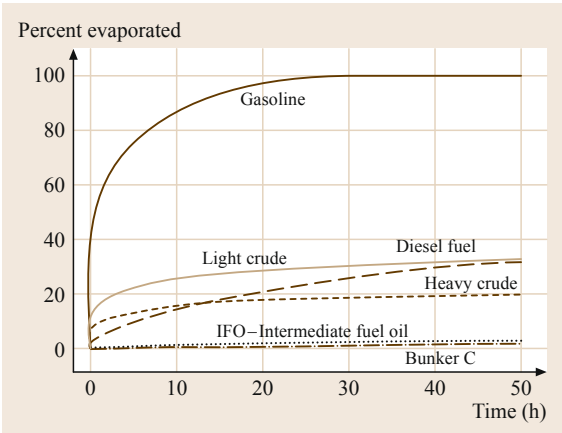


Fig. 43.2 Evaporation of typical oils at 15 °C (after [43.2])

43.4.2 Water Uptake

Water can enter oil through several processes. Emulsification is the process by which one liquid is dispersed into another one in the form of small droplets. To be called an emulsion, the product must have a certain amount of stability. Otherwise, the process is called water. Water droplets can remain in an oil layer in a stable form and the resulting material is completely different from the starting oil. Table 43.3 lists the five ways oil uptakes water and the resulting properties after water is in the oil [43.8].

The understanding of emulsion formation is in the early phases, but it starts with sea energy forcing the entry of small water droplets, about 10–25 μm in size, into the oil. If the oil is only slightly viscous, these small droplets will not leave the oil quickly. On the other hand, if the oil is too viscous, droplets will not enter the oil to any significant extent. Once in the oil, any asphaltenes and resins in the oil will interact with the water droplets to stabilize them. Depending on the quantity and type of asphaltenes and resins, an emulsion may be formed. The conditions required for emulsions of any stability to form may only be reached after a period of evaporation. Evaporation lowers the amount of low-molecular-weight compounds and increases the viscosity to the critical value.

Water can be present in oil in five ways:

1. First, some oils contain about 1% water as soluble water. This water does not significantly change the physical or chemical properties of the oil.

Table 43.3 Five methods by which oil uptakes water (after [43.8])

Type	Mechanism	Starting oil characteristics	Requirements	After water uptake			
				Color	Typical lifetime	Typical viscosity increase[a]	Typical water uptake
Soluble oil	Solubility	Most		Same	Years	1	< 1%
Unstable or does not uptake water	None	Many oils		Same		1	
Meso-stable	Viscous entrainment and A/R interaction	Moderate viscosity and some A/R	Sea energy	Reddish until broken	3–6 days	50	50–70%
Stable	*Viscous entrainment and A/R interaction*	*Moderate viscosity and some A/R*	*Sea energy*	*Reddish*	*Months*	800–1000	60–80%
Entrained	Viscous entrainment		Sea energy	As oil	2–10 days	2–5	30–40%

A/R: asphaltenes and resins, [a] viscosity increase from starting oil

2. The second way is when water droplets are not held in the oil long enough to form an emulsion. These are called oils that do not form any type of water-in-oil mixtures or unstable emulsions. These result when water droplets are incorporated into oil by the sea's wave action and there are not enough asphaltenes and resins in the oil and/or if there is an insufficient viscosity to prevent droplets from quickly leaving the oil mass. Unstable emulsions break down into water and oil within minutes or a few hours at most, once the sea energy diminishes. The properties and appearance of the unstable emulsion are the same as those of the starting oil.
3. Meso-stable emulsions represent the third way in which water can be present in oil. These are formed when small droplets of water are stabilized to a certain extent by a combination of the viscosity of the oil and the interfacial action of asphaltenes and resins. The viscosity of meso-stable emulsions is 20–80 times higher than that of the starting oil. These emulsions generally break down into oil and water or sometimes into water, oil, and emulsion remnants within a few days. Meso-stable emulsions are viscous liquids that are reddish-brown in color, until broken.
4. The fourth way in which water exists in oil is in the form of stable emulsions. These form in a way similar to meso-stable emulsions except that the oil contains sufficient asphaltenes and resins to stabilize water droplets. The viscosity of stable emulsions is 800–1000 times higher than that of the starting oil and the emulsion will remain stable for weeks and even months after formation. Stable emulsions are reddish-brown in color and appear to be nearly solid.
5. The fifth way in which oil can contain water is by viscosity entrainment. If the viscosity of the oil is such that water droplets can penetrate, the oil can contain about 30–40% water as long as it is in an energetic sea. Once the sea calms or the oil is removed, the water slowly drains. Typically most of the water would be gone before about 2 days. Such water uptake is called entrained water. This is not an emulsion type such as the meso-stable or stable emulsions which are stabilized by the chemical action of resins and asphaltenes.

These increases in volume and viscosity make cleanup operations more difficult. Stable emulsified oil is difficult or impossible to disperse, to recover with skimmers or to burn. Emulsions can be broken down with special chemicals in order to recover the oil with skimmers or to burn it. It is thought that emulsions break down into oil and water by further weathering, oxidation, by dilution with un-emulsified oils and freeze-thaw action. Meso-stable emulsions are relatively easy to break down, whereas stable emulsions may take months or years to break down naturally.

43.4.3 Natural Dispersion

Natural dispersion occurs when fine droplets of oil are produced by wave action or turbulence. Small oil droplets (less than about 30 μm) are somewhat stable in water for short periods of time. Large droplets (lager than about 100 μm) tend to rise and will not stay in the water column for more than a few seconds. Depending on oil conditions and the amount of sea energy available, natural dispersion can be insignificant or it can temporarily remove the bulk of the oil. Natural dispersion is dependent on both the oil properties and the amount of sea energy. Heavy oils such as Bunker C or a heavy crude will not disperse naturally to any significant extent, whereas light crudes and diesel fuel can disperse significantly. In addition, significant wave action is needed to disperse oil. In 30 years of monitoring spills on the oceans, those spills where oil has dispersed naturally have all occurred in very energetic seas sometimes with winds as high as 80 kn (40 m s^{-1}). The long-term fate of dispersed oil is that most of the dispersed oil may also rise and form another surface slick or it may become associated with sediment and be precipitated to the bottom.

43.4.4 Dissolution

Through the process of dissolution, some of the most soluble components of the oil are lost to the water under the slick. These include some of the lower molecular weight aromatics and some of the polar compounds. As only a small amount, usually much less than a fraction of a percent of the oil, actually enters the water column, dissolution does not measurably change the mass balance of the oil in the environment. The significance of dissolution is that the soluble aromatic compounds are particularly toxic to fish and other aquatic life. If a spill of oil containing a large amount of soluble aromatic components occurs in shallow water and creates a high localized concentration of compounds, then significant numbers of aquatic organisms can be killed. Gasoline, diesel fuel, and light crude oils are the most likely to cause aquatic toxicity. On open water, the concentrations of hydrocarbons in the water column are unlikely to kill aquatic organisms.

43.4.5 Photooxidation

Photooxidation can change the composition of an oil. The photooxidized products may be somewhat soluble

and dissolve into the water. It is not well understood how photooxidation specifically affects oils, although certain oils are susceptible to the process, while others are not. For most oils, photooxidation is not an important process in terms of changing their fate or mass balance after a spill.

43.4.6 Sedimentation and Oil–Mineral Particle Interaction

Sedimentation is the process by which oil is deposited on the bottom of the sea or other water body. While the process itself is not fully understood, certain facts about it are. Most sedimentation noted in the past has occurred when oil droplets reached a higher density than water after interacting with mineral matter in the water column. This interaction sometimes occurs on the shoreline or very close to the shore. Once oil is on the bottom, it is usually covered by other sediment and degrades very slowly. In a few well-studied spills, a significant amount (about 10%) of the oil was sedimented on the sea floor [43.9]. Such amounts can be very harmful to a biota which may come in contact with the oil on the sea bottom.

Oil slicks and oils on shorelines sometimes interact with mineral fines suspended in the water column and the oil is thereby transferred to the water column [43.9]. Particles of mineral with oil attached may be heavier than water and sink to the bottom as sediment or the oil may detach and refloat. Oil–fines interaction does not generally play a significant role in the fate of most oil spills in their early stages, but can have an impact on the rejuvenation of an oiled shoreline over the long term.

43.4.7 Biodegradation

A large number of microorganisms are capable of degrading petroleum hydrocarbons. Many species of bacteria, fungi, and yeasts metabolize petroleum hydrocarbons as a food energy source [43.10, 11]. Bacteria and other degrading organisms are very abundant in areas where there have been petroleum seeps, although these microorganisms are found everywhere in the environment. As each species can utilize only a few related compounds at most, broad-spectrum degradation does not occur. Hydrocarbons metabolized by microorganisms are generally converted to an oxidized compound, which may be further degraded, may be soluble, or may accumulate in the remaining oil. The aquatic toxicity of the biodegradation products is sometimes greater than that of the parent compounds.

The rate of biodegradation depends primarily on the nature of the hydrocarbons and then on the temperature. Generally, rates of degradation tend to increase as the temperature rises. Some groupings of bacteria, however, function better at lower temperatures and others function better at higher temperatures. Indigenous bacteria and other microorganisms are often the best adapted and most effective at degrading oil as they are acclimatized to the temperatures and other conditions of the area. Adding super-bugs to the oil does not necessarily improve the degradation rate.

The rate of biodegradation is greatest on saturates, particularly for those oils containing approximately 12–20 carbons. Aromatics and asphaltenes, which have a high molecular weight, biodegrade very slowly, if at all. This explains the durability of roof shingles containing tar and roads made of asphalt, as both tar and asphalt consist primarily of aromatics and asphaltenes. On the other hand, diesel fuel is a highly biodegradable oil as it is largely composed of biodegradable saturates. Light crude oils are also biodegradable to some extent. While gasoline contains biodegradable components, it also contains some compounds that are toxic to some microorganisms. Gasoline will evaporate before it can degrade. Heavy crude oils contain little material that is readily biodegradable and Bunker C contains almost none.

The rate of biodegradation is also highly dependent on the availability of oxygen. On land, oils such as diesel can degrade rapidly at the surface, but very slowly if at all only a few centimeters below the surface, depending on oxygen availability. In water, oxygen levels can be so low that degradation is limited. It is estimated that it would take all the dissolved oxygen in approximately 400 000 l of sea water to completely degrade 1 l of oil [43.1]. The rate of degradation also depends on the availability of nutrients such as nitrogen and phosphorus, which are most likely to be available on shorelines or on land. Oil degrades significantly at the oil–water interface at sea and on land, mostly at the interface between soil, oil, and air.

Biodegradation can be a very slow process for some oils. It may take weeks for 50% of a diesel fuel to biodegrade under optimal conditions and years for 10% of a crude oil to biodegrade under less optimal conditions. For this reason, biodegradation is not considered an important weathering process in the short term.

43.4.8 Tar Ball Formation

Tar balls are agglomerations of heavy oil less than about 10 cm in diameter. Larger accumulations of the same material ranging from about 10 cm to 1 m in diameter are called tar mats. Tar mats are pancake-shaped. Their formation is still not completely understood, but it is known that they are formed from the residuals of heavy crudes and Bunker fuels. After these oils weather at sea

and slicks are broken up, the residuals remain in tar balls or tar mats. The formation of droplets into tar balls and tar mats has also been observed, with adhesion being the binding force.

The formation of tar balls is the ultimate fate of many spilled oils at sea. These tar balls are then often deposited on shorelines. The oil may come from spills, but it is also residual oil from natural oil seeps or from deliberate operational releases such as from ships. Tar balls are regularly recovered by machine or by hand from recreational beaches by various agencies or resort owners.

43.4.9 Spreading and Movement

Oil spilled on water spreads relatively rapidly [43.12]. After an oil spill on water, the oil tends to spread into a slick over the water surface. This is especially true for the lighter petroleum products such as gasoline, diesel fuel, and light crude oils, which rapidly form very thin slicks. Heavier crudes and Bunker C spread to slicks as thick as several millimeters. Heavy oils may also form tar balls and tar mats and thus may not go through progressive stages of thinning.

Oil spreads horizontally over the water surface even in the complete absence of wind and water currents. This spreading is caused by the force of gravity and the interfacial tension between oil and water. The viscosity of the oil opposes these forces. As time passes, the effect of gravity on the oil diminishes, but the force of the interfacial tension continues to spread the oil. The transition between these forces takes place in the first few hours after the spill occurs.

Oil slicks will elongate in the direction of the wind and currents and as spreading progresses, take on many shapes depending on the driving forces. Oil sheens often precede heavier or thicker oil concentrations. If the winds are high (more than $20\,\mathrm{km\,h^{-1}}$), the sheen may separate from thicker slicks and move further downwind.

In addition to their natural tendency to spread, oil slicks on water are moved along the water surface, primarily by surface currents and winds. The slick generally moves at a rate that is 100% of the surface current and approximately 3% of the wind speed.

43.4.10 Submergence/Sinking

When oil is denser than the surface water, it may sink. Some rare types of heavy crudes and Bunker fuels can reach these densities and sink. When this occurs, the oil may sink to a denser layer of water rather than to the bottom. It is important to note that sinking of any form, whether to the bottom or to the top of a layer of dense sea water, is rare. When oil does sink, it complicates cleanup operations as the oil can be recovered only with specialized underwater suction devices or special dredges.

Over-washing is another phenomenon that occurs quite frequently. At moderate sea states, a dense slick can be over-washed with water. When this occurs, the oil can disappear from view if the spill is being observed from an oblique angle, as would occur if someone is observing from a ship. Over-washing causes confusion about the fate of an oil spill as it can give the impression that the oil has sunk and then resurfaced.

43.5 Analysis, Detection, and Remote Sensing of Oil Spills

43.5.1 Laboratory Analysis

Laboratory analysis can provide information to help identify an oil if its source is unknown and a sample of the starting oil is available. With a sample of the source oil, the degree of weathering and the amount of evaporation or biodegradation can be determined for the spilled oil. Through laboratory analysis, the more toxic compounds in the oil can be measured and the relative toxicity of the oil at various stages of the spill can be determined. This is important information to have as the spill progresses.

Taking a sample of oil and then transporting it to a laboratory for subsequent analysis is common practice [43.13]. The simplest and the most common form of analysis is to measure the quantity of oil in a sample. Such analysis results in a value known as total petroleum hydrocarbons (TPH). The TPH measurement can be obtained in many ways, including extracting the oil from the soil or by further chemical analysis.

A typical form of analysis is to use a gas chromatograph (GC). The system is calibrated by passing known amounts of standard materials through the unit. The amount of many individual compounds in the oil is thereby measured. The compounds that pass through the detector can also be totaled and a TPH value determined. While it is highly accurate, this TPH value does not include resins, asphaltenes, and some other components of the oil with higher molecular weight which do not pass through the column.

A typical chromatogram of a light crude oil with some of the more prominent components of the oil

identified is shown in Fig. 43.3. The two most common types of detector used on a gas chromatogram are flame ionization detectors (FID) and mass spectrometers (MS). The latter method is usually called GC-MS and can be used to quantify and identify many components in oil. The mass spectrometer provides information about the structure of the substance so compounds can be more positively identified. This information can then be used to predict how long the oil has been in the environment and what percentage of it has evaporated or biodegraded. The same technique can be used to fingerprint an oil and positively identify its source. Certain compounds are consistently distributed in oil, regardless of weathering, and these are used to identify the specific oil.

43.5.2 Detection and Surveillance

Oil spills are often located and surveyed from helicopters or aircraft using only human vision [43.14]. There are some conditions, however, such as fog and darkness, in which oil on the surface cannot be seen. Very thin oil sheens are also difficult to detect as is oil viewed from an oblique angle (less than 45°) especially in hazy or other conditions that limit vision. Oil can also be difficult to see in high seas and among debris or weeds and it can blend into dark backgrounds, such as water, soil, or shorelines.

In addition, many naturally occurring substances or phenomena can be mistaken for oil. These include weeds and sunken kelp beds, whale sperm, biogenic or natural oils such as from plants, glacial flour (finely ground mineral material), sea spume (organic material), wave shadows, sun glint and wind sheens on water and oceanic and riverine fronts where two different bodies of water meet, such as a river entering the sea.

A very thin oil sheen as it appears on water is shown in Fig. 43.4. This figure shows the thickness and appearance of oil that could be present under such circumstances. It is important to note that color and slick appearance are only roughly related to thickness and that this relationship only applies to very thin slicks.

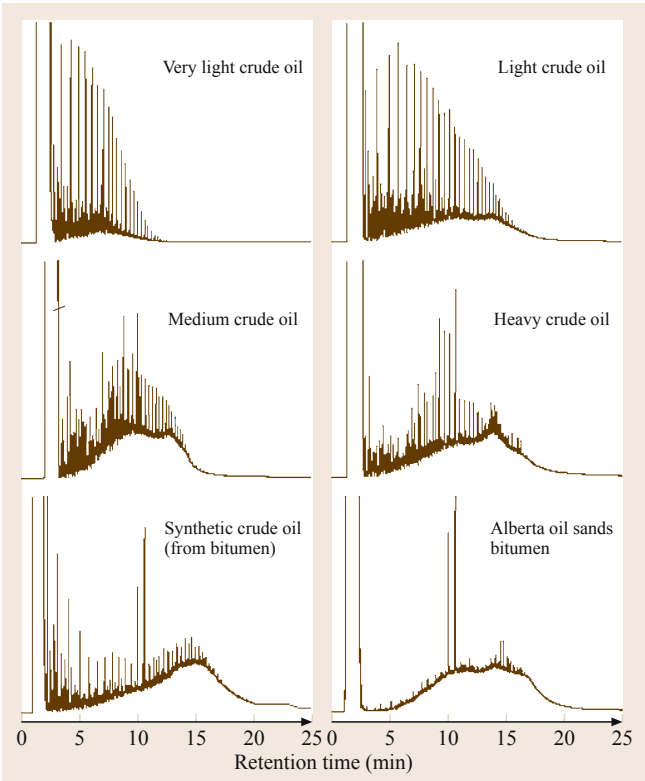


Fig. 43.3 Chromatograms of several types of oil. The detector used was a flame ionization detector (FID) (courtesy of Environment Canada)

43.5.3 Remote Sensing

Remote sensing of oil involves the use of sensors to detect or map oil spills. Oil often cannot be detected in certain conditions using visual means. Remote sensing provides methods to map out the locations of spills under many conditions. Remote sensing is usually carried out with instruments onboard aircraft or from satellite. Table 43.4 shows a summary of sensors, their characteristics, and their suitability to various missions [43.14].

Many devices employing the visible spectrum, including the conventional camera and video camera, are available at a low cost. As these devices are subject

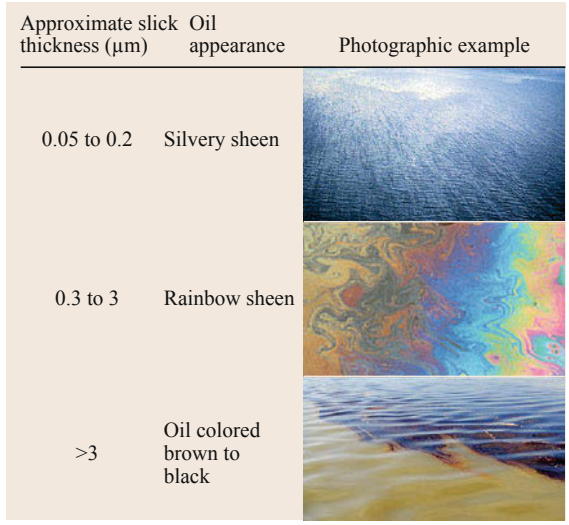


Fig. 43.4 The appearance of thin films of oil on water (courtesy of Merv Fingas)

Table 43.4 Sensor characteristics and suitability (after [43.14])

Sensor	Mission suitability					Typical coverage width [km]	Acquisition cost range k$	State of development
	Support for cleanup	Night & fog operation	Oiled shoreline survey	Spill mapping	Ship discharge surveillance			
Still camera	2	n/a	2	2	2	0.25–2	1–5	High
Video	2	n/a	2	2	2	0.25–5	1–10	High
Night time vision camera	3	4	n/a	2	2	0.25–2	5–20	Medium
IR camera (8–14 μm)	4	2	n/a	3	3	0.25–2	20–50	High
UV camera	2	n/a	n/a	3	2	0.25–2	4–20	Medium
Multispectral scanner	1	n/a	1	2	1	0.25–2	100–200	Medium
Radar	n/a	4	n/a	4	3	5–50	1200–8000	High
Microwave radiometer	1	3	n/a	2	2	1–5	400–1000	Medium
Laser fluorosensor	4	3	5	1	5	0.01–0.1	300–1000	Medium

Key: n/a = not applicable; IR: infrared; UV: ultraviolet; numerical values represent a scale from 1 = poorly suited to 5 = *ideally suited*

to the same interferences as visual surveillance, they are used primarily to document the spill or to provide a frame of reference for other sensors.

Thick oil on water absorbs infrared radiation from the sun and thus appears in infrared imagery as hot on a cold ocean surface. Unfortunately, many other false targets such as weeds, biogenic oils, debris, and oceanic and riverine fronts can interfere with oil detection. Infrared sensors are relatively inexpensive and widely used for supporting cleanup operations and directing cleanup crews to the nonsheen portions of an oil spill.

Oils that contain aromatic compounds, as most oils do, will absorb ultraviolet light and give off visible light in response. Since very few other compounds respond in this way, this can be used as a positive method of detecting oil at sea or on land. Laser fluorosensors use a laser in the ultraviolet spectrum to trigger this fluorescing phenomenon and a sensitive light detection system to provide an oil-specific detection tool. There is information in the visible light return that can be used to determine whether the oil is a light or heavy oil or a lubricating oil. In a sense, using laser fluorosensors is like performing chemistry from the air. Laser fluorosensors are the most powerful remote sensing tools available because they are subject to few interferences. Laser fluorosensors work equally well on water and on land and are the only reliable means of detecting oil in certain ice and snow situations.

Few types of sensors can be used to measure the thickness of an oil slick. The passive microwave sensor can be calibrated to measure the relative thickness of an oil slick. Absolute thickness cannot be measured for the multiple reasons. Other factors such as atmospheric conditions also change the radiation levels; the signal changes in cyclical fashion with spill thickness and the signal must be averaged over a relatively wide area and the slick can change throughout this area. The cyclical nature of the microwave radiation variance with

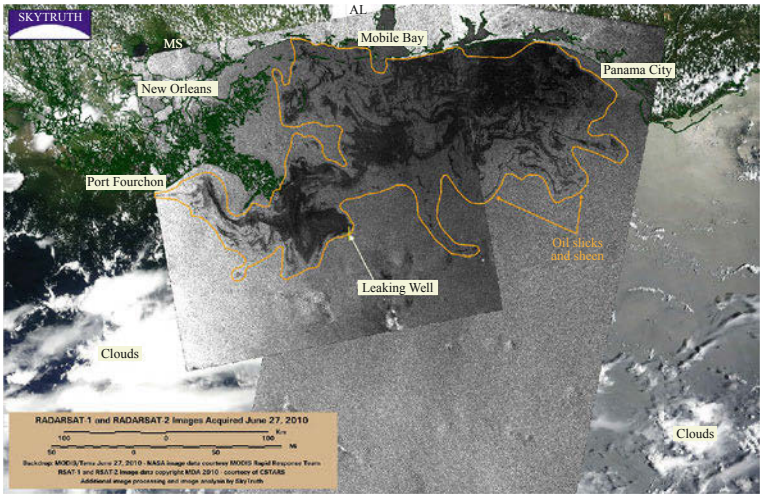


Fig. 43.5 A RADARSAT-1 and -2 combined image of the Gulf oil spill. The radar images are laid upon a satellite visible image of the coast. Oil slicks and sheen are *outlined in color* (Photo courtesy of Canadian Space Agency)

thickness has been overcome by some suppliers by using multi-frequency units – for example, five different frequencies.

Oil on the sea calms smaller waves (waves on the order of a few centimeters in length), therefore radar can detect oil on the sea as a calm area. The technique is highly prone to false targets, however, and is limited to a narrow range of wind speeds (approximately 2–6 m s^{-1}). Below this wind speed, there are not enough small waves to yield a difference between the oiled area and the sea. At higher winds, the waves can propagate through the oil and the radar may not be able to see into the troughs between the waves. Radar is not useful near coastlines or between head lands because the wind shadows look like oil. There are also many natural calms on the oceans that can resemble oil. Despite its large size and expense, radar equipment is particularly well suited for searches of large areas and for work at night or in foggy or other bad weather conditions.

Several radar satellites are available which operate in the same manner as airborne radar and share their many limitations [43.14]. Despite these limitations, radar imagery from satellite is particularly useful for mapping large oil spills. Much modern spill response depends on radar satellite data. Figure 43.5 shows a major spill as imaged in a satellite-borne radar sensor.

While many satellites provide images in the visible spectrum, oil cannot be seen in these images unless the spill is very large or rare sea conditions are prevalent that provide a contrast to the oil. Oil has no spectral characteristics that allow it to be enhanced from the background.

43.6 Containment on Water

Containment of an oil spill is the process of confining the oil, either to prevent it from spreading to a particular area, to divert it to another area where it can be recovered or treated, or to concentrate the oil so it can be recovered or burned. Containment booms or simply booms are the most frequently used piece of equipment for containing an oil spill on water. Booms are generally the first equipment mobilized at a spill and are often used throughout the operation.

43.6.1 Containment Booms

A boom is a floating mechanical barrier designed to stop or divert the movement of oil on water. Booms resemble a vertical curtain with portions extending above and below the water line. Most commercial booms consist of four basic components:

- A means of flotation
- A freeboard member to prevent oil from splashing over the top of the boom
- A skirt to prevent oil from being swept underneath the boom
- Connectors and one or more tension members to support the entire boom.

Booms are constructed in sections, usually 15 or 30 m long, with connectors installed on each end so that sections of the boom can be attached to each other, towed, or anchored. A section of a typical boom is shown in Fig. 43.6 [43.1].

Boom fabrics consist of various fiber types such as nylon, polyester, aramid (kevlar), or blends of the same. These are then saturated or coated with polymeric coatings of various types ranging from polyvinyl chloride (PVC), polyurethane, nitrile, or blends of those materials to provide an impervious layer over the woven base fabric.

Most booms are also fitted with one or more tension members which run along the bottom of the boom and reinforce it against the horizontal load imposed by waves and currents. Tension members are usually made of steel cables or chains but sometimes consist of nylon or polyester webbing or ropes. The boom fabric itself is not strong enough to withstand the powerful forces to which booms are subjected, except in protected waters. For example, the force on a 100 m long section of boom could be as much as 10 000 kg, de-

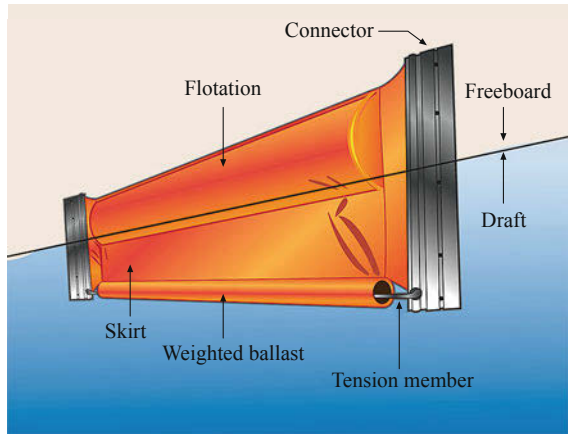


Fig. 43.6 Boom components (after [43.1])

pending on sea conditions and the construction of the boom.

The three basic types of booms are fence and curtain booms, which are common, and shoreline seal booms, which are relatively rare. Booms are also classified according to where they are used – offshore, inshore, harbor, and river booms, based on their size and ruggedness of construction. The fence boom is constructed with a freeboard member above the float. Although relatively inexpensive, these booms are not recommended for use in high winds or strong water currents. Curtain booms are constructed with a skirt below the floats and no freeboard member above the float.

The characteristics of booms that are important in determining their operating ability are the buoyancy-to-weight ratio or reserve buoyancy, the heave response, and the roll response [43.15–17]. The buoyancy-to-weight ratio or reserve buoyancy is determined by the amount of flotation and the weight of the boom. This means that the float must provide enough buoyancy to balance the weight of the boom with the force exerted by currents and waves, thereby maintaining the boom's stability. The greater a boom's reserve buoyancy, the greater its ability to rise and fall with the waves. The heave response is the boom's ability to conform to steep waves. It is indicated by the reserve buoyancy and the flexibility of the boom. A boom with good heave response will move with the waves on the surface of the water and not be alternately submerged and thrust out of the water by the wave action. The roll response refers to the boom's ability to remain upright in the water and not roll over.

Booms are used primarily to contain oil, although they are also used to deflect oil. When used for containment, booms are often arranged in a U, V, or J configuration. The U configuration is the most common and is achieved by towing the boom behind two vessels, anchoring the boom, or by combining these two techniques. The U shape is created by the current pushing against the center of the boom. The critical requirement is that the current in the apex of the U does not exceed $0.5\,\mathrm{m\,s^{-1}}$ or 1 kn, which is referred to as the critical velocity. This is the speed of the current flowing perpendicular to the boom, above which oil will be lost from the boom. Booms are also used in a *sweep* configuration to either deflect oil or contain it for pickup by skimmers. The sweep is held away from the vessel by a fixed arm and the boom allowed to form a U shape. A skimmer is usually placed in the U or is sometimes fixed in the vessel's hull and the oil is deflected to this position.

If used in areas where the currents are likely to exceed $0.5\,\mathrm{m\,s^{-1}}$ or 1 kn, such as in rivers and estuaries, booms are often used in the deflection mode. The boom is then deployed at various angles to the current so that the critical velocity is not exceeded. The oil can then be deflected to areas where it can be collected.

43.6.2 Boom Failures

A boom's performance and its ability to contain oil are affected by water currents, waves, and winds. Either alone or in combination, these forces often lead to boom failure and loss of oil. Eight common ways in which booms fail are discussed here. Some of these are illustrated in Fig. 43.7 [43.1].

Entrainment Failure

This type of failure is caused by the speed of the water current and is more likely to happen with a lighter oil. When oil is being contained by a boom in moving water, if the current is fast enough, the boom acts like a dam and the surface water being held back is diverted downward and accelerates. The resulting turbulence causes droplets to break away from the oil that has built up in front of the boom, referred to as the oil headwave, pass under the boom and resurface behind it. The water speed at which the headwave becomes unstable and the oil droplets begin to break away is referred to as the critical velocity. It is the speed of the current flowing perpendicular to the boom, above which oil losses occur. For most booms riding perpendicular to the current, this critical velocity is approximately $0.5\,\mathrm{m\,s^{-1}}$ (about 1 kn).

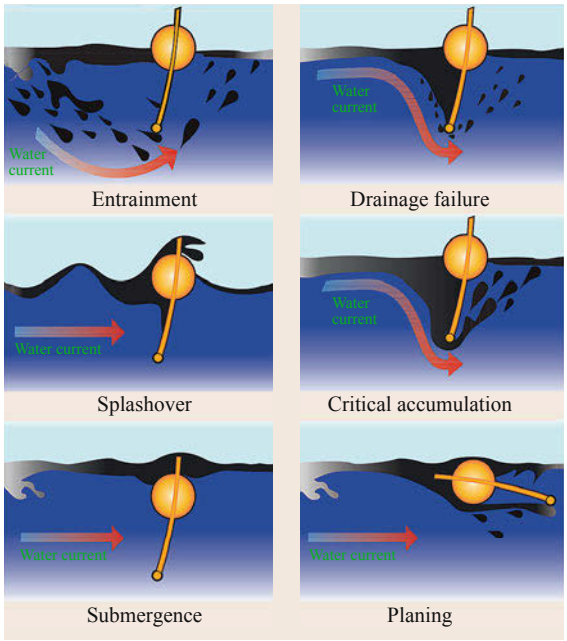


Fig. 43.7 Boom failure modes (after [43.1])

Drainage Failure

Similar to entrainment, this type of failure is related to the speed of the water current, except that it affects the oil directly at the boom. After critical velocity is reached, large amounts of the oil contained directly at the boom can be swept under the boom by the current.

Critical Accumulation

This type of failure usually occurs when heavier oils, which are not likely to become entrained in water, are being contained. Heavier oils tend to accumulate close to the leading edge of the boom and are swept underneath the boom when a certain critical accumulation point occurs. This accumulation is often reached at current velocities approaching the critical velocities, but can also be reached at lower current velocities.

Splashover

This failure occurs in rough or high seas when the waves are higher than the boom's freeboard and oil splashes over the boom's float or freeboard member. It can also occur as a result of extensive oil accumulation in the boom compared to the freeboard.

Submergence Failure

This type of failure occurs when water goes over the boom. Often the boom is not buoyant enough to follow the wave motion and some of the boom sinks below the water line and oil passes over it. Submergence failure is usually the result of poor heave response. Failure due to submergence is not that common as other forms of failure, such as entrainment, usually occurs first.

Planing

Planing occurs when the boom moves from its designed vertical position to almost a horizontal position on the water. Oil passes over or under a planing boom. Planing occurs if the tension members are poorly designed and do not hold the boom in a vertical position or if the boom is towed in currents exceeding the critical velocity.

Structural Failure

This occurs when any of the boom's components fail and the boom lets oil escape. Sometimes structural failure is so serious that the boom is carried away by the current. Floating debris, such as logs and ice, can contribute to structural failure.

Shallow Water Blockage

This type of failure occurs when rapid currents form under a boom when it is used in shallow waters. With the boom acting like a dam, the flow of water under it increases and oil is lost in several of the ways already described. Shallow water is probably the only situation in which a smaller boom might work better than a larger one. It should be noted, however, that booms are not often used in shallow water.

43.6.3 Sorbent Booms and Barriers

Sorbent booms are specialized containment and recovery devices made of porous sorbent material such as woven or fabric polypropylene, which absorbs the oil while it is being contained. Sorbent booms are used when the oil slick is relatively thin, that is, for the final polishing of an oil spill, to remove small traces of oil or sheen, or as a backup to other booms. Sorbent booms are often placed off a shoreline that is relatively unoiled or freshly cleaned to remove traces of oil that may re-contaminate the shoreline. They are not absorbent enough to be used as a primary countermeasure technique for any significant amount of oil.

Sorbent booms require considerable additional support to prevent breakage under the force of strong water currents [43.18]. They also require some form of flotation so they do not sink once saturated with oil and water.

43.7 Oil Recovery on Water

43.7.1 Skimmers

Skimmers are mechanical devices designed to remove oil from the water surface. Skimmers are classified according to the area where they are used, for example, inshore, offshore, in shallow water, or in rivers and by the viscosity of the oil they are intended to recover, that is heavy or light oil [43.19]. Skimmers are available in a variety of forms, including independent units built into a vessel or containment device and units that operate in either a stationary or mobile (advancing) mode. Some skimmers have storage space for the recovered oil and some of these also have equipment such as separators to treat the recovered oil.

The effectiveness of a skimmer is rated according to the amount of oil that it recovers, as well as the amount of water picked up with the oil. Removing water from the recovered oil can be as difficult as the initial recovery. Effectiveness depends on a variety of factors including the type of oil spilled, the properties of the oil such as viscosity, the thickness of the slick, sea conditions, wind speed, ambient tempera-

ture, and the presence of ice or debris. Most skimmers function best when the oil slick is relatively thick. The oil is usually collected in booms before skimmers are effective.

Weather conditions at a spill site have a major effect on the efficiency of skimmers. All skimmers work best in calm waters. Depending on the type of skimmer, most will not work effectively in waves greater than 1 m or in currents exceeding 1 kn. Most skimmers do not operate effectively in waters with ice or debris such as branches, seaweed, and floating waste. Very viscous oils, tar balls, or oiled debris can clog the intake or entrance of skimmers and make it impossible to pump oil from the skimmer's recovery system. Skimmers are also classified according to their basic operating principles [43.1]:

- Oleophilic surface skimmers
- Weir skimmers
- Suction skimmers or vacuum devices
- Elevating skimmers.

Oleophilic Surface Skimmers

Oleophilic surface skimmers, sometimes called sorbent surface skimmers, use a surface to which oil can adhere to remove the oil from the water surface. This oleophilic surface can be in the form of a disk, drum, belt, brush, or rope, which is moved through the oil on the top of the water. A wiper blade or pressure roller removes the oil and deposits it into an onboard container or the oil is directly pumped to storage facilities on a barge or on shore. The oleophilic surface itself can be steel, aluminum, fabric, or plastics

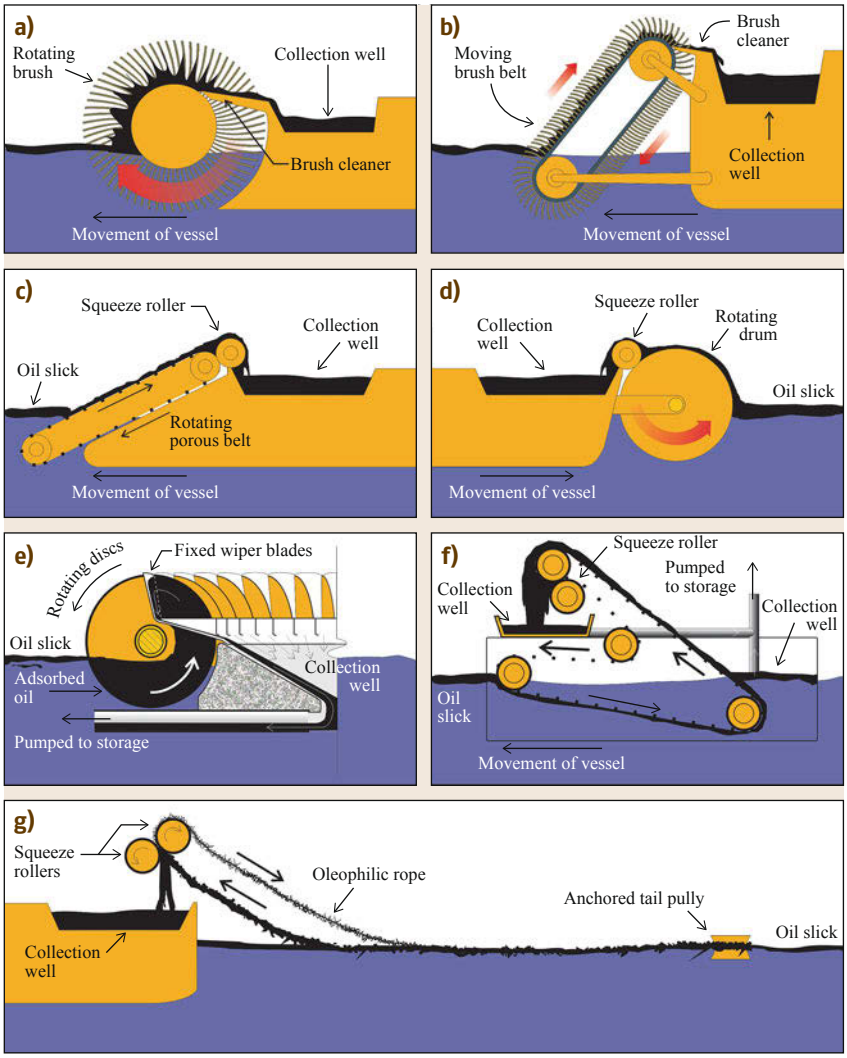


Fig. 43.8a–g Schematics of oleophilic skimmers: **(a)** brush drum skimmer, **(b)** brush belt skimmer, **(c)** adsorbent belt skimmer, **(d)** drum skimmer, **(e)** disk skimmer, **(f)** inverted belt skimmer, **(g)** rope skimmer

Fig. 43.9 Photograph of an oleophilic drum skimmer operating in a nearshore area (courtesy of Lamor corporation)

such as polypropylene and polyvinyl chloride. Figure 43.8 shows schematics of several types of oleophilic skimmers.

Oleophilic skimmers pick up very little water compared to the amount of oil recovered, which means they have a high oil-to-water recovery ratio. They can operate efficiently on relatively thin oil slicks and are not as susceptible to ice and debris as some other types of skimmers.

The disk skimmer is a common type of oleophilic surface device. The disks are usually made of either polyvinyl chloride or steel. Disc skimmers work best with light crude oil and are well suited to working in small waves and among weeds or debris. New grooved disks have proved to be useful for recovering heavier oils.

The drum skimmer is another type of oleophilic surface skimmer. The drums are made of either a proprietary polymer or steel. The drum skimmer works relatively well with fuels and light crude, but is less effective with heavy oils. Drum skimmers are often smaller in size like the disk skimmer. One such skimmer is shown in Fig. 43.9.

Belt skimmers are constructed of a variety of oleophilic materials ranging from fabric to conveyor belting. Most belt skimmers function by lifting oil up from the water surface to a recovery well. Belt skimmers of all types work best with heavier crudes and some are specially constructed to recover tar balls and very heavy oils. Belt skimmers are large and usually built into specialized cleanup vessels.

Brush skimmers use plastic brushes attached to drums, chains, or belts to recover the oil from the water surface. The oil is usually removed from the brushes by wedge-shaped scrapers. Brush skimmers are particularly useful for recovering heavier oils, but are less effective for fuels and light crudes.

Rope skimmers remove oil from the water surface with an oleophilic rope of a polymer material, usually polypropylene. Some skimmers have one or two long ropes which are held in the slick by a floating, anchored pulley. Others use a series of small ropes that hang down to the water surface from a suspended skimmer body.

Weir Skimmers

Weir skimmers are a major group of skimmers that use gravity to drain the oil from the surface of the water into a submerged holding tank. In their simplest form, these devices consist of a weir or dam, a holding tank, and a connection to an external or internal pump to remove the oil. Many different models and sizes of weir skimmers are available.

Most weir skimmers are mounted at the center of 3 or 4 floats which keep the weir edge at the water–oil interface. Weir skimmers do not work well in ice and debris or in rough waters and they are not effective for very heavy oils or tar balls. Weir skimmers are economical, however, and they can have large oil recovery capacities.

Suction or Vacuum Skimmers

Suction or vacuum skimmers use a vacuum to remove oil from the water surface. Often the skimmer is only a small floating head connected to an external source of vacuum, such as a vacuum truck. Suction skimmers are similar to weir skimmers in that they sit on the water surface, generally use an external vacuum pump system such as a vacuum truck and are adjusted to float at the oil–water interface.

Elevating Skimmers

Elevating skimmers use conveyors to lift oil from the water surface into a recovery area. A paddle belt or wheel or a conveyor belt with ridges is adjusted to the top of the water layer and oil is moved up the recovery device on a plate or another moving belt. The operation is similar to removing liquid from a floor with a squeegee. The oil is usually removed from the conveyer by gravity.

Other Devices

Many other devices are used to recover oil. Several skimmers combine some of the principles of operation already discussed. For example, one skimmer uses an inverted belt both as an oleophilic skimmer and a submersion skimmer.

43.7.2 Skimmer Performance

A skimmer's performance is affected by a number of factors including the thickness of the oil being recov-

ered, the extent of weathering and emulsification of the oil, the presence of debris and weather conditions at the time of recovery operations [43.20].

A skimmer's overall performance is usually determined by a combination of its recovery rate and the percentage of oil recovered. The recovery rate is the volume of oil recovered under specific conditions. It is measured as volume per unit of time, $m^3\,h^{-1}$ and is usually given as a range. If a skimmer takes in a lot of water, it is detrimental to the overall efficiency of an oil spill recovery operation. The summary results of performance testing on various types of skimmers are given in Table 43.5.

In addition to these characteristics, other important measures of a skimmer's performance include the amount of emulsification caused by the skimmer, its ability to deal with debris, ease of deployment, ruggedness, applicability to specific situations, and reliability.

43.7.3 Sorbents

Sorbents are materials that recover oil through either absorption or adsorption. They play an important role in oil spill cleanup and are used in the following ways:

- To clean up the final traces of oil spills on water or land
- As a backup to other containment means, such as sorbent booms
- As a primary recovery means for very small spills
- As a passive means of cleanup.

An example of such passive cleanup is when sorbent booms are anchored off lightly oiled shorelines to absorb any remaining oil released from the shore and prevent further contamination or re-oiling of the shoreline.

Sorbents can be natural or synthetic materials. Natural sorbents are divided into organic materials, such as peat moss or wood products and inorganic materials, such as vermiculite or clay. Sorbents are available in a loose form, which includes granules, powder, chunks, and cubes, often contained in bags, nets, or socks. Sorbents are also available formed into pads, rolls, blankets and pillows. Formed sorbents are also made into sorbent booms and sweeps.

The use of synthetic sorbents in oil spill recovery has increased in the last few years. These sorbents are often used to wipe other oil spill recovery equipment, such as skimmers and booms, after a spill cleanup operation. Sheets or rolls of sorbent are often used for this purpose. The capacity of a sorbent depends on the amount of surface area to which the oil can adhere as well as the type of surface. A fine porous sorbent with

Table 43.5 Performance of typical skimmers

Skimmer type	Recovery rate [$m^3\,h^{-1}$] for given oil type[a]				Percent oil[b]
	Diesel	Light crude	Heavy crude	Bunker C	
Oleophilic skimmers					
Small disk	0.4–1	0.2–2			80–95
Large disk		10–20	10–50		80–95
Brush drum	0.2–0.8	0.5–20	0.5–2	0.5–2	80–95
Brush belt	0.4–1	15–30	1–10	1–10	80–95
Large drum		10–30			80–95
Small drum	0.5–5	0.5–5			80–95
Large belt	1–5	1–20	3–20	3–10	75–95
Inverted belt		10–30			85–95
Rope mop		2–20	2–10		80–95
Weir Skimmers					
Small weir	0.2–10	0.5–5	2–20		20–80
Large weir		30–100	5–10	3–5	50–90
Advancing weir	1–10	5–30	5–25		30–70
Elevating skimmers					
Paddle conveyer		1–10	1–20	1–5	10–40
Suction skimmers					
Small	0.3–1	0.3–2			3–10
Large trawl unit		2–40			20–90
Large vacuum unit		3–20	3–10		10–80

[a]Recovery rate depends on the thickness of the oil, type of oil, sea conditions, etc.
[b]This is the percentage of oil in the recovered product

many small capillaries has a large amount of surface area and is best for recovering light crude oils or fuels. Sorbents with a coarse surface are best for cleaning up a heavy crude oil or Bunker.

The performance of sorbents is measured in terms of total oil recovery and water pickup, similar to skimmers [43.21]. Oil recovery is the weight of a particular oil recovered compared to the original weight of the sorbent. For example, highly efficient synthetic sorbent may recover up to 30 times its own weight in oil and an inorganic sorbent may recover only twice its weight in oil. The amount of water picked up is also important. Ideally a sorbent would not recover any water.

43.8 Separation, Pumping, Decontamination, and Disposal

43.8.1 Temporary Storage

After oil is recovered, sufficient storage space must be available for the recovered product. The recovered oil often contains large amounts of water and debris which increase the amount of storage space required.

Several types of specially built tanks are available to store recovered oil [43.1]. Flexible portable tanks, often constructed of plastic sheeting and a frame, are the most common type of storage used for spills recovered on land and from rivers and lakes. These are available in a range of sizes from approximately 1 to 100 m^3 and require little storage space before assembly. Most of these types of tanks do not have a roof, however, so rain or snow can enter the tank and vapors can escape. Rigid tanks, which are usually constructed of metal, are also available and are often used at sea. Pillow tanks, constructed of polymers and heavy fabrics, are usually used to store oil recovered on land. Oil recovered on land is often stored in stationary tanks built for other purposes, and in dump trucks and modular containers, lined with plastic. Recovered oil can also be temporarily stored in pits or berms lined with polymer sheets, although this open type of storage is not suitable for volatile oils.

Oil recovered at sea is often temporarily stored in barges. Many cleanup organizations have barges that are used solely for storing recovered oil and lease barges for use at larger spills.

43.8.2 Pumps

Pumps play an important role in oil spill recovery. They are an integral part of most skimmers and are also used to transfer oil from skimmers to storage tanks. Pumps used for recovered oil differ from water pumps in that they must be capable of pumping very viscous oils and dealing with water, air, and debris [43.1]. The three basic types of pumps used for pumping oil recovered from spills are centrifugal pumps, vacuum systems, and positive displacement pumps.

Centrifugal pumps have a spinning vane which moves the liquid out of a chamber by centrifugal force. These pumps, which are regularly used for pumping water and wastewater, are not designed for use with oil and are generally not capable of dealing with material more viscous than light crudes. They are economical and universally available.

Vacuum systems consist of vacuum pumps and tanks mounted on a skid or truck. The vacuum pump creates a vacuum in the tank and the oil moves directly through a hose or pipe to the tank from the skimmer or the source of the oil. Vacuum systems can handle debris, viscous oils, and the intake of air or water.

Positive displacement pumps are often built directly into skimmers to recover more viscous oils. These pumps have a variety of operating principles, all of which have a common scheme. Oil enters a chamber in the pump where it is pushed by a moving blade, shoe, or piston to the exit of the pump. The oil and other material with it must move through the chamber because there is no alternative passage, thus the name *positive displacement*.

The screw or auger pump is a common type of positive displacement pump. The oil drops into part of the screw and is carried to an output. Wiper blades remove oil from the auger flights to prevent it from remaining on the auger shaft. The screw pump can deal with very viscous oils and is often built into skimmers.

The diaphragm pump uses a flexible plate or diaphragm to move oil from a chamber. This type of pump usually requires a valve, which limits its use to material that can pass through the valve making it unsuitable for oil containing debris.

A vane pump, which uses a movable metal or polymer plate to move oil in a chamber, functions in a manner similar to a centrifugal pump, but with positive displacement.

A piston-like plunger in the sliding shoe pump moves oil along between the input and output ports. This pump does not require valves, although certain models do include them. The piston pump is similar to the sliding shoe pump except that oil is simply pushed out of the cylinder from the input valve to the output valve. Both sliding shoe and piston pumps can handle viscous oils, but generally cannot handle debris.

The progressive cavity pump uses a rotating member inside a molded cylinder that together form a cavity that moves from input to output as the center is rotated.

The performance of pumps is usually measured in terms of the volume displaced per unit of time at a given viscosity, suction head, and pressure head. The suction head is the maximum height that a pump can draw the target liquid and the pressure head is the maximum height that a pump can push the liquid. Other important factors to consider are the pump's ability to deal with emulsions and debris and the degree of emulsion formation that takes place in the pump itself.

43.8.3 Separators

As all skimmers recover some water with the oil, a means to separate oil and water is usually required [43.22]. The oil must be separated from the recovery mixture for disposal, recycling, or direct reuse by a refinery. Sometimes settling tanks or gravity separators are incorporated into skimmers, but separators are more often installed on recovery ships or barges. Portable storage tanks are often used as separators, with outlets installed on the bottom of the tanks so that water that has settled to the bottom of the tank can be drained off, leaving the oil in the tank. Vacuum trucks are also used in this way to separate oil and water. Screens or other devices for removing debris are also incorporated into separators.

A gravity separator is the most common type of separator. In its simplest form, this consists of a large holding tank in which the oil and water mixture is held long enough for the oil to separate by gravity alone. When inflow volumes are large, it can be difficult to find large enough separators to provide the long residence times required.

Separators are often made with baffles or other interior devices that increase the residence time and thus the degree of separation. The parallel plate separator is a form of gravity separator. Many parallel plates are placed perpendicular to the flow, creating areas of low water turbulence where drops of oil can re-coalesce from the water and rise to the surface.

Centrifugal separators have spinning members that drive the heavier water from the lighter oil, which collects at the center of the vessel. These separators are very efficient but have less capacity than gravity separators and cannot handle large debris. Centrifugal separators are now becoming more common. Centrifugal separators are best suited to constant amounts of oil and water.

As emulsions are not broken down in separators, emulsion-breaking chemicals are often added to the recovered mixture before it enters the separator. Heating the emulsions to 80 or 90 °C usually results in separation and the water can then be removed, although this process uses a lot of energy.

Separator performance is measured by the water removal efficiency and the throughput volume. Important factors affecting performance include the ability to handle small debris (larger debris is usually removed) and a wide variety of oil and water ratios, with oil content and flow rate sometimes changing suddenly.

43.8.4 Decontamination

Equipment and vessels used during spills often become contaminated with oil. Before transporting this equipment further, it is decontaminated. This typically involves removal to a lined area, a high-pressure wash and treatment of the recovered water. Special areas are prepared for the decontamination of vessels, booms or skimmers. Large vessels, must of necessity, be decontaminated on the sea and this involves circling the vessel with booms and recovering the oil released from the vessel. Often lightly contaminated vessels are cleaned by hand using sorbent cloths. The primary tool for oil removal is high-pressure water. The water released from decontamination is treated as recovered oil would be.

Workers must also decontaminate their boots and clothing if covered with oil. Stations are often set up very close to embarkation points to avoid carrying contamination further.

43.8.5 Disposal

Disposing of the recovered oil and oiled debris is one of the most difficult aspects of an oil spill cleanup operation [43.23]. Any form of disposal is subject to local, provincial, or state and federal legislation. Unfortunately, most recovered oil consists of a wide range of contents and cannot be classified as simply liquid or solid waste. The recovered oil may contain water which is difficult to separate from the oil and many types of debris, including vegetation, sand, gravel, logs, branches, garbage, and pieces of containment booms. This debris may be too difficult to remove and thus the entire bulk material may have to be disposed of.

Spilled liquid oil is sometimes directly reused either by reprocessing in a refinery or as a heating fuel. Some power plants and even small heating plants can use a broad spectrum of hydrocarbon fuels. Often the equipment at refineries cannot handle oils with debris, excessive amounts of water, or other contaminants and the cost of pretreating the oils can far exceed the value that might be obtained from using them.

Heavier oils are sometimes sufficiently free of debris to be used as road cover when mixed with regular asphalt. Recovered material from cleaning up beaches can sometimes be used in this way. If the material is of the right consistency, usually sand, the entire mixture might be mixed with road asphalt.

Incineration is a frequent means of disposal for recovered oil as large quantities of oil and debris can be disposed of in a relatively short time. Disadvantages are the high cost, which may include the cost of transporting the material to the facility. In addition, approval must be obtained from government regulatory authorities. Several incinerators have been developed for disposing of either liquid or solid materials, but these all require special permits or authority to operate. In remote locations, it may be necessary to burn oiled debris directly on the recovery sites without an incinerator because it is too bulky to transport to the nearest incinerator.

Contaminated beach material is difficult to incinerate because of the sand and gravel content. There are also machines to wash oily sand or gravel. The oil recovered from this process must be separated from the wash water and then disposed of separately. Burning debris on site is usually only applied to lightly oiled driftwood and special permission must be obtained from the appropriate authorities.

Oiled debris, beach material, and sorbents are sometimes disposed of at landfill sites. Legislation requires that this material not contain free oil that could migrate from the site and contaminate groundwater. Most governments have standard leachability test procedures which determine whether the material will release oil in a given period of time.

43.9 Spill-Treating Agents

43.9.1 Dispersants

A dispersant is a term used for chemical spill-treating agents that promote the formation of small droplets of oil which disperse into the top layer of the water column [43.24]. Dispersants contain surfactants that have molecules with both a water-soluble and oil-soluble component. Surfactants or surfactant mixtures used in dispersants have approximately the same solubility in oil and water, which temporarily stabilizes oil droplets in water so that the oil will disperse into the water column.

Two major issues associated with the use of dispersants – their effectiveness and the toxicity of the resulting oil dispersion in the water column – have generated controversy in the last 40 years. Some products used in the late 1960s and early 1970s were highly toxic and severely damaged the marine environment [43.25]. The effectiveness of a dispersant is determined by measuring the amount of oil that it puts into the water column and comparing it to the amount of oil that remains on the water surface. When a dispersant is working, a coffee-colored plume of dispersed oil appears in the water column and can be seen from ships and aircraft. This plume can take up to half an hour to form. If there is no such plume, it indicates little or no effectiveness. If only a white plume forms, this is the dispersant alone, also indicating little or no effectiveness.

Effectiveness is influenced by many factors, including the composition and degree of weathering of the oil, the amount and type of dispersant applied, sea energy, salinity of the water, and water temperature. The composition of the oil is the most important of these factors, followed closely by sea energy and the amount of dispersant applied. Below a certain oil thickness, the applied dispersant will interact with the water and not the oil.

Upon forming a dispersion in the water, the plume spreads. Some of the surfactant in the dispersant forming the dispersion, slowly leaches into the water. This slowly destabilizes the dispersion. After about 1–3 h some of the heavy components of the oil may resurface. During this time, there is a competition between re-dispersal of the droplets by wave action and the slow, continual rise of droplets. A significant portion of the slick may resurface within about a day. As the transport of the subsurface and the surface slick often differs, the resurfacing slick may be too thin to be seen and may be in a different location than the un-dispersed slick.

43.9.2 Surface-Washing Agents

Surface-washing agents or beach cleaners are different from dispersants although historically both products were sometimes referred to as dispersants. Surface-washing agents are effective in some situations, but they have not been widely accepted, partially because of the confusion with dispersants [43.26]. While toxicity has been a problem with some dispersants in the past in surface applications, testing has shown that the better surface-washing agents have less aquatic toxicity than dispersants and their use could prevent damage to shoreline species.

While both products contain surfactants, those in dispersants are equally soluble in both water and oil,

whereas in surface-washing agents, the surfactants are more soluble in water than in oil. Surface-washing agents operate by a different mechanism than dispersants. This mechanism is known as detergency and is similar to the use of detergents for washing clothes. Rather than causing the oil to disperse, surface-washing agents are intended to be applied to shorelines to release the oil from the surface. During low tide, the oil is sprayed with the surface-washing agent that is then left to soak for as long as possible. The oil is then washed off with a low-pressure water stream in an area that has been isolated using booms. Skimmers are typically used to remove the released oil.

43.9.3 Solidifiers

Solidifiers are intended to change liquid oil to a solid compound that can be collected from the water surface with nets or mechanical means. Solidifiers consist of cross-linking chemicals that couple two molecules or more, or polymerization catalysts that cause molecules to link to each other. Solidifiers usually consist of powders that rapidly react with and fuse the oil. Depending on the agent, about 10–40% by weight of the agent is required to solidify the oil, under ideal mixing conditions.

Solidifiers have not been used in the past for a number of reasons. Most importantly, if oil is solidified at sea, it makes recovery more difficult as skimming equipment, pumps, tanks, and separators are built to deal with liquids. Secondly, such a large amount of agent is required to solidify oil that it would be impossible to treat even a moderate spill.

43.9.4 Biodegradation Agents

Biodegradation agents are primarily used to accelerate the biodegradation of oil in the environment. They are used primarily on shorelines or land. They are not effective when used at sea because of the high degree of dilution and the rapid movement of oil. Biodegradation agents include bioenhancement agents which contain fertilizers or other materials to enhance the activity of hydrocarbon-degrading organisms, bioaugmentation agents which contain microbes to degrade oil and combinations of these two.

43.10 In-Situ Burning

In-situ burning is an oil spill cleanup technique that involves controlled burning of the oil at the spill site. The technique has been used at actual spill sites for some time, especially in ice-covered waters where the oil is contained by the ice. During the 2010 oil spill in the Gulf of Mexico, it was extensively used and contributed greatly to the removal of oil from the surface [43.27].

43.10.1 Advantages

Burning has some advantages over other spill cleanup techniques, the most significant of which is its ability to be a final solution and its capacity to rapidly remove large amounts of oil [43.28]. Burning can prevent oil from spreading to other areas and contaminating shorelines and biota. Burning oil is a final, one-step solution. When oil is recovered mechanically, it must be transported, stored, and disposed of, which requires equipment, personnel, time, and money. Often not enough of these resources are available when large spills occur. Burning generates a small amount of burn residue which can be recovered or further reduced through repeated burns.

In ideal circumstances, in-situ burning requires less equipment and much less labor than other cleanup techniques. It can be applied in remote areas where other methods cannot be used because of distance and lack of infrastructure. In some circumstances, such as when oil is mixed with or on ice, it may be the only available option for dealing with an oil spill.

Finally, while the efficiency of a burn varies with a number of physical factors, removal efficiencies are generally greater than those for other response methods such as skimming and the use of chemical dispersants. During several test and actual burns, efficiency rates as high as 95% were achieved.

43.10.2 Disadvantages

The most obvious disadvantage of burning oil is the large black smoke plume. The concerns revolve around toxic emissions from the smoke plume. These emissions are discussed in this section. The second disadvantage is that the oil will not ignite and burn quantitatively unless it is thick enough. Most oils spread rapidly on water and the slick quickly becomes too thin for burning to be feasible. Fire-resistant booms must typically be used to concentrate the oil into thicker slicks so that the oil can be burned.

43.10.3 Ignition and What Will Burn

Early studies of in-situ burning focused on ignition as being the key to successful burning of oil on wa-

ter [43.28]. It has since been found that ignition can be difficult, but only under certain circumstances. Studies have shown that almost any type of oil will burn on water or land if the slick can be ignited. Ignition may be difficult, however, at winds greater than 20 m s^{-1} (40 kn).

An important fact of in-situ burning is that oils can be more easily ignited if they are at least 1–3 mm thick and will continue to burn down to slicks about 0.5–1 mm thick. Sufficient heat is required to vaporize material so the fire will continue to burn. In very thin slicks, most of the heat is lost to the water and vaporization/combustion is not sustained. In general, heavy oils and weathered oils take longer to ignite than lighter oils. Often a primer, such as diesel fuel, may be needed to start the combustion of heavy oils. This is also the case for oil that contains water, although oil that is completely emulsified with water may not ignite at all. While the ignitability of emulsions with varying water concentrations is not well understood, oil containing some emulsion can be ignited and burned.

43.10.4 Burn Efficiency and Rates

Burn efficiency is measured as the percentage of starting oil removed compared to the amount of residue left. The amount of soot produced is usually ignored as it is a small amount and difficult to measure. Burn efficiency is largely a function of oil thickness. Oil thicker than about 2–3 mm can be easily ignited and will burn down to about 0.5–1 mm. If a 2 mm thick slick is ignited and burns down to 1 mm, the maximum burn efficiency is 50%. If a 20 mm thick pool of oil is ignited, however, and burns down to 1 mm, the burn efficiency is about 95%. Higher efficiency is usually achieved when towing a fire-resistant boom as the oil is continually driven to the rear, burned, and leaving only a small amount of residue unburned at the end.

Most of the residue from a burn is unburned oil with some lighter or more volatile products removed. The residue is adhesive and therefore can be recovered manually. Residue from burning heavier oils and from very efficient burns may sometimes sink in water, although this rarely happens as the residue, when cooled, is only slightly denser than sea water. Most oil pools burn at a rate of about 2–4 mm per minute which means that the depth of oil is reduced by 2–4 mm a minute. This burn rate varies with the type of oil, the degree of weathering, and the water content of the oil.

An optimal burn rate for diesel fuel and light crudes is about 5000 l m^{-2} of oil per day (100 gal ft^{-2} per day).

43.10.5 Use of Containment

For most crude oils, however, a sufficient thickness is only maintained for a few hours, at most, after the spill occurs [43.28]. Most oil on the open sea rapidly spreads to an equilibrium thickness, which is about 0.01–0.1 mm for light crude oils and about 0.05–0.5 mm for heavy crudes and residual oils. Containment is required to concentrate the oil so it is thick enough to ignite and burn efficiently.

Fire-resistant booms are also used by spill responders to isolate the oil from the source of the spill. When considering burning as a spill cleanup technique, the integrity of the source of the spill and the possibility of further spillage is always a priority. If there is any possibility that the fire could flash back to the source of the spill, such as an oil tanker, the oil is not ignited.

Special fire-resistant booms are available to contain oil when using burning as a spill cleanup technique. As they must be able to withstand heat for long periods of time, these booms are tested for fire resistance and for containment capability and designs are modified in response to test results. Fire-resistant booms require special handling because of their size and weight.

Fire-resistant booms are made of a variety of materials including ceramic, stainless steel, and water-cooled fibre glass. Fire booms must withstand the high temperatures, high heat-flux, as well as the mechanical forces during an oil spill burn. In addition, it is expected that a particular fire-resistant boom should withstand a multi-hour burn and be able to be re-used several times.

During the Deepwater Horizon spill in the Gulf of Mexico, more than 400 burns were carried out using fire-resistant booms. In this spill, some time had passed between the rise of the oil to the surface and its movement to the zone designated for in-situ burning. Fire resistant booms were both used to collect sufficient oil for a burn and also to isolate the burns from adjacent areas.

One approximately 200 m length of fire-resistant boom can contain about 50 000 l (11 000 gal) of oil, which, if it were a light crude, takes about 45 min to burn. In total, it would take about 3 h to collect this amount of oil, tow it away from the slick and burn it. One burn team, consisting of two tow vessels and one fire-resistant boom, could burn about three lots of oil per working shift. If there were two shifts each day, about 300 000 l of oil could be burned by each team in one day.

Lastly, it should be noted that thick oils, typically greater than about 5–10 mm, can be burned without containment on water. Such thick oils are typically

high-weathered crude oils or heavy fuel oils. Opportunities to perform uncontained burning are not frequent.

43.10.6 Emissions from Burning Oil

The possibility of releasing toxic emissions into the atmosphere or the water has created the biggest barrier to the widespread use and acceptance of burning oil as a spill countermeasure. Some atmospheric emissions of concern are particulate matter precipitating from the smoke plume, combustion gases, and unburned hydrocarbons [43.28]. While soot particles consist primarily of carbon particles, they also contain a number of absorbed chemicals. The residue left at the burn site is also a matter of concern. Possible water emissions include sinking or floating burn residue and soluble organic compounds.

Extensive studies have been conducted recently to measure and analyze all these components of emissions from oil spill burns [43.28]. Most burns produce an abundance of particulate matter. Particulate matter at ground level is a health concern close to the fire and under the plume, although concentrations decline rapidly downwind from the fire. The greatest concern is the smaller or respirable particles which are 2.5 μm or less in size. Concentrations at ground level (1 m) can still be above normal health concern levels (35 $\mu g\,m^{-3}$) as far downwind as 500 m from a small crude oil fire, such as from the amount of oil that could be contained in a 500 m long boom.

Polyaromatic hydrocarbons, or PAHs, are a primary concern in the emissions from burning oil, both in the soot particles and as a gaseous emission. All crude oils contain PAHs, varying from as much as 9% down to about 0.001%. Most of these PAHs are burned to fundamental gases except those left in the residue and the soot. The amount of residue left from a crude oil fire varies but generally ranges from 1 to 10%. It has been found that, compared to the original oil, the soot from several experimental burns contained a similar concentration of some PAHs of higher molecular weight and lower concentrations of PAHs of lower molecular weight. This could be a concern as the higher molecular weight PAHs are generally more toxic. This is offset, however, by the fact that in all cases the overall amount of PAHs in the soot and residue is much less than in the original oil.

The amount of soot produced by in-situ oil fires is not known exactly, although estimates vary from 0.5 to 3% of the original oil volume.

43.11 Shoreline Cleanup and Restoration

43.11.1 Fate and Behavior on Shorelines

The fate and behavior of oil on shorelines are influenced by many factors [43.29]. These factors include the type and amount of oil, the degree of weathering of the oil, both before it reaches the shoreline and while on the shoreline, the temperature, the state of the tide when the oil washes onshore, the type of beach substrate, its material composition, the type and sensitivity of biota on the beach, and the steepness of the shore.

The extent that an oil penetrates and spreads, its adhesiveness and how much the oil mixes with the type of material on the shoreline are all important factors in terms of cleanup. Cleanup is more difficult if the oil penetrates deeply into the shoreline. Penetration varies with the type of oil and the type of material on the shoreline. For example, oil does not penetrate much into fine beach material such as sand or clay, but will penetrate extensively on a shore consisting of coarse boulders. A very light oil such as diesel on a cobble beach can penetrate to about a meter under some conditions and is difficult to remove. On the other hand, a weathered crude deposited on a fine sand beach can remain on the surface indefinitely and is removed fairly easily using mechanical equipment.

The extent of oil coverage often depends on the stage of the tide when the oil is deposited on the shoreline. At high tide, oil can be deposited above the normal tide line and often spreads over a broad intertidal area. The least amount of oiling occurs when the oil is deposited on the shoreline during the falling tide, although this is less likely to occur as the water is moving away from the shoreline. The nature of the intertidal zone, for example, its composition and slope, will often dictate the fate of the oil. If large amounts of oil are not retained in the intertidal zone, then the oil will have less impact on the area.

The fate of oil on shorelines also depends on the wave regime. Oil can be removed and carried away by energetic waves within days whereas it can remain for decades in sheltered areas. Beaches are dynamic environments that change in profile during seasonal storms. This can result in oil being buried on the beach in layers, often as deep as 1 m, or buried oil can be brought to the surface.

Oil stranded on shorelines, especially above the high tide line, weathers with time and becomes more

adhesive, viscous, and difficult to remove [43.30, 31]. As oil stranded above the high tide line is above the limit of normal wave action, physical removal can occur only during storm events.

Another mechanism that can significantly affect the fate of oil on shorelines is the mixing of the oil with beach material. Oil often mixes with sand and gravel on beaches and then weathers to form a hard, resilient material called asphalt pavement which is difficult to remove. This material may be only 1–30% oil by weight, which greatly increases the amount of material to be removed. Sometimes this stranded oil causes no environmental concerns because the oil is entirely bound and none is lost to the water or refloated, but there may be a concern with this oil being visible on the shoreline, depending on the location of the shoreline.

43.11.2 Types of Shorelines and Their Sensitivity to Oil

The type of shoreline is crucial in determining the fate and effects of an oil spill as well as the cleanup methods to be used [43.29]. In fact, the shoreline's basic structure and the size of material present are the most important factors in terms of oil spill cleanup. The types summarized here are:

- Bedrock
- Man-made solid structures
- Boulder beaches
- Pebble-cobble
- Mixed sand-gravel beaches
- Sand beaches
- Sand tidal flats
- Mud tidal flats
- Marshes.

Bedrock shorelines consist of rock that is largely impermeable to oil, although oil can penetrate through crevices or fractures in the rock. Oil is more likely to be deposited in the upper tidal zone. If the shore is exposed to wave action, a significant amount of oil is likely to be removed after each tidal cycle.

Shorelines consisting of man-made solid structures include retaining walls, harbor walls, breakwaters, ramps, and docks and are generally made of rocks, concrete, steel, and wood. This type of shoreline is usually considered impermeable to oil, although there are some types that are permeable and these may be considered similar to their natural counterparts.

Boulder beaches consist primarily of materials that are more than 256 mm in diameter. These beaches are not altered by any conditions other than ice, human activity, or extreme wave conditions. Boulder beaches often give way to mud or sand tidal flats in the lower intertidal zone. Because of the large spaces between individual boulders, oil can be carried down to the sediments and remain there for years.

Pebble-cobble beaches consist of materials ranging in size from 2 to 256 mm. Pebbles range in size from 4 to 64 mm and cobbles from 64–256 mm. Some fine materials may be present in the interstitial areas between pebbles and there may also be large boulders in the area. Oil readily penetrates pebble-cobble beaches through the open spaces between the rocks.

A mixed sediment beach consists of a variety of materials from 0.1 to 64 mm in size and possible cobbles up to 256 mm. These are sometimes called sand-gravel beaches and sometimes gravel beaches, because the larger gravel appears to predominate. Only lighter oils can penetrate sand-gravel beaches. For this reason, these beaches are not considered to be particularly sensitive to oil spills. Oil from spills can form asphalt pavement at the upper-tidal reaches. Oil residence times vary, but are generally shorter than other types of beaches.

Sand beaches are what most people envision as a beach. Sand is defined as a particle 0.1–2 mm in diameter, consisting of several different sizes and types of minerals. Coarse sand is usually defined as 0.5–2 mm in size and fine sand is less than 0.5 mm in size. On many coasts, sand beaches are often located between other types of beaches. Only lighter oils penetrate sand beaches and the residence time is likely to be short, except when oil is buried or carried to the upper tidal areas. Oil can easily become buried in sand and over time this can result in layers of sand and oil. In recreational areas, sand beaches are given a high cleanup priority if oiling of any type occurs.

Sand flats consist of material similar to sand beaches but are at shallow angles and never drain completely. They contain a lot of silt or very fine material. The surface layer of sand flats, which consists of a few centimeters, is dynamic and unstable. This surface layer is usually water-saturated and thus impermeable to oils. Sand tidal flats are difficult to access and cleanup is therefore limited.

Mud flats are similar to sand tidal flats in that they are at shallow angles and have a thin, mobile surface layer consisting of water-saturated mud that is impermeable to oil, although oil can penetrate through holes made by burrowing animals. Oil is likely to concentrate on the upper tidal zones. Mud flats are not accessible to vehicles or response personnel and thus cannot be readily cleaned.

Marshes are important ecological habitats that often serve as nurseries for marine and bird life in the area. Marshes range from fringing marshes, which are narrow areas beside a main water body, to wide salt marsh

Table 43.6 Cleanup techniques and shoreline types (after [43.29])

Shoreline type	Condition of the oil	Natural recovery	Flooding	Low-pressure cold water	Low-pressure warm water	Manual removal	Vacuums	Mechanical removal	Sorbents	Tilling/ aeration	Sediment reworking/ surf-washing	Cleaning agents
Bedrock	Fluid	+	+	+	▲	▲	▲		▲			■
	Solid				■	▲						+
Man-made	Fluid	+	+	+	▲	▲			▲			■
	Solid				■	▲						+
Boulder	Fluid	+	+	+	▲	▲			▲			■
	Solid				■							+
Pebble–cobble	Fluid	+	+	+		▲			▲	▲	▲	
	Solid					▲		▲				■
Mixed sand–gravel	Fluid	+	+	+		▲			▲	▲	▲	
	Solid					▲		+				■
Sand beach	Fluid	+	+	+		▲		+	▲	▲	▲	
	Solid					▲		+				
Sand tidal flats	Fluid	+	+	+		▲	▲		▲			
	Solid					▲		▲				
Mud tidal flats	Fluid	+	+	+			▲		▲			
	Solid							▲				■
Marshes	Fluid	+	+	+					▲			
	Solid	+	+	+		▲						■

+ Acceptable method, ▲ suitable method for small quantities, ■ conditional method, may only work under special circumstances

meadows. Salt marsh meadows often flood only during high tides in spring or during storm surges. Marshes are rich in vegetation that traps oil. Light oils can penetrate into marsh sediments through animal burrows or cracks. Heavier oils tend to remain on the surface and smother plants or animals. Oiled marshes, fresh or salt, may take years or even decades to recover. Marshes are difficult to access and entering them by foot or by vehicle can cause more damage than the oil itself.

43.11.3 Cleanup Methods

Many methods are available for removing oil from shorelines. Most of them are costly and take a long time to carry out. The selection of the appropriate cleanup technique is based on the type of substrate, the depth of oil in the sediments, the amount and type of oil and its present form/condition, the ability of the shoreline to support traffic, the environmental, human, and cultural sensitivity of the shoreline and the prevailing ocean and weather conditions. Some cleanup techniques suitable for use on the various types of shoreline are listed in Table 43.6 [43.29].

The primary objective of cleanup operations is to minimize the effects of the stranded oil and accelerate the natural recovery of affected areas. The length of time required to complete the cleanup is another important criterion when selecting a cleanup technique. The longer oil is on a beach, the harder it is to clean up. A method that removes most of the mobile oil rapidly is much better, in many circumstances, than a more thorough one that takes weeks to carry out. Time often dictates the cleanup method used.

43.11.4 Recommended Cleanup Methods

Some recommended shoreline cleanup methods are natural recovery, manual removal, flooding or washing, use of vacuums, mechanical removal, tilling and aeration, sediment reworking or surf washing and the use of sorbents or chemical cleaning agents [43.29]. Sometimes the best response to an oil spill on a shoreline may be to leave the oil and monitor the natural recovery of the affected area. This would be the case if more damage would be caused by cleanup than by leaving the environment to recover on its own.

Manual removal is the most common method of shoreline cleanup. Teams of workers pick up oil, oiled sediments, or oily debris with gloved hands, rakes, forks, trowels, shovels, or sorbent materials. It may also include scraping or wiping with sorbent materials or sifting sand to remove tar balls.

Flooding or washing shorelines are also common cleanup methods. Low-pressure washing with cool or lukewarm water causes little ecological damage and removes oil quickly. Warmer water removes more oil, but causes more damage. High pressures and temperatures cause severe ecological damage but might be safely used on man-made surfaces.

Low-pressure cool or warm water washing uses water at pressures less than about 200 kPa (50 psi) and temperatures less than about 30 °C. Flooding is a process in which a large flow or deluge of water is released on the upper portion of the beach.

Low-pressure washing and flooding are often combined to ensure that oil is carried down the beach to the water, where it can be recovered with skimmers.

References

43.1 M.F. Fingas: *The Basics of Oil Spill Cleanup* (Taylor and Francis, New York 2012)

43.2 ITOPF: *Response to Marine Oil Spills* (Witherby, Edinburgh 2012)

43.3 J. Hilyard (Ed.): *International Petroleum Encyclopedia* (Pennwell, Tulsa 2011)

43.4 D.S. Etkin: Spill occurrences: A world overview. In: *Oil Spill Science and Technology*, ed. by M. Fingas (Gulf, New York 2011) pp. 7–48

43.5 B. Hollebone: Measurement of oil physical properties. In: *Oil Spill Science and Technology*, ed. by M. Fingas (Gulf, New York 2011) pp. 63–86

43.6 National Research Council: *Oil in the Sea* (National Academy Press, Washington 2003)

43.7 M. Fingas: Introduction to oil spill modeling. In: *Oil Spill Science and Technology*, ed. by M. Fingas (Gulf, New York 2011) pp. 187–200

43.8 B. Fieldhouse, M. Fingas: Studies on water-in-oil products from crude oils and petroleum products, Mar. Pollut. Bull. **64**, 272–283 (2011)

43.9 J. Sun, A. Khelifa, X. Zheng, Z. Wang, S. Wong, L.L. So: Formation of oil-SPM aggregates under various mixing intensities, Proc. 33rd AMOP Tech. Sem., Vol. 1 (2010) pp. 145–158

43.10 R.C. Prince, K.M. McFarlin, J.D. Butler, E.J. Febbo, T.J. Nedwed: The primary biodegradation of dispersed crude oil in the sea, Chemosphere **90**(2), 521–526 (2013)

43.11 S.Z. Yang, H.J. Jin, Z. Wei, R. He, Y.J. Li, S.P. Yu: Bioremediation of oil spills in cold environments: A review, Pedosphere **19**, 371–381 (2009)

43.12 B. Lehr, D.S. Etkin: Ecological risk assessment modeling in spill response decisions, Proc. 35th AMOP Tech. Sem., Vol. 1 (2012) pp. 667–674

43.13 Z. Wang, M.F. Fingas: Oil spill identification, J. Chromatogr. A **843**, 369–411 (1999)

43.14 M. Fingas, C.E. Brown: Oil spill remote sensing: A review. In: *Oil Spill Science and Technology*, ed. by M. Fingas (Gulf, New York 2011) pp. 111–169

43.15 A. Amini, E. Bollaert, J.L. Boillat, A.J. Schleiss: Dynamics of low-viscosity oils retained by rigid and flexible barriers, Ocean Eng. **35**, 1479–1491 (2008)

43.16 R. Chebbi: Profile of oil spill confined with floating boom, Chem. Eng. Sci. **64**, 467–473 (2009)

43.17 S.-P. Zhu, D. Strunin: A numerical model for the confinement of oil spill with floating booms, Spill Sci. Technol. Bull. **7**, 249–255 (2002)

43.18 D. Cooper, K. Flood, C.E. Brown: Multi-track sorbent boom testing with loose sorbent material, Proc. 29th AMOP Tech. Sem., Vol. 1 (2006) pp. 173–194

43.19 N.P. Ventikos, E. Vergetis, H.N. Psaraftis, G. Triantafyllou: A high-level synthesis of oil spill response equipment and countermeasures, J. Hazard. Mater. **107**, 51–58 (2004)

43.20 A. Guarino, J.E. Delgado, W. Schmidt, M. Crickard, B.N. Midkiff: Development of skimmer testing protocol based on ASTM standards by Minerals Management Service and US Coast Guard at Ohmsett facility, Proc. 30th AMOP Tech. Sem., Vol. 1 (2007) pp. 371–386

43.21 D. Cooper, I. Gausemel: Oil spill sorbents: Testing protocol and certification listing program, Proc. Int. Oil Spill Conf., Vol. 1 (1993) pp. 549–551

43.22 A.B. Nordvik, J.L. Simmons, K.R. Bitting, A. Lewis, T. Strøm-Kristiansen: Oil and water separation in marine oil spill clean-up operations, Spill Sci. Technol. Bull. **3**, 107–122 (1996)

43.23 M. McDonagh, J. Abbott, R. Swannell, E. Gundlach, A. Nordvik: Handling and disposal of oily waste from oil spills at sea, Proc. Int. Oil Spill Conf., Vol. 1 (1995) pp. 589–593

43.24 M. Fingas: Oil spill dispersants: A technical summary. In: *Oil Spill Science and Technology*, ed. by M. Fingas (Gulf, New York 2011) pp. 435–582

43.25 J. Wise, J.P. Wise Sr.: A review of the toxicity of chemical dispersants, Rev. Environ. Health **26**, 281–300 (2011)

43.26 M. Fingas, B. Fieldhouse: Surface-washing agents. In: *Oil Spill Science and Technology*, ed. by M. Fingas (Gulf, New York 2011) pp. 683–711

43.27 N. Mabile: Controlled in-situ burning: Transition from alternative technology to conventional spill response option, Proc. 35th AMOP Tech. Sem., Vol. 1 (2012) pp. 584–605

43.28 M. Fingas: In-situ burning. In: *Oil Spill Science and Technology*, ed. by M. Fingas (Gulf, New York 2011) pp. 737–903

43.29 E.H. Owens, G. Sergy: *Field Guide for the Protection and Cleanup of Oiled Shorelines* (Environment Canada, Ottawa 2009)

43.30 E. Taylor, D. Reimer: Oil persistence on beaches in Prince William Sound – A review of SCAT surveys conducted from 1989 to 2002, Mar. Pollut. Bull. **56**, 458–474 (2008)

43.31 Y. Xia, M.C. Boufadel: Lessons from the Exxon Valdez oil spill disaster in Alaska, Disaster Adv. **3**, 270–273 (2010)

Part E

Ocean Re

Part E Ocean Renewable Energy

Ed. by Nikolaos I. Xiros

44 Marine Hydrokinetic Energy Resource Assessment
Manhar R. Dhanak, Dania Beach, USA
Alana E.S. Duerr, Englewood, USA
James H. VanZwieten, Boca Raton, USA

45 Ocean Wave Energy Conversion Concepts
Nikolaos I. Xiros, New Orleans, USA
Manhar R. Dhanak, Dania Beach, USA

46 Ocean Current Energy Conversion
Howard P. Hanson, Boca Raton, USA
James H. VanZwieten, Boca Raton, USA
Gabriel M. Alsenas, Boca Raton, USA

47 Harvesting Energy by Flow Included Motions
Michael M. Bernitsas, Ann Arbor, USA

48 Ocean Thermal Energy Conversion
Muthukamatchi Ravindran, Chennai, India
Raju Abraham, Chennai, India

49 Offshore Wind Energy
Mareike Strach-Sonsalla, Bremerhaven, Germany
Matthias Stammler, Hannover, Germany
Jan Wenske, Bremerhaven, Germany
Jason Jonkman, Golden, USA
Fabian Vorpahl, Osnabrück, Germany

In Part E, a range of contemporary ocean energy technologies are covered in six chapters, encompassing a resource assessment and an overview of wave energy, current energy, hydrokinetic energy, thermal energy, and last but not least, offshore wind.

Specifically, in Chap. 44, ocean energy resource assessment is given. Marine hydrokinetic energy includes energy due to waves, tides, and ocean currents. The characteristics of these forms of energy and the assessment of their potential for extraction are discussed briefly. Detailed consideration is given to the assessment of ocean current energy, including a case study of the resource assessment of the Florida Current. Estimates of global and local open ocean current resources are obtained based on data from an ocean model. The power densities of major western boundary currents are estimated, and the potential for development of ocean currents globally is assessed. The principal factors that govern economic viability of harnessing an ocean current at a location include the in-situ energy density, the distance of the location from the shore, and the local depth of the seafloor. A metrics based on these considerations is developed. Once potential sites have been identified, considerations would need to be given to regulatory and permitting requirements, including assessment of the potential impact on the environment and its ecosystems, and the impact on the energy resource itself; development of optimal design of devices for high performance; and mitigation of deployment and maintenance costs.

In Chap. 45 the most important ocean wave energy conversion concepts are presented. Basic concepts in primary ocean wave energy capture and its conversion to electricity using power takeoff systems are described. Several patented inventions in harnessing wave energy have been reported over the past 4 decades and many continue to be filed. They are typically based on these basic concepts in wave energy capture and appropriate power takeoff systems for conversion to electricity. An introduction to these methods is provided, and recent developments are discussed. Finally, an account of some of the well-known practical field applications of ocean wave energy technologies is presented in an appendix, with emphasis on commercial technologies.

In Chap. 46, ocean current energy conversion is investigated. Ocean currents in tidal flows and at the western boundaries of the ocean basins, as well as currents in rivers, can have as much potential for power production as the wind. With the increased interest in sources of renewable energy, realizing this potential has undergone a renaissance in recent years. This overview provides an introduction to ocean current energy conversion by emphasizing the contrasts between winds and currents and discussing some of the unique challenges associated with this untapped renewable energy source.

In Chap. 47, a detailed account of hydrokinetic energy conversion is given. Marine hydrokinetic (MHK) energy is clean, renewable, and available worldwide. It comes in two forms: vertical in waves and horizontal in currents, tides, and rivers. Apart from a few major ocean currents, most ocean currents have flow speeds of less than 3 kn, and most rivers have speeds of less than 2 kn, making harvesting of their MHK energy by steady-lift technologies (turbines) challenging. Horizontal MHK energy can also be harnessed using alternating-lift technologies. Fish utilize alternating lift to propel efficiently in water either as individuals or in schools. Engineered structures – bluff bodies such as circular cylinders and prisms, or slender bodies like hydrofoils – may generate alternating lift in quasi-steady uniform flows. When these structures have scale-relevant flexibility, severe flow-structure interaction (FSI) phenomena may be induced. In typical engineering applications, FSI phenomena are destructive and, thus, avoided by design or suppressed using excessive damping or appendages. If, instead, FSI are enhanced, they may result in vigorous flow-induced motion (FIM) of the body, leading to conversion of MHK energy to potential and kinetic energy in a mechanical oscillator. Hydrofoils can harvest MHK energy through flutter – a well-studied and understood form of instability. On the other hand, bluff bodies such as circular or rectangular cross-section cylinders may exhibit several forms of FIM, individually or in *schools*; these have been studied extensively but are still not well understood for either suppression or enhancement. Those FIMs are vortex-induced vibration (VIV), galloping, buffeting, and gap-flow in multi-body interactions. When enhanced, they convert MHK energy to mechanical energy with a high power density (power-to-weight ratio), even from low-speed horizontal flows. This chapter presents an overview of the concepts of alternating-lift technologies, the underlying physical principles, the available experimental and computational methods for studying the relevant FIM, the research challenges that have been overcome and those lying ahead, field-deployment progress, technology development, and benchmarking.

In Chap. 48, ocean thermal energy conversion is visited. Ocean thermal energy conversion (OTEC) utilizes the thermal gradient available in the ocean to operate a heat engine to produce work output. Even though the concept is simple and has been used for almost one century, during the last three decades it has gained momentum due to a worldwide search for

a clean, continuous energy source to replace the fossil fuels. There are technological hurdles to overcome to tap the immense potential of OTEC. However, the technology is still mature enough to establish commercial power plants. This chapter presents the technological and economical aspects of OTEC, with an overview of the current status. It also suggests some further research and technological guidance based on the Indian OTEC experiences for this promising technology for future.

Finally, in Chap. 49, an insight of offshore wind energy conversion is presented. In 1991, the Vindeby Offshore Wind Farm, the first offshore wind farm in the world, started feeding electricity to the grid off the coast of Lolland, Denmark. Since then, offshore wind energy has developed from this early experiment to a multibillion dollar market and an important pillar of the worldwide renewable energy production. Unit sizes grew from 450 kW at Vindeby to the 7.5 MW-class offshore wind turbines that are currently (by October 2014) in the prototyping phase. This chapter gives an overview of the state-of-the-art in offshore wind turbine technology and introduces the principles of modeling and simulating an offshore wind turbine (OWT). The OWT components rotor, nacelle, support structure control system, and power electronics are introduced, and current technological challenges are presented. The OWT system dynamics, as well as the environment (wind and ocean waves), are described from the perspective of OWT modelers and designers. Finally, an outlook on future technology is provided. The descriptions in this chapter are focused on a single OWT – more precisely, a horizontal-axis wind turbine – as a dynamic system. Offshore wind farms and wind farm effects are not described in detail in this chapter but an introduction and further references are given.

44. Marine Hydrokinetic Energy Resource Assessment

Manhar R. Dhanak, Alana E.S. Duerr, James H. VanZwieten

Marine hydrokinetic energy includes that due to waves, tides, and ocean currents. The characteristics of these forms of energy and the assessment of their potential for extraction are discussed briefly herein. Detailed consideration is given to the assessment of ocean current energy, including a case study of the resource assessment of the Florida Current. Estimates of global and local open ocean current resources are obtained based on data from an ocean model. The power densities of major western boundary currents are estimated and the potential for development of ocean currents globally is assessed. Principal factors that govern economic viability of harnessing an ocean current at a location include the in-situ power density, the distance of the location from the shore, and the local depth of the seafloor. A metrics based on these considerations is discussed. Once potential sites are identified, considerations would need to be given to regulatory and permitting requirements, including assessment of potential impact on the environment and its ecosystems, marine spatial planning and the impact on the energy resource itself; development of optimal design of devices for high performance; and mitigation of deployment and maintenance costs.

44.1 **Wave Energy Resource** 1100
44.2 **Tidal and Ocean Current Energy Resource** 1101
44.2.1 Tidal Currents 1102
44.2.2 Ocean Currents 1102
44.2.3 Power and Power Density of Tidal and Ocean Currents 1104
44.3 **Assessment of Global Ocean Current Resources** 1106
44.3.1 Gulf Stream Case Study 1110
44.4 **Other Considerations** 1113
References 1114

The energy in the oceans may be classed as hydrokinetic when it is in the form of waves, and tidal and ocean currents, and as potential energy when it in the forms of thermal and salinity gradients, and tidal water levels. The prevalence of these forms of energy is geographic specific. Major ocean currents are typically western boundary currents on the eastern coasts of the continents. Tidal energy typically represents significant ocean energy resources in the estuaries and channels at a number of locations worldwide, such as the northeast of United States or the western coast of Europe. Significantly high wave energy sources are typically found in mid to high latitudes (40–60°) along coastlines facing large open oceanic regions with significant wind fetch, such as the northwest coast of the US. Thermal gradients are significant at certain locations in low to mid latitudes (0–35°). The feasibility of cost-effectively harnessing any of these forms of energy from a specific region depend on a number of factors, including the distance of the region from the shore, the local water depth, and the potential impact of harnessing the energy on the environment and on the local stakeholders. Developing just a small fraction of the available wave energy resources could allow for millions of homes to be powered.

In this chapter, we provide a brief assessment of the various forms of marine hydrokinetic energy and provide details of the methodology for estimating the potential for energy extraction in the particular case of the hydrokinetic energy of an open ocean current, together with an overview of the world's major ocean currents as significant ocean energy resources. The ocean current energy potential at identified global locations is estimated using a computational ocean model. The model's accuracy in predicting the actual current conditions and hence the resource potential is discussed. In Sect. 44.1, the potential for wave energy is discussed; in Sect. 44.2, the potential for tidal and ocean currents is discussed; in Sect. 44.3, practical considerations in extracting ocean and tidal current energy are discussed; and in Sect. 44.4, a case study for assessing extractable ocean current energy resource is discussed.

44.1 Wave Energy Resource

Waves are significant sources of hydrokinetic energy. Typically, waves generated by winds are of interest here. Waves form in a complex evolving process involving fluctuations in air pressure over a range of frequencies induced by unsteady turbulent winds, shear layer instabilities, and interaction between waves [44.1–3]. The height, period, wavelength, and direction of propagation of waves are governed by the wind velocity, wind duration, fetch – the distance over water that the wind blows along a particular direction – and water depth and bottom topography. Large waves correspond to high winds operating over long duration over considerable fetch. The absence of significant wind strength, duration, or fetch curtails the growth of waves. Waves in a given area typically have a spectrum of heights, periods, and wavelengths, from very small, short capillary waves of periods $T < 0.1\,\mathrm{s}$ to chops of wave heights of $O(0.1-10\,\mathrm{m})$ and periods $1\,\mathrm{s} < T < 10\,\mathrm{s}$ to swells of wave heights of the same order but periods in the range $10\,\mathrm{s} < T < 30\,\mathrm{s}$. While chops are associated with local winds, have relatively short wavelength and are of short duration, swells are typically generated by distant storms, are narrow banded, have longer wavelengths and travel long distances; as waves disperse from the region of the storm, shorter waves get dissipated, while longer $O(100\,\mathrm{m})$, faster traveling, waves persist over great $O(1000-10\,000\,\mathrm{km})$ distances. The wave height of a group of waves is characterized by the significant wave height, H_s. For specific wind velocity, duration, and fetch, a threshold, referred to as *fully developed seas*, is reached, whereby additional action of the wind results in breaking of wave tops and formation of *whitecaps*. Wave-induced oscillatory motion in the water column is highest at the surface and diminishes rapidly with depth.

Wave energy propagates with the group velocity. The wave energy flux per unit along crest width (wave power density), which is the mean rate of transport of the wave energy through a vertical plane of unit span along a wave crest, for unidirectional waves is given in units of W/m by

$$P_\mathrm{h} = \rho g \int_0^\infty C_\mathrm{g} S(f)\,\mathrm{d}f\,, \tag{44.1}$$

where $S(f)$ is the wave spectrum and $C_\mathrm{g}(f,d)$ is the group velocity of the harmonic wave component of frequency f Hz in waters of depth d. In the case of directional spread, we consider $S(f) = \int_{-\pi}^{\pi} S(f,\theta)\,\mathrm{d}\theta$. According to linear wave theory,

$$C_\mathrm{g} = \frac{g \tanh kd}{4\pi f}\left(1 + \frac{2kd}{\sinh 2kd}\right)\,,$$
$$(2\pi f)^2 = gk \tanh kd\,. \tag{44.2}$$

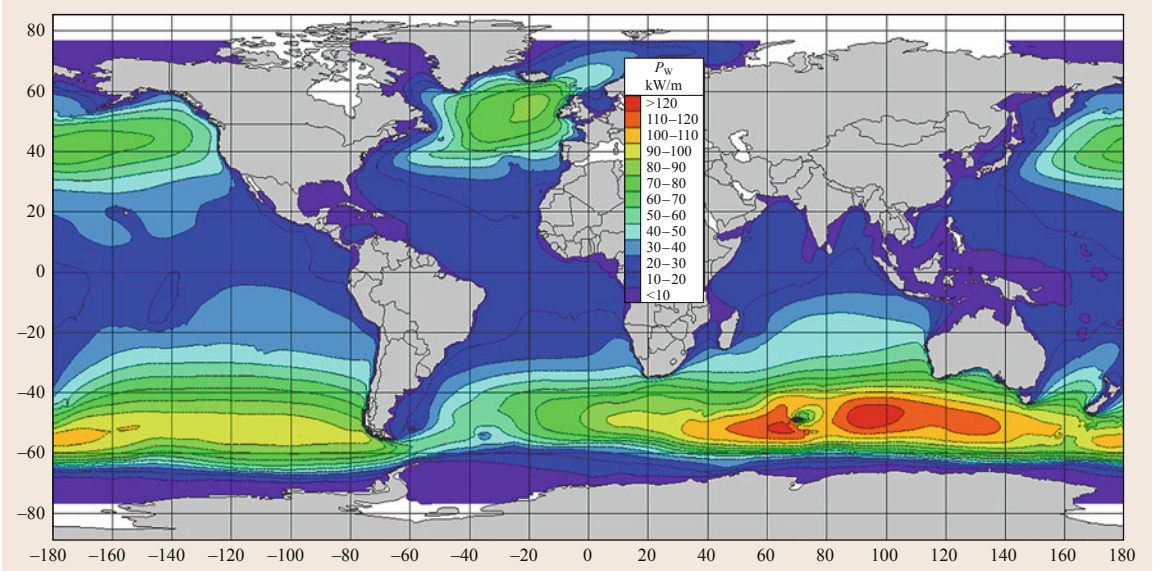


Fig. 44.1 Global distribution of annual mean wave power. Wave power density in kW/m is shown, based on NOAA WAVEWATCH-III (NWW3) wind–wave model (after [44.4])

In deep water, $C_g = \frac{g}{4\pi f}$, so that

$$P_h = \frac{\rho g^2}{4\pi} \int_0^\infty \frac{S(f)}{f}\,df = \frac{\rho g^2}{4\pi} m_{-1} = \frac{\rho g^2 H_0^2 T_e}{64\pi}\,, \tag{44.3}$$

where

$$H_0 = 4\sqrt{m_0} = 4\sqrt{\int_0^\infty S(f)\,df}$$

is the significant wave height and $T_e = \frac{m_{-1}}{m_0}$ is the wave energy period.

In shallow water, $C_g = \sqrt{gd}$ and $S(f) \to S_s(f)$, the shallow water spectrum, so that

$$P_h = \rho g \sqrt{gd} \int_0^\infty S_s(f)\,df = \frac{\rho g \sqrt{gd} H^2}{16}\,, \tag{44.4}$$

where H is the shallow-water wave height. As the waves approach the shore they undergo shoaling and refraction. Equating (44.3) and (44.4) gives the shoaling coefficient $\kappa_s = H/H_0 = (gT_e^2/(16\pi^2 d))^{\frac{1}{4}}$. In practice, some of the wave energy may be lost on approach to the shoreline through wave breaking.

As an example, consider a swell of wave energy period $T_e = 10\,s$ and wave height $H_0 = 0.5\,m$. For such a wave, (44.3) implies that the available energy is

$$P_h = \frac{\rho g^2 (2.5)}{64\pi} \simeq 1.2\,kW/m\,.$$

Figure 44.1 shows the global distribution of the annual mean available wave energy resource. As can be seen, major sources of wave energy along the continental coastlines lie on the western boundaries of the land masses in mid to high (40–60°) latitudes north and south of the Equator. In particular, northwest USA, western Europe, southern Chile, and south and west Australia have high wave power densities along their coastlines. *Gunn* and *Stock-Williams* [44.5] use NOAA Wavewatch III model [44.6] data to estimate that worldwide the total mean wave energy resource is over 2 TW or 17 532 TWh/year, with the mean resource being higher in winter than in summer in either hemisphere; they estimate that accounting for the effect of directional spread implies around 15% reduction in this estimate. Other estimates vary between 1.3 TW or 11 396 TWh/year [44.7] and 3 TW or 26 297 TWh/year [44.8]. *Gunn* and *Stock-Williams* [44.5] found that Australia (280 GW), USA (223 GW) and Chile (194 GW) have the highest potential for wave energy conversion.

The fraction of this available wave energy resource that can be extracted is based on a number of factors, including available conversion technologies, their efficiencies and packing capacities, economic viability, and environmental impact considerations. Using an example of an array of Pelamis devices, *Gunn* and *Stock-Williams* [44.5] suggest that based on current technologies, 4.6% of the 2 TW of available global wave energy may be extractable. An Electric Power Research Institute (EPRI) report [44.9] finds that technically recoverable resources for electric generation from waves around the US coastline is approximately 1170 TWh/year (134 GW), which is almost one third of the 4000 TWh of annual electricity usage in the USA.

44.2 Tidal and Ocean Current Energy Resource

Tidal currents are globally ubiquitous and together with major ocean currents represent important marine hydrokinetic sources of renewable energy. Importantly, tidal currents are highly predictable, while major ocean currents, such as the Gulf Stream and the Kuroshio Current, in a number of places around the world flow fairly persistently as part of ocean circulation systems. Therefore, tidal and ocean current energy conversion, along with solar, wind, waves, and other renewables, is an imperative consideration for a future energy portfolio for sustainability. Industry and governments worldwide are moving to invest in developing these forms of ocean energy; for example, the European Marine Energy Center (EMEC) has developed a major at-sea test facility for tidal current energy conversion devices in Orkney. A number of studies have been conducted worldwide, and various energy conversion devices have been designed and built, borrowing much of the technology from the wind industry. Since seawater is 800 times denser than air, ocean and tidal currents are energy dense, such that at low speeds they have kinetic energy per unit area comparable to that of winds of speed nearly an order of magnitude higher. For example, currents of speed 1 m/s have the same power capacity per unit area as winds of speed 9.3 m/s. Because of this physical property, ocean currents contain an enormous amount of energy that can be captured and converted to a usable form. Despite this potential, challenges remain

in research and development of efficient energy extraction devices, assessment of environmental impact, development of best practices and operating procedures, and establishment of policies and regulations.

44.2.1 Tidal Currents

Tides are very long waves, made up of harmonic constituents whose periods range from $O(1-10\times10^4\,\mathrm{hr})$, that flow to and fro between the oceans and coastlines in response to the gravitational forces exerted by the moon and the sun on the earth in combination with inertial forces, leading to the rise and fall of the sea level at the coastlines. At any coastal location, high and low tides correspond to the crest and trough of the tidal wave, respectively, reaching the location, the difference in sea levels between the two states being the tidal range at the location. The tidal range provides a good measure of the tidal energy resources at a location.

The tidal motion of the water is primarily a balance between the gravitational forces of the moon and the sun and the forces of inertia. In view of its relative proximity to the earth, the moon plays a greater role than the sun in generating the tides. The gravitational force of the moon on earth overcomes the inertial forces on the side closest to it and pulls the associated ocean into a bulge, while on the side furthest from the moon, inertial forces dominate and pull the water column away in the opposite direction, leading to a similar bulge furthest away from the moon. As the earth rotates and the moon moves around the earth and the alignment between the earth and the moon shift accordingly, the bulges maintain position with respect to the alignment causing cyclic variation in water levels over a lunar day, which is 24 hr 50 min long. If the earth were perfectly spherical and fully covered with water, this would result in semi-diurnal tides, involving two approximately equal high and two approximately equal low tides a day, everywhere. However, since the earth is not perfectly spherical and tidal flows of the oceans are interrupted by landmasses, tidal patterns are complicated. Many places, such as the east coast of the US, have semi-diurnal tides with the harmonic component M_2, of period of 12 hr, 25 min, being the major constituent of the tidal wave; other major constituents are the principal solar semi-diurnal S_2 component (12 hr period) and larger lunar elliptic semi-diurnal component (12 hr, 40 min period). Other places, such as in the Gulf of Mexico, have diurnal, or one high and one low tide a day, primarily governed by diurnal lunar components K_1 (23 hr, 56 min period), and O_1 (23 hr, 49 min period), and solar diurnal component S_1 (24 hr period). Yet other places, such as the western coast of the US, experience mixed semi-diurnal tides, involving two unequal high and two unequal low tides a day. Regionally, tidal flows are influenced by the shape of coastlines and opening to bays as well as by interactions with local riverine flows in estuaries, by weather patterns and winds as well as other currents. The tidal cycles and their amplitudes vary with time and position along a coast. During full moon and at new moon, the earth, the moon and the sun are aligned and the solar tide adds to the lunar tide, resulting in occurrence of maximum tidal ranges or *spring tides*. At first and third quarter moon, when the moon and the sun are at right angles, the lunar and solar tides act to dampen the tidal amplitudes, resulting in minimal tidal range or *neap tides*. Figure 44.2 shows the principal tidal component M_2, which is determined from satellite altimetry [44.10]. The associated tidal ranges are typically maximum at certain places near coastlines, particularly along the coast of northwest Europe, the northeast and northwest coasts of North America, the east coast of Africa, the northwest coast of Australia, southern and eastern parts of Asia, and northern and southern parts of South America. *Charlier* and *Justus* [44.11] (see also [44.12]) estimate that the theoretical mean global tidal power resource is around 3 TW, of which around 1 TW is available in shallow waters.

44.2.2 Ocean Currents

The mean flow in the ocean basins, separated by landmasses, comprises of large-scale subtropical gyres or circulation of surface currents that are set up by persistent action of prevailing winds, typically by the trade winds near the Equator and by the Westerlies at higher latitudes (see, for example, [44.14]). The subtropical gyres (Fig. 44.3), two in the Atlantic, two in the Pacific, and one in the Indian Ocean, rotate clockwise in the northern hemisphere and anticlockwise in the southern hemisphere and transport warmer waters from lower latitudes to higher latitudes and cooler waters from higher latitudes to lower latitudes. In the Atlantic, the North Equatorial Current, the Caribbean Current, the Gulf Stream (including the Loop Current in the Gulf of Mexico and the Florida Current), the North Atlantic Current, and the Canary Current mark the north Atlantic gyre, while the South Equatorial Current, the Brazil Current, the Antarctic Circumpolar Current, and the Benguela Current mark the south Atlantic gyre. In the Pacific, the North Equatorial Current, the Kuroshio Current, the North Pacific Current, and the California Current mark the north Pacific gyre, while the South Equatorial Current, the East Australian Current, the Antarctic Circumpolar Current, and the Peru Current mark the south Pacific gyre. In the Indian Ocean, the South Equatorial Current, the Agulhas Current, the

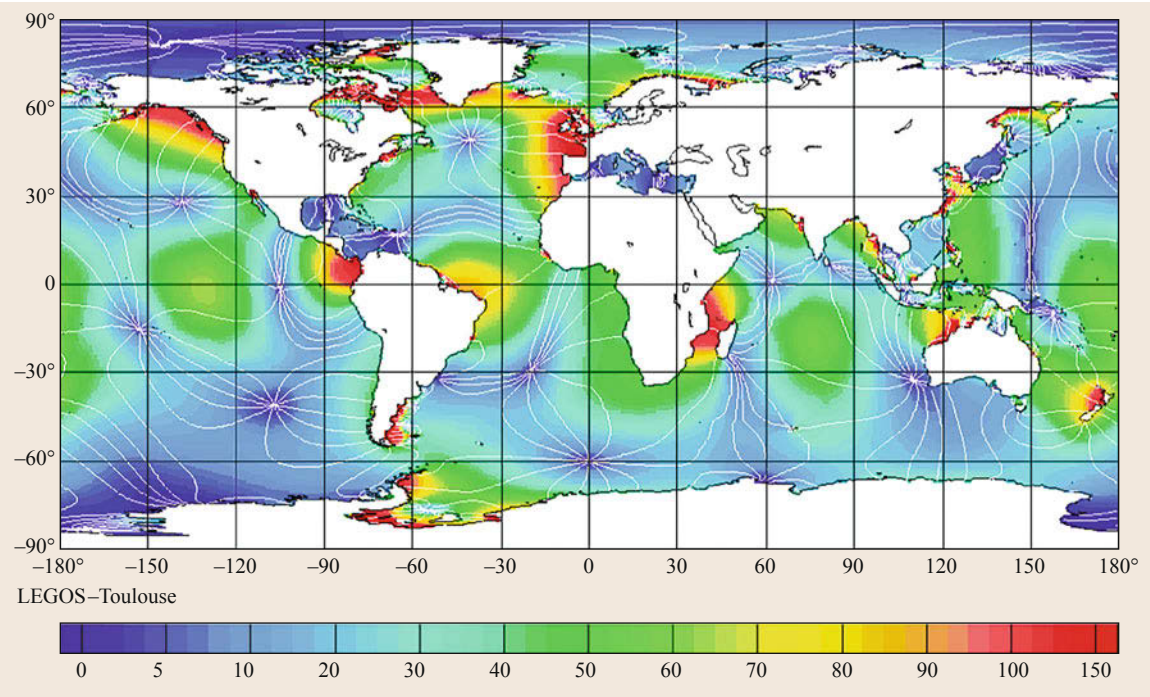


Fig. 44.2 M_2 tide (after [44.10])

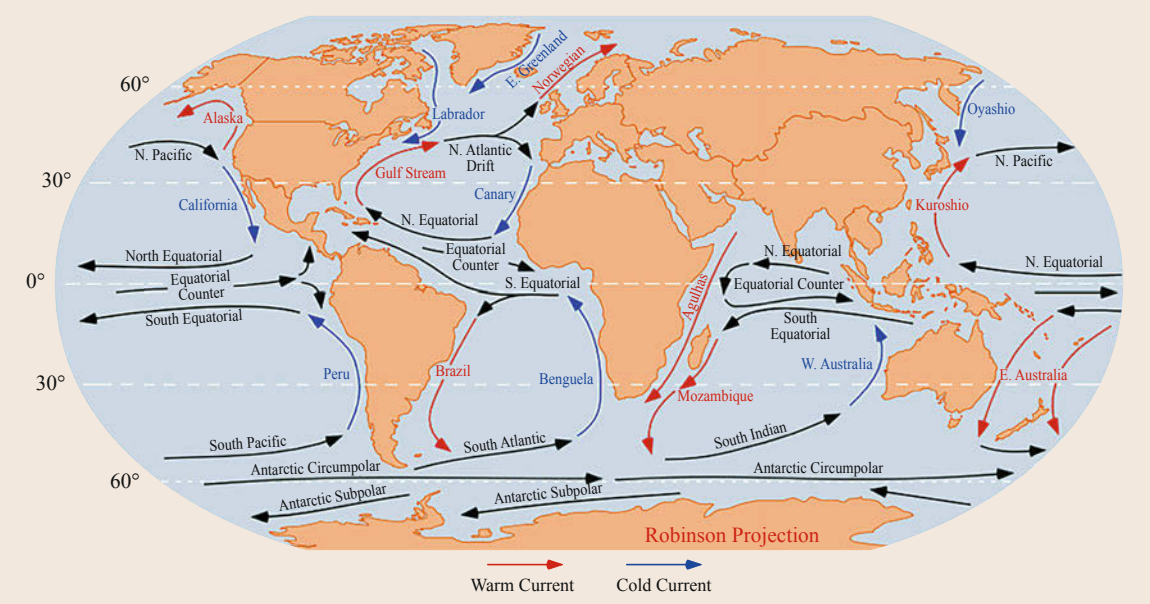


Fig. 44.3 Subtropical gyres and mean ocean currents (after [44.13])

Antarctic Circumpolar Current, and the West Australian Current mark the Indian Ocean gyre in the southern hemisphere. The landmass of the Indian subcontinent prevents full development of a gyre in the northern Indian Ocean; instead reversible currents associated with monsoon winds mark the region. In between the north and south subtropical gyres in the Atlantic and the Pacific lie Equatorial Counter Currents. Finally, smaller subpolar gyres occur, one in the North Atlantic and the other in the North Pacific.

The western arms of the subtropical gyres, referred to as *western boundary currents*, are narrow, fast, with mean speeds in the range 1–2.5 m/s, and extend deep, up to 400 m, while the eastern arms, or *eastern boundary currents*, are wide, slow, typically with mean speeds < 0.2 m/s, and shallow. The Gulf Stream and the Kuroshio Current are the most intense western boundary currents. The intensification of the western boundary currents is linked to the increasing strength of the Coriolis force as one moves away from the Equator [44.15]. The stronger Coriolis force deflects currents driven eastwards by the Westerlies south to form eastern boundary currents, which are broad so that the center of the gyre lie to the west of the basin. At lower latitudes, weak Coriolis force and the trade winds act to deflect the eastern boundary currents and transport its surface waters westwards. At the western boundary of the basin, the waters are driven towards the poles in a narrower region west of the center of the gyres, with the ocean surface sloping down westwards, thereby leading to significant contribution from the mean geostrophic flow. The time-averaged (over several days) velocity is given by the approximate equations [44.15], see also [44.16]

$$\begin{aligned}
\frac{1}{\rho}\frac{\partial p}{\partial x} &= \mathrm{g}\frac{\partial \zeta}{\partial x} = fv + \frac{\partial}{\partial z}\left(A_z \frac{\partial u}{\partial z}\right) \\
&\quad + A_H\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right) , \\
\frac{1}{\rho}\frac{\partial p}{\partial y} &= \mathrm{g}\frac{\partial \zeta}{\partial y} = -fu + \frac{\partial}{\partial z}\left(A_z \frac{\partial v}{\partial z}\right) \\
&\quad + A_H\left(\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2}\right) , \\
\frac{1}{\rho}\frac{\partial p}{\partial z} &= -\mathrm{g} ,
\end{aligned} \tag{44.5}$$

where positive x and y are along east and north directions, z points vertically upwards, and u and v are the respective horizontal components of the current velocity; p is the pressure, A_z and A_H are coefficients of eddy friction, and ζ is the mean free surface elevation.

44.2.3 Power and Power Density of Tidal and Ocean Currents

Within a column of moving fluid, the power density P_u and the available power P or hydrokinetic energy flux across a cross-sectional area A within the current may be estimated using the following equations

$$P_u(y,z,t) = \frac{1}{2}\rho U^2 (\boldsymbol{U}\cdot\boldsymbol{n}) , \tag{44.6}$$

$$P = \iint_A P_u \, \mathrm{d}A , \tag{44.7}$$

where ρ is the density of the fluid, $\boldsymbol{U}$ is the velocity of the fluid, $\boldsymbol{n}$ is the unit vector normal to the area A through which the fluid flows. If the cross-stream components of velocity can be neglected, $P_u(y,z,t) \simeq \frac{1}{2}\rho(\boldsymbol{U}\cdot\boldsymbol{n})^3$.

The power that can be extracted by a single hydrokinetic turbine may be expressed as

$$P_E = \eta P , \tag{44.8}$$

where $\eta(<1)$ is efficiency of the extraction process. In open oceans, the maximum value of η is typically estimated using the ideal, frictionless, steady incompressible flow model due to *Betz* [44.17] (see also [44.18]), where the turbine rotor is represented by an actuator disk of area A_D (Fig. 44.4). We consider the axial speed upstream, at the rotor, and downstream of the rotor to be U_1, U_D, and $U_2 = \alpha U_1 (\alpha < 1)$, respectively, each being uniform across the appropriate section within the stream tube.

Conservation of mass requires that

$$\iint_{CS} \rho\,(\boldsymbol{U}\cdot\boldsymbol{n})\, \mathrm{d}A = 0 , \tag{44.9}$$

for the fluid enclosed by the control surface CS, shown by the dashed line in Fig. 44.4; here the normal vector $\boldsymbol{n}$ points out of the volume enclosed by CS. Since $(\boldsymbol{U}\cdot\boldsymbol{n}) = 0$ along the curved sections of CS, we have

$$U_1 A_1 = U_2 A_2 = U_D A_D , \tag{44.10}$$

where A_1 and A_2 are the areas of the upstream and downstream sections of CS.

Applying the energy equation to the control volume between sections 1 and 2, we have

$$-\dot{W} = \iint_{CS} \left(\frac{1}{2}\rho U^2 + p + \rho g z\right)(\boldsymbol{U}\cdot\boldsymbol{n})\, \mathrm{d}A , \tag{44.11}$$

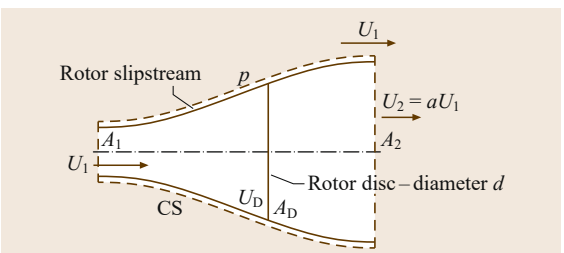


Fig. 44.4 Schematics of actuator disc model

where $\dot{W} = P_E$, is the extracted power or the work done by the fluid on the system enclosed by CS and p is the pressure. For the case of uniform flow considered here and the control surface as in Fig. 44.4, $p + \rho g z = C$, a constant over CS (using Bernoulli's equation outside of the stream tube), and (44.11) becomes

$$\begin{aligned}\dot{W} = P_E &= \left(\frac{1}{2}\rho U_1^2 + C\right) U_1 A_1 \\ &\quad - \left(\frac{1}{2}\rho U_2^2 + C\right) U_2 A_2 \,, \\ &= \frac{1}{2}\rho \left(U_1^2 - U_2^2\right) U_D A_D \,, \end{aligned} \tag{44.12}$$

using (44.10). Finally, conservation of momentum of the system enclosed by CS implies

$$\boldsymbol{F}_D = \iint_{CS} \frac{1}{2}\rho \boldsymbol{U}(\boldsymbol{U}\cdot\boldsymbol{n})\,\mathrm{d}A - \iint_{CS} (p + \rho g z)\boldsymbol{n}\,\mathrm{d}A \,, \tag{44.13}$$

where $\boldsymbol{F}_D = (-F_D, 0, 0)$ is the reaction force on the rotor to maintain it in position. Applying (44.13) to the CS in Fig. 44.4, we note that the pressure term integrates to zero so that, as in (44.10),

$$F_D = \rho(U_1 - U_2)U_D A_D \,. \tag{44.14}$$

Then extracted power P_E is also given by

$$P_E = F_D U_D = \rho(U_1 - U_2)U_D^2 A_D \,. \tag{44.15}$$

Equating (44.12) and (44.15), we obtain that the flow speed at the disk is given by

$$U_D = \frac{U_1(1+\alpha)}{2} \,, \tag{44.16}$$

where $\alpha = U_2/U_1$. Hence from (44.15), the extracted power is given by

$$P_E = \frac{1}{4}\rho U_1^3 A_D (1-\alpha)(1+\alpha)^2 \,. \tag{44.17}$$

If we regard the available reference power as $P = \frac{1}{2}\rho U_1^3 A_D$, corresponding to the energy flux through a circle of diameter A_D upstream of the disk, then the ideal efficiency of the extraction is given by

$$\eta = \frac{P_E}{\frac{1}{2}\rho U_1^3 A_D} = \frac{1}{2}(1-\alpha)(1+\alpha)^2 \,, \tag{44.18}$$

which has a maximum value of $\eta_{max} = 16/27 \approx 0.59$, when $\alpha = 1/3$ and hence $U_D = \frac{2}{3}U_1$. This maximum value is referred to as the Betz limit. Thus, maximum power that can be extracted is given by

$$P_{Emax} = \frac{1}{2}\rho U_1^3 A_D \eta_{max} \,. \tag{44.19}$$

Maximum efficiencies are lower than 0.59 when allowance is made for the finite number of blades and rotational speeds; the Betz limit is achieved only for higher values of the tip speed ratio $U_D/\Omega R_D$, where Ω is the rate of rotation and R_D is the radius of the rotor [44.19]. For nonuniform flows, the integral equations (44.9)–(44.13) must be used. Most designs for hydrokinetic turbines allow the turbine to yaw so that the turbine may be normal to the direction of the current; however, if the turbine is unable to yaw, then the full integral equations should account for the associated directional flow. Additionally, if the flow is unsteady, unsteady versions of the momentum and energy equations are needed.

Typically, tidal energy would be harnessed from tidal streams in channels and openings in coastlines or in estuaries. Therefore in making tidal energy resource assessment, consideration is given to the characteristics of the flow in such channels. *Garrett* and *Cummins* [44.20] consider an extension of (44.19) to flow in a finite tidal channel. They show that Betz limit applies when the rotor size is small compared with the channel's width. If the rotor size is comparable to the channel width, then they show that

$$P_{Emax} = \frac{\frac{1}{2}\rho U_1^3 A_D \eta_{max}}{\left(1 - \frac{A_D}{A_c}\right)^2} \,, \tag{44.20}$$

where A_c is the cross-sectional area of the channel – the efficiency of the turbine is increased due to the confining effects of the channel and associated increase in the pressure drop across the turbine. In practice, for larger blockage ratios A_D/A_c this increased efficiency needs to be balanced against the effect of increased blockage or drag on the flow that will lead to reduction in hydrokinetic energy flux [44.21]. Indeed, *Garrett* and *Cummins* [44.22] (see [44.23]), on the basis of an analysis of tidal flow in a constricted channel of variable cross-section connecting two large bodies of water, showed that if the pressure gradient in the channel is primarily balanced by frictional forces, the extractable mean power (averaged over a tidal cycle) is given by

$$P_E = 0.556\rho g a Q_m \left[1 - \left(\frac{Q_m}{Q_{0m}}\right)^2\right] \,, \tag{44.21}$$

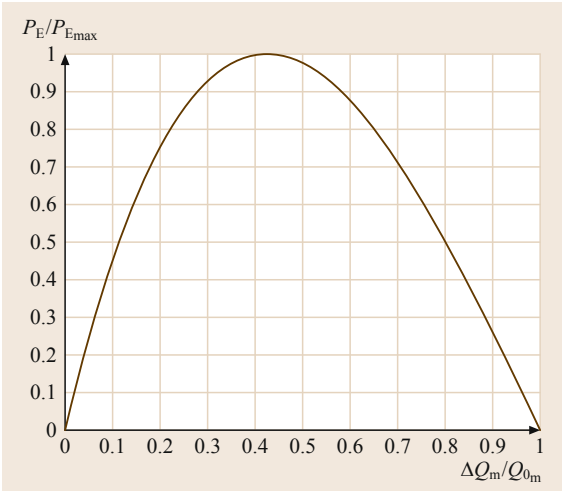


Fig. 44.5 Extracted power vs. reduction in flow rate, $\Delta Q_m = Q_{0m} - Q_m$, in the channel

where ρ is the density of water, Q_m and Q_{0m} are, respectively, the peak volume flow rates through the channel in the presence of/in the absence of power extraction, and a is the amplitude of a sinusoidal tidal head $a\cos(2\pi t/T)$ of period T between the two ends of the channel; the tidal volume flow rate is considered to be $Q = Q_m|\cos(2\pi t/T)|^{\frac{1}{2}}$. The result (44.21) is shown plotted in Fig. 44.5.

Extracting power from the tidal flow results in reduction $\Delta Q_m = Q_{0m} - Q_m$ in volume flow rate. Extractable power increases as ΔQ_m increases until

$$\Delta Q_m = \left(1 - \frac{1}{\sqrt{3}}\right) Q_{0m} = 0.42 Q_{0m}$$

(i. e., until there is 42% reduction in the flow rate), which marks the condition under which maximum mean power can be extracted from the channel,

$$P_{Emax} = 0.21\rho g a Q_{0m} \,. \tag{44.22}$$

Beyond this value of ΔQ_m, the extractable power drops to zero as $\Delta Q_m \to Q_{0m}$ or $Q_m \to 0$. If, on the basis of environmental impact considerations, a maximum of 5% reduction in channel flow rate (i. e., $\frac{\Delta Q_m}{Q_{0m}} = 0.05$) is permitted, then extraction of only 23% of the maximum extractable tidal power in the channel may be permitted. That is, $P_{EPermitted} = 0.23 P_{Emax}$.

On the basis of these considerations, and using (44.22) in a numerical simulation, *Haas* [44.24] assessed the theoretically available tidal energy resource in streams around the US coast, estimating the resource potential for tidal generation in USA to be 250 TWh/year. He showed that Alaska contains the largest number of locations with high kinetic power density, followed by Maine, Washington, Oregon, California, New Hampshire, Massachusetts, New York, New Jersey, North and South Carolina, Georgia, and Florida. The average tidal stream power density at a number of these locations exceeds 8 kW/m^2.

In practice, significant power extraction from tidal channels will require arrays involving thousands of turbines [44.25]. Theoretically, the power of an array of turbines can be estimated using (44.9)–(44.13); depending on the spacing of the turbines within the array, and the impact of the presence of the array and associated blockage of the overall flow of the current, the efficiency of each individual turbine will be affected [44.20, 21]. Additionally, if turbines are spaced not only cross-stream, but also upstream and downstream of each other, the wake effects of the upstream turbines will impact the inflow to the downstream turbine. This could potentially have a significant impact on the array's overall performance.

44.3 Assessment of Global Ocean Current Resources

In this section, we discuss the methodology for assessing ocean current energy resources and make an assessment of the energy potential of major ocean currents worldwide, with particular emphasis on the Florida Current [44.26–29], based on flow simulated using the Hybrid Coordinate Ocean Model (HYCOM) [44.30]. HYCOM is a data assimilative hybrid isopycnal-sigma-pressure coordinate ocean model aimed at depicting the state of the oceans at fine resolution in real time.

The 3-year time-averaged surface speeds of the global ocean currents computed using HYCOM are shown in Fig. 44.6, from which the strength of each of the major currents identified in Fig. 44.3 may be discerned. Correspondingly, regions with potential high ocean current energy resource potential can be identified from Fig. 44.6. Practical considerations, however, require that these regions be in accessible waters, generally close to shore, for cost-effective development and placement of the necessary infrastructure. On this basis, Fig. 44.6 suggests that the following regions, lying in the vicinity of the major western boundary currents, likely have good potential for development of ocean current energy: the east coast of the USA, the northeast coast of Brazil, the east and southeast coasts of Africa,

the east coasts of Japan and Taiwan, and the southeast coast of Australia.

The geostrophic components of western boundary currents may also be discerned from satellite altimetry data [44.31, 32].

As discussed in Sect. 44.2.2, major ocean currents are wind-driven surface currents that extend in a layer beneath the sea surface. The plot of time-averaged surface currents in Fig. 44.6 provides good indication of the global locations of high potential for development of ocean current energy. The actual levels of this potential also depend on the vertical extent and structure of the velocity field, which vary between the major currents and with associated geographic locations. In deep waters, the flow structure is primarily governed by baroclinic geostrophic flows associated with balance between pressure gradient forces and Coriolis force, whereas in shallower waters, frictional forces may become important, giving rise to steeper velocity gradients. The surface currents, therefore, penetrate relatively deeper in deep waters than in shallow waters; the ocean current energy resource is correspondingly higher at a site in deeper waters. The greater water depth, on the other hand, may pose challenges for installation of the infrastructure for energy extraction.

The ocean current energy flux per unit area or the power density at each node of the HYCOM computational grid may be estimated using

$$P_{\mathrm{u},i} = \overline{\frac{\rho_i}{2} U_i^3}\,, \tag{44.23}$$

where ρ_i is the seawater density and U_i the water speed at the i-th node (44.6). $P_{\mathrm{u},i}$ is averaged over the 3-year period considered. Here we compute and examine the characteristics of the power density at constant depths. In Figs. 44.7–44.11, the time-averaged power density $P_{\mathrm{u},i}$, for $P_{\mathrm{u},i} > 500\,\mathrm{W/m^2}$, is plotted globally at specified depths; $P_{\mathrm{u},i} < 500\,\mathrm{W/m^2}$ is likely to be too low to support commercial scale ocean energy extrac-

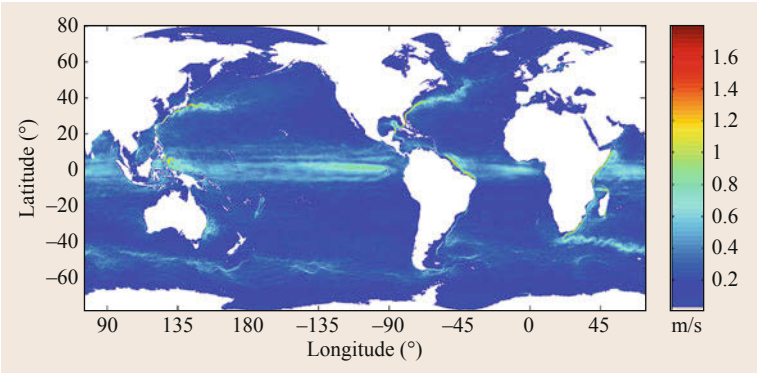


Fig. 44.6 3-year surface current average, HYCOM global model, 2009–2011

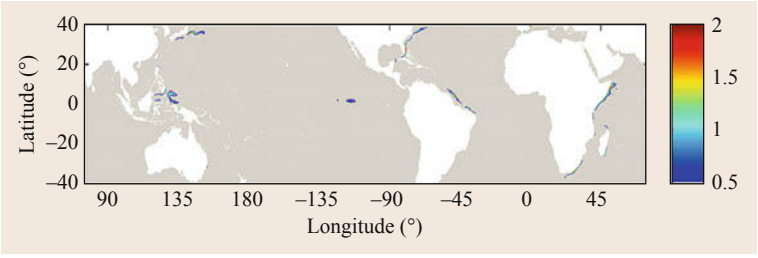


Fig. 44.7 3-year average ocean current power density at the sea surface in $\mathrm{kW/m^2}$

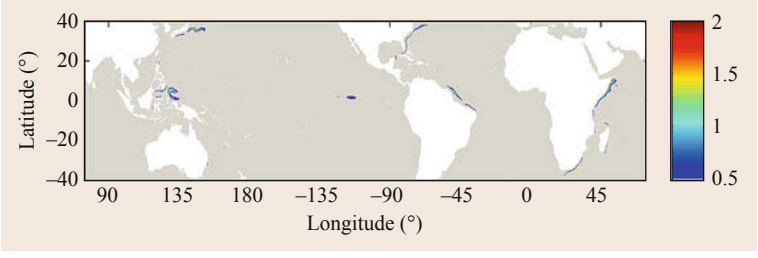


Fig. 44.8 3-year average ocean current power density at 30 m in $\mathrm{kW/m^2}$

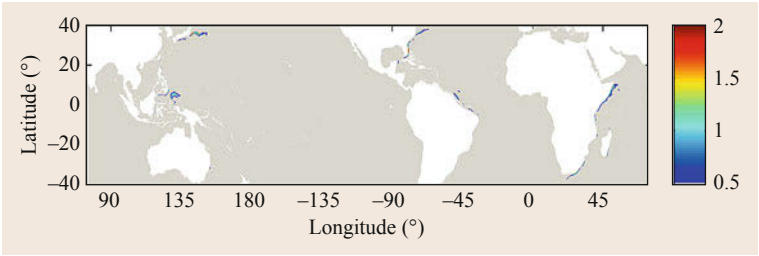


Fig. 44.9 3-year average ocean current power density at 50 m in kW/m^2

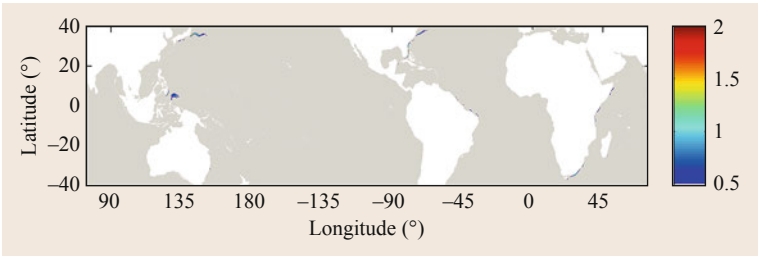


Fig. 44.10 3-year average ocean current power density at 75 m in kW/m^2

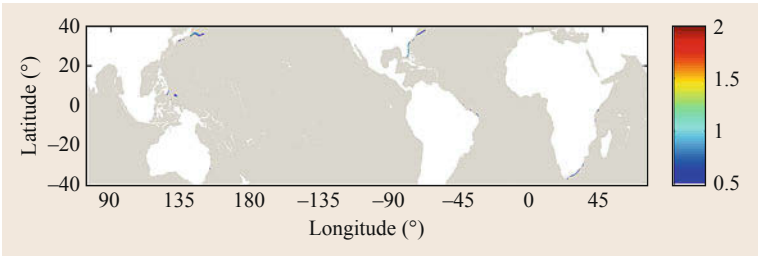


Fig. 44.11 3-year average ocean current power density at 100 m in kW/m^2

tion [44.24]. At the surface, where the current is the strongest, the power density is most dense. Since the surface currents decay with depth, the power densities decrease correspondingly, as is evident from Figs. 44.7–44.11.

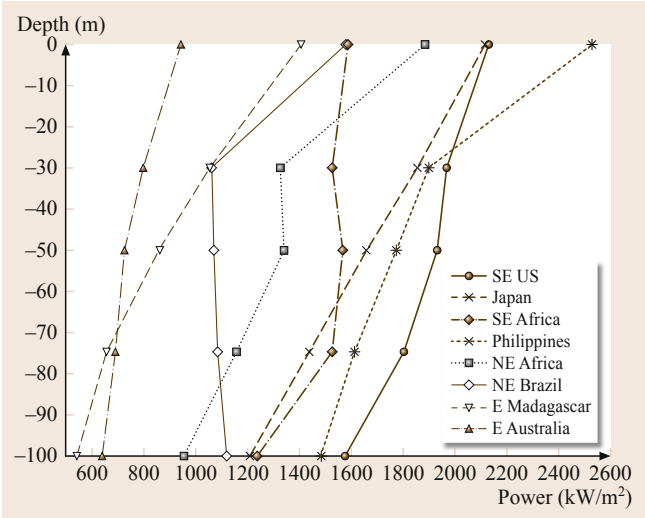


Fig. 44.12 Time average maximum power density in each region as a function of depth

Based on these plots, locations worldwide that have best ocean current energy potential are the east coast of the USA, the northeast and southeast coasts of Africa, the southeast coast of the Philippines, the eastern coast of Japan and Taiwan, the northeast coast of Brazil, the southeast coast of Africa, and the east coast of Madagascar; the corresponding currents are, respectively, the Gulf Stream (USA), the Agulhas Current (Africa), a combination of the North Equatorial, South Equatorial, and Equatorial Counter Current (the Philippines), the Kuroshio Current (Japan and Taiwan), the Southern Equatorial Current (Brazil), and the Mozambique Current (Madagascar).

The estimated maximum value of the time-averaged mean power density at various depths for each of the major ocean currents is provided in Fig. 44.12. The depth-specific maximum for each current is over the respective area where $P_{u,i} > 500\,\mathrm{W/m^2}$. The maxima are not necessarily at the same geographic points at each depth. At many of these locations, the ocean current energy potential extends to a depth of 100 m. This facilitates placement of turbine infrastructure at depths

Table 44.1 Ocean current energy extraction potential at 50 m depth

Region	Max power density [kW/m^2]	Sea floor depth [m]	Distance from shore [km]
US East Coast	1.93	500–600	57
Southeast Africa	1.66	700–800	28
The Philippines	1.57	3000–3500	32
Japan	1.78	3500–4000	116
Northeast Africa	1.34	400–500	12
Brazil	1.08	300–400	57
Madagascar	0.86	300–400	29
Australia	0.73	1400–1500	38

where impact on shipping is minimized but where the resource is still viable. It also facilitates placement of turbines with relatively large rotors.

The value of the power density is an important driving factor in choosing a viable location for turbine placement. The distance from the shore and the depth of the seafloor at the location are other factors, since they significantly determine the overall cost of designing, building, installing, and maintaining the turbine infrastructure. The length of power transmission cables and the associated costs increase the further the location is from shore. Additionally, operations and maintenance (OM) costs as well as installation costs of the turbine array at a distance from the shore increase accordingly. The depth of the ocean floor at the installation location significantly governs the type of installation required. The deeper the ocean floor, the more complicated the installation system and higher the OM costs, so that the overall costs are higher. For an installation depth of 50 m that avoids potential interference with shipping and interaction between the turbines and the free surface, the location of maximum power density, together with the associated distance from shore and the sea floor depth are compared in Table 44.1.

In Table 44.1, the highest maximum power density is 1.93 kW/m^2 and is off the US East Coast; however, it is 57 km offshore and the sea floor depth is 500–600 m. The choice of the installation location will, therefore, be a compromise between these three factors, as well as other considerations such as the state of the local electric grid infrastructure that will impact the overall economics of the installation. It is important to note that in several of these regions HYCOM predicted slightly lower power densities in much shallower water, a much more accessible resource for development; therefore a more rigorous regional site assessment, which is beyond the scope of this global analysis, is necessary. It is also important to note that HYCOM has been shown to significantly under predict the power density in some regions. For example, a measurement-based study estimates average power density of 2.2 kW/m^2 at 20 m-depth level at a location off South Africa, where the sea floor depth is approximately 100 m [44.33]. A second measurement-based study estimates average power density of 3.0 kW/m^2 at a 50 m-depth level off Florida [44.34]. These studies highlight the importance of conducting corroborative measurement based analyses.

Another factor of interest is the areal extent at an identified region over which viable hydrokinetic energy potential exists. In Table 44.2, average statistics for the area over which a given power density threshold is met are presented for the global currents 50 m-depth level. Additionally, the average power density over the region in which energy extraction may be viable – namely where the average power density is greater than 500 W/m^2 – is estimated.

An economics metric can be created to compare the energy development potential at these sites. A average power density and maximum power density would be positive factors in the economics metric of current energy development, while depth and distance from shore

Table 44.2 Area available for ocean current energy development at various locations at an operating depth-level of 50 m (1000 km^2)

Region	Area with power density greater than threshold (in ×1000 km^2)					
	> 500 W/m^2	> 750 W/m^2	> 1000 W/m^2	> 1500 W/m^2	> 2000 W/m^2	Average power density in area with $P_{u,i}$ > 500 W/m^2 [W/m^2]
Gulf Stream	147	51	25	7	0	776
SE Africa	71	28	14	0	0	765
The Philippines	194	89	9	1	0	744
Japan	172	72	37	6	0	792
NE Africa	183	37	2	0	0	661
Brazil	57	3	0	0	0	599
Madagascar	9	1	0	0	0	645
Australia	4	0	0	0	0	582

Table 44.3 Example global ocean current energy potential economics metric F, with all weighting factors = 1

Location	F
Gulf Stream	7058
SE Africa	4276
The Philippines	2190
Japan	558
NE Africa	28911
Brazil	1861
Madagascar	466
Australia	33

would be negative factors. As the positive factors increase, so does the economic outlook of the array, and as the negative factors increase, the economic outlook of energy development at the site decreases. An example metric may be

$$F = \frac{(A\alpha)(P_{\text{umax}}\beta)(P_{\text{umean}}\gamma)}{(D\zeta)(d\xi)} , \qquad (44.24)$$

where A is the area over which the average power density is greater than 500 W/m^2 in units of 1000 km^2, α is the weighting factor for A in km^{-2}, $P_{\text{umax}} = \overset{\max}{i}\{P_{\text{u},i}\}$ is the maximum power density in W/m^2, β is the weighting factor for P_{umax} in m^2/W, $P_{\text{umean}} = \frac{1}{N}\sum_{i=1}^{N} P_{\text{u},i}$ is the mean power density within area A in W/m^2, γ is the weighting factor for P_{umean} in m^2/W, D is the distance from shore in km, ζ is the weighting factor for D in km^{-1}, d is the depth of the sea floor in m, and ξ is the weighting factor for d in m^{-1}. Using this example metric, and assuming that the weighting factors all have a numerical value of 1, the results for the global currents at 50 m are given in Table 44.3.

With all the weighting factors set to 1, the metric suggests that Northeast Africa is the best global location for hydrokinetic energy development, the metric being heavily biased by the distance of the potential from the shore (12 km). This metric can be modified not only by selecting other weighting factors, which were set to 1 in the case considered above, but also by changing the factors considered in the metric. Additional factors can be added or removed from this metric depending on the initial economic evaluation of the turbine array.

44.3.1 Gulf Stream Case Study

As an example, the Gulf Stream off the east coast of the United States is considered further for its hydrokinetic resources using the HYCOM Gulf of Mexico basin model data [44.26, 27]. The time-averaged power densities for various depths are plotted in Figs. 44.13–44.17. The figures highlight the power density decay with

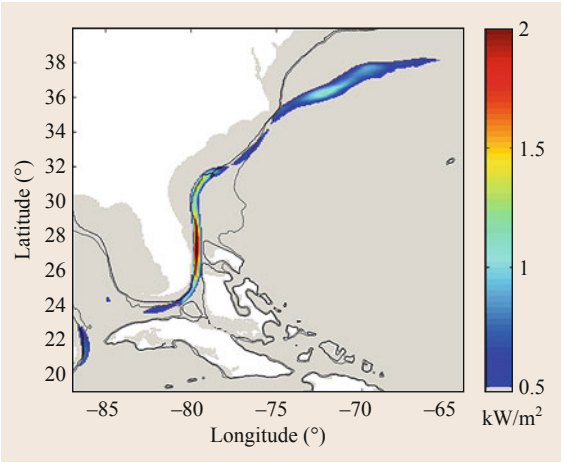


Fig. 44.13 3-year average power density of the Florida Current at the surface

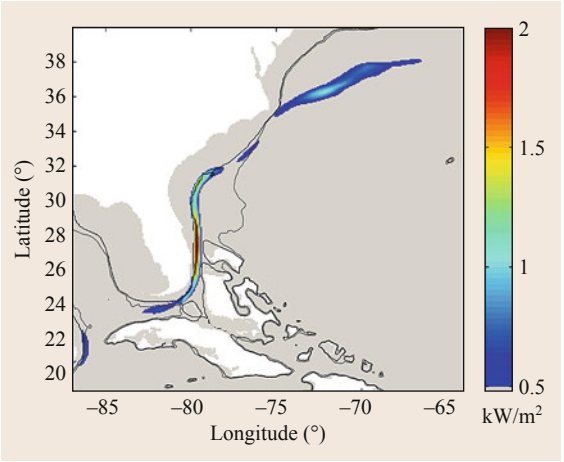


Fig. 44.14 3-year average power density of the Florida Current at 30 m

depth. This is especially apparent in comparing the plot for the power density at the surface (Fig. 44.13) with that at 100 m (Fig. 44.17). As seen in the figures, the highest power density in the Gulf Stream is off the east coast of Florida, corresponding to the Florida Current – the portion of the Gulf Stream that flows off the east coast of Florida. The proximity to large population centers, including Miami, Fort Lauderdale, and West Palm Beach, make the Florida Current an attractive energy resource for the state of Florida. Figure 44.18 highlights the relationship between latitude and power density in this region, suggesting that the latitude with the highest power density is approximately 27°N.

The factors that govern resource assessment are shown in Table 44.4. It may be noted that the location of the maximum power density at the surface is

Table 44.4 Gulf Stream hydrokinetic resource assessment factors (based on HYCOM data [44.26, 27])

Depth [m]	Location		Power density [kW/m^2]	Average power density in area with $P_{u,i} > 500$ W/m^2 [W/m^2]	Sea floor depth [m]	Distance from shore [km]
	Latitude, °N	Longitude, °E				
0	27.09	−79.76	2.13	797	400–500	35
30	27.45	−79.68	1.97	782	500–600	57
50	27.45	−79.68	1.93	776	500–600	57
75	27.45	−79.68	1.81	762	500–600	57
100	27.45	−79.68	1.58	739	500–600	57

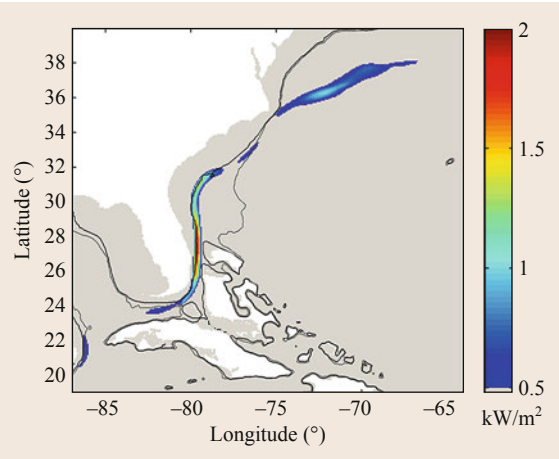


Fig. 44.15 3-year average power density of the Florida Current at 50 m

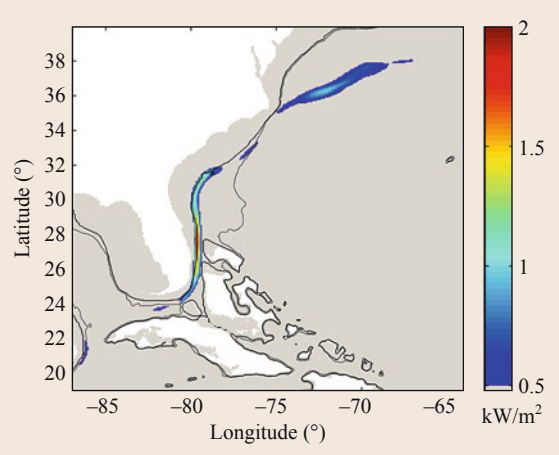


Fig. 44.16 3-year average power density of the Florida Current at 75 m

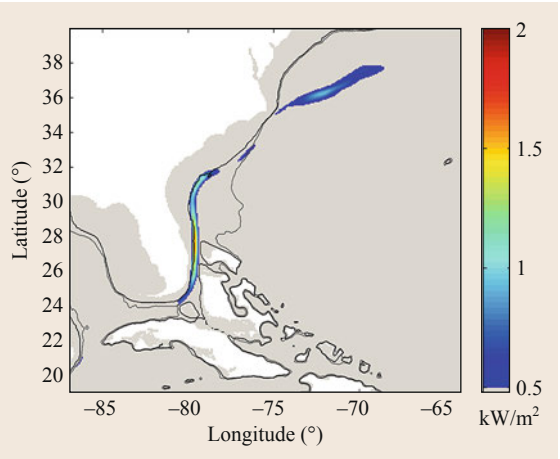


Fig. 44.17 3-year average power density of the Florida Current at 100 m

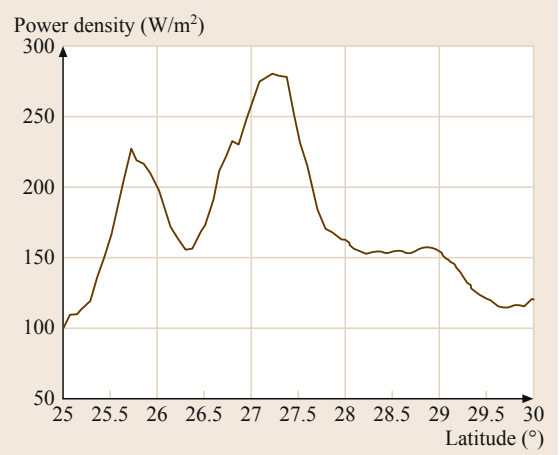


Fig. 44.18 Variation in Florida Current power density with latitude

different from the location of the corresponding values at depths between 20 and 100 m. This suggests that use of power density based on surface current data in site selection may be misleading, since turbine arrays would be installed at some depth below the water surface. The maximum power densities at 30, 50, 75, and 100 m depth are at a location 40 km away. The maximum power densities at these depths are at the same geographic location within the Florida Current; however, this may not happen in every global current. Care

is needed in making the resource assessment and deciding on optimal locations. The type and size of devices used and the water depth at which they operate optimally will dictate the depth at which a detailed resource assessment would be required.

Based on this overall assessment, the best location in the Florida Current for hydrokinetic energy extraction seems to be at 27.45°N, 79.76°W. However, use of the metric (44.21) may suggest an alternative location as being more desirable. A method for finding an ideal location in the Florida Current is suggested by *Duerr* [44.26], based on consideration of the variability of the hydrokinetic resource with latitude rather than water depth. The Florida Current is divided into constant latitude cross-sections to evaluate the hydrokinetic power over the entire cross-section, as in (44.7). In addition to the hydrokinetic power, the hydrokinetic power density, the depth at the core of the current, and the distance of the core from the shore are evaluated at each cross-section from 25 to 30°N.

The economics metric in this case is slightly modified from (44.24)

$$F = \frac{(A\alpha)(P\beta)(P_{\text{umean}}\gamma)}{(D\zeta)(d\xi)} , \tag{44.25}$$

where the P_{umax} term in (44.24) has been replaced with P, the average power at the cross-section in GW, and the unit of β is adjusted accordingly to GW^{-1}. The values of the metric F based on the HYCOM data, with the weighting factors set to 1 as before and with D now given by the location of the core of the current, are shown in Table 44.5.

This economics metric suggests that 26.95°N is the best latitude for hydrokinetic energy extraction. This

Table 44.5 Florida hydrokinetic resource assessment factors based on constant latitude cross-section analysis (using HYCOM data)

Latitude [°N]	Power [GW]	Mean power density [W/m^2]	Depth [m]	Distance [km]	Area with power density > 500 W/m^2 [km^2]	Metric F
25.51	8.1	164.8	700	30	5.28	0.33
25.66	9.0	209.1	700	32	6.15	0.51
25.80	9.5	218.4	700	33	6.61	0.59
25.95	9.6	207.9	600	34	6.73	0.66
26.09	9.6	183.3	500	35	6.70	0.68
26.23	9.7	163.0	500	36	6.71	0.59
26.38	10.2	156.3	500	29	6.89	0.76
26.52	10.3	173.5	400	29	6.91	1.08
26.66	10.6	211.5	400	29	7.03	1.36
26.81	11.0	233.1	400	29	7.27	1.60
26.95	11.5	248.3	400	29	7.58	1.89
27.09	11.9	275.7	400	37	7.84	1.75
27.23	12.2	280.4	500	46	8.06	1.21
27.37	12.3	278.6	500	54	8.19	1.05
27.52	12.4	232.5	500	54	8.18	0.88
27.66	12.2	199.6	500	62	8.10	0.64
27.80	12.3	169.8	600	70	7.97	0.40
27.94	12.4	163.6	500	77	7.94	0.41
28.08	12.4	157.4	500	77	7.88	0.40
28.22	12.4	153.1	400	84	7.87	0.44
28.36	12.6	154.1	400	83	7.91	0.46
28.51	12.7	153.9	400	75	7.92	0.52
28.65	12.8	153.3	400	82	8.07	0.49
28.79	13.0	155.7	500	96	8.15	0.35
28.93	13.1	156.4	500	102	8.20	0.33
29.07	12.9	149.5	600	109	8.09	0.24
29.21	12.7	141.2	600	115	7.98	0.21
29.35	12.4	127.5	600	121	7.80	0.17
29.49	12.2	121.0	700	127	7.63	0.13
29.62	12.1	115.7	700	126	7.46	0.12
29.76	11.9	115.1	700	132	7.28	0.11
29.90	11.7	115.4	700	132	7.12	0.10
30.04	11.7	123.6	600	140	7.02	0.12

is only one method of calculating a suitable metric, and other factors can be introduced into the equation. Overall, based on the constant-latitude economics metric (44.25) and the constant-depth economics metric (44.21), Florida Current's viable resource seems to be located between 26.45 and 27.45°N. These results, based on the ocean model, would need to be complemented by comprehensive in-situ surveys and in-depth studies to determine any discrepancies between the ocean models and in-situ data.

Overall, present-day ocean models based on data assimilation, such as HYCOM, provide excellent predictions of oceanic flows. However, care is needed when using ocean models. For instance, coarseness of the model grid may result in under or over prediction of, say, the speed of a current or the associated volume flow rate and correspondingly under or over-prediction of its hydrokinetic energy resource. Comparing the data from an ocean model with in-situ measurements is, therefore, essential in validating the predictions of the model. Once such deficiencies in the ocean model have been identified and characterized, the resource prediction can be better qualified. Local models with refined grids may be considered for improvement of the model.

While HYCOM is one of the most widely used ocean models, the velocity predictions of the global model are not perfect for ocean energy resource assessment. In both the Florida Current [44.28] and off the coast of Brazil [44.35], it has been shown that the HYCOM global model under-predicts the current velocities. While the overall trend in the difference between the in-situ data and the HYCOM is the same in both studies, the *Jeans* et al. [44.35] study suggests that HYCOM's prediction off Brazil is highly uncorrelated with the in-situ data, while the *Duerr* et al. [44.28] study suggests that HYCOM predicts the overall trend in the velocity data well – that is, the shape of the velocity profile predicted by HYCOM is very similar to the in-situ observations – but under-predicts the value of maximum velocity.

The applicability of an ocean model of sufficiently fine resolution to different global locations and oceanic conditions needs to be studied in order that hydrokinetic energy predictions based on the model can be suitably qualified. In *Duerr* et al. [44.28], it is suggested that a gain term – calculated using in-situ and model data – may be applied to the HYCOM data in order to improve the agreement between the model and physical conditions. This process is straightforward if in-situ field observations of currents are available. However, in the absence of such data, the details of the ocean model should be considered in order to understand possible implications in using the model for hydrokinetic resource assessment.

44.4 Other Considerations

A number of other considerations are involved in assessment of marine hydrokinetic energy resources. These include:

- *Design of devices and technologies*. To optimize performance while minimizing environmental impact at a specific site, special consideration is needed in designing devices and arrays of devices in light of prevailing local conditions. These factors could also be introduced in the metrics (44.21), (44.25). Particular considerations include:
  - Material selection, corrosion and bio-fouling control, and advanced coatings for devices and their components
  - Assessment of hydrodynamic loads and cavitation
  - Effect of shear and turbulence on hydrodynamic performance
  - Device machinery, generators and power takeoff systems
  - Optimization in design of single and arrays of devices through modeling and simulation
  - Flow and wake interactions in arrays of devices.
- *Deployment and operations*. Robust device deployment systems and low OM costs are required for viable ocean energy development. Special considerations are needed for:
  - The choice of mooring systems and offshore platforms
  - Access requirements for the devices for planned and unplanned maintenance
  - Implementation of smart machine condition monitoring for system reliability
  - Modeling and simulation studies to assess maintenance costs.
- *Environmental impact*. Regulatory agencies require that before permission to begin marine hydrokinetic ocean energy development can be granted, an assessment of the environmental impact the activity may cause needs to be made. It is expected that the activity will not have zero impact, but will require consideration of ways of mitigating the impact. Typical issues that arise are:
  - Impact on aquatic life and mortality around devices
  - Operational and construction noise

- Electromagnetic field emissions from device infrastructure and transmission cables
- Impact on shipping, protected sites and naval operations
- Impact on road traffic and other shore activities in the vicinity ocean energy development
- Impact on energy resource. In the case of ocean currents, determination of the potential effects of energy extraction on the ocean circulation systems.

- *Socio-economics of marine hydrokinetic energy*. Economic and social issues associated with the energy development, marine spatial planning and the social acceptability of the energy extraction technologies and their implementation, including their sustainability, are significant considerations in determining viability of the resource. These include:
  - Cost of electricity compared with other sources
  - Concerns with impact of the energy technologies on stakeholder activities and their livelihood
  - Political and social will.

Based on these considerations, an iterative process may result in making the final selection of a suitable location that has the desired attributes.

References

44.1 O.M. Phillips: On the generation of waves by turbulent wind, J. Fluid Mech. **2**(5), 417–445 (1957)

44.2 J.W. Miles: On the generation of surface waves by shear flows, J. Fluid Mech. **3**(2), 185–204 (1957)

44.3 K. Hasselmann, T.P. Barnett, E. Bouws, H. Carlson, D.E. Cartwright, K. Enke, J.A. Ewing, H. Gienapp, D.E. Hasselmann, P. Kruseman, A. Meerburg, P. Müller, D.J. Olbers, K. Richter, W. Sell, H. Walden: Measurements of wind-wave growth and swell decay during the Joint North Sea Wave Project (JONSWAP), Dtsch. Hydrogr. Z **1**(8), 1–95 (1973)

44.4 A. Cornett: A global wave energy resource assessment, Proc. 18th ISOPE Conf. (2008)

44.5 K. Gunn, C. Stock-Williams: Quantifying the global wave power resource, Renew. Energy **44**, 296–304 (2012)

44.6 H.L. Tolman: User manual and system documentation of WAVEWATCH-III, Version 2.22, NOAA/NWS/NCEP/MMAB Technical Note (2002)

44.7 R. Boud, T.W. Thorpe: Wavenet: Results from the work of the European thematic network on wave energy, ERK5-CT-1999-20001, European Community (2003) pp. 307–308

44.8 G. Mørk, S. Barstow, A. Kabuth, T. Pontes: Assessing the global wave energy potential, Proc. 29th Int. Conf. Ocean Offshore Mech. Arct. Eng. (ASME) (2010)

44.9 G. Hagerman, G. Scott: *Mapping and Assessment of the United States Ocean Wave Energy Resource, Tech. Rep. 1024637* (Electric Power Research Institute, Palo Alto 2011)

44.10 AVISO: Sun and Moon shape tides on Earth, Published online October, 2000. http://www.aviso.oceanobs.com/en/news/idm/2000/oct-2000-sun-and-moon-shape-tides-on-earth/index.html

44.11 R.H. Charlier, J.R. Justus: *Ocean Energies: Environmental, Economic and Technological Aspects of Alternative Power Sources*, Elsevier Oceanography Series (Elsevier, Amsterdam 1993)

44.12 A. Lewis, S. Estefen, J. Huckerby, W. Musial, T. Pontes, J. Torres-Martinez: Ocean energy. In: *IPCC Special Report on Renewable Energy Sources and Climate Change Mitigation*, ed. by O. Edenhofer, R. Pichs-Madruga, Y. Sokona, K. Seyboth, P. Matschoss, S. Kadner, T. Zwickel, P. Eickemeier, G. Hansen, S. Schlömer, C. von Stechow (Cambridge Univ. Press, Cambridge 2011)

44.13 M. Pidwirny: Surface and subsurface ocean currents: Ocean current map. In: *Fundamentals of Physical Geography*, 2nd edn. (eBook) http://www.physicalgeography.net/fundamentals/8q_1.html (2006)

44.14 A.E. Gill: *Atmosphere-Ocean Dynamics*, International Geophysics Series, Vol. 30 (Oxford Academic Press, Oxford 1982)

44.15 W.H. Munk: On the wind-driven ocean circulation, J. Meteorol. **7**(2), 79–93 (1950)

44.16 R.H. Stewart: *Introduction to Physical Oceanography*, http://oceanworld.tamu.edu/resources/ocng_textbook/PDF_files/book_pdf_files.html

44.17 A. Betz: Das Maximum der theoretisch möglichen Ausnützung des Windes durch Windmotoren, Z. Gesamte Turbinenwesen **26**, 307–309 (1920)

44.18 G.A.M. Van Kuik: The Lanchester–Betz–Joukowsky limit, Wind Energy **10**, 289–291 (2007)

44.19 V.L. Okulov, J.N. Sørensen: Refined Betz limit for rotors with a finite number of blades, Wind Energy **11**, 415–426 (2008)

44.20 C. Garrett, P. Cummins: The efficiency of a turbine in a tidal channel, J. Fluid Mech. **588**, 243–251 (2007)

44.21 T. Nishino, R.H.J. Willden: The efficiency of an array of tidal turbines partially blocking a wide channel, J. Fluid Mech. **708**, 596–606 (2012)

44.22 C. Garrett, P. Cummins: The power potential of tidal currents in channels, Proc. R. Soc. A. **461**, 2563–2572 (2005)

44.23 G. Sutherland, M. Foreman, C. Garrett: Tidal current energy assessment for Johnstone Strait, Vancouver Island, Proc. Inst. Mech. Eng, J. Power Energy A. **221**(2), 147–157 (2007)

44.24 K.A. Haas, H.M. Fritz, S.P. French, B.T. Smith, V. Neary: *Assessment of Energy Production Poten-*

tial from Tidal Streams in the United States: Final Report (Georgia Tech Research Corporation, Savannah 2011)

44.25 T.A.A. Adcock, S. Draper, G.T. Houlsby, A.G.L. Borthwick, S. Serhadlıoğlu: The available power from tidal stream turbines in the Pentland Firth, Proc. R. Soc. A **469**(2157), 20130072 (2013)

44.26 A.E.S. Duerr: A Hydrokinetic Resource Assessment Of the Florida Current, Ph.D. Thesis (Florida Atlantic University, Boca Raton 2012)

44.27 A.E.S. Duerr, M.R. Dhanak: An assessment of the hydrokinetic energy resource of the Florida Current, IEEE J. Ocean. Eng. **37**, 281–293 (2012)

44.28 A.E. Duerr, M.R. Dhanak, J.H. Van Zwieten: Utilizing the hybrid coordinate ocean model for the assessment of Florida Current's hydrokinetic renewable energy resource, Mar. Technol. Soc. J. **46**(5), 24–33 (2012)

44.29 J.H. Van Zwieten Jr., A.E.S. Duerr, G.M. Alsenas, H.P. Hanson: Global ocean current energy assessment: An initial look, Proc. 1st Mar. Energy Technol. Symp. (METS) (2013)

44.30 E.P. Chassignet, H.E. Hurlburt, E.J. Metzger, O.M. Smedstad, J. Cummings, G.R. Halliwell, R. Bleck, R. Baraille, A.J. Wallcraft, C. Lozano, H.L. Tolman, A. Srinivasan, S. Hankin, P. Cornillon, R. Weisberg, A. Barth, R. He, F. Werner, J. Wilkin: US GODAE: Global ocean prediction with the hybrid coordinate ocean model (HYCOM), Oceanography **22**(2), 64–75 (2009)

44.31 N. Maximenko, P. Niiler, M.-H. Rio, O. Melnichenko, L. Centurioni, D. Chambers, V. Zlotnicki, B. Galperin: Mean dynamic topography of the ocean derived from satellite and drifting buoy data using three different techniques, J. Atmos. Oceanic Technol. **26**(9), 1910–1919 (2009)

44.32 P. Knudsen, R. Bingham, O. Andersen, M.-H. Rio: A global mean dynamic topography and ocean circulation estimation using a preliminary GOCE gravity model, J. Geod. **85**, 861–879 (2011)

44.33 J.H. VanZwieten Jr., I. Meyer, G.M. Alsenas: Evaluation of HYCOM as a tool for ocean current energy assessment, Proc. 2nd Mar. Energy Technol. Symp. (METS) (2014)

44.34 J.H. VanZwieten Jr., W.E. Baxley, G.M. Alsenas, I. Meyer, M. Muglia, C. Lowcher, J. Bane, M. Gabr, R. He, T. Hudon, R. Stevens, A.E.S. Duerr: Ocean current turbine mooring considerations, Proc. Offshore Technol. Conf. (2015), OTC-25965-MS

44.35 G. Jeans, L. Harrington-Missin, C. Herry, M. Prevosto, C. Maisondieu, J.A.M. Lima: Deepwater current profile data sources for riser engineering offshore Brazil, Proc. 31st Conf. Ocean Offshore Arct. Eng. (ASME) (2012), OMAE2012-83400

45. Ocean Wave Energy Conversion Concepts

Nikolaos I. Xiros, Manhar R. Dhanak

The purpose of this chapter is to present the basic concepts of ocean wave energy conversion as an extension to the supporting subjects presented earlier in this handbook so that they can be useful to scientists, engineers and inventors. Although new research regarding all wave energy conversion aspects has been unveiling continuously over the past few decades and is expected to remain so in the foreseeable future, they are based on only a few fundamental wave energy conversion techniques. These methods are described and their uses and performance are illustrated. Finally, an account of several recent developments and advances in ocean wave energy research is presented as an appendix with an emphasis to commercialization of different technologies in conjunction with their financial viability, techno-economic, and environmental impact.

45.1 **Basic Concepts in Primary Energy Capture** 1118
45.1.1 Point Absorbers 1118
45.1.2 Oscillating Water Column 1121
45.1.3 Submerged Pressure Differential Devices 1122
45.1.4 Oscillating Wave Surge Convertors 1123
45.1.5 Attenuator and Terminator 1125
45.1.6 Overtopping Devices 1126
45.1.7 Other Advanced Concepts 1126
45.1.8 Control 1126
45.2 **Power Takeoff Systems** 1127
45.2.1 Mechanical Transmission PTOs 1127
45.2.2 Pneumatic and Hydro Turbines 1129
45.2.3 Hydraulic Power Takeoff Systems 1130
45.2.4 Linear Generator 1131
45.2.5 Other Considerations 1132
45.A **Appendix: Practical Applications of Wave Energy Conversion Technologies** 1133
45.A.1 Point Absorber Systems 1133
45.A.2 Submerged Pressure Differential Devices 1136
45.A.3 Oscillating Water Column Systems 1137
45.A.4 Oscillating Wave Surge Convertors 1141
45.A.5 Attenuators 1142
45.A.6 Overtopping Devices 1143
References 1143

Energy of progressive ocean surface waves transported by the waves across the ocean to the shores is a significant source of renewable energy worldwide. For a regular progressive wave of small height H, circular frequency ω, and wavelength λ propagating along the direction x, given by the surface elevation

$$\eta(x,t) = \frac{H}{2}\cos(\omega t - kx)\,, \tag{45.1}$$

where $k = 2\pi/\lambda$ is the wave number, the rate of transport of energy per unit length of the wave frontage J is given by

$$J = \frac{\mathrm{d}\hat{E}}{\mathrm{d}t} = \frac{1}{8}\rho g H^2 c_\mathrm{g}\,. \tag{45.2a}$$

Here, $\hat{E}$ is the mean energy density and c_g is the group speed of the waves; in deep water, using

$$c_\mathrm{g} = \mathrm{g}/2\omega\,,$$

we have

$$J = \frac{\rho g^2 H^2}{16\omega}\,. \tag{45.2b}$$

Equations (45.2) represent the available power per unit length of the wave frontage that can potentially be harnessed. A wave energy converter (WEC) is the name given to a device that captures and converts wave energy for electricity generation or other useful work.

45.1 Basic Concepts in Primary Energy Capture

A number of WEC devices have been designed and developed, although there is no clear best device and to date only a few devices have been deployed in the field. Several factors influence concept designs, including whether the device is located offshore, near-shore, or onshore or whether the device is to be floating, submerged, or bottom standing, and whether the energy conversion machinery is mechanical, hydraulic, or pneumatic. The choice of a particular WEC method of primary energy capture depends on these and other factors, including environmental impact, economic, and marine spatial planning considerations. *Grimwade* et al. [45.1], *Falcão* [45.2], and *Drew* et al. [45.3] provide a detailed review of WEC technologies. The basic concepts of primary energy capture of the typical WEC devices are typically classified as:

- *Point absorber* is typically a floating buoy of size small compared with the length of prevailing waves and captures wave energy through wave-induced heaving or pitching motion.
- *Oscillating water column* involves a tube-like structure with air in the tube above a column of water that oscillates and acts as a piston in compressing and decompressing the air and ducting it through a pneumatic turbine.
- *Submerged pressure differential device* is based on exploiting the dynamic pressure differential induced underwater through the motion of waves over the device.
- *Oscillating wave surge converter* extracts energy from the surge motion of the waves. They are generally seabed mounted devices located in near shore sites.
- *Attenuator and terminator* are generally long floating structures aligned in parallel with the wave direction, which then absorbs the motion caused by the waves. Its motion can be damped to produce energy.

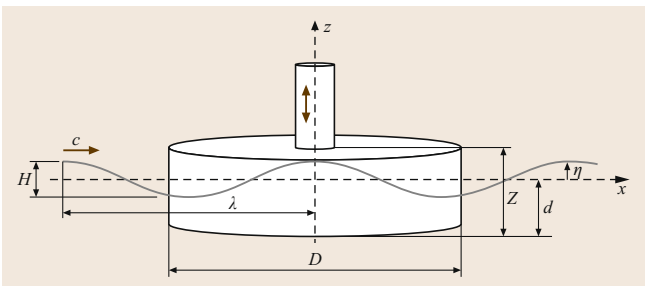


Fig. 45.1 Floating body undergoing purely heaving motion in waves

- *Overtopping devices*are composed of a wave surge/focusing system, and contains a ramp over which waves travel into a raised storage reservoir.

These concepts are discussed in detail below.

45.1.1 Point Absorbers

Point absorbers are small floating or submerged buoys that absorb wave energy through their heaving or pitching motion in response to the action of the waves. The size of the buoy is typically small compared with the wavelength λ of the dominant wave, or wave with the most energy, at the site. The heaving/pitching motion of the buoy in turn excites an electromechanical energy generator via a power takeoff (PTO) system.

A floating system such as the one shown in Fig. 45.1 undergoes heaving motion according to the following equation

$$(m + a_{33})\,\ddot{z} + b_{33}\dot{z} + Kz = F_Z + F_d + F_R\,, \qquad (45.3)$$

where z is the vertical displacement of the floating system, m is the mass of the system, a_{33} is the added mass, b_{33} is the coefficient of radiation damping associated with waves generated by the heaving response, K is the hydrostatic restoring force coefficient, given by $K = \rho g A_{wp}$ for a freely floating body with ρ the mass density of seawater and A_{wp} the water plane area of the float, F_Z is the wave exciting force, F_d is the viscous (nonlinear) damping force, and F_R is the force due to an associated PTO and other resisting forces and is represented here as mechanical damping, given by $F_R = -R_m\dot{z}$. In practice, the buoy in Fig. 45.1 would be connected to a large damping plate (Fig. 45.2) so that F_R includes additional contributions with spring-like restoring force, proportional to z and added-mass-type inertia force proportional to $\ddot{z}$. See [45.4] and (45.42) for the inclusion of these additional contributions to F_R.

The natural heaving frequency of the floating body is

$$f_z = \frac{1}{T_z} = \frac{\omega_z}{2\pi} = \frac{1}{2\pi}\sqrt{\frac{K}{m + a_{33}}}\,, \qquad (45.4)$$

where T_z is the natural heaving period, and ω_z is the natural circular heaving frequency; see Sect. 45.2.4 for the modification to ω_z to include restoring force and inertia forces associated with the PTO system. The design of a point absorber WEC is based on approximately matching this natural frequency with the frequency of the dominant waves at the site, and inducing resonant

or near resonant heaving motion of the floating body. A WEC is designed to ensure that f_z lies in the range of frequencies of waves at the site; this is achieved through the selection of its mass and restoring force.

For purely heaving motion in regular waves of frequency ω, given by (45.3), the corresponding induced wave force F_Z (choosing $x = 0$ to coincide with body center) may be expressed as

$$F_Z = F_0 \cos(\omega t + \gamma) \,, \tag{45.5}$$

where γ is a phase angle associated with the wave force ($\gamma = 0$ for a body that is symmetric in the x-z and y-z planes). Then from (45.3), on choosing $F_R = -R_m \dot{z}$ and ignoring the viscous forces F_d, the body displacement z is given by

$$\begin{aligned} z &= \frac{\left(\frac{F_0}{K}\right)\cos(\omega t + \gamma - \sigma_z)}{\sqrt{\left(1 - \frac{\omega^2}{\omega_z^2}\right)^2 + \left[2\frac{(\Delta_m + \Delta_D)\omega}{\omega_z}\right]^2}} \\ &= Z_0 \cos\left(\omega t + \gamma - \sigma_z\right) \,, \end{aligned} \tag{45.6}$$

where Z_0 is the amplitude of the motion, $\sigma_z = \tan^{-1}\left(2(\Delta_m + \Delta_D)\omega\omega_z/(\omega_z^2 - \omega^2)\right)$ is the phase lag between the wave motion and the heave response, and $\Delta_m = R_m\omega_z/2K$ and $\Delta_D = b_{33}\omega_z/2K$ are the dimensionless mechanical and radiation damping coefficients, respectively. If, for simplicity, the frequency dependence of $\Delta_Z = (\Delta_m + \Delta_D)$ and a_{33} can be neglected and these quantities considered as constants, then the amplification factor $Z_0/(F_0/\rho K)$ and the phase angle σ_z may be depicted as in Fig. 45.3 as functions of ω/ω_z. At resonance, $\sigma_z = \pi/2$, $\omega/\omega_z = 1$ and $Z_0 = F_0/(2(\Delta_m + \Delta_D)K) = F_0/((R_m + b_{33})\omega_z)$.

At resonance, the heave amplitude achieves a maximum value, the magnitude of which is governed by the damping factor. Minimizing the damping factor and the restoring force, as part of designing a WEC, contributes to maximizing the resonance response. The heave velocity and acceleration of the body are, respectively, given by

$$\begin{aligned} \frac{\mathrm{d}z}{\mathrm{d}t} &= -\omega Z_0 \sin(\omega t + \gamma - \sigma_z) \quad \text{and} \\ \frac{\mathrm{d}^2 z}{\mathrm{d}t^2} &= -\omega^2 Z_0 \cos(\omega t + \gamma - \sigma_z) \,. \end{aligned} \tag{45.7}$$

The exciting mechanical power is given by $F_Z\frac{\mathrm{d}z}{\mathrm{d}t}$, where F_Z is given by (45.5). Thus, choosing $\gamma = 0$ for a symmetrical body, the average exciting power over one wave period, $\hat{P}_{Ez}$, is given by

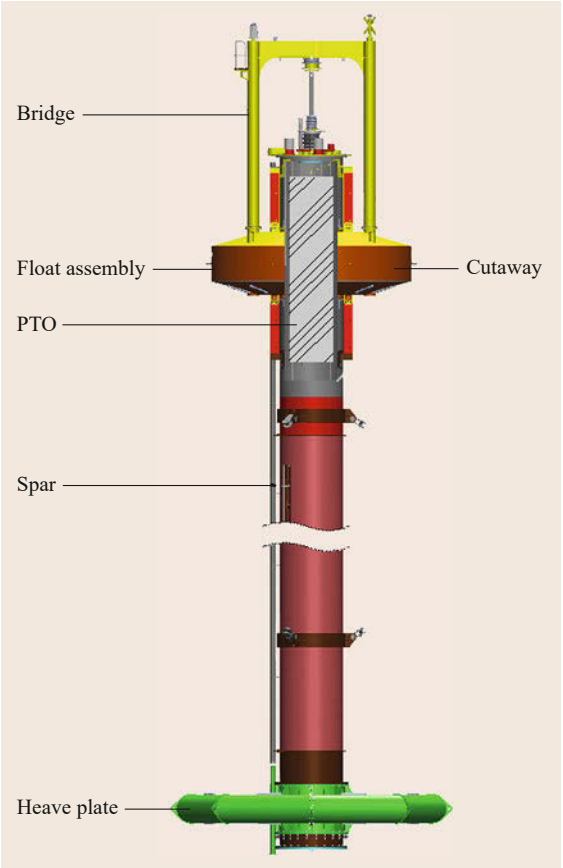


Fig. 45.2 Practical floating WEC buoy, PowerBuoy (from Ocean Power Technologies)

$$\begin{aligned} \hat{P}_{Ez} = \frac{1}{T_z}\int_0^{T_z} F_0\omega Z_0[-\cos\sigma_z \sin\omega t\cos\omega t \\ + \sin\sigma_z \cos^2(\omega t)]\mathrm{d}t = \frac{F_0\omega Z_0 \sin\sigma_z}{2} \,. \end{aligned} \tag{45.8}$$

The radiated power and the absorbed or mechanical power are, respectively, given by $b_{33}\dot{z}^2$ and $R_m\dot{z}^2$, so that the corresponding average radiated and absorbed power over one wave period, $\hat{P}_{Rz}$ and $\hat{P}_{az}$ are given by

$$\hat{P}_{Rz} = \frac{b_{33}(\omega Z_0)^2}{2} \,; \quad \hat{P}_{az} = \frac{R_m(\omega Z_0)^2}{2} \,. \tag{45.9}$$

$\hat{P}_{az}$ is the absorbed power that is utilized. It is optimized if $R_m = R_{m_{opt}}$ is selected so that $(\partial\hat{P}_{az}/\partial R_m)_{R_{m_{opt}}} = 0$

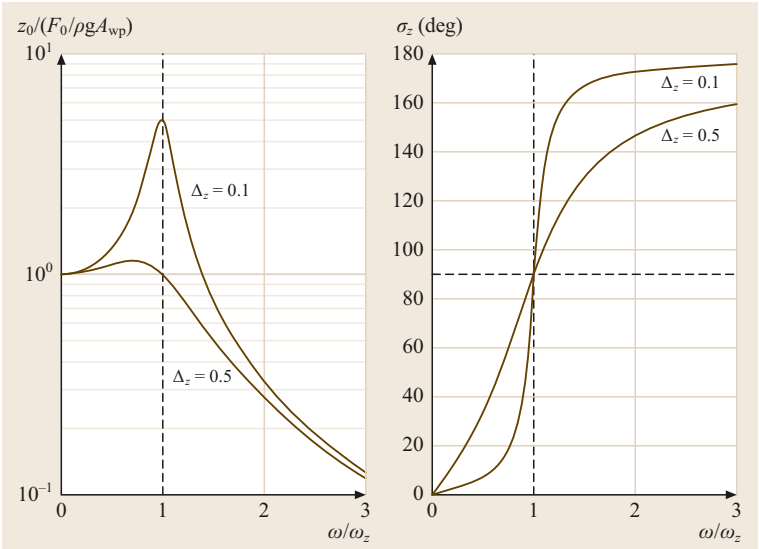


Fig. 45.3 Amplification factor and phase angle variation in heaving motion of a floating body; $\Delta_Z = \Delta_m + \Delta_D = (R_m + b_{33})\omega_z/2K$

[45.5]. Thus, for $R_m^2 \equiv R_{m_{opt}}^2 = K^2/\omega^2\left(1-\omega^2/\omega_z^2\right)^2 + b_{33}^2$, we obtain

$$\hat{P}_{az_{opt}}(\omega) = \frac{\frac{\omega F_0^2}{4}}{\omega b_{33} + \sqrt{K^2\left(1-\frac{\omega^2}{\omega_z^2}\right)^2 + \omega^2 b_{33}^2}} \,. \tag{45.10}$$

$\hat{P}_{az_{opt}}$ attains its maximum value when $\omega = \omega_z$, for which $R_m = b_{33}$ and

$$\hat{P}_{az_{opt}}(\omega_z) = \frac{F_0^2}{8b_{33}} \quad \text{and} \quad \hat{P}_{Ez_{opt}} = \frac{F_0^2}{4b_{33}} \,. \tag{45.11}$$

Equations (45.11) imply a maximum 50% efficiency $\eta_{eff_{max}} = \hat{P}_{az_{opt}}/\hat{P}_{Ez_{opt}}$ in the case of an axisymmetric body. For nonsymmetric bodies, involving destructive interference between the incident wave and waves generated by the heaving float downstream of the body, higher efficiencies may be expected [45.6, 7].

It may be noted from (45.5) and (45.7) that at the optimal resonance condition, with $\sigma_z = \pi/2$, $\omega = \omega_z$, and $R_m = b_{33}$, the heaving velocity of the buoy, $dz/dt = \omega_z Z_0 \cos(\omega_z t + \gamma) = F_Z/2b_{33}$, is in phase with the incident wave forcing.

Neglecting the frequency dependence of a_{33} and b_{33} (typically, this is so at high frequencies) the ratio $[\hat{P}_{az_{opt}}(\omega)/F_0^2(\omega)]/[\hat{P}_{az_{opt}}(\omega_z)/F_0^2(\omega_z)]$ is plotted in Fig. 45.4 for two different values of Δ_D.

Inspection of (45.10) and Fig. 45.3 suggests that while a smaller value of Δ_D means a larger value of $\hat{P}_{az_{opt}}(\omega_z)$, it implies a narrower range of frequencies over which wave energy can be harnessed. *Falnes* [45.5] shows that the maximum absorbed power $\hat{P}_{az}$ that can be utilized may alternately be given by

$$\hat{P}_{az_{opt}}(\omega) = \left(\hat{P}_{Ez} - \hat{P}_{Rz}\right)|_{R_m = R_D} = \frac{1}{2}\left[F_0 \omega Z_0 \sin\sigma_z - b_{33}(\omega Z_0)^2\right],$$

which is maximum when $\omega Z_0 = F_0 \sin\sigma_z/2b_{33}$, giving

$$\hat{P}_{az_{max}} = \frac{F_0^2 \sin^2\sigma_z}{8b_{33}} \,, \tag{45.12}$$

which at the condition of resonance, when $\sigma_z = \pi/2$, is the same as $\hat{P}_{az_{opt}}(\omega_z)$ in (45.11).

An absorption or capture width, $d_a = \hat{P}_{az}/J$, where J given by (45.1), is defined as the width of the wave front across which passes an average amount of power equal to that absorbed by the point absorber; note that J is the power density per unit length of the wave frontage. Thus, the maximum absorption width for a point absorber is $d_{a_{max}} = F_0^2/(8Jb_{33})$. For a heaving cylinder

$$b_{33} = \frac{2\pi F_0^2}{\lambda \rho g H^2 c_g} \tag{45.13}$$

using reciprocity [45.8]. Then we have from (45.12) with $\sigma_z = \pi/2$,

$$\hat{P}_{az_{max}} = \frac{F_0^2}{8b_{33}} = \frac{\lambda \rho g H^2 c_g}{16\pi} = \frac{\lambda}{2\pi} J \quad \text{and}$$
$$d_{a_{max}} = \frac{\lambda}{2\pi} \,. \tag{45.14a}$$

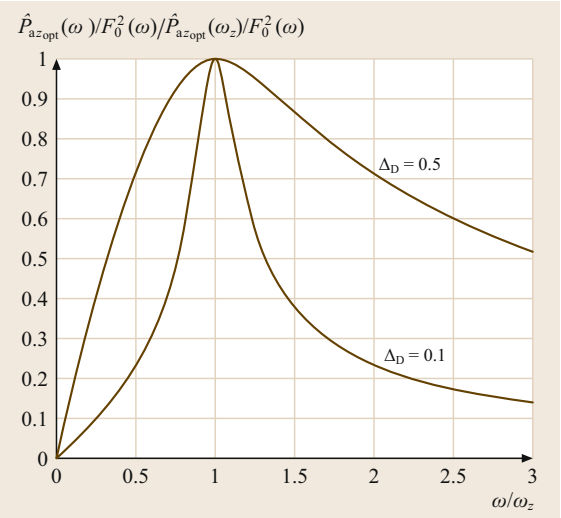


Fig. 45.4 Characteristics of optimal power absorption for two values of Δ_D

In deep water, using $c_g = g/2\omega$, and $\lambda = 2\pi g/\omega^2$, this becomes

$$\hat{P}_{az_{max}} = \frac{\rho g^3 H^2}{16\omega^3} \,. \tag{45.14b}$$

Thus a point absorber absorbs much more energy than it is incident on its cross-section [45.9–11]. Since the wave energy is typically available in a narrow spectral frequency band, a properly tuned point absorber can be an effective WEC. *Budal* [45.12] has shown that for two parallel rows of linear arrays in normally incident waves, the spacing can be as large as λ for perfect absorption, owing to the interference of adjacent buoys.

Similar considerations apply for pitching and combined heaving and pitching motions of a floating body [45.13, 14]. *Falnes* [45.7] provides consideration for the effect of inclusion of a PTO system (see [45.4] and Sect. 45.2.4), as well as consideration of arrays of point absorbers.

Well-known examples of point absorber systems include the Power Buoy, Seabased, Wavebob, AquabuOY, CETO III, and Wave Star.

45.1.2 Oscillating Water Column

An oscillating water column (OWC) is a hollow structure that is partially submerged, being open to the sea below the water surface. It encloses a column of air above the column of water (Fig. 45.4). Waves induce the water level in the structure to rise and fall, leading to compression and expansion of the air column. The trapped air flows to and from the open air through a turbine that can rotate due to airflow in either direction. The mechanical conversion of energy via the rotation of the turbine is used to generate electricity. *Heath* [45.15], and *Falcão* and *Henriques* [45.16] provide reviews of OWC WECs.

The instantaneous air pressure in the air chamber in Fig. 45.4 is commonly modeled as being uniform, taking into account the compressibility of the air in the chamber [45.5, 17–19]. Let the uniform air pressure in the chamber be given by, $p(t) = p_a(t) + p_{atm}$, where $p_a(t)$ represents the variation in pressure in the chamber in response to the oscillation of the water column from the equilibrium pressure p_{atm}, which is considered the same as the external ambient pressure and is assumed constant. For isentropic flows,

$$\frac{p}{p_{atm}} = \left(\frac{\rho}{\rho_{atm}}\right)^{\gamma} , \tag{45.15a}$$

where $\rho(t) = \rho_a(t) + \rho_{atm}$ and ρ_{atm} are the respective instantaneous air densities in the chamber corresponding to pressures $p(t)$, and equilibrium pressure p_{atm}, and $\gamma = 1.4$ under adiabatic conditions. $p_a(t)$ is the excess pressure in the chamber over the ambient. On the basis of linear theory with $|p_a| \ll p_{atm}$, (45.15a) can be shown using a Taylor expansion, to imply

$$\frac{d\rho_a}{dt} = \frac{1}{c_0^2}\frac{dp_a}{dt} \,; \quad c_0 = \sqrt{\frac{dp}{d\rho}} \approx \sqrt{\frac{\gamma p_{atm}}{\rho_{atm}}} \,. \tag{45.15b}$$

Here, c_0 is the speed of sound that is constant to leading order in the linearized analysis. The instantaneous rate of change of mass in the chamber in response to the OWC is then given by

$$\begin{aligned} \dot{m} &= \frac{d(\rho_a \Omega)}{dt} = \rho_a \frac{d\Omega}{dt} + \Omega \frac{d\rho_a}{dt} \\ &\approx \rho_{atm} \frac{d\Omega}{dt} + \Omega_0 \frac{1}{c_0^2}\frac{dp_a}{dt} \,, \end{aligned} \tag{45.16}$$

where $\Omega(t)$ is the instantaneous volume of the air chamber, and Ω_0 is the volume of the air chamber under equilibrium conditions. Then by the conservation of mass, we have

$$\dot{m} = -\rho_{atm} Q_{out} \,, \tag{45.17}$$

where Q_{out} is the airflow rate across the turbine. Considering that Q_{out} is primarily a function of the pressure difference across the turbine. Thus, linearizing about the equilibrium conditions,

$$Q_{out}(p_a) \approx Q_{out}(0) + p_a \left.\frac{\partial Q_{out}}{\partial p_a}\right|_{p_a=0} \,, \tag{45.18}$$

where $Q_{out}(0) \approx 0$ is the mean flow through the turbine for flow driven purely by the oscillatory motion; $(\partial Q/\partial p_a)|_{p_a=0}$ depends on the characteristics of the generator. *Sarmento* and *Falcão* [45.18] suggest that for a Wells turbine, $\dot{m} \approx (KD_2/n)p_a$, where D_2 is the turbine diameter (Fig. 45.4), n is the rotational speed of the turbine, p_a is the pressure drop across the turbine, and K is a dimensionless design constant that depends on the turbine characteristics; thus, from (45.18), $(\partial Q_{out}/\partial p_a)|_{p_a=0} = -KD_2/(\rho_{atm} n)$. Then from (45.16), we have

$$\frac{d\Omega}{dt} \approx \frac{KD_2}{\rho_{atm} n} p_a - \frac{\Omega_0}{\rho_{atm} c_0^2} \frac{dp_a}{dt} . \tag{45.19}$$

If the structure is fixed, $d\Omega/dt$ is given by the rate of volume of water that is pumped into the chamber by the wave motion,

$$\frac{d\Omega}{dt} = \iint_{S_1} w_1 dS \approx \iint_{S_1} \frac{d\eta_1}{dt} dS , \tag{45.20}$$

where $w_1(r, \theta, t)$ is the vertical velocity at the internal free surface, given by the small displacement $z = \eta_1$ of the surface as shown in Fig. 45.4, and the integral is over the area S_1 of the internal water surface; the last term on the right is based on the linearized kinematic boundary condition at the free surface. *Falnes* [45.5], and *Martins-Rivas* and *Mei* [45.17] provide a detailed analysis of the hydrodynamics associated with the OWC to enable determining Q and p_a [45.16], posing a linearized potential flow problem and decomposing it into diffraction and radiation problems. The full details of the analysis are fairly complex and can be found in these papers.

The power output rate, averaged over an oscillatory period, at which the chamber pressure does work in pushing the air through the turbine is

$$\begin{aligned} P_{out} &= \overline{\dot{m} \frac{p_a}{\rho_{atm}}} = \frac{KD_2}{n\rho_{atm}} \overline{p_a^2} \\ &\approx \left(\overline{\frac{d\Omega}{dt} p_a} + \frac{1}{2c_0^2 \rho_{atm}} \overline{\frac{dp_a^2}{dt}} \right) = \overline{\frac{d\Omega}{dt} p_a} , \end{aligned} \tag{45.21}$$

since $\overline{dp_a^2/dt}$ averages to zero for oscillatory variation of the pressure.

If the hollow structure is considered to be in the form of a fixed vertical circular pipe shown in Fig. 45.5, then the water column in the pipe has a natural frequency of oscillation [45.13]

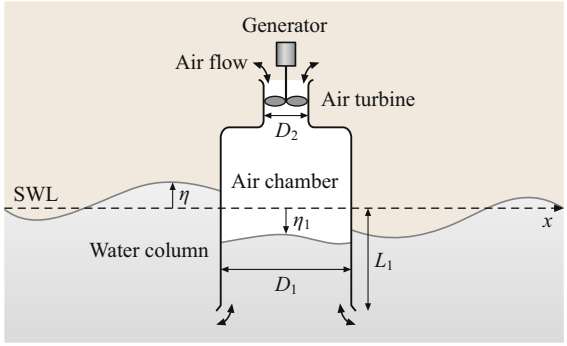


Fig. 45.5 Schematics of an OWC WEC

$$\begin{aligned} f_c = \frac{1}{T_c} = \frac{\omega_c}{2\pi} &\approx \frac{1}{2\pi} \sqrt{\frac{\rho g S_1}{\rho S_1 (L_1 + L_1')}} \\ &= \frac{1}{2\pi} \sqrt{\frac{g}{L_1 + L_1'}} , \end{aligned} \tag{45.22}$$

where L_1 is the undisturbed length of the water column and L_1' represents a contribution due to the added mass associated with the motion of the water column. When the frequency of a sinusoidal wave matches ω_c, resonance occurs and $|d\Omega/dt|$ is maximum. If the structure also undergoes heaving motion, a peak value in the air velocities can also be expected at the heaving resonant frequency ω_z of the structure given by (45.3) [45.20]. The air flow is optimized if the structure is designed so that $\omega_c = \omega_z$.

45.1.3 Submerged Pressure Differential Devices

For a surface wave of small amplitude in water of depth h, described by (45.1), the pressure in the water column is given by

$$p = \rho g \frac{\cosh[k(z+h)]}{\cosh(kh)} \eta(x,t) - \rho g z , \tag{45.23}$$

so that it varies in tune with the free-surface elevation. That is, the dynamic pressure is high at a location under a crest and low under a tough. Submerged pressure differential devices, which are typically located on the seabed near shore, exploit this variation for wave energy conversion to electricity. A well-known device in this category is the Archimedes wave swing (AWS), which heaves in response to the passage of a wave over it. A schematic of the device is shown in Fig. 45.6. AWS is an air-filled cylindrical chamber that consists of two parts, a lid, known as the

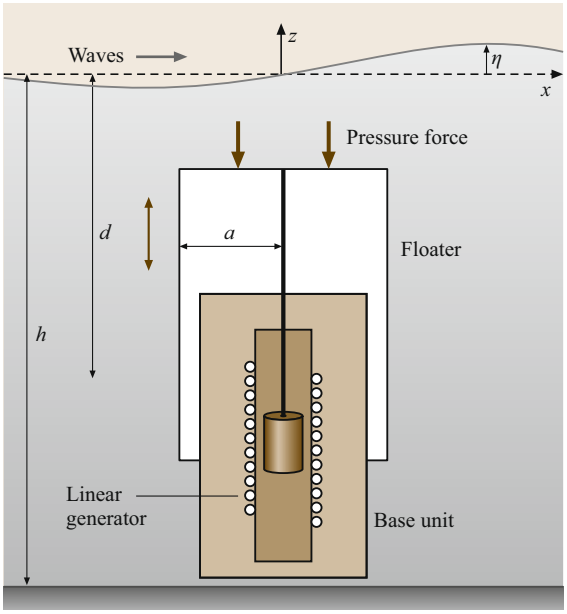


Fig. 45.6 Schematics of submerged pressure differential device, AWS (after [45.2])

floater, which heaves in response to wave motion, and a fixed bottom unit that houses a linear electric generator [45.21]. The floater, as it heaves, is closest to the base unit when under a crest and furthest from it when under a trough. By adjusting natural heaving frequency to match the average wave frequency, optimal energy harvesting conditions can be attained. The motion of the floater is governed by an equation similar to (45.3), with the spring constant K being given by $K = -(\mathrm{d}F_\mathrm{s}/\mathrm{d}z)|_{z_0}$, where F_s is the resisting force with contributions from the air pressure inside the chamber, the hydrostatic pressure, and the force associated with the generator [45.21] and the wave-induced force F_z contains contribution to the pressure force from the exciting and the scattered waves. Other forces are similar to the those defined in (45.3). The exciting force can be estimated by the Froude–Krylov force on the top of the cylindrical floater is given by

$$F_\mathrm{KR} = \int_0^a \int_0^{2\pi} pr\mathrm{d}r\mathrm{d}\theta \approx \frac{\pi\rho g H a}{k} \frac{\cosh(-kd + kh)}{\cosh(kh)} J_1(ka)\cos(\omega t)\,, \tag{45.24}$$

where a is the radius of the floater and $J_1(x)$ is Bessel's function of the first kind of order 1 and the force is in phase with the wave with the device located at $x = 0$.

Alternatively, *McCormick* [45.13] describes a system where the floater in Fig. 45.5 essentially is replaced by a piston of radius a and the linear motor is replaced by a hydraulic pump with a smaller piston of radius a_2. The wave-induced pressure sets the large piston into heaving motion and it in turn drives the smaller piston of the pump. The pressure force on the large piston surface transfers through to the smaller piston, resulting in dynamic hydraulic pressure whose amplitude is larger by the factor $(a/a_2)^2$. The hydraulic pump may be part of a PTO system for the WEC.

Examples of submerged pressure differential devices include the PYSIS, CETO I, and AWSI.

45.1.4 Oscillating Wave Surge Convertors

An oscillating wave surge converter (OWSC) typically comprises of a body that surges back and forth in response to horizontal particle velocity in the water column associated with a surface wave, either (a) as a hinged flap (Fig. 45.7a) or (b) as a piston (Fig. 45.7b). *Gomes* et al. [45.22] provide a review of OWSCs. A commercial form of a hinged body is the Aquamarine Power Oyster [45.23], while a patented device by

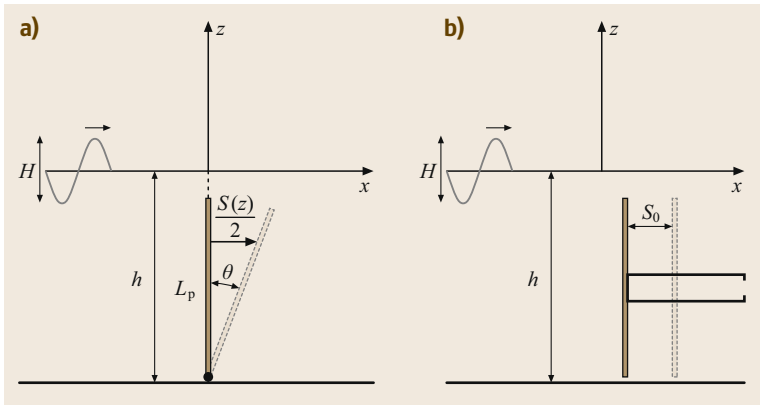


Fig. 45.7 (a) Hinged-flap-type OSWC. (b) Piston-type OSWC

E. A. Wall (1911) [45.13] provides an example of the horizontal piston-based system.

The stroke $S(z)$ of the flap or piston is given by

$$S(z) = \begin{cases} S_0\left(1+\frac{z}{h}\right) & \text{Flapcase}\,, \\ S_0 & \text{Pistoncase}\,. \end{cases} \tag{45.25}$$

For a regular progressive wave of small height H, circular frequency ω, and wavelength λ propagating along the direction x, given by the surface elevation (45.1), the horizontal particle velocity is given by

$$u = \frac{\omega H}{2}\frac{\cosh k(z+h)}{\sinh kh}\cos(\omega t - kx)\,. \tag{45.26}$$

Following *Milgram* [45.24], *Dean* and *Dalrymple* [45.25] suggest that in constrained motion, the reverse of wavemaker theory could be used to determine S_0. Thus, on the basis of linear theory and using the boundary condition on the flap/piston, $u(0,z,t) = S(z)/2\omega\cos(\omega t)$, they show, with reference to Fig. 45.7, that the stroke needed to absorb the wave energy incident on the flap/piston is given by

$$S_0 = \begin{cases} \frac{H}{4}\frac{kh}{\sinh(kh)}\frac{\sinh 2kh + 2kh}{kh\sinh kh - \cosh kh + 1} \\ \quad \text{Flapcase}\,, \\ \frac{H}{2}\frac{\sinh 2kh + 2kh}{(\cosh 2kh - 1)} \\ \quad \text{Pistoncase}\,. \end{cases} \tag{45.27}$$

In practice, incident waves will be diffracted at the finite-width WEC and secondary waves will be radiated from the oscillating body. Further, some of the energy will be lost to damping associated with the PTO system.

In the case of the hinged flap, if we model the torque associated with the mechanical PTO system as being proportional to the angular speed of the flap, $M_\mathrm{m} = -Q_\mathrm{m}\dot{\theta}$, then the equation for the angular displacement θ may be written as

$$\left(I_\mathrm{y} + I_{55}\right)\ddot{\theta} + (b_{55} + Q_\mathrm{m})\dot{\theta} + C\theta = M_\theta + M_\mathrm{d}\,, \tag{45.28}$$

where I_y is the moment of inertia of the flap about the hinge, I_{55} is the added moment of inertia about the hinge, b_{55} is the coefficient of moment of radiation damping, M_θ is the moment of the incident-wave-induced force, M_d is the torque due to viscous forces, C is the hydrostatic restoring moment. If the thickness t_p of the flap is small compared with its length L_f and width B of the flap, and the mass of the plate is uniformly distributed with mass density of the plate given by ρ_f, then

$$I_\mathrm{y} \approx \frac{1}{3}\rho_\mathrm{f} t_\mathrm{f} B L_\mathrm{f}^3\,,$$

$$C = \rho g t_\mathrm{f} B L_\mathrm{f} r_\mathrm{b} - \rho_\mathrm{f} t_\mathrm{f} B L_\mathrm{f} r_\mathrm{g} = \frac{1}{2}\rho g t_\mathrm{f} B L_\mathrm{f}^2\left(1 - \frac{\rho_\mathrm{f}}{\rho}\right)\,,$$

where r_b and r_g are the distance from the hinge of the center of buoyancy and center of mass of the flap, respectively, and for uniform mass distribution, $r_\mathrm{b} = r_\mathrm{g} = L_\mathrm{f}/2$. Then for a regular wave described by (45.1), θ is given by

$$\theta = \frac{\frac{|M_\theta|}{C}\cos(\omega t + \gamma - \sigma_\mathrm{z})}{\sqrt{\left(1 - \frac{\omega^2}{\omega_\theta^2}\right)^2 + \left(\frac{b_{55}+Q_\mathrm{m}}{C}\right)^2\omega^2}}\,, \tag{45.29}$$

where $\omega_\theta = \sqrt{C/(I_\mathrm{y} + I_{55})}$ is the natural frequency of oscillation of the system, γ is a phase angle associated with the moment of the wave force,

$$\sigma_\mathrm{z} = \tan^{-1}\left(2\frac{(\Delta_\theta + \Delta_\mathrm{m})\,\omega\omega_\theta}{\omega_\theta^2 - \omega^2}\right)\,,$$

$$\Delta_\theta = \frac{b_{55}\omega_\theta}{2C}\,,$$

$$\Delta_{\theta m} = \frac{Q_\mathrm{m}\omega_\theta}{2C}\,.$$

The average power generated by the flap is given by

$$P_\mathrm{PTO} = \frac{1}{2}Q_\mathrm{m}\omega^2\,|\theta|^2\,. \tag{45.30}$$

P_PTO has a maximum or optimal value $P_{\mathrm{PTO}_\mathrm{opt}}$ when

$$\omega^2 \mathrm{Q}_\mathrm{m}^2 = (\omega_\theta^2 - \omega^2)^2\left(I_\mathrm{y} + I_{55}\right)^2 + \omega^2 b_{55}^2\,, \tag{45.31}$$

given by

$$P_{\mathrm{PTO}_\mathrm{opt}} = \frac{\frac{1}{4}\omega\,|M_\theta|^2}{\omega b_{55} + \sqrt{(\omega_\theta^2 - \omega^2)^2\left(I_\mathrm{y} + I_{55}\right)^2 + \omega^2 b_{55}^2}} \tag{45.32}$$

or

$$P_{\mathrm{PTO_{opt}}} = \frac{1}{4C}\omega_\theta \left|M_\theta\right|^2 \times \frac{\frac{\omega}{\omega_\theta}}{2\Delta_\theta \frac{\omega}{\omega_\theta} + \sqrt{\left(1-\frac{\omega^2}{\omega_\theta^2}\right)^2 + 4\Delta_\theta^2 \frac{\omega^2}{\omega_\theta^2}}} . \tag{45.33}$$

Gomes et al. [45.22] provide numerical computations of the problem for regular and irregular waves.

Examples of OSWC include the Oyster, Wave Roller, and C-Wave.

45.1.5 Attenuator and Terminator

An extension of the point absorber is a line absorber, involving a number of modular array of oscillating units connected using mechanical joints and oriented so as to optimize the oscillations of the units. In one arrangement, the line absorber is aligned perpendicular to the direction of wave propagation and is referred to as a *terminator*. In another arrangement, it is aligned parallel to the prevailing direction of the wave propagation and is referred to as an *attenuator*. The size of a line absorber may be comparable to the wavelength of typical waves. An example of a terminator WEC is the Salter Duck, involving several pear-shaped ducks pitching with respect to a common *spine*, consisting of one or more horizontal cylindrical sections hinged together that accommodates a bending motion [45.6]. Each duck rocks and turns over in response to the varying dynamic pressure on its paunch as well as the varying hydrostatic pressure, both in phase, converting the oscillating motion to rotational motion of gyroscopes located within the duck (Fig. 45.8). The duck displaces the same volume of water as occupied by orbiting particles beneath the wave crest upstream of the duck. The natural frequency of motion of the duck is given by

$$\omega_\theta = \sqrt{\frac{C}{I_\mathrm{y} + I_{55}}} , \tag{45.34}$$

where C is the hydrostatic restoring moment, I_y is the moment of inertia, and I_{55} is the added moment of inertia about the axis of the spine (Fig. 45.9). The aim is to match this frequency to the frequency of the prevailing waves for resonant response. Wave-induced pitch moments react against a common cylindrical spine of the line absorber. Mathematical analysis and laboratory tests suggest a potential for up to 90% efficiency in capturing the incident wave energy [45.27, 28]; reduced efficiencies are expected in realistic ocean environment.

Examples of attenuator WECs include the Hagen–Cockerell raft [45.29], the Pelamis [45.30], the McCabe Wave Pump, Sea Power Platform, and Wello. The Pelamis involves a similar spine arrangement as the terminator, but oriented parallel to the wave direction. The attenuator follows the motion of the passing waves. Each element of the device is set into motion by the incident waves from crest to trough. The floating elements of the attenuator are usually located either side of suitable PTO module. As the waves pass, they generate relative motion between each element. The relative motion is typically converted to mechanical power in the PTO module, through either a hydraulic circuit or a suitable mechanical gear train. Motion of the hinged joints of the Pelamis is restrained by hydraulic rams that pump fluid into high-pressure accumulators, which absorb the wave power. The absorption of power is maximized when waves are small and the response is minimized in storms through control of the pumping action.

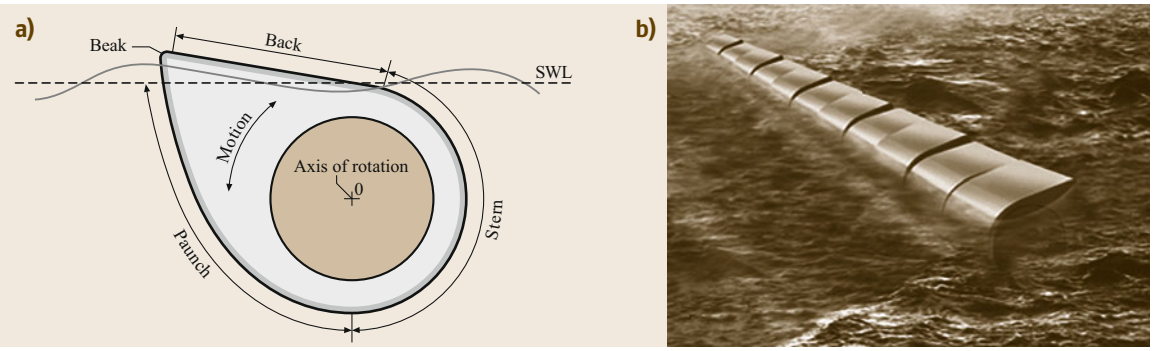


Fig. 45.8a,b Salter's Duck **(a)** individual unit (after [45.13]), **(b)** as a terminator line absorber (after [45.26])

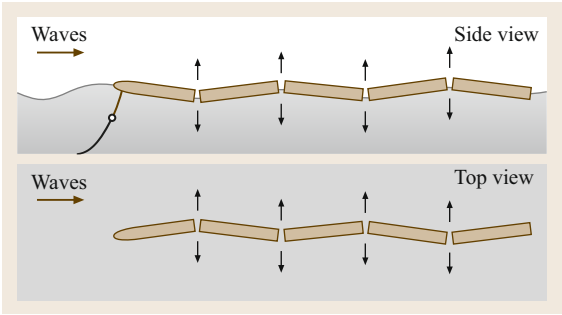


Fig. 45.9 Pelamis motion in waves (after [45.31])

45.1.6 Overtopping Devices

Overtopping devices [45.32] capture seawater driven by incident waves into a reservoir above sea level and through turbines before releasing the water back to sea (Fig. 45.10). Examples of such devices include the Wave Dragon, TAPCHAN, Seawave Slot-Cone Generator, and Floating Wave Power Vessel. Wave Dragon uses a pair of large curved reflectors to channel the flow associated with the waves to flow up a ramp and over the top into a raised reservoir, from which the water passes through a number of low-head turbines and is returned to the sea.

45.1.7 Other Advanced Concepts

An application of piezoelectric crystals to convert wave-induced pressure fluctuations to electrical energy was proposed by *Burfoot* and *Taylor* [45.33] and *Taylor* [45.34], utilizing the property that compressional piezoelectric material produces a voltage difference in response to a pressure force. The concept consisted of a system of piezoelectric films stacked together. Further, *Salomon* and *Harding* [45.35] proposed a *protonic conduction* method for WEC, involving pumping of hydrogen gas by wave-induced motion from one chamber to another across a porous-conducting electrodes, the gas deposits electrons on the electrodes, thereby generating an electrical current in an external circuit. Further details on these and related topics are provided by *McCormick* [45.13]; see also [45.36]. *Alam* [45.37] has proposed the development of a synthetic-seabed carpet for wave energy absorption on the basis of observations that muddy seafloors effectively overpassing ocean waves within a few wavelengths of the waves. The carpet drives hydraulic generators for conversion of wave energy to electricity. Professor *Ronald Yeung* and his team at Berkeley [45.38] report the development of a highly efficient tethered *Berkeley Wedge* point absorber that exploits its nonsymmetric shape to absorb wave energy, achieving around 90% efficiency in regular waves.

Several commercial WEC systems are described in Appendix 45.A.

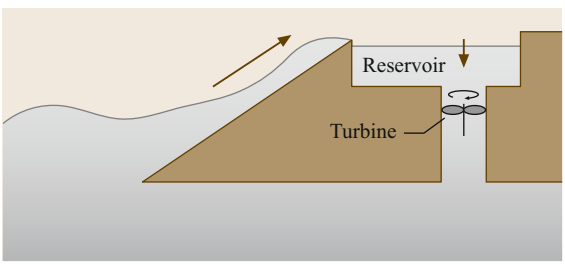


Fig. 45.10 Schematics of an overtoppping device

45.1.8 Control

As considered above, matching the natural frequency of oscillation of a WEC to the incident wave frequency is important for optimal energy capture. Otherwise the wave energy capture capacity typically decays rapidly. Therefore, the ability to control the WEC natural frequency and phase lag with respect to the incident wave is needed to optimize the power captured under various sea states as the wave frequency changes over time. The process of changing the natural frequency of a WEC is known as tuning, and involves adjusting one or more of the following: device shape, size, mass, stiffness, or damping. There are three types of tuning [45.39]: *fixed tuning*, *slow tuning*, and *fast tuning*.

Fixed tuning is a passive control that involves optimization during the design process of those properties of the WEC that may be impractical to change following construction. The optimization aims to provide a good overall match of the WEC response to the prevailing wave spectrum in the region.

Slow tuning is exercised over a period of minutes to hours and may involve altering WEC properties to match existing wave conditions or sea states. For example, slow tuning of a point absorber buoy may be achieved through changing its buoyancy through ballasting, thereby altering its mass and stiffness. Slow tuning is effective in the long term and typically involves, through real-time measurements, identification of the peak wave power characteristics.

Fast tuning is an ideal control system that can change the WEC response characteristics from wave to wave. It involves predicting the height and frequency of incident waves in order that WEC properties can be altered sufficiently rapidly to optimize the response when the waves reach the device. Fast tuning presents significant challenges both in terms of predicting wave characteristics in real time and minimizing the time needed to implement required changes on the device.

In practice, a combination of fixed, slow, and fast tuning need to be implemented in the WEC control system, in view of the challenges, to optimize WEC performance.

As discussed in Sect. 45.1.1, optimum energy absorption corresponds to resonant conditions with the amplitude of oscillation being such that the absorbed power is equal to the power radiated by the device to the sea as it oscillates. Since an idealized optimal control system itself may involve radiating energy to the sea as it is being implemented, the system is likely to be coupled, making it difficult to match the absorbed power to the radiated power. A suboptimal control strategy initially proposed by *Budal* and *Falnes* [45.40] that has been studied significantly is *latching control*; a variant called *declutching* and combinations of latching and declutching have also been studied [45.41, 42]. With reference to a heaving point absorber buoy, latching control involves controlling the phase between wave forcing and the buoy response velocity. It is based on the recognition, as pointed out in Sect. 45.1.1, that when the natural frequency of the buoy matches that of the wave and leads to resonance, the velocity of the buoy is in phase with the wave forcing. The objective is to achieve approximate optimum phase when the natural frequency of the WEC does not match the wave forcing frequency. This is met by locking the WEC in a fixed position at the instant when the device's oscillating velocity becomes zero and effectively the buoy displacement is maximum and releasing it after a time interval; this is facilitated through use of brakes. The length of the time interval is a single parameter in a discrete problem whose value is determined as part of the control strategy of the system. In particular, latching aims to slow down the natural oscillatory response of the buoy when its natural frequency is higher than the wave forcing frequency so that the heaving velocity of the buoy and the wave force reach their maxima at the same time [45.42]. Latching is a discrete control strategy. Latching considerably increases absorbed power [45.1].

Clément and *Babarit* [45.42] show how the analysis of monochromatic regular waves can be extended to real irregular seas, assuming linearity and considering the wave surface as a sum of harmonic components. However, the approach required the prediction of the incident irregular waves. The better the wave characteristics can be predicted, the closer the converted power approaches the optimum power determined for regular seas in Sect. 45.1.1.

45.2 Power Takeoff Systems

The methods of energy capture for WEC devices are generally classed as either *mechanical*, *turbine*, *hydraulic*, *pneumatic* or *linear generator*. The typical PTO components for WEC devices have been listed in [45.1] and are provided in Fig. 45.11. The mechanically driven method typically involves high-speed rotary electrical generators [45.3]. Point absorbers and line absorbers are not directly compatible with such rotary electrical machines and a hydraulic transmission system is typically used to interface the devices with electrical generators [45.43]. These different types of PTOs are described in the following.

Each WEC system involves a particular method of wave–device interaction (such as heave, pitch, or surge) and may need mechanical, pneumatic, hydraulic, or linear generator PTO. Some PTO systems are better suited for a particular primary wave energy capture system than others. There are a number of variations. A tentative guide is provided by *Grimwade* et al. [45.1] as shown in Table 45.1

45.2.1 Mechanical Transmission PTOs

A mechanical transmission system, involving gears, for a point absorber WEC, such as considered in Sect. 45.1.1, may be illustrated [45.13] using Fig. 45.12 which shows a heaving cylindrical point absorber attached to a vertical shaft outfitted with gearing on its right. These gears couple with a sprocket that is attached to an electrical generator axle. In regular monochromatic waves, the vertical velocity of the heaving buoy and the geared axle under resonance condition is given by (45.7), with $\sigma_z = \pi/2$ and setting $\gamma = 0$ for simplicity, as

$$\frac{\mathrm{d}z}{\mathrm{d}t} = \omega Z_0 \cos(\omega t) \,, \tag{45.35}$$

Table 45.1 Preferred PTO systems for various forms of primary energy capture (after [45.1]

WEC primary energy capture system	PTO: Transmission of mechanical energy
OWC	Pneumatic turbine
Attenuator	Hydraulic circuit
Point absorber	Hydro/hydraulic circuit, direct drive
Overtopping device	Low-head hydro turbine
OWSC	Hydro/hydraulic circuit
Submerged pressure differential	Hydraulic circuit/direct drive

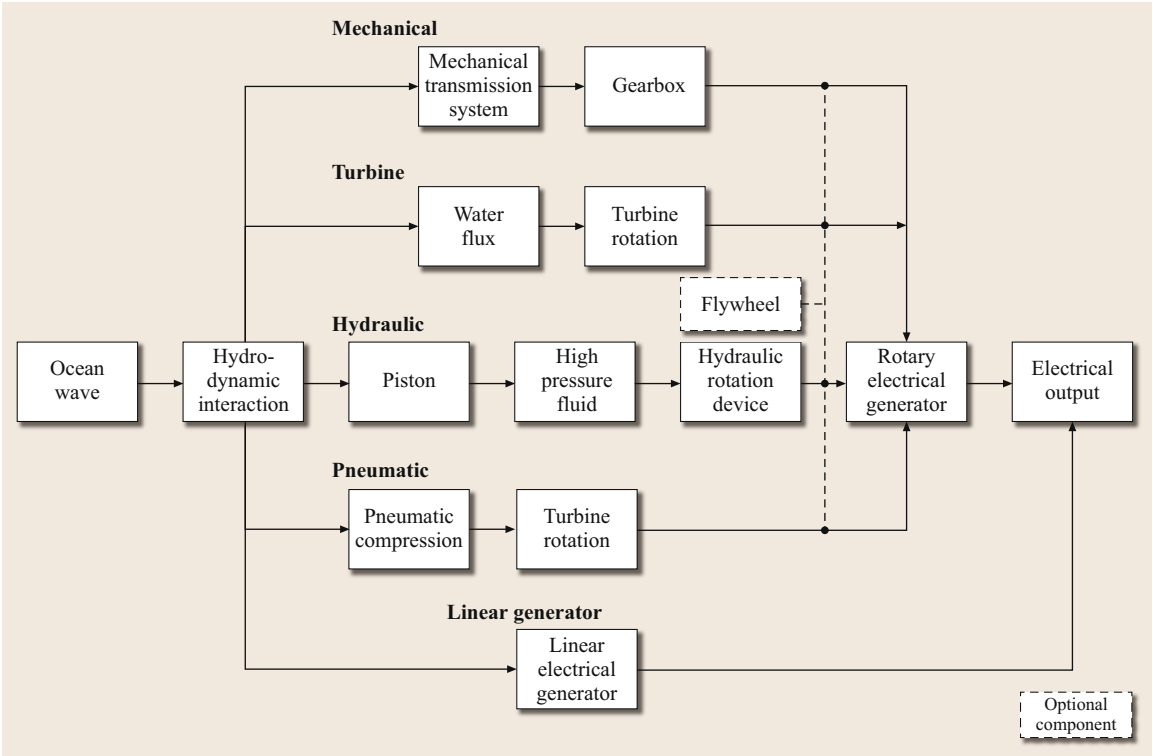


Fig. 45.11 PTO systems for WEC devices (after [45.1])

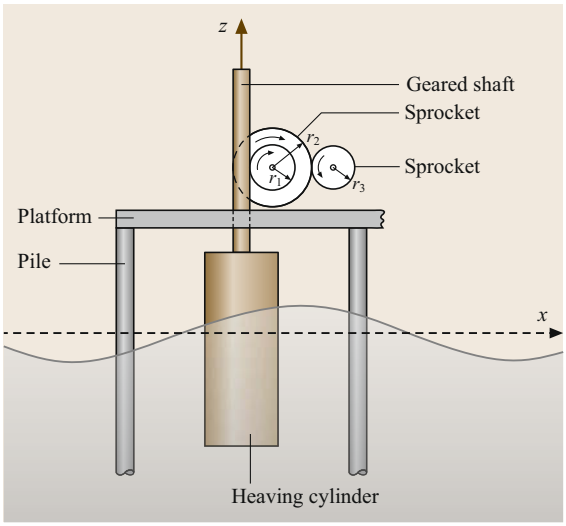


Fig. 45.12 Gearing-sprocket generator powertrain to harvest heaving wave motion (after [45.13])

where Z_0 is the heaving amplitude and ω is the angular frequency of the waves.

Then, assuming no slip between the gear and sprocket, the angular velocity ω_1 of the sprocket of radius r_1 is given by

$$\omega_1 = \frac{\omega Z_0}{r_1}\cos(\omega t)\,, \tag{45.36}$$

which varies with time. Let a sprocket of radius r_2 be rigidly interlocked with the one of radius r_1, so that they turn with the same angular speed. Then the angular speed ω_3 of the sprocket with radius r_3 is given

$$\omega_3 = \frac{r_2}{r_3}\omega_1 = \left(\frac{r_2}{r_1 r_3}\right)\omega Z_0 \cos(\omega t)\,, \tag{45.37}$$

where ω_3 exceeds ω_1 by a factor $(r_2/r_1 r_3)$. The sprocket with radius r_3 is connected to the axle of the electrical generator, which will provide the instantaneous power output given by

$$P = EI = E_0 I_0 \,|\sin(\omega_3 t)|\,|\sin(\omega_3 t + \phi)|\,, \tag{45.38a}$$

where $E = E_0\,|\sin(\omega_3 t)|$ is the instantaneous rectified voltage, $I = I_0\,|\sin(\omega_3 t + \phi)|$ is the instantaneous electrical current, and ϕ is the phase lag between voltage

and current. Using (45.37),

$$P = E_0 I_0 \left|\sin\left[\left(\frac{r_2}{r_1 r_3}\right)\omega Z_0 t\cos(\omega t)\right]\right| \times \left|\sin\left[\left(\frac{r_2}{r_1 r_3}\right)\omega Z_0 t\cos(\omega t)+\phi\right]\right| . \quad (45.38b)$$

The instantaneous power output therefore varies in amplitude and frequency. It is therefore quiet complex and needs to be processed and conditioned for transmission to a grid. The average frequency of the supply is given by

$$\hat{f} = \frac{\hat{\omega}}{2\pi} = \frac{1}{2\pi}\frac{1}{T}\int_0^T \left(\frac{r_2}{r_1 r_3}\right)\omega Z_0 \left|\cos\left(2\frac{\pi t}{T}\right)\right| \mathrm{d}t = \frac{1}{\pi^2}\left(\frac{r_2}{r_1 r_3}\right)\omega Z_0 . \quad (45.39)$$

45.2.2 Pneumatic and Hydro Turbines

The flow of fluid (seawater or air) induced by wave motion in typical WEC systems drives a turbine, which is directly coupled to a generator. The types of WEC devices using turbines include OWC and overtopping devices. Use of air in pneumatic turbines as the working fluid facilitates increasing the low speeds associated with wave motion to high rates of airflow. Disadvantages of pneumatic turbines are its low efficiency (though this can be improved through pitch control), poor starting, radiated noise, and high axial thrust in comparison with conventional turbines [45.41, 44]. Typical pneumatic turbines that are used are self-rectifying turbines, that is, they deliver a uni-directional torque for bi-directional airflow. They include the Wells turbine, and the Impulse turbine [45.1]. A comprehensive review of various turbine technologies is provided by *Takao* and *Setoguchi* [45.45].

Pneumatic Turbines

The basic Wells turbine has a rotor consisting of radial airfoils (Fig. 45.13). The airflow through the rotor generates a tangential force, which extracts the power, and an axial force related to drag on the airfoils. The rotor blades are symmetrical about the chord line of the blade cross-section. The air flows in the axial direction, normal to the rotor blades. However, the flow induced by the rotation of the rotor leads to an incident angle of attack at the blades. The angle of attack is small enough for the blades to perform without stall. The flow over

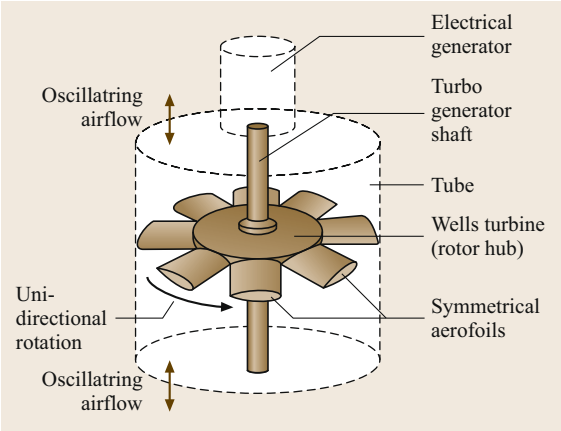


Fig. 45.13 Schematics of a Wells turbine (after [45.1])

the airfoils creates a lift force normal to the apparent flow direction and a drag force in the apparent direction of the flow. The tangential component of the lift force exceeds that of the drag force, resulting in a torque that drives the rotor. In view of the symmetry of the turbine, the sign of the torque is maintained even when the flow through the turbine reverses, so that an oscillating airflow is converted into a rectified, unidirectional, torque, whose value depends on the tip speed ratio, $\lambda_{TSR} = \pi n D/V$, where n is the rotor angular speed in s^{-1}, D is the rotor diameter, and V is the axial speed of the airflow. The turbines are not self-starting, and require an external source to initiate the rotation. For details of the Wells turbine, see [45.46]. The disadvantages of the Wells turbine are: limited operational flow range, poor starting conditions, induced noise, and high axial thrust. Examples of application of Wells turbine for OWC are the Pico and Limpet. A variant of the Wells turbine is the Dennis-Auld turbine [45.47]. In contrast to the Wells turbines, the blades are symmetrical in cross-section about the position of the maximum blade thickness rather than about the chord line. The turbine is still self-rectifying, but facilitates a variable pitch angle to enable optimal angle of attack. Other variants of the Wells turbine are discussed by *Takao* and *Setoguchi* [45.45].

In an impulse air turbine [45.22, 48], the airflow is deviated tangentially by guide vanes (Fig. 45.14) so that a tangential force on the rotor blades is directly produced by transferring the kinetic energy of the airflow to the rotor. The Impulse turbine can generally operate over a wider range of flow speeds than a Wells turbine. It has better starting characteristics and is less noisy [45.1]. The guide vanes can either be fixed or pitched. The pitching mechanism of the latter can either be controlled by the airflow or by another active mechanism. A disadvantage of the impulse turbine with

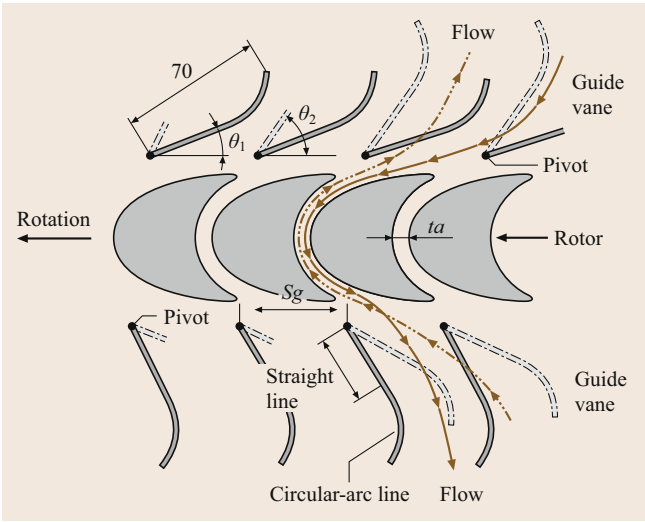


Fig. 45.14 Impulse turbine with self-pitch-controlled guide vanes (after [45.45])

movable guide vanes is that it has more moving parts that have to withstand a large number of oscillations per day. An example of the use of an impulse turbine for an OWC WEC is a plant at Vizhinjam, Kerala, India.

Hydro Turbines

For applications in overtopping WEC devices, various well-developed hydro turbines have been used that are classed as either hydro reaction turbines or hydro impulse turbines. Francis and Kaplan designs are examples of hydro reaction turbines while Pelton wheel is a hydro impulse turbine [45.1]. A hydro reaction turbine involves a rotating element that is fully immersed in water and is enclosed within a pressure vessel. A hydro impulse turbine runner, on the other hand, operates in air, and is driven by high-speed water jets that are developed by allowing low-velocity water to pass through a nozzle; the rotating blades of the turbine deflect the jets to maximize the force on the blades. The Kaplan turbine is suitable for overtopping WEC devices (Fig. 45.10), where the wave-induced flow is captured and directed over a ramp to a storage reservoir that is at a higher location above sea level, creating a head of water pressure that is released in a controlled manner via the Kaplan turbine, generating power in the process. In a Pelton wheel, nozzles direct streams of water against its spoon-shaped buckets mounted around the edge of the wheel. As the water flows into the bucket, its direction follows its contours, thereby exerting pressure on the bucket and the wheel. The water then decelerates and flows out on the other side of the bucket at low velocity. In the process, the water's momentum is transferred to the turbine.

45.2.3 Hydraulic Power Takeoff Systems

Hydraulics enable converting mechanical energy of linear motion into hydraulic energy and hydraulic energy into mechanical rotation, and thereby conversion to electrical power. As a result, hydraulic PTO systems are used in a number of WEC devices, most notably in Pelamis WEC. A possible medium for such a system may be seawater or other fluid such as fresh water or oil in a sealed system. Figure 45.15 illustrates a typical Hydraulics PTO system, adapted from the one used for Pelamis for a heaving buoy point absorber. The heaving motion of the buoy induced by surface waves drives a low-speed piston or pump back and forth, pumping hydraulic oil through the pipeline circuitry around the power generation system. Electronically controlled valves in the control manifold control the flow of fluid between the hydraulic pistons and the hydraulic circuit. The valves are controlled to facilitate passage of oil between different combinations of chambers to provide optimum benefit. For example, valves in the control

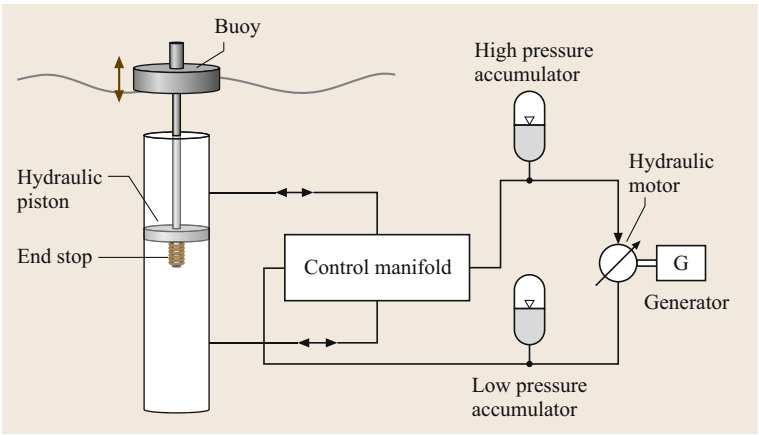


Fig. 45.15 Schematics of a typical hydraulic PTO system

manifold rectify the flow of oil from the low-speed piston so that it is unidirectional, even though the linear motion in the piston is oscillatory [45.3]. The oil flows through smoothing accumulators where it may be stored under pressure. It is then supplied at a controlled rate to hydraulic motors that drive electrical generators to produce electricity. Accumulators in the circuit not only facilitate energy storage and maintain constant flow to the hydraulic motor, but also decouple the hydraulic piston and the motors, thereby damping out the pulsatile nature of the energy captured by the hydraulic pumps and having a smoothing effect on power transmission; this facilitates the variation of the pump torque, in support of capturing optimal energy from the fluctuating pressure delivery, while the generator is driven at a steady rate. The low-pressure accumulator provides a small boost pressure to reduce the risk of cavitation on the low-pressure side [45.3]. In the case of the Pelamis, each individual unit of the attenuator has its own sealed PTO system, including power generation, which utilizes biodegradable transformer oil as the hydraulic fluid. The waves excite the motion of each joint of the attenuator, which in turn excites a hydraulic pump. The PTO hydraulic circuit is similar to that shown in Fig. 45.15. Pressurized chambers in the circuit exchange oil directly with storage accumulators [45.49].

Multiple WEC devices can be connected to a single sealed hydraulic circuit and generator. Likewise, a single hydraulic circuit can supply hydraulic power to multiple hydraulic motors and electrical generators, which can be engaged or disengaged from service as required [45.1]. Importantly, for floating devices, hydraulic circuits provide the flexibility in placement of motors and generators, including onshore.

Hydraulic PTO systems present some challenges [45.1, 3]:

1. Containment of oil in the hydraulic circuit is important. Special measures would therefore be needed to minimize leakages. Use of biodegradable hydraulic fluid would minimize risks to the marine environment on the basis that in the event of a leakage the fluid will biodegrade.
2. Duty-cycle maintenance of the hydraulic system involving several moving parts, pipes carrying hydraulic fluid, and seals that wear under friction, components that could fail due to fatigue under significant cyclic loads (order of thousands of cycles per day depending on the wave climate), coupled with the maintenance requirements posed by the harsh marine environment present a significant challenge [45.50]. Special design considerations, in terms of placement, material selection, use of corrosion resistant coatings, choice of electronic control systems, and other factors impacting maintenance requirements, would be needed. Locating the hydraulic PTO system above water on a barge or buoy or onshore may mitigate some of the maintenance requirements. An example of this is the Oyster WEC by Aquamarine in which the PTO system is located onshore [45.23].
3. Compressibility, bearing and seal friction of the hydraulic cylinders, and flow losses through valves and pipes are sources of inefficiencies. The impact could be minimized through careful design [45.49], including the development of optimal control strategies.
4. Extreme conditions. Special measures are needed to minimize potential risk of damaging systems under extreme loads, including use of end stops [45.3] on the hydraulic pumps or pistons that absorb the wave energy so that the hydraulic rams do not damage the pumps.

Hydraulic motors facilitate variable displacement operation in response to available power, and provide high-speed output in driving the electrical generator [45.1]. They are not required to be at the same location as the primary energy devices and can be placed at the best location for weight considerations. Types of hydraulic motors that may be considered are an axial or radial piston motors, vane motors for high-speed application, and gear motors. For further details, refer to the MARINET report [45.1] and reference provided therein.

45.2.4 Linear Generator

Use of an electrical machine as a PTO system for a WEC provides the possibility of directly converting mechanical energy into electrical energy. Linear mechanical motion associated with primary wave energy absorbing devices, as opposed to rotary motion induced by typical wind turbines, implies the need for a *linear generator* direct drive electrical machines for WECs. Such a machine involves power generation through one or more permanent magnets set to move through a coil by the primary energy capture device. The basic concept is schematically illustrated in Fig. 45.16. A series of magnets with alternating polarity are mounted on a *translator* that is directly coupled to a heaving buoy and that moves through a support structure, a *stator*, containing windings of the coil. The support structure may be fixed to the seabed or attached to a large drag or inertia plate. As the translator oscillates in response to the wave-induced motion of the primary wave energy capture device (the heaving point absorber in the

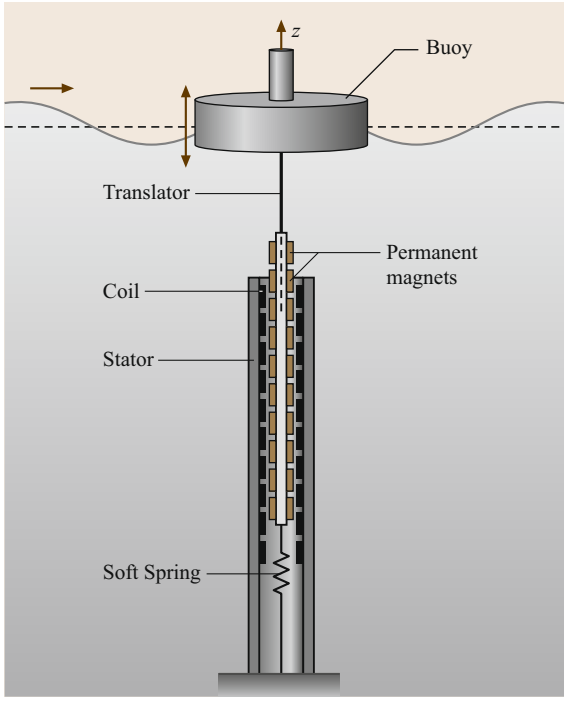


Fig. 45.16 Schematics of a typical Linear generator PTO system

figure), an electric current is induced in the stator windings.

If the translator moves at vertical velocity V_T and the stator is fixed to the seabed, the electrical power produced may be estimated by

$$P_e = \mathbb{Z} V_T^2 \,, \tag{45.40}$$

where $\mathbb{Z}$ depends on the number of turns in the windings, the strength of the magnetic field, the length of the wire and the resistance of the wire loop [45.51, 52]. For a buoy purely heaving in regular waves, since the translator moves in unison with the buoy, V_T may be estimated by (45.7)

$$\frac{dz}{dt} = -\omega Z_0 \sin(\omega t + \gamma - \sigma_z) \,, \tag{45.41}$$

where Z_0 is given by (45.3)–(45.6) with

$$F_R = -(m_{LG} + m_{33})\ddot{z} - R_m \dot{z} - K_{LG} z \,, \tag{45.42}$$

with $R_m \dot{z}$ being the mechanical damping force partly associated with the electromagnetic force resisting the motion of the translator, m_{LG} and K_{LG} are, respectively, the mass and the stiffness of the translator in the heaving system and m_{33} is the additional added mass due to its presence. The natural frequency of oscillation

$$\omega_z = \sqrt{\frac{\rho g A_{wp} + K_{LG}}{m + m_{LG} + a_{33} + m_{33}}}$$

of the system is accordingly modified so that [45.51].

The time-average electrical power is then estimated as

$$\hat{P}_e = \frac{1}{T}\int_0^T P_e dt = \frac{1}{2} Z \omega^2 Z_0^2 \,. \tag{45.43}$$

Use of a direct drive linear generator implies a brushless machine, fewer moving parts, potentially lower maintenance requirements, and higher efficiency. The linear generator appears particularly suitable for point absorber and submerged pressure differential devices, and its application to several WEC devices has been considered and demonstrated, including the AWS submerged pressure differential device [45.53], and the Seabased AB [45.54] device. There are, however, a number of challenges. This includes need for long-life lubrication. If the system is operated fully flooded to avoid requirements for tight seals and provides improved cooling, prevention of corrosion on the magnets needs to be addressed [45.43]. Output voltage varies both in frequency and amplitude. Mechanical strains are very high with low power factor. For a peak speed of $V_{Tmax} = 2\,\text{m/s}$, the air (or water) gaps that the generators need to overcome shear stress can be very large. Furthermore, possibilities of power storage and smoothing in the energy chain need to be addressed [45.1].

45.2.5 Other Considerations

Other considerations and requirements common with other sources of marine energy such as ocean thermal energy conversion (OTEC) include power transmission, including underwater power cables, cable power conditioning, and energy storage. These requirements are significant drivers of costs. Power transmission over long distances is one option, and involves consideration of configuration, choice of cables and transmission losses utilizing power to develop products offshore, such as hydrogen, ammonia, synthetic fuels and aluminum, are other. Storage and transport of electricity in large lithium batteries have been proposed. Further design considerations include simulation and experimental testing of systems. Details of these and related considerations can be found in [45.1, 13, 16, 55–57].

45.A Appendix: Practical Applications of Wave Energy Conversion Technologies

An important advantage in wave energy conversion (WEC) is that the supporting technologies can be located onshore, nearshore, or offshore. Several patented WEC methods and techniques have been developed over the last four decades. Practical and commercial applications of the concepts discussed in Sect. 45.1 and 45.2, resulting from these patents are described here and are grouped together accordingly as:

1. Point absorber systems
2. OWC systems
3. Attenuators
4. *Oscillating wave surge convertors*
5. Overtopping systems.

45.A.1 Point Absorber Systems

The principle of behind the point absorber system is discussed in Sect. 45.1.1. Here are some of some well-known point absorber systems that have been developed. Here are some well-known commercial point-absorber devices [45.1] (Table 45.2).

Power Buoy

The Power Buoy is composed of three main parts: a float, a spar, and a heave plate as shown in Fig. 45.17. The wave-induced motion of the float causes the spar to oscillate up and down. The heave plate ensures that the spar maintains a relatively stationary position. This relative oscillation of the float with respect to the spar acts as prime mover for a mechanical subsystem within the spar itself. This mechanical subsystem acts as a crank mechanism in a reciprocating engine since it transforms the rectilinear motion of the float into a rotational one. The spinning mechanical subsystem transmits motion to electric power generators that can power on-site payloads or for nearby maritime uses by a submerged cable.

OPT's PowerBuoy can accommodate two substantially different output ranges: up to 350 W and up to 15 kW. OPT's APB350 may play the role of an uninterruptable power supply (UPS) which can even recharge itself from the waves. The APB350 supplies uninterrupted power to either onboard payloads or to instrumentation on the seabed. It also maintains a real

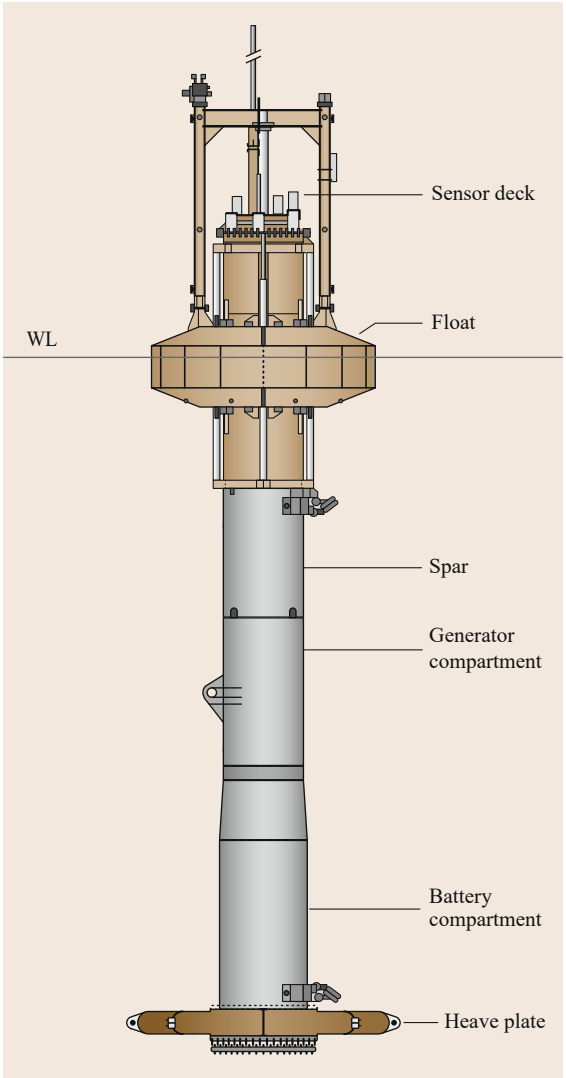


Fig. 45.17 Schematic representation of PowerBuoy device developed by Ocean Power Technologies

Table 45.2 Important wave energy conversion technologies employing point absorbers

Technology/plant name	Company	Country	Primary energy capture	PTO system
Power Buoy	Ocean Power Technologies	USA	Point absorber	Hydraulic circuit
Seabased	Seabased AB	Sweden	Point absorber	Linear generator
CETO III	Carnegie Wave Energy Limited	Ireland	Point absorber	Hydraulic circuit
Wave Star	Wave Star Energy	Denmark	Point absorber (multiple points)	Hydraulic circuit
AquabuOY	Finavera (Aquaenergy)	Canada	Point absorber	Hydraulic circuit

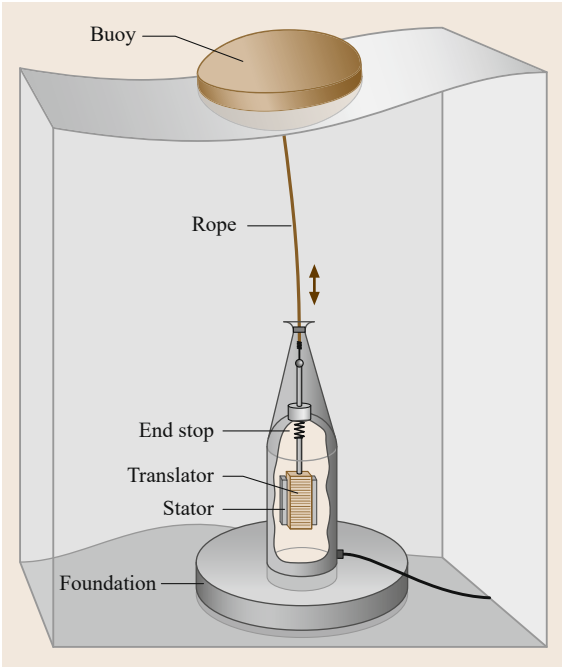


Fig. 45.18 Uppsala Seabased AB WEC

time data transfer stream and telecommunication with remote coastal locations.

PB40 is a system for applications with higher power demand.

Seabased WEC

Seabased wave energy generators (Fig. 45.18) are point absorbers that utilize linear generator PTO that deployed on the seabed. These generators have been developed Seabased AB of Sweden. The linear generator is connected to a wave energy capturing surface buoy via a line. The series of WECs are connected to a central substation, from where electricity can be transmitted to a grid in the form of an AC current. The linear generators are encapsulated and anchored to the seabed.

CETO

Named after a Greek sea goddess, CETO offers the potential to harvest the energy content of ocean waves.

The CETO system (Fig. 45.19) is not similar to other wave energy harvesting concepts since it functions submerged where it is safely protected from the elements, for example, storms, and minimally visible from the coast. The buoys, positioned underwater, act as prime movers for pumping and generator outfits that are also located offshore, as part of the system package; power is transmitted to the coast by subsea cables to; finally, power is transmitted to the coast by subsea cables.

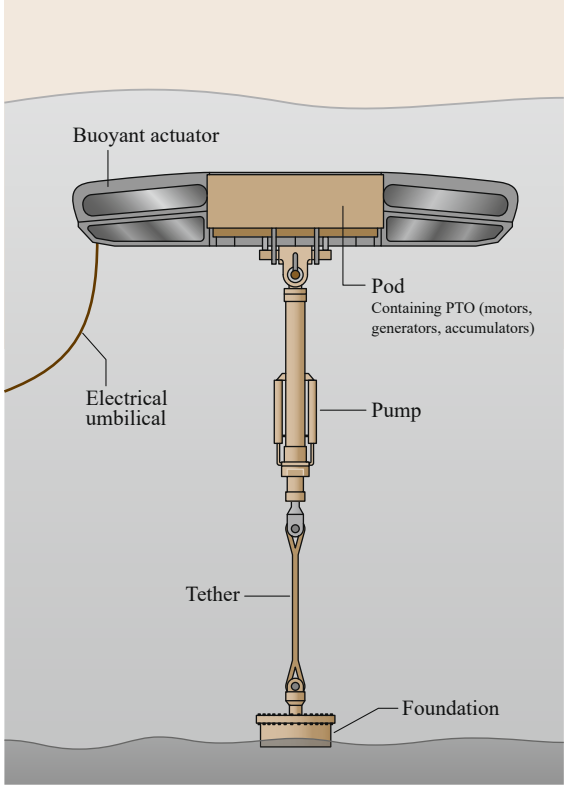


Fig. 45.19 The CETO 6 system developed by Carnegie Wave Energy Ltd.

The first effort toward CETO was made in 1999. Development of a CETO system demonstrator begun in 2003 with the first prototype. The CETO I model validated that it can deliver clean energy and freshwater from ocean waves.

Between 2006 and 2008, CETO 2 prototypes were developed and tested in the waters of Fremantle in Western Australia at Carnegie's purpose built Wave Energy Research Facility. These were approximately 1 kW prototypes and provided new commercial design concepts. Since then more prototypes have been built and further advances have been made.

Wavestar

This system was conceived by sailing enthusiasts Niels and Keld Hansen in 2000. The objective was to develop a stable power regular driven by ocean swells and crests 5–10 s apart. This is possible by a line of half-submerged buoys that emerge and sink in turn as each wave passes through, giving rise to the characteristic Wavestar design as shown in Fig. 45.20. Such design enables energy to be generated in a continuous manner in spite of the periodicity demonstrated by incident waves.

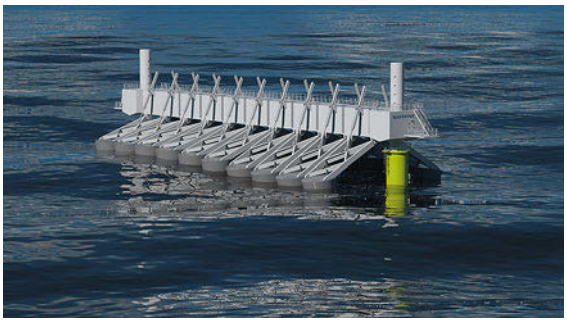


Fig. 45.20 The center bridge of Wavestar with the two buoy rows

The machine's unique storm protection system, one of the multitudes of patented elements in the design, and ensures its survivability. Furthermore it is in its own right a seminal point in the design of wave energy harvesting systems.

The Wavestar system can be built alongside with wind turbines, of for example, advanced design [45.58]. Such a concept can enhance efficiency and reduce initial costs.

In 2004, a tank trial campaign was carried out on a scale of 1 : 40 model. The objective was optimization of the fundamental conceptual architecture of the device and to record the electrical power output under common North Sea wave climate conditions.

In 2005, the first 1 : 10 prototype was procured, installed and connected to a power utility grid in waters at Nissum Bredning. At this location, waves are approx. 1 : 10 of North Sea waves. Research continues and 1 : 2 scale model was built and deployed for trials in Hanstholm during the second half of 2009.

LabBuoy

LabBuoy is a floating system for wave energy harvesting that has power transmission and conversion subsystems attached on a breakwater or a pier. In this manner, higher power output is enabled because of wave reflection developed at the breakwater face as well as being more reliable and safe to operate.

An entire EU-funded project has been carried out to investigate the LabBuoy concept. The main objectives of the project were evaluation of feasibility, socioeconomic and environmental footprint of the proposed system for a range of sea states commonly encountered in Europe. Furthermore, the LabBuoy project concerned model testing of an on-shore operational wave energy harvesting system. The system is floating, while power transmission and conversion subsystems are located above sea surface on a solid foundation (breakwater or pier), Fig. 45.21.

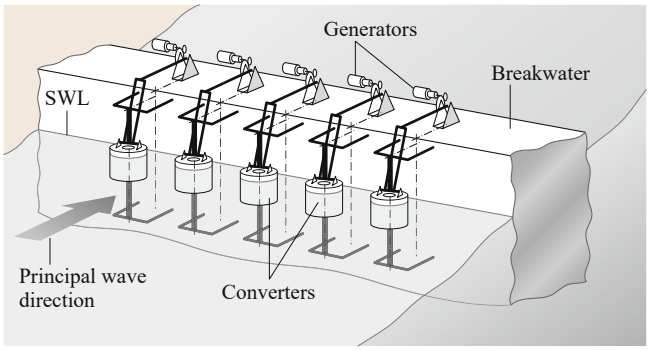


Fig. 45.21 The LabBuoy concept developed at the National Technical University Athens

A 1 : 15 scale model consisting of five converters in series was manufactured and deployed in the wave tank of the National Technical University of Athens, and a number of experiments under various wave conditions of interest were conducted. The central aim of the scale experiments was the validation of the numerical simulation models of the device.

A DC motor is attached to the rear end of the arm of each model, which provides a controllable resistance torque to the arm motion via a toothed belt. Each motor is controlled by a controller with encoder feedback. The motors can be controlled to excite the floats to specific motions, such as, for example, forced sinusoidal oscillations. Each model converter is equipped with the force sensor for the measurement of the vertical buoy force; angular position sensor for the measurement of the instantaneous position of the model arm, and; optical encoder for the measurement of the instantaneous rotational speed of the model arm.

In addition, simultaneously to the model properties, the instantaneous wave height was sensed by wave sensors in six significant locations.

The data acquisition and data management system is capable of real-time measurements, that is, it works independently of the operating system of the host computer, thus being not affected by limitations regarding data transfer rate. The system consists of a controller, a D/A converter, an analog signal conditioner, a strain gauge signal conditioner, an A/D converter, and a counter card.

A *virtual instrument* (vi) has been developed in the LabView s/w environment to control several functions of the data acquisition and data management system, in particular:

- Sensor calibration
- Adjustment of the hardware profile (scan rate, channel identification, amplification and filtering, etc.)

- Adjustment of the profile of the PTO (power-take-off) control system (operational mode, PID parameters etc.)
- Calibration and zero adjustment of the sensors; data conversion to physical units
- On-line data display
- Data storage
- vi's HMI (human machine interface)

The vi can be made available on the Internet to remote partners, allowing monitoring and operation. Data transfer to a remote location is also possible with an additional FTP module, also developed in the course of the project.

With the equipment described above a number of experiments were conducted, in particular:

- Diffraction measurements with fixed buoys under various wave conditions of interest. This series of experiments includes measurements of the vertical forces acting on the buoys at regular waves of various heights H and periods T, as well as irregular waves of various significant wave heights H_s and energy periods T_e. The experiments were conducted at two different wave incident angles (0° and 30°).
- Free oscillation measurements; this series of experiments includes measurement of the buoy motion and forces during the free oscillation of the buoy after its release from a submerged position, and parallel measurements of the waves radiated by the buoys. This experimental series provides information about the system inertia and the radiation coefficients of the buoys.
- Forced oscillation measurements: In this experimental series, the buoys are forced to harmonic oscillations about their state of equilibrium. The motion is driven by the DC motors. The experimental series includes measurement of motion, forces, and radiated waves at various oscillation frequencies and amplitudes.

Additional studies conducted in the project concerned the technical and economic feasibility of the device, the perspectives of the large-scale implementation of the technology in Europe and its expected economic and environmental impact. Some general conclusions are summarized in the following.

With some modifications in the original design, the LabBuoy technology is forecasted to provide an efficient, economic, safe, and environment-friendly source of energy. The main areas of environmental concern of the LabBuoy technology are typical for on/near-shore operating wave power devices: optical and acoustic intrusion, shore alteration, and impact on fauna.

Optical and acoustic intrusion is expected to be easy manageable. As with most wave power conversion technologies, the latter two impacts have to be studied carefully in subsequent projects. *LabBuoy breakwaters* are likely to find easy acceptance in built-up, coastal industrial areas, such as harbors, ports, shipyards, refineries etc. The noise level created in the energy conversion phase might also become a low priority consideration in this case.

Finally, Fig. 45.22 shows the European coastline exposed to exploitable wave power climate (dotted lines). The total length of this coastline is estimated to be at least ≈ 30 000 km. Assuming an initially available coastline in the EU (incl. Iceland and Norway) of 180 km in total for device deployment, this would correspond to a market of ≈ 23 500 devices with floater diameters of 2–4 m. This market corresponds to a job-market of ≈ 19 200 man-years of skilled technical personnel, not including operation, maintenance, and repair. The labor effort for project development would be approx. 700–1000 man-years of engineers and administrators.

45.A.2 Submerged Pressure Differential Devices

As described in Sect. 45.1.3, these devices are driven by pressure differential set-up by passage of waves

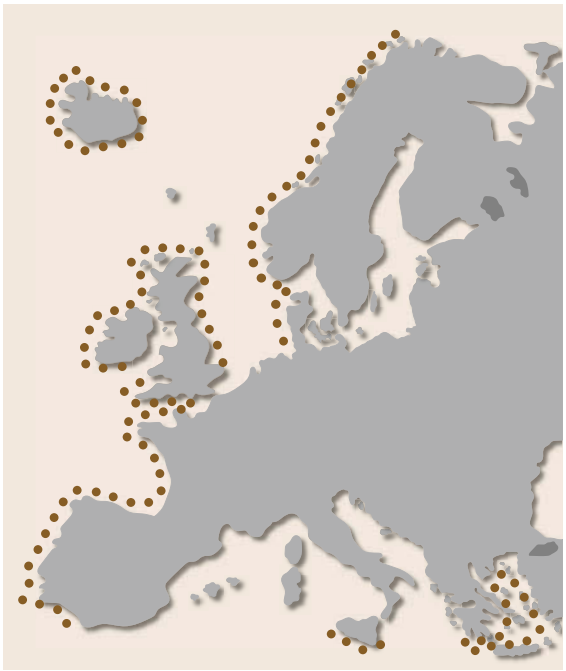


Fig. 45.22 Potential European market, candidate areas for device deployment are indicated by *dotted lines*

over the device. Here are some well-known commercial submerged pressure differential devices [45.1] (Table 45.3).

AWS

The patented AWS is an underwater wave energy harvesting buoy. It can provide reliable and affordable power for coastal settlements and offshore installations. The waves wing responds to fluctuation in seawater column pressure driven by passing waves. This motion is converted to electric power by a direct-drive generator. The concept works in seawater depths in excess of 25 m and is available for ratings between 25 and 250 kW by matching and choosing the proper scale.

The AWS is practically an underwater piston that entraps air, which in turn expands or contracts as seawater column pressure fluctuates due to waves. The relative motion of the two components of the system is converted to electric power by a rectilinear generator. Therefore, the system exhibits a tunable resonance that can be matched to dominant wave spectra in a region even in real time.

The concept was also put to the test offshore Portugal in 2004. Since that campaign, the waves wing has been further improved to match customer requirements closer.

45.A.3 Oscillating Water Column Systems

As described in Sect. 45.1.2, the principle of operation of this approach is an almost fully enclosed chamber, open at the bottom so that seawater waves can enter and exit and with openings toward its top where air turbines are placed. Specifically, incident ocean waves force the seawater column in the system's chamber to vertically oscillate causing air to flow out and in, respectively, and the turbine to turn. The following are known commercial OWC devices [45.1] (Table 45.4).

Limpet

In 1998, Queen's University Belfast in cooperation with Wavegen Ireland Ltd., Charles Brand Ltd., Kirk McClure Morton, and I.S.T. Portugal undertook the development and testing of an onshore power plant based on the OWC concept for the electricity generation. The system was named LIMPET an acronym for land installed marine power energy transmitter. It was set up on the island of Islay offshore the Scottish west coast. It started operations in November 2000. Since then the plant is running and operated by remote control. It supplies power to the UK. The successful unattended operation of this power plant has publicly shown what the potential of ocean wave energy can be toward a national energy portfolio.

The system employs three seawater columns that oscillate in concrete-built chambers with inner dimensions of 6×6 m and slope of 40° with respect to the calm sea surface plane, in effect, a total of 169 m^2 of water surface as shown in Fig. 45.23. The upper subpart of all three chambers, where atmospheric air is entrapped, is interconnected to each other so that electromechanical energy conversion can be performed by a single turbine-generator set instead of three.

The air turbine type employed in the system is that of Wells so that it can turn continuously in the same sense independent of the air-flow direction. LIMPET power plant consists of two such Wells turbines mounted in cascade; both turbines are made of stainless steel and have diameter of 2.6 m. The blades of each Wells turbine are directly mounted on an induction generator with specially modified design. Since each generator has a capacity of 250 kW, the total nominal power output of the LIMPET power plant is 500 kW.

Table 45.3 Important wave energy conversion technologies employing submerged pressure differentials

Technology/plant name	Company	Country	Primary energy capture	PTO system
AWS I	TeamWork	Portugal	Submerged pressure differential	Linear generator
CETO I	PIPO Systems SL	Spain	Submerged pressure differential	Hydro turbine
PYSIS	Wavebob AB	Ireland	Submerged pressure differential/point absorber	Mechanical

Table 45.4 Important wave energy conversion technologies employing oscillating water column(s)

Technology/plant name	Company	Country	Primary energy capture	PTO system
Limpet	WaveGen	United Kingdom	OWC onshore	Air turbine
Pico	OWC Wave Energy Centre	Portugal	OWC onshore	Air turbine
WECA system	Daedalus Informatics Ltd.	Greece	OWC nearshore	Air turbine
AWS III	AWS Ocean Energy Ltd.	United Kingdom	OWC floating	Air turbine
Mutriku	WaveGen EVE	Spain	OWC onshore	Air turbine
Port Kembla	Oceanlinx	Australia	OWC nearshore	Air turbine
OE Buoy	Ocean Energy Ltd.	Ireland	OWC floating	Air turbine
OWEL	Offshore Wave Energy Ltd.	United Kingdom	OWC floating	Air turbine

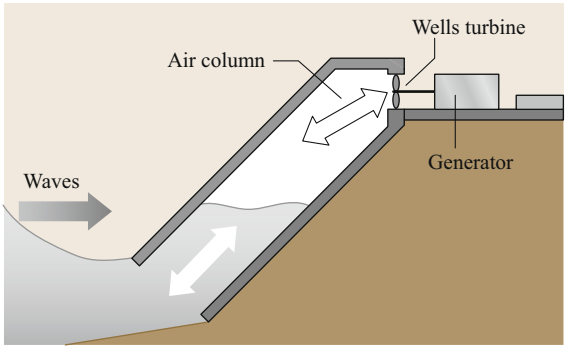


Fig. 45.23 Principle of operation of the LIMPET power plant

The use of induction generators allows for variable speed operation while at the same time minimizing maintenance requirements and maximizing reliability. The speed range of the system is between 700 and 1500 rpm and in order to connect with the power grid that requires constant speed and frequency, rectification, and then inversion is required; these processes are part of a power electronic drive of the VFD (variable frequency drive) type that is used in standard induction motors for motion control applications as well.

The operation of the LIMPET plant is continuously monitored in real time by a SCADA (supervisory control and data acquisition system) Also, the acoustic noise generated by the plant mainly due to the restricted air flows through the turbines is mitigated by intermitting a specially designed acoustic reverberation chamber between the power-plant core and the ambient.

By using the SCADA system and pressure transducers at various positions on the seabed oceanic wave data have been collected. Also, the wave loads on the walls of the facility's chambers have been recorded in conjunction with the seawater column level oscillations by employing both pressure transducers and sounding surface sensors.

Despite the fact that all electrical power generation technologies have some footprint on surrounding and environment at large, it is widely recognized that onshore wave energy harvesting is one of the friendliest to the environment. Essentially, they cause no problem to flora or fauna; just a minor visual distraction.

Pico Plant

Wave energy research in Portugal started as early as in 1978 in many national foundations and in cooperation with the country's government agency for industry and technology. In 1991, through program Joule, the European Commission began looking for a site to have some qualified wave energy harvesting along the European coastline. One of the sites selected was Porto Cachorro on Pico Island of the Azores archipelago of Portugal. In 1993 and 1995, the contracts with the European Commission were undersigned and the order for the studies, design and construction of the project to be carried out. The site was selected thanks to its demonstrated potential of 13.4 kW/m as well as it is easily accessible from land.

The technology employed is again that of OWC capable of delivering up to 400 kW. The chamber has dimensions of 12 m × 12 m and is founded on rocky seabed with a water depth of 8 m. In addition, one of the major objectives in developing the power plant was to be used as test site for PTO gear like turbines, generators, control equipment, and valves.

The facility of a concrete structure built on top of rocky seabed, spanning a small gully making the area a focal point of sea waves. The OWC has a 12 m × 12 m cross-sectional area. A horizontal-axis Wells turbine is driving a generator. There is space available for another turbine-generator set to be installed if desired. A relief valve is employed in order to mitigate positive or negative pressure excursions in the air chamber. In this manner, aerodynamic losses due to rotor blade stalling are greatly reduced in the system's turbine.

Simulation results for the powertrain from wave to wire were employed to specify the turbine as well as the relief valve. Design nominal value of 120 Pa s/m^3 was selected for the turbine damping coefficient. The turbine was expected to deliver 560 kW of shaft maximum instantaneous power with a maximum efficiency of 80%. An effective cross-sectional area of 0.8 m^2 was specified for the relief valve.

The air turbine, electric generator, and power electronics package of the plant are located in a 10 m × 12 m room on the rear of plant's main entrapped air chamber. The rest of the electrical, control, and monitoring outfit is installed in two rooms underneath the turbine housing.

A substantial civil engineering contract was put in place to get rid of rocks and boulders from the sea bottom on the side of the facility facing the sea. However damage to the front wall of the structure was identified during the summer of 1997. The damage was attributed to a poor underwater concrete pouring process that was used in 1996.

Some repair works were performed over the summer of 1998. The structural aspect encompassed the construction of the spaces housing outfit and gear as well as the topside of the chamber; works were finished in August of 1998.

The aerodynamic aspects of the Wells turbine development were performed during 1996. The turbine, associated ducts, and two of the air control valves were commissioned by Applied Research and Technology

Ltd. (ART, Inverness, Scotland) during 1997. Two different valves were used to handle airflow through the turbine's structure. A fast-acting valve positioned in proximity to the turbine controlled by a PLC (programmable logic controller) is regulates start and stop of the machine. A second valve, positioned toward the far end of the air duct where the plant's chamber is meeting with the turbine, can shut off the duct and machinery subsystems in the case of severe storms or during repairs.

Employing a variable speed electric generator spanning a range between 750 and 1500 rpm is considered a critical design element of this system. The configuration of such an unconventional electrical subsystem, encompassing advanced power electronics and controls, required substantial development effort and design work. The supplier of the conventional and unconventional electrical gear was EFACEC Engenharia SA (Portugal).

The concept of the Pico OWC plant is shown in Fig. 45.24. The integrated power plant, was completed during the summer of 1998. Accidental flood by seawater took place on September 5, 1998 when commissioning of some standard electrical gear was being installed in the lower housing space of the facility. Affected outfit like the 630 kVA transformer and a main AC panel was later shipped back to manufacturing facilities for repair or was replaced. This accident did not allow the facility to become functional before October 1998. Finally, first trials were not performed till 2005 and revealed major limitations in the original design. Improvements led to increased availability and power output of the plant from 2006 to 2008. Further improvements in 2009 led to significant reduction of the vibration level experienced by the turbine. After adding a secondary security circuit allowing for safer operation of the fully automated plant in 2010, the total output recorded over the period September–December 2010 was 45 MWh in 1450 h of operation.

In 2011, preparations to build a second platform for testing turbines were made. In the framework of EU-funded project MARINET pointed at the Pico power plant as an important wave energy-harvesting infrastructure. However, additional investment of at least 1.5–2 million Euros will be required.

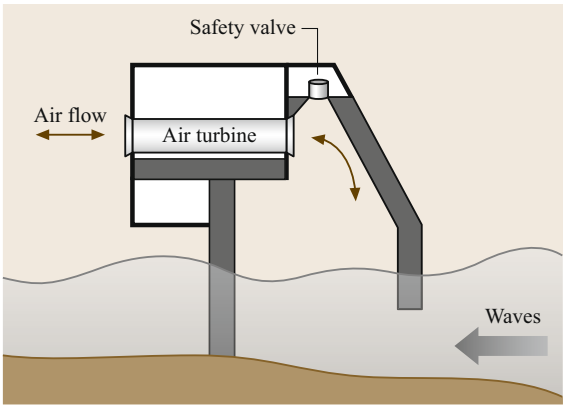


Fig. 45.24 Schematic representation of the Pico power plant (after [45.2])

The WECA System (Wave Energy Conversion Actuator)

Daedalus Informatics Ltd. in Greece has developed a technology employing the OWC concept that can be attached on existing infrastructures located on the coast, in shallow or even deep waters. The system employs a wedge-like form to focus the incident wave field to the interior chamber; there the enhanced wave action causes air pressure fluctuations that are harnessed by an accumulator. The full scale system will be 6 m high and 7 m wide and is expected to produce 20 kW at full capacity.

Oceanlinx Bluewave – greenWave

Oceanlinx is a company based in Australia since 1997. After many experimental campaigns, they have boiled down their designs to two major proposals employing the OWC concept. Even though the two proposed plants are of similar design (Fig 45.25), greenWave, is more suitable for shallow-water applications; in contrast, blueWave is for deeper water applications.

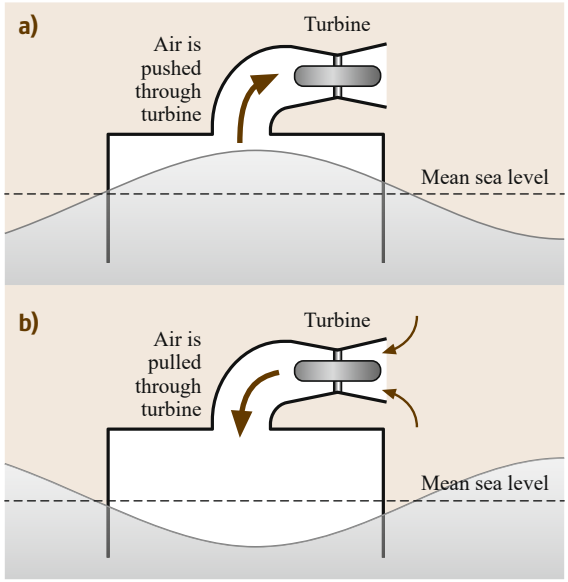


Fig. 45.25a,b Schematic concept of operation for **(a)** greenWave and **(b)** blueWave, working with the airWAVE turbine patended by Oceanlinx

A greenWave system can be installed in waters of depth not exceeding 10 m and founded on the seabed; in most cases use of concrete or steel is qualified. The foundation configuration is chosen to match the morphology of the seabed at the installation site. Since the system is originally waterproof and buoyant by design it can float. After being towed to the installation site, waterproofing is removed and the system ends up partially submerged. The distance from shore depends on the slope of the seabed. A typical greenWave system includes one OWC unit. Power output depends greatly on the wave potential of the installation site; a greenWave module with 20 m of width can output up to about 1 MW of power to be used either for electricity generation or seawater desalination or both.

Some of the major differences of a blueWave module when compared to a greenWave one, are that the former has to be built exclusively by using stainless steel and that the system includes six OWCs. It can be used for water depths between 40 and 80 m with appropriate mooring and anchoring. The power output can reach about 2.5 MW.

Both concepts of Oceanlinx employ the patented design air turbine airwave. This concept comes as an amendment to the pre-existing Denniss-Auld turbine design. Because of this, airWAVE is situated at a level significantly higher than the water line and has a smaller number of moving parts. It has significantly enhanced efficiency index when assessed vis-à-vis with other relevant turbine designs; in this manner, the overall wave-to-wire efficiency of Oceanlinx powertrains are also superior.

The majority of air turbine designs can work only with unidirectional flow. When looking at the concept of the OWC, unidirectional-flow designs simply fail to fully harness the resource. Early turbine design efforts to surpass this conceptual mismatch regularly come with grave efficiency reduction and other technical difficulties. The airWAVE design in contrast has met the objectives with higher success because it utilizes air pressure no matter how high or low the sea state is.

The major design modifications found in an airWAVE turbine include firstly two-ways action, 100%. Then pressure drop takes place across one single stage leading to significantly improved efficiency. Finally, an appropriate latching control and timing scheme in conjunction with the fact that the machine runs at almost constant RPM enables for optimum power generation.

As far as environmental footprint is concerned, the system does not include any contaminants or treatment in order to mitigate maritime life from growing. Also, note that since no moving components are needed underwater, it will be quite easy and safe for maritime life to move in and about the system. Moreover, the system behaves like a natural underwater cave and is expected that its impact on its immediate ecosystem will be rather small and indeed under conditions beneficial by playing the role of a manmade reef that allows living organisms to thrive. Given that the systems are developed to be installed at a distance quite far offshore, an insignificant visual impact is expected. The system's industry noise level emissions are not expected to raise any risk for sensitive mammals and birds of the maritime.

SPERBOY

SPERBOY is a system that has been conceived and patented by Embley Energy. It is a floating powertrain enabled for wave energy conversion using the OWC mechanism. Air entrapped in chamber(s) is driven by one or more OWCs, in turn driven by incoming waves, and then led through turbine-generator sets. Its design requires deployment in substantial arrays 8 to 12 miles offshore so that SPERBOY can output significant amounts of energy at rather competitive cost. Embley Energy Limited was established in 1998 and has, since 2002, implemented an ocean wave energy research project that may lead to commercialization of the system.

SPERBOY is practically an OWC system with a flotation buoy on top and an underwater fully confined column. On board the buoy and OWC, a superstructure holds the machinery, for example, generators and air turbines, as well as the rest of the powertrain outfit. The working principle is practically the same with that of stationary OWC's deployed on the coastline and permanent structures. However, the following critical differences were introduced: (a) the system can be deployed in deep water to maximize power output, and (b) the overall power plant is floating, maintaining optimal conditions for hydrodynamic interaction with the dominant yet rather volatile wave field, and generating power at low cost.

The procedure and requirements to deploy a SPERBOY are rather simple. The major subsystem can be manufactured close by the installation site. Then it can be carried and set up by appropriate watercraft. The actual installation site needs to be known to decide what infrastructure will be required. Research so far has investigated the development of a farm with 10 systems. However, this number or size is not optimal for commercial power generation. Spacing between adjacent units is about 350 m, depending on bathymetry. A full scale farm may include 1000 systems spanning up to $10-15\,km^2$.

Minimal to even zero routine maintenance is the objective of the modular design concept adopted for

SPERBOY. Such characteristics provide SPERBOY with a competitive advantage within the OWC type of power plants since the system is simple, easy to build, low in maintenance requirement, designed to operate over a long period and enable for competitively low costs of power generation. It is noted here that along the Atlantic seaboard of Europe each system is estimated to deliver a rated capacity of about 1 MW. The expected range of power generation cost from SPERBOY is £ 50–140 per MWh with estimates currently at hand. The simplicity of the SPERBOY wave energy harvesting concept, along with a highly competitive cost of generation, puts it at the cutting edge of wave energy harvesting technologies.

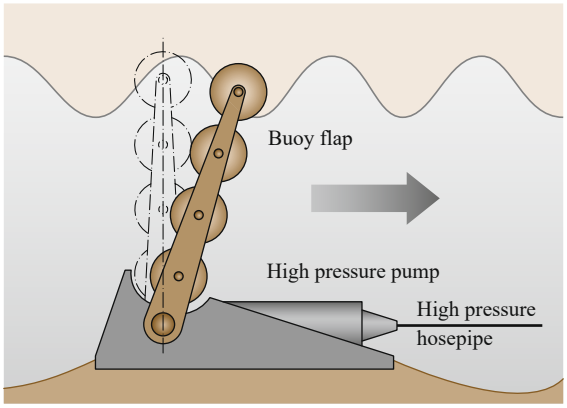


Fig. 45.26 Working principle of the Oyster wave energy system

45.A.4 Oscillating Wave Surge Convertors

This category includes systems that may float or are placed on the seabed and as sea waves, or even water currents, pass through the induced horizontal motion (surge) drives pumping of some hydraulic fluid acing as the medium for transfer of power. The concept is described in Sect. 45.1.4. The following are some the well-known commercial OWSC devices [45.1] (Table 45.5).

Oyster

Aquamarine Power's Oyster system can harvest energy from waves close to the coast. The captured wave energy is then converted into clean electric power. Practically, Oyster is a pump driven by waves that can thrust water at high pressure to run an onshore water turbine. The pumping concept is shown in Fig. 45.26.

Some of Oyster's main advantages include the following:

Simplicity due to being a mechanical offshore device with a minimum number of submerged moving parts. No automatic controller, gearbox, or shut-down mode is required and no complicated electronic modules offshore.

The system also demonstrates remarkable survivability due to the ability of the hinged flap to duck under waves exceeding a certain height in conjunction with the system's nearshore location and robustly designed supporting structure. An experimental Oyster system has been tested offshore in Orkney Islands in Scotland.

WaveRoller

WaveRoller is a system that can harvest wave energy and generate electric power or other forms of usable energy. The system can work in close-to-coast locations (approximately 0.3–2 km from the shore) and at depths between 8 and 20 m. Depending on locality prevailing conditions it is partly or entirely underwater and moored to the bottom. A standalone WaveRoller module (one panel) can generate between 500 and 1000 kW.

The concept of WaveRoller was inspired by a Finnish professional diver Rauno Koivusaari while exploring a shipwreck. He observed that a rather massive flat part of the ship's body was reciprocating back and forth, excited by the powerful water motion beneath the sea surface surge of water induced by incoming waves.

A WaveRoller responds practically in an identical manner as the part of the shipwreck that originally caught Rauno's attention. The reciprocating motion of the water triggered by waves puts a composite panel, into motion. As the module oscillates due to sea wave action, hydraulic piston pumps attached to the panel thrust hydraulic fluids inside a closed circuit. All the components of the system are sealed watertight in appropriate housing. Finally, the high-pressure fluid flow are driving a hydraulic motor that in turn acts as prime mover to an electric power generator.

bioWAVE

The bioWAVE is a system by BioPower Systems, Australia, that allows power generation from sea waves.

Table 45.5 Important wave energy conversion technologies employing oscillating water surge

Technology/plant name	Company	Country	Primary energy capture	PTO system
Oyster	AquaMarine Power Ltd.	United Kingdom	OWSC	Hydraulic circuit/hydro turbine
Wave Roller	AW-Energy Oy	Finland	OWSC	Hydraulic circuit
C-Wave	C-Wave Ltd.	United Kingdom	OWSC	Hydraulic circuit

The bioWAVE can work in swell waves, harvesting power both at sea surface and underwater. It is a pitching structure mounted on the seabed. The bioWAVE prototype under construction is expected to run in waters about 30 m deep. An envisioned 1 MW commercial unit will be able to run at water depths of 40–45 m.

The bioWAVE is attached to the seabed, with a pivot close to the bottom. The arrangement of buoyant floats, or *blades*, interacts with the oscillating surface (potential energy) and the sub-surface reciprocating water movement (kinetic energy). In effect, the pivoting module oscillates horizontally in synchronicity with incident waves. This motion is converted to electric power through an onboard self-contained hydraulic system, called O-Drive.

45.A.5 Attenuators

The principle of operation of this category is discussed in Sect. 45.1.5. These converters are built around a modular floating structure, the segments of which are connected using mechanical joints. Wave-induced motions compress hydraulic fluid in a closed circulation system in order to transfer the power to hydraulic motors that may be used to drive loads like electric generators. The following are some of the well-known attenuator WEC devices [45.1] (Table 45.6).

Pelamis

Pelamis Wave Power's Agucadoura Wave Farm isa commercial wave energy harvesting power plant situated 5 km off the Agucadoura coast in Portugal. The farm has been producing 2.25 MW by employing three Pelamis modules (Fig 45.27) since September 2008.

Wave energy harvesting of the technology that led to Pelamis was initiated by *Salter's Duck* (discussed in Sect. 45.1.5) invented by Professor Stephen Salter of the University of Edinburgh [45.6, 27, 28, 59, 60]). Pelamis is of cylindrical shape and consists of core tube segments held together by hinged joints. The system is 140 m long, 3.5 m in diameter. It can generate up to 750 kW rated power at a capacity factor 0.25–0.4.

During the design and development of Pelamis, special care has been taken to maximize survivability. The system's small cross-sectional area and streamlined hull form as well as compliant moorings in conjunction with the self-limiting hydrostatic loading, mitigate power absorption when wave height becomes too big to safely operate.

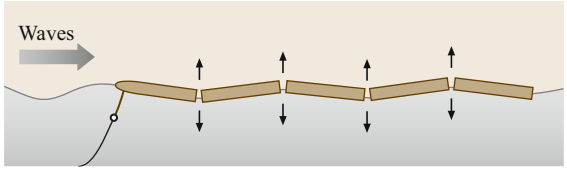


Fig. 45.27 Schematic representation of the Pelamis system, an attenuator WEC device

The system modules for the site in Portugal were manufactured in Stonehaven by Ross Deeptech. After commissioning of the electro-hydraulic power take-off systems was complete, the modules were transferred to Lewis. At this location, they were connected with the main tube segments made by Camcal. The components were transported in March 2006 to the Port of Peniche. There they were assembled, commissioned and installed at a location 5 km offshore north Portugal, near Póvoa de Varzim.

DEXA

The DEXAWAVE (or DEXA for short) (by DEXAWAVE, Denmark) wave energy harvesting concept is simple to build construction. It includes two rigid pontoons, hinged together using a patented hinge. Each one of those pontoons can pivot with respect to the other. A hydraulic PTO system on top of the device, can deliver up to 250 kW.

Just like Pelamis, the origins of DEXA can be traced to an earlier contraption, namely Cockerel's Raft that was developed in the early 1980s [45.61]. However, the DEXAWAVE development team came up with an improved design. In specific, the idea in Cockerel's Raft is that a floating body perturbed away from equilibrium will attempt to recover by a force equal to 44% of its total weight if one of its ends is fully emerged out of the water. By splitting the weight to two equal floating buoys connected with an almost weightless beam, the restoring force can be increased to 50% the total weight value while the amount of building material is also reduced significantly.

Table 45.6 Important wave energy conversion technologies employing attenuators

Technology/plant name	Company	Country	Primary energy capture	PTO system
McCabe wave Pump	Hydam Technology Ltd.	Ireland	Attenuator	Hydraulic circuit
Pelamis	Pelamis Wave Power Ltd.	United Kingdom	Attenuator	Hydraulic circuit
Sea Power Platform	Sea Power Ltd.	Ireland	Attenuator	Hydraulic circuit
Wello	Wello Finland	Portugal	Attenuator	Hydraulic circuit

Table 45.7 Important wave energy conversion technologies employing overtopping devices

Technology/plant name	Company	Country	Primary energy capture	PTO system
Wave Dragon	Wave Dragon Ltd.	Denmark	Overtopping	Hydro turbine
TAPCHAN	Norwave A.S.	Norway	Overtopping	Hydro turbine
Seawave Slot-Cone Generator	WAVEnergy AS	Norway	Overtopping	Hydro turbine
Floating Wave Power Vessel	Sea Power	Sweden	Overtopping	Hydro turbine

Owing to the simple mechanical construction of the DEXA converter, as a structure of floating pontoons, the converter can be installed relatively simple, and fast. Only a tug boat, and float with a simple chain crane is required. No divers, ROVs, crane platforms, etc. are required.

A scale model of the DEXAWAVE WEC is currently being tested in Danish waters and at Malta.

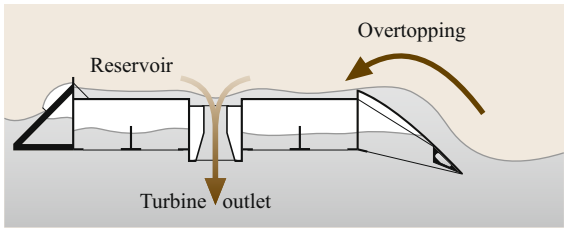


Fig. 45.28 Operation principle of Wave Dragon power plant (graphic by Erik Friis-Madsen)

45.A.6 Overtopping Devices

The principle of operation of this category is discussed in Sect. 45.1.6. Overtopping systems come with a seawater pool or tank that, when waves break, fills up. The seawater entrapped in the pool is then fed back to the ocean through a passage where a low-head water turbine produces electricity. Additional contraptions are often employed in overtopping systems, like e.g., collectors, in order to focus or otherwise enhance the energy harvested from the wave field. The following are some of the well-known commercial overtopping devices [45.1] (Table 45.7).

Wave Dragon

Wave Dragon is a floating wave energy harvesting overtopping system with a slack mooring. The concept allows deployment in either standalone modules or farms of Wave Dragon modules; the latter can be built up to match the power output of conventional power stations working, for example, with coal or hydrocarbon fuels.

A facility of this type, part of a utility grid, exists in Nissum Bredning, Denmark. Long-term trials are being conducted to assess performance including availability and power generation under the full range of operating conditions.

The Wave Dragon enables an incident wave to focus in the direction of a ramp. Located past the ramp, a tank of substantial size collects the seawater overtopping developing thus a head value. Seawater in the tank is then discharged from the tank through channels where water turbines are mounted in order to take advantage of the head between the level in the tank and the sea surface.

The size of a certain Wave Dragon is the outcome of an optimization procedure to match it with the wave climate at the installation site. The design variables include main body width, reflector length, overall weight, system number and turbine dimensions etc. The Wave Dragon at Nissum Bredning is built to match a rather modest wave climate at about 0.4 kW/m. This system is developed for trial purposes only and not meant for full-scale deployment. To maximize seawater overtopping capacity a joint system consisting of wave reflectors and a ramp with double curving has been developed and patented as shown in Fig. 45.28.

References

45.1 J. Grimwade, D. Hails, E. Robles, F. Salcedo, J. Bard, P. Kracht, J.-B. Richard, D. Schledde, A.R. Årdal, J.I. Marvik, N.A. Ringheim, H. Svendsen, M. Molinas: D2.03 Review of Relevant PTO Systems, Tech. Rep., MARINET (2012) http://www.fp7-marinet.eu/public/docs/D2.03_Review_of_relevant_PTO_systems.pdf

45.2 A.F.O. Falcão: Wave energy utilization: A review of the technologies, Renew. Sustain. Energy Rev. **14**(3), 899–918 (2010)

45.3 B. Drew, A.R. Plummer, M.N. Sahinkaya: A review of wave energy converter technology, Proc. IMechE A: J. Power Energy **223**, 887–902 (2009)

45.4 R.W. Yeung, A. Peiffer, N. Tom, T. Matlak: Design, analysis, and evaluation of the UC-Berkeley wave-energy extractor, Proc. ASME 29th Int. Conf. Ocean, Offshore Arct. Eng. (OMAE), Shanghai (2010)

45.5 J. Falnes: *Ocean Waves and Oscillating Systems: Linear Interactions Including Wave-Energy Extraction* (Cambridge Univ. Press, Cambridge 2002)

45.6 S.H. Salter, D. Jeffrey, J. Taylor: The architecture of nodding duck wave power generators, The Naval Architect **Jan**, 21–24 (1976)

45.7 J. Falnes: *Ocean Waves and Oscillating Systems: Linear Interactions Including Wave-Energy Extraction* (Cambridge Univ. Press, Cambridge 2004)

45.8 J.N. Newman: The exciting forces on bodies in waves, J. Ship Res. **6**(3), 10–17 (1962)

45.9 K. Budal, J. Falnes: A resonant point absorber of ocean waves, Nature **256**, 478–479 (1975)

45.10 D.V. Evans: A theory for wave-power absorption by oscillating bodies, J. Fluid Mech. **77**, 1–25 (1976)

45.11 J.N. Newman: The interaction of stationary vessels with regular waves, Proc. 11th Symp. Naval Hydrodyn. (Mechanical Engineering Publications, London 1976) pp. 491–501

45.12 K. Budal: Theory for absorption of wave power by a system of interacting bodies, J. Ship Res. **21**, 248–253 (1977)

45.13 M.E. McCormick: *Ocean Wave Energy Conversion* (Dover Publications, Mineola 2007)

45.14 A.P. McCabe, A. Bradshaw, J.A.C. Meadowcroft, G. Aggidis: Developments in the design of the PS Frog Mk 5 wave energy converter, Renew. Energy **31**, 141–151 (2006)

45.15 T.V. Heath: A review of oscillating water columns, Phil. Trans. R. Soc. A **370**, 235–245 (2012)

45.16 A.F.O. Falcão, J.C.C. Henriques: Oscillating-water-column wave energy converters and air turbines: A review, Renew. Energy **85**, 139–424 (2016)

45.17 H. Martin-Rivas, C.C. Mei: Wave power extraction from an oscillating water column along a straight coast, Ocean Eng. **36**, 426–433 (2009)

45.18 A.J.N.A. Sarmento, A.F.O. Falcão: Wave generation by an oscillating surface-pressure and its application in wave-energy extraction, J. Fluid Mech. **150**, 467–485 (1985)

45.19 D.V. Evans: Wave power absorption by systems of oscillating surface-pressure distributions, J. Fluid Mech. **114**, 481–499 (1982)

45.20 M.E. McCormick: A modified linear analysis of a wave-energy conversion buoy, Ocean Eng. **3**(3), 133–144 (1976)

45.21 M.G. De Sousa Prado, F. Gardner, M. Damen, H. Polinder: Modelling and test results of the archimedes wave swing, Proc. Inst. Mech. Eng. **220**, 855–868 (2006)

45.22 R.P.F. Gomes, M.F.P. Lopes, J.C.C. Henriques, L.M.C. Gato, A.F.O. Falcão: The dynamics and power extraction of bottom-hinged plate wave energy converters in regular and irregular waves, Ocean Eng. **96**, 86–99 (2015)

45.23 T. Whittaker, M. Folley: Nearshore oscillating wave surge converters and the development of Oyster, Phil. Trans. R. Soc. A **370**, 345–364 (2012)

45.24 J.H. Milgram: Active water-wave absorbers, J. Fluid Mech. **42**, 845–859 (1970)

45.25 R.G. Dean, R.A. Dalrymple: *Water Wave Mechanics for Engineers and Scientists* (World Scientific, Singapore 1991)

45.26 J. Falnes: A review of wave-energy extraction, Mar. Struct. **20**, 185–201 (2007)

45.27 A. Mynett, D. Serman, C. Mei: Characteristics of Salter's Cam for extracting energy from ocean waves, Appl. Ocean Res. **1**(1), 13–20 (1979)

45.28 A.D. Carmichael: An Experimental Study and Engineering Evaluation of the Salter Cam Wave Energy Converter. Report No. MITSG 72–22 (MIT, Cambridge, 1978)

45.29 P. Haren, C.C. Mei: An array of Hagen–Cockerell wave power absorbers in head seas, Appl. Ocean Res. **4**, 51–56 (1982)

45.30 R. Yemm, D. Pizer, C. Retzler, R. Henderson: Pelamis: Experience from concept to connection, Phil. Trans. R. Soc. A **370**, 365–380 (2012)

45.31 Pelamis Wave Power Downloads. Retrieved from Pelamis Wave Power Ltd. (2012) http://www.pelamiswave.com/upload/document/PWP-brochure-online.pdf

45.32 M. Jasinski, W. Knapp, M. Faust, E. Fris-Madsen: The power takeoff system of the Multi-MW Wave Dragon Wave Energy Converter, Eur. Wave Tidal Energy Conf. (2007)

45.33 J.C. Burfoot, G.W. Taylor: *Polar Dieletrics and Their Applications* (Macmillan, London 1979)

45.34 G.W. Taylor: *Piezoelectric Power Generation from Ocean Waves*, Tech. Report. (Princeton Resources, Princeton 1979)

45.35 R.E. Salomon, S.M. Harding: Gas concentration cells for the conversion of ocean wave energy, Ocean Eng. **6**(3), 317–327 (1979)

45.36 N.I. Xiros: Nonlinear control modeling for arrays of coupled mechatronic transducers, Proc. IMECE2012, Houston (2012), doi:10.1115/IMECE2012-89424

45.37 M.R. Alam: Nonlinear analysis of an actuated seafloor-mounted carpet for a high-performance wave energy extraction, Proc. R. Soc. A **468**, 3153–3171 (2012)

45.38 F. Madhi, M.E. Sinclair, R.W. Yeung: The "Berkeley Wedge": An asymmetrical energy-capturing floating breakwater of high performance, Mar. Syst. Ocean Technol. **9**(1), 5–16 (2014)

45.39 J.P. Ruiz-Minguela, M. Santos, P. Ibañez, J.L. Villate, F. Salcedo: Control techniques of ocean wave energy converters, Int. Conf. Energy (IADAT-ICE), Bilbao (2008)

45.40 K. Budal, J. Falnes: Interacting point absorbers with controlled motion. In: *Power from Sea Waves*, ed. by B. Count (Academic Press, London 1980)

45.41 S.H. Salter, J.R.M. Taylor, N.J. Caldwell: Power conversion mechanisms for wave energy, Proc. Inst. Mech. Eng. M: J. Eng. Maritime Environ. **216**, 1–27 (2002)

45.42 A.H. Clément, A. Babarit: Discrete control of resonant wave energy devices, Phil. Trans. R. Soc. A **370**, 288–314 (2012)

45.43 M.A. Mueller, N.J. Baker: Direct drive electrical power take-off for offshore marine energy converters, Proc. IMechE A: J. Power Energy **219**(A3), 223–234 (2005)

45.44 T.-H. Kim, M. Takao, T. Setoguchi, K. Kaneko, M. Inoue: Performance comparison of turbines for wave power conversion, Int. J. Therm. Sci. **40**(7), 681–689 (2001)

45.45 M. Takao, T. Setoguchi: Air turbines for wave energy conversion, Int. J. Rotating Mach. **2012**, 717398 (2012) doi:10.1155/2012/717398
45.46 S. Raghunathan: The Wells air turbine for wave energy conversion, Prog. Aerosp. Sci. **31**, 335–386 (1995)
45.47 T. Finnigan, D. Auld: Model testing of a variable-pitch aerodynamic turbine, Proc. ISOPE, ISOPE-I-03-053, Honolulu (2003) pp. 25–30
45.48 T.W. Kim, K. Kaneko, T. Setoguchi, M. Inoue: Aerodynamic performance of an impulse turbine with self pitch-controlled guide vanes for wave power generator, Proc 1st KSME-JSME Thermal Fluid Eng. Conf., Vol. 2 (1988) pp. 133–137
45.49 R. Henderson: Design, simulation, and testing of a novel hydraulic power take-off system for the Pelamis wave energy converter, Renew. Energy **31**(2), 271–283 (2006)
45.50 Carbon Trust Report: Guidelines on design and operation of wave energy converters, Tech. Rep. (Carbon Trust, Det Norske Veritas 2005) http://www.gl-group.com/pdf/WECguideline_tcm4-270406.pdf
45.51 T. Omholt: A wave activated electric generator, Proc. Oceans '78, Mar. Technol. Conf., Washington (1978) pp. 585–589
45.52 N.I. Xiros, C.S. Edrington, S. Balathandayuthapani: Nonlinear modeling of linear induction machines for analysis and control of novel electric warship subsystems, ASNE Electr. Mach. Technol. Symp. EMTS 10.1.21 (2010)
45.53 H. Polinder, M.E.C. Damen, F. Gardner: Linear PM generator system for wave energy conversion in the AWS, IEEE Trans. Energy Convers. **19**(3), 583–589 (2004)
45.54 L. Jonsson, M. Krell: Evaluating the Potential of Seabased's Wave Power Technology in New Zealand, Tech. Rep. UPTEC ES11 002 (Uppsala University, Uppsala 2011)
45.55 N.I. Xiros, I.K. Chatjigeorgiou: Nonlinear identification and input-output representation of the modal dynamics of marine slender structures, J. Offshore Mech. Arct. Eng. **129**(3), 188–200 (2007)
45.56 D.C. Kazangas, N.I. Xiros, I.K. Chatjigeorgiou: Reduced-order, nonlinear approximation of catenary riser dynamics using frequency domain identification, Proc. IMechE M **227**(4), 343–356 (2013)
45.57 N.I. Xiros, E. Logis, E. Gasparis, S. Tsolakidis, K. Kardasis: Theoretical and experimental investigation of unmanned boat electric propulsion system with PMDC motor and waterjet, J. Mar. Eng. Technol. **8**(2), 27–43 (2009)
45.58 M.I. Xiros, N.I. Xiros: Remarks on wind turbine power absorption increase by including the axial force due to the radial pressure gradient in the general momentum theory, Wind Energy **10**(1), 99–102 (2007)
45.59 S.H. Salter: Wave power, Nature **249**(5459), 720–724 (1974)
45.60 D. Serman: Theory of Salter's Energy Wave Energy Device in a Random Sea, MS Thesis (Department of Civil Engineering, MIT Cambridge 1978)
45.61 P. Haren: Optimal design of Haren-Cockerell Raft, MS Thesis (Department of Civil Engineering, MIT Cambridge 1878)

46. Ocean Current Energy Conversion

Howard P. Hanson, James H. VanZwieten, Gabriel M. Alsenas

Ocean currents in tidal flows and at the western boundaries of the ocean basins, as well as currents in rivers, can have as much potential for power production as the wind. With increased interest in sources of renewable energy, realizing this potential has undergone a renaissance in recent years. This chapter provides an introduction to ocean current energy conversion by emphasizing the contrasts between winds and currents and discussing some of the unique challenges associated with this untapped renewable energy source.

46.1 **Fundamentals** 1148
46.2 **The Betz Limit** 1149
46.3 **Conversion Systems** 1151
46.3.1 Axial-Flow Systems 1151
46.3.2 Cross-Flow Systems 1154
46.3.3 Tidal/Open-Ocean/In-Stream Differences 1155
46.4 **Supporting Infrastructure** 1156
46.4.1 Moorings 1156
46.4.2 Masts, Bases, and Platforms 1158
46.4.3 Other Components 1159
46.5 **Beyond Engineering** 1160
46.6 **Summary** 1160
References 1160

In addition to surface waves, oceanic motions that can be used as sources for renewable energy include ocean current systems on a variety of time and space scales. These sources of *marine hydrokinetic energy*, and the most common rotating technologies used to harness this energy, are the topic of this chapter.

Ocean currents have three principal forcing mechanisms [46.1].

Seawater density differences associated with the global distributions of temperature and salinity drive a thermohaline circulation, which is sometimes called the *oceanic conveyer belt*; the Atlantic basin component of this circulation is the Atlantic meridional overturning circulation (AMOC) [46.2]. The currents of the thermohaline circulation are, by themselves, too slow to be of interest for renewable energy purposes.

Winds in the atmosphere exert stress on the ocean surface, and the latitudinal differences associated with the transition from easterly trade winds in the tropics to the mid-latitude westerlies translate into a torque on the surface of the oceans. This torque, combined with earth's rotation and nearly spherical shape, results in large-scale, *wind-driven gyres* in the various ocean basins, in which the effects of friction, negligible in the interior, cause western boundary currents on the bounding continents' east coasts – the Gulf Stream in the Atlantic and the Kuroshio in the Pacific are familiar examples. These currents, with speeds of $2\,\mathrm{m\,s^{-1}}$ and more in some locations, do have potential for marine renewable energy (MRE) extraction.

Finally, the gravitational attraction of earth's moon as the earth rotates under it causes the oceanic tides, and these changes of sea level can cause strong tidal currents, particularly at the inlets to large bays, of $5\,\mathrm{m\,s^{-1}}$ and even more. These are also of interest to energy developers. Quantifying the potential of these strong currents provides a useful starting point for a discussion of energy-extraction technology.

In the context of currents that can be used to produce power, it is also relevant to note that many rivers offer the potential for renewable energy production. Although the focus of this chapter is on MRE, most of the background here also applies to in-stream, riverine cases.

Much of what follows in this first section is familiar territory in the atmosphere, given the maturity of the wind-power industry [46.3]. Because MRE is a relatively young field, however, it is presented here as an introductory tutorial into some of the fundamental issues of extracting power from currents using rotating equipment. Ocean–atmosphere contrasts will be emphasized.

46.1 Fundamentals

Like the atmosphere, the ocean behaves as a Newtonian fluid, provided that the effects of earth's shape and rotation are taken into consideration. These complications, however, are not needed for the analysis here, and so they will be ignored in the following.

Fluid behavior is described by the equations of motion, for the ocean a closed set of nonlinear partial differential equations that includes Newton's second law as written for fluids (the Navier–Stokes equations), conservation principles for mass and its constituents as appropriate, and, if needed, heat energy via the first law of thermodynamics [46.1, 103ff.]. In the present case, we need only the three Navier–Stokes equations (one for each orthogonal direction, x, y, z, with z taken here to be opposite to the attraction of gravity, meaning, generally, up) and the conservation of mass as expressed by the continuity equation. In full, three-dimensional (3-D) form, these are, respectively,

$$\rho\frac{DU}{Dt} \equiv \rho\left(\frac{\partial U}{\partial t} + U\frac{\partial U}{\partial x} + V\frac{\partial U}{\partial y} + W\frac{\partial U}{\partial z}\right) = -\frac{\partial p}{\partial x} + \frac{\partial}{\partial x}\left(\mu\frac{\partial U}{\partial x}\right), \tag{46.1a}$$

$$\rho\frac{DV}{Dt} \equiv \rho\left(\frac{\partial V}{\partial t} + U\frac{\partial V}{\partial x} + V\frac{\partial V}{\partial y} + W\frac{\partial V}{\partial z}\right) = -\frac{\partial p}{\partial y} + \frac{\partial}{\partial y}\left(\mu\frac{\partial V}{\partial y}\right), \tag{46.1b}$$

$$\rho\frac{DW}{Dt} \equiv \rho\left(\frac{\partial W}{\partial t} + U\frac{\partial W}{\partial x} + V\frac{\partial W}{\partial y} + W\frac{\partial W}{\partial z}\right) = -\frac{\partial p}{\partial z} + \frac{\partial}{\partial z}\left(\mu\frac{\partial W}{\partial z}\right) - \rho g, \tag{46.1c}$$

$$\frac{D\rho}{Dt} \equiv \frac{\partial \rho}{\partial t} + U\frac{\partial \rho}{\partial x} + V\frac{\partial \rho}{\partial y} + W\frac{\partial \rho}{\partial z} = -\rho\left(\frac{\partial U}{\partial x} + \frac{\partial V}{\partial y} + \frac{\partial W}{\partial z}\right) = 0. \tag{46.2}$$

Here, the velocities in the (x, y, z) directions are (U, V, W), respectively, t is time, ρ is the fluid's density, p is pressure, g is the gravitational constant ($9.81\,\mathrm{m\,s^{-2}}$), and μ is the fluid's viscosity, which will be ignored here.

In (46.1) the terms are, left to right, the change of the velocity component following the flow ($D[\,]/Dt$ represents the material, or total, derivative); which is defined as the sum of the *local* change of the velocity component with respect to time and three terms representing the advection of velocity up its gradient by the flow field; which equal the sum of forces acting on the fluid, in this case the pressure gradient force and the effects of friction expressed through viscosity. The equation for vertical motion (46.1c) also includes the gravitational force on the right-hand side. The terms in the continuity equation (46.2) are the time rate of change of density following the flow; which is defined by the four terms of the total derivative representing the local time rate of change and advection of density (all of these apply to compressible fluids), which equal the convergence (i. e., the opposite of the divergence) of the flow field multiplied by the density.

We will assume frictionless, incompressible flow, which is appropriate for the oceanic cases under consideration here – the friction terms in (46.1) will be dropped, and the derivatives of ρ in (46.2) vanish.

It is useful to define a kinetic energy per unit volume, $E \equiv \frac{\rho}{2}(U^2 + V^2 + W^2)$. Its conservation equation can be derived from (46.1) by multiplying (46.1a), (46.1b), (46.1c) by U, V, W, respectively, adding the three equations, and rearranging terms. This results (for frictionless flow with no body forces beyond gravity) in

$$\frac{DE}{Dt} \equiv \frac{\partial E}{\partial t} + U\frac{\partial E}{\partial x} + V\frac{\partial E}{\partial y} + W\frac{\partial E}{\partial z} = -\left(U\frac{\partial p}{\partial x} + V\frac{\partial p}{\partial y} + W\frac{\partial p}{\partial z}\right) - \rho g W. \tag{46.3a}$$

Changes of kinetic energy following the flow, then, are associated with the work of the pressure gradient force, or a change of potential energy in earth's gravity field. Multiplying the incompressible form of (46.2) by E/ρ and adding it to (46.3a) gives the *flux form* of the kinetic energy equation

$$\frac{\partial E}{\partial t} = -\left\{\frac{\partial (UE)}{\partial x} + \frac{\partial (VE)}{\partial y} + \frac{\partial (WE)}{\partial z}\right\} - \left\{U\frac{\partial p}{\partial x} + V\frac{\partial p}{\partial y} + W\frac{\partial p}{\partial z}\right\} - \rho g W. \tag{46.3b}$$

This shows that the local change of kinetic energy is the sum of the energy flux convergence and the work of the pressure field and of gravity.

This kinetic energy flux warrants further discussion. If the x direction is aligned with the flow (x positive downstream), then V and W can be ignored, and the magnitude of kinetic energy flux becomes

$$UE = \frac{\rho}{2}U^3 \equiv \Psi, \tag{46.4}$$

a very important quantity that represents the *power density* of the flow, that is, the power ($\mathrm{W\,m^{-2}}$) of the flow per unit area orthogonal to the flow direction. Note

the distinction between power $P \equiv \frac{DE}{Dt}$ and power density Ψ. Power is the (total) time derivative of energy, just as acceleration is the time derivative of velocity, while power density requires a different interpretation. In terms of (46.3b), the local change of energy with time is equal to the (opposite of) downstream gradient of power density, plus other work terms. In physical terms, the power density in an area A orthogonal to the flow is that power available to increase flow's kinetic energy by an amount (per unit time) $P = A\Psi$. However, other considerations prevent all of this power from being used, as will be seen in the next section.

It is useful to compare, using a simple order-of-magnitude analysis, power densities in the atmosphere and the ocean, because both the media flow (in winds and currents, respectively), and both are available for energy generation using turbines and other methods. Near the earth's surface, an order of magnitude for the density of air is $\rho_a \approx 1\,\mathrm{kg\,m^{-3}}$, while the density of seawater is much larger, $\rho_w \approx 1000\,\mathrm{kg\,m^{-2}}$. In contrast, characteristic values for winds and currents are the opposite, with winds occurring at factors of 10 or so higher than currents, in locations of interest to energy developers. The cubic dependence of Ψ on flow speed in (46.4) means, then, that the power densities in the two media are commonly of the same order of magnitude. More specifically, using actual values for the densities, it can be said that a $2\,\mathrm{m\,s^{-1}}$ (or 4 kt) ocean current has the power density of a gale-force (38 kt) wind. Because observations of flows in tidal channels and in western boundary currents routinely approach $2\,\mathrm{m\,s^{-1}}$, interest in the oceans as a source of renewable energy is high. And because these flows in the ocean can be harnessed with much the same equipment as the atmospheric winds (suitably modified for use underwater), much of the basis for ocean current energy conversion comes from the many decades of experience in the wind-power industry. An example is the fundamental limitation on energy extraction from a flow.

46.2 The Betz Limit

From a physical perspective, it is impossible to extract all the kinetic energy from a flow, because that would bring the flow to a halt, in which case it would not be able to continue to turn the device (the turbine) that is doing the extraction. Instead, there must be some optimal point at which the power extraction is maximized with respect to the flow's slowing down.

This was first considered by the German engineer *Albert Betz* in 1926 in a monograph (in German) called *Wind energy and its extraction through windmills* [46.4]; the topic has been discussed at length in numerous other publications since (a useful translation is [46.5]). Betz showed that the optimal extraction amount – now called the *Betz limit*, as it is impossible to extract more power than this – is about 60% of the inflow power. For completeness, we will re-derive this here. It is worth noting that the many claims of power generation above the Betz limit are all driven by marketing considerations rather than science and engineering, as they inevitably involve funnel-like shrouds and cowlings that allow the effective area of a given turbine to be increased and the flow to be accelerated accordingly.

Figure 46.1 provides a schematic of the flow under consideration, a steady, incompressible, 1-D (one-dimensional) flow (in its undisturbed state) in which an idealized *turbine* is introduced. The flow through the device is bounded by the two solid streamlines (in the plane of projection in the figure) that bound an upstream area A_u and the other areas downstream. Far upstream, the flow occurs at ambient pressure p_0 and velocity U_u. Far downstream, where the flow has returned to ambient pressure p_o, the streamlines bound an area A_d, through which the flow velocity is U_d. At the position of the turbine, which sweeps an area A_t, the flow velocity is U_t; the pressure field undergoes a pressure drop of δp from its (just) upstream value of p_t across the rotor – δp is therefore the force per unit area (downstream) across the face of the idealized turbine.

Continuing the notation that x represents the downstream direction (and therefore the areas in Fig. 46.1

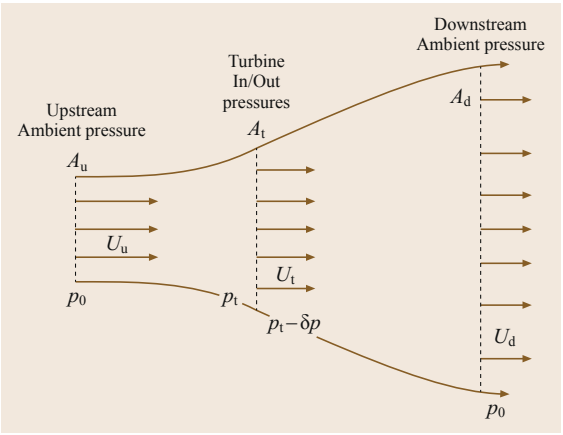


Fig. 46.1 Schematic of flow through an idealized turbine (not to scale in the streamwise direction)

are in the $Y-Z$ plane), (46.1a) becomes, for time-independent flow,

$$\rho\frac{DU}{Dt}=\rho\left(U\frac{\partial U}{\partial x}\right)=\frac{\rho}{2}\frac{\partial U^2}{\partial x}=\frac{\partial E}{\partial x}=-\frac{\partial p}{\partial x}\,. \tag{46.5}$$

The last equality implies that, along an undisturbed streamline, $E+p=$ const., thus recovering the Bernoulli principle, in which context E is often interpreted as the *dynamic pressure*. (Energy per unit volume [$\mathrm{J\,m^{-3}}=\mathrm{N\,m\,m^{-3}}$] has the same units as pressure [$\mathrm{N\,m^{-2}}$], so there is no inconsistency here). Applying this to the flow upstream of the turbine and, separately, to the flow downstream of the turbine, and then subtracting the resulting expressions results in

$$E_\mathrm{u}-E_\mathrm{d}=\delta p\,. \tag{46.6}$$

Note that this recovery of the Bernoulli principle comes directly from the Navier–Stokes equations, the momentum budget along streamlines – unlike many derivations, no energetic arguments were involved here.

It is also useful to consider the momentum budget of the entire flow structure in Fig. 46.1. A physical interpretation of the momentum budget is that, in any domain, the difference between the flux of momentum into and out of the domain is equal to the sum of the forces within the domain that are opposing the flow (because these forces are sinks of momentum). This implies that

$$U_\mathrm{u}\rho A_\mathrm{u}U_\mathrm{u}-U_\mathrm{d}\rho A_\mathrm{d}U_\mathrm{d}=A_\mathrm{t}\delta p\,. \tag{46.7a}$$

Continuity (i. e., no loss or gain of mass) in the system requires that $AU=$ const., and making use of this in (46.7a) allows it to be rewritten as

$$\rho U_\mathrm{t}(U_\mathrm{u}-U_\mathrm{d})=\delta p\,. \tag{46.7b}$$

By substituting the definition of E into (46.6) and eliminating δp between (46.6) and (46.7b), it is possible to show that $U_\mathrm{t}=(U_\mathrm{u}+U_\mathrm{d})/2$; that is, the velocity at the face of the turbine rotor is the average of the upstream and downstream velocities. In the following, it is convenient to define the ratio of these as $\eta\equiv U_\mathrm{d}/U_\mathrm{u}$.

Now, (46.3a) and (46.5) suggest that

$$P=U\frac{\partial E}{\partial x}=UF\,. \tag{46.8}$$

At the turbine, then, the power is

$$\begin{aligned}P_\mathrm{t}&=U_\mathrm{t}(A_\mathrm{t}\delta p)=\rho A_\mathrm{t}U_\mathrm{t}^2(U_\mathrm{u}-U_\mathrm{d})\\&=\frac{\rho}{4}A_\mathrm{t}U_\mathrm{u}^3(1+\eta)^2(1-\eta)\,,\end{aligned} \tag{46.9}$$

which makes use of (46.7b). Maximizing (46.9) with respect to η gives two roots, one negative, and therefore unphysical, and the other at $\eta=\frac{1}{3}$.

It is useful to compare this maximum power with that of the undisturbed flow through this area by defining a power coefficient (this one with the subscript B to denote Betz) as

$$C_\mathrm{B}\equiv\frac{P_\mathrm{t}}{A_\mathrm{t}\Psi_\mathrm{u}}=\frac{1}{2}(1+\eta)^2(1-\eta)=\frac{16}{27}\cong 0.593\,. \tag{46.10a}$$

Thus, the maximum power an idealized turbine can produce is just less than 60% of that available in the undisturbed flow. In a more general sense, it is useful to define a turbine power coefficient C_P as the ratio of the actual power produced by a device to that of the undisturbed flow. Due to internal friction and blade inefficiencies, $C_\mathrm{P}<C_\mathrm{B}$ in all cases without shrouds, with experimental device tests calculating values for C_P that approach 0.5 [46.6]. With this definition, the generalized expression for the power generated by a given device is

$$P=C_\mathrm{P}A\Psi, \tag{46.10b}$$

where C_P and A are properties of the device and Ψ is the power density of the undisturbed flow.

Note that substitution of the results for U_t and η into (46.7a) shows that $\delta p=\frac{4}{9}\rho U_\mathrm{u}^2$; that is, the pressure drop across the *turbine* in Fig. 46.1, which is the force on that turbine, is proportional to the flow velocity squared (the factor of 4/9 applies only at the Betz limit). Re-applying the order-of-magnitude argument above concerning power density to this result provides further insight. A similar comparison of atmosphere and ocean shows that, for identical power densities, the forces on objects in the ocean are about an order of magnitude larger than in the atmosphere. This has obvious implications for design of oceanic energy-recovery equipment, turbine rotors in particular.

46.3 Conversion Systems

As mentioned at the outset, this chapter focuses on rotating energy conversion systems, devices falling under the general description of ocean-current turbines (OCTs). Although MRE is a young field, progress is being made in many of its constituent areas; in particular, design requirements for OCTs are beginning to become standardized [46.7]. The purpose of this chapter is not to discuss such standards in detail; rather the discussion here is meant as background for further advancements in MRE development.

An important concept related to all OCTs, as well as wind turbines, is the *cut-in speed* of a particular device. This is simply the minimum flow speed, generally defined in terms of the undisturbed flow, or U_u in the terminology here, that will turn the system's rotor and generate power. Efficient rotor blades and minimal internal friction of the device will decrease cut-in speed, thus allowing an OCT to function at lower current speeds and to have a higher capacity factor, or fraction of the system's designed output, for the location in question.

At the same time, in real-world ocean currents there is a point of diminishing returns, as shown in Fig. 46.2. In this figure, the *production factor* is the amount of energy that hypothetical systems of various cut-in speeds installed at three hub depths would produce relative to the potential amount if cut-in speed were not a factor (Production factor differs from capacity factor in that capacity factor uses the system's designed power output as the denominator and does not take into account current variations above the cut-in speed, while production factor does). Note that, at this location, there is little to be gained by potentially expensive design and manufacturing improvements to lower cut-in speed below $1\,\mathrm{m\,s^{-1}}$. This underscores the importance of careful site characterization studies prior to choosing what equipment to install.

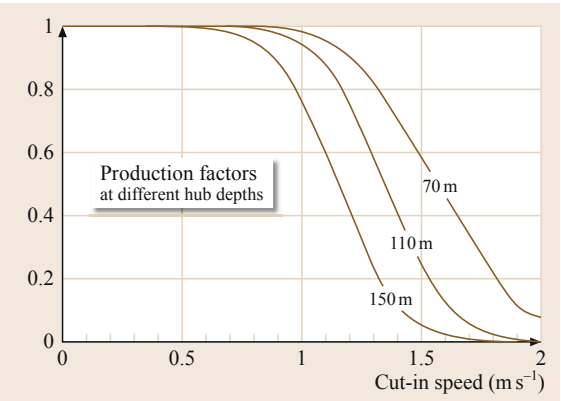


Fig. 46.2 Fraction of potential energy production as a function of system cut-in speed; this is an example from the Florida current off Fort Lauderdale. Based on acoustic Doppler current profiler ADCP data from the Florida current at 26° N, March 2009–March 2010

Such a site characterization would naturally include an estimate of the amount of power available at that location and some secondary estimate of what can be extracted. Cut-in speed is also useful in this context. For example, oceanographic measurements of the Florida current obtained using Pegasus floats in the 1980s across the straits of Florida at 27° N [46.8] were processed by *Hanson* et al. [46.9] to show that there is some 20 GW ($= 10^9$ W) of total power in the flow, but that the power available at and above a particular cut-in speed decreases markedly with cut-in speed (Fig. 46.3).

With this background, it is possible to discuss devices that can make use of this power.

46.3.1 Axial-Flow Systems

Figure 46.4 shows the main components of an axial-flow OCT designed and fabricated at the Southeast National Marine Renewable Energy Center (SNMREC) for the purpose of conducting at-sea research of OCT features and components. The designation of *axial-flow* for this and other such systems is fitting because such a concept aligns with the flow along that the rotor's axis of rotation. Note that the system is designed such the

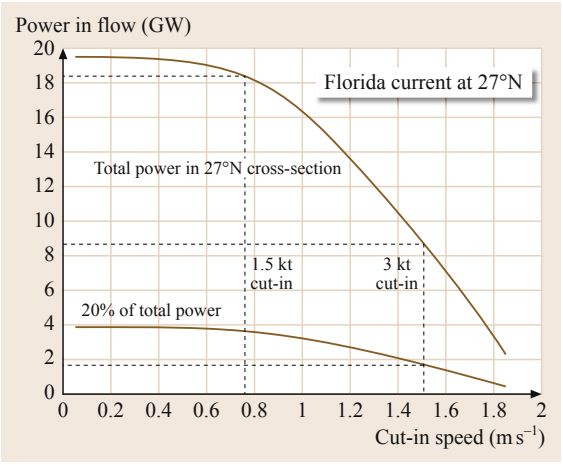


Fig. 46.3 Power in a Florida current cross-section at 27° N depicted as a cumulative function of cut-in speed. The extractable power is assumed to involve devices of 40% net efficiency ($C_P = 0.4$) that occupy half of the cross-sectional area at and above a particular speed (after [46.8])

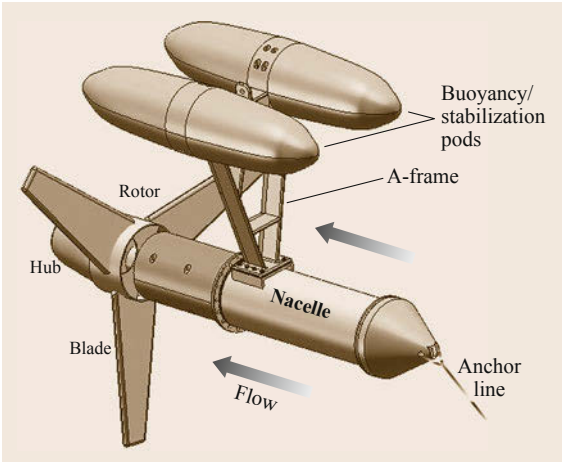


Fig. 46.4 Positively buoyant version of 20 kW SNMREC test turbine, showing main components

rotor is aft, unlike many wind systems. Internal components, including a gear-box and generator, along with other electrical components, are housed inside the nacelle (Fig. 46.5).

This system is designed to float at 50−100 m while moored to an anchor. A later variant is negatively buoyant, as shown in Fig. 46.5, to be deployed from a ship. In both versions, the two pods act as stabilization, with a moment arm created by the A-frame to provide resistance against the torque of the rotor to prevent the turbine from rolling. As an alternative to such stabilization, some systems are being designed with two counter-rotating rotors connected by a framework resembling a wing, such as that shown in Fig. 46.6. In this concept, a wing can generate lift for the system, providing a means for depth control.

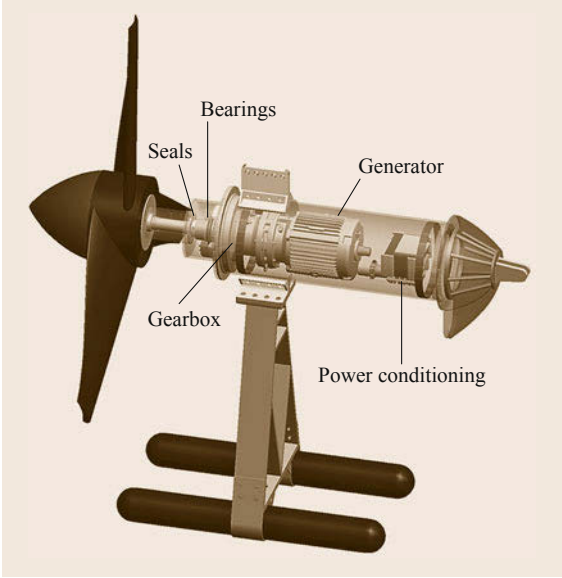


Fig. 46.5 Negatively buoyant version of Fig. 46.4 turbine system, with internal components depicted. This version is to be lowered from a work boat for testing purposes

An important concept in axial-flow rotor design is the *tip-speed ratio*, or TSR. This is the ratio of the speed of the tips of the rotor blades to the undisturbed flow speed U_u. Given a rotor of diameter D rotating at a frequency f (often expressed as revolutions-per-minute, or RPM, but formally having units of s^{-1}), the TSR is

$$\lambda = \frac{2\pi f(D/2)}{U_u} \,. \tag{46.11}$$

Typically, for a 3-blade rotor such as those depicted in Figs. 46.4–46.6 the optimum TSR is in the range $6 \lesssim \lambda \lesssim 7$. A physical interpretation of this provides insight into the importance of this parameter.

Rotor blades are most hydrodynamically efficient when undisturbed flow acts on them. It must be the case, therefore, that λ cannot be too large, because, for a given streamline, the turbulence generated by a given blade would not have time to clear the disk of the rotor before the arrival of the next blade. On the other hand, the smaller the value of λ, the less power will be generated (if $\lambda = 0$, the rotor is not turning and no power is generated).

Consider a highly efficient turbine for which the overall coefficient of power approaches the Betz limit, $C_P \approx C_B$. From the analysis in the previous section, it is easily shown that $U_t = \frac{2}{3}U_u$. The time it takes for an n-bladed rotor to turn the distance between two of the blade tips is $\Delta t_r = (nf)^{-1}$. During this time, the undisturbed water just upstream of the turbine in Fig. 46.1 will travel a distance Δx, which is

$$\Delta x = U_t \Delta t_r = \frac{2}{3}\frac{U_u}{nf} = \frac{2}{3}\frac{\pi D}{n\lambda} \,.$$

This means, for a 3-blade rotor turning at $\lambda = 6.5$, that $\Delta x \cong 0.1D$. If the size of the rotor hub is neglected in comparison to the blade length, then the blades are approximately $D/2$ in length, so that Δx represents approximately 20% of a blade length in this case. Because the cross-section of rotor blades is foil-shaped, blade length is much greater than blade width, which is called the *chord* of the blade. Consequently, Δx is much larger than the width of the blade tip. This means that, near the blade tips, at least, the turbulence from a blade passage can clear the rotor plane before the arrival of the next blade if the tip's chord is smaller than at the root, and relatively undisturbed water is available to lift each blade and turn the rotor. Note that maintaining a similar

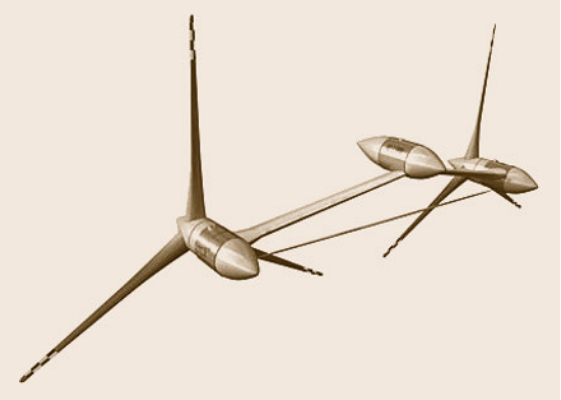


Fig. 46.6 The Aquantis C-plane dual rotor design; (courtesy of Aquantis, Inc.)

value for Δx requires that the TSR decrease in proportion to the number of rotor blades.

The order-of-magnitude analysis used previously is of interest here as well. Recall that, for wind speeds that are 10 times current speeds, the power densities of air and water are about the same, but the forces on structures in the water are a factor of 10 greater. On the other hand, blade-tip speeds, for identical TSRs, are a factor of 10 *less* in the water because, from (46.11) with U_u a factor of 10 less than the wind speed, rotors of the same diameter turn 10 times more slowly in the ocean. In addition, slow blade-tip speeds reduce the potential for cavitation, as resulting cavitation numbers lie in the range of 5–15.

The design of rotor blades, particularly for wind power, is a well-developed field with a rich literature that includes entire textbooks, and a full treatment of it here is beyond the scope of this chapter. In a fundamental sense, a rotor blade functions in much the same fashion as an airplane wing, as depicted in Fig. 46.7. The cross-section shown, a foil that is symmetric about its centerline, is positioned at a moderate angle of attack with respect to the flow. In response, the flow distorts asymmetrically, and the upper streamlines converge while the lower ones diverge. Why the flow converges above and diverges below the wing is a surprisingly complex phenomenon, well beyond the scope here. Suffice it to say that it is *not* due to the common explanation of equal transit times for air above and below but rather to more complex fluid mechanics. In any case, the convergence and divergence are well-documented from observations.

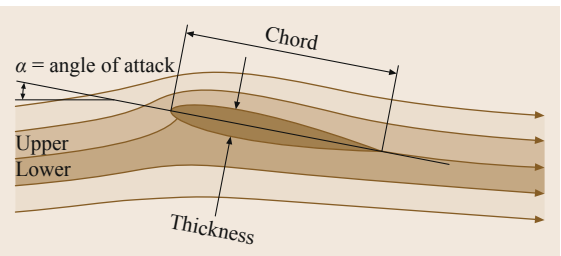


Fig. 46.7 Cross-section of an airplane wing – or a rotor blade – and its local flow field

By the continuity equation (46.2), this convergence and divergence imply that the flow accelerates above the wing and decelerates below, and, by the Bernoulli principle, implied by (46.5), the pressure above the wing decreases and that below increases – thus generating lift, a force at right angles to the flow (up, in Fig. 46.7, and opposite to gravity, for an airplane wing).

If, however, Fig. 46.7 is interpreted as a downward view of a rotor blade foil when the blade is at the twelve-o'clock position, the lift is seen to generate an anti-clockwise force (reckoned looking downstream) that would make the rotor turn in that direction. While flat plates will produce these same angle-of-attack effects, careful foil design provides improved lift as well as better structural integrity.

For an airplane wing, the flow is created by the forward motion (to the left, in Fig. 46.8) of the airplane plus any motion of the air itself. In general, for airplanes the former is much greater than the latter. However, in the case of a rotor blade, the background flow causes the blade to move, and, given TSRs of 6–7 or so, at a significant velocity. The combined relative motion of the fluid with respect to the moving blade is therefore what the blade sees as the flow. Further, because the speed of any section of a blade depends on its radial distance

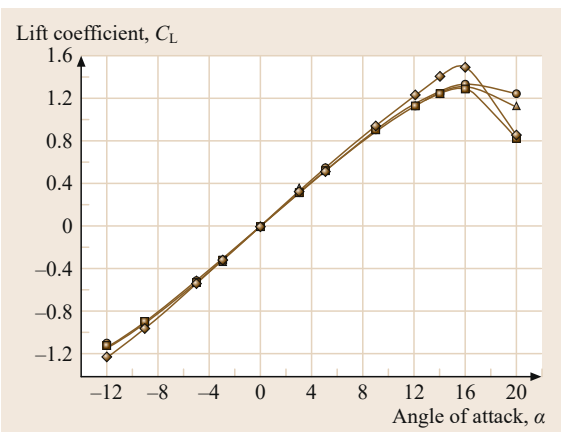


Fig. 46.8 Coefficient of lift as function of angle of attack for airfoil shown in Fig. 46.6. The airfoil begins to stall at about 16°, the point at which the experimental curve reaches the sharpest peak. The other curves represent results from computer simulations (after [46.10])

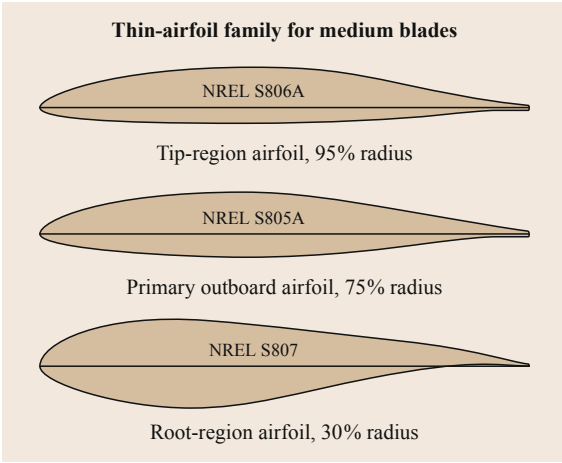


Fig. 46.9 Airfoils as a function of position along a wind-turbine blade from top to bottom, near the tip to near the blade root, shown with a standard chord. The blade thickness is increased near the root for strength; blades constructed on this pattern would decrease the chord toward the blade tip, and twist could be added as well. This family is for blades 5–10 m in length (after [46.11])

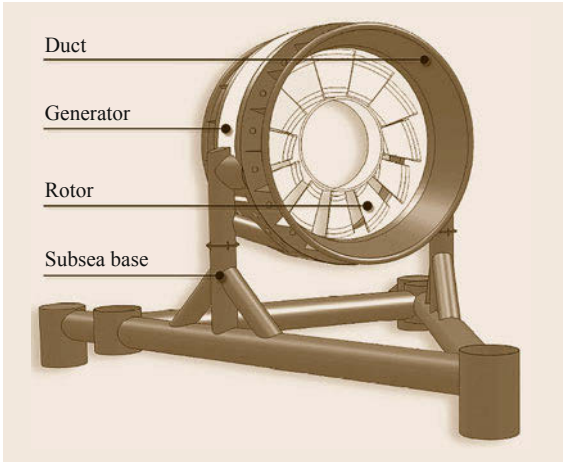


Fig. 46.10 OpenHydro, Ltd. Open Centre tidal turbine. The bearings are in the duct, and the center of the rotor is open. For tidal applications, the blades are designed to accommodate bi-directional flows (courtesy of OpenHydro & Nova Scotia Power, Inc.)

from the hub, this relative motion varies along the blade length.

To analyze blade performance, it is useful to define a lift coefficient C_L as the lift force per unit blade length divided by the dynamic pressure E (where the velocity relative to the blade is used) times the chord of the blade. Now, as shown in Fig. 46.8, lift increases with the angle of attack (up to the point that the foil stalls) as well as the square of the relative flow speed and the chord. Consequently, rotor blades are often twisted, so that the angle of attack is a function of distance from the hub, and tapered as well, in order to distribute the lifting force along the blade length in a way to accommodate the structural characteristics of the blade.

The wind power industry is sufficiently mature that standard blade shape sets exist. MRE, on the other hand, is relatively new, and OCTs, at this writing, either select standard blades or develop new, proprietary shapes. The US Department of Energy's (DOE) National Renewable Energy Laboratory (NREL) has taken a lead role in the wind industry to develop families of wind-turbine blades that are based on modified versions of US National Advisory Committee for Aeronautics (which transitioned to become the National Aeronautics and Space Administration (NASA) in 1958) standard airfoils. An example of how foil shape changes along the length of the blade is shown in Fig. 46.9, which is taken from [46.11].

In addition to systems with open rotors such as those in Figs. 46.4–46.6, other axial-flow systems with ducted rotors have been designed and fielded. The rotors of these systems tend to resemble those of high-pressure turbines rather than aircraft propellers. For example, the OpenHydro Group, Ltd. deployed its Open Centre system (Fig. 46.10) in several locations in Europe and North America for testing. As noted previously, the ducting can also be designed to increase the effective area of the rotor's sweep, thus increasing power production.

These examples of axial-flow systems are but three of a wide range of designs. The US Department of Energy maintains a database of technology describing systems that have been reported in the literature and online [46.12].

46.3.2 Cross-Flow Systems

Also included in the DOE database are examples of other rotating current energy conversion devices that can be considered cross-flow systems. In these systems, the axis of rotation of the turbine is at right angles to the flow; generally the axis is oriented either horizontally or vertically. The simplest such rotors, such as the Savonius rotor (Fig. 46.11a), are turned by the drag force of the moving fluid. These systems are not particularly efficient because the net drag available to turn the rotor is the difference between the drag on the two curved surfaces – the concave surface will *capture* the moving fluid and induce more drag than the convex surface, so that the rotor works much like a cup anemometer.

Efficiency can be improved by replacing the curved surfaces of the Savonius rotor with lifting surfaces in

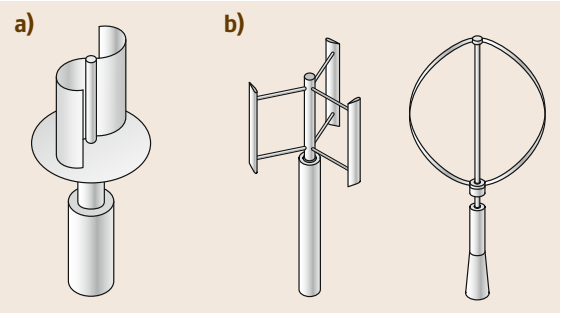


Fig. 46.11a,b Simple vertical-axis cross-flow rotors. **(a)** Savonius rotor; **(b)** Darrieus rotors

the form of blades, which then becomes a Darrieus rotor (Fig. 46.11b), with either straight or curved blades. Efficiency can be improved further by using helical blades, and these are the basis for some of the more promising designs today. For example, Ocean Renewable Power Corporation uses a helical-bladed, horizontal cross-flow design for its TideGen system (Fig. 46.12).

Although these cross-flow systems appear to be radically different from axial-flow systems, the fundamental principles discussed previously still apply. In particular, the power capability of a specific design depends on its size (the area A swept by the rotor), its efficiency (embodied in the power coefficient C_P) and the power density of the flow Ψ, and (46.10b) applies.

In wind power systems, the traditional axial-flow design dominates installed capacity, in part because they are more efficient ([46.13] and Fig. 46.13). There are, however, other important considerations for MRE, including structural requirements, site restrictions, and flow variations.

Recall, for example, the order-of-magnitude analysis comparing atmospheric and oceanic energy recovery. For nearly identical power densities, the ocean currents are about a factor of 10 less than the wind speeds, but in this circumstance, the forces on oceanic equipment are a factor of 10 higher than in the atmosphere – indeed, the previous results imply that the force/power ratio is proportional to U_u^{-1}. This suggests that vertical-rotor Darrieus turbines, such as those in Fig. 46.12, have far more challenging design requirements for oceanic applications, as all of the force on the device is transmitted through the single bearing at the base of the unit. Even if a cage is constructed around the rotor to allow bearings at the top and bottom, there remains the issue of affixing the system to the seabed. A mast-mounted system would present the same challenges as the single-bearing systems of Fig. 46.11.

Moreover, the horizontal configuration can be more advantageous from the perspectives of site and flow. Tidal channels can be shallow, and a horizontal axis-system would occupy less of the water column than a vertical system, thereby minimizing potential impacts on surface traffic. And vertical shear in the ocean tends to be much larger than horizontal shear, so systems such as that in Fig. 46.11 would tend to operate in a more homogeneous flow environment.

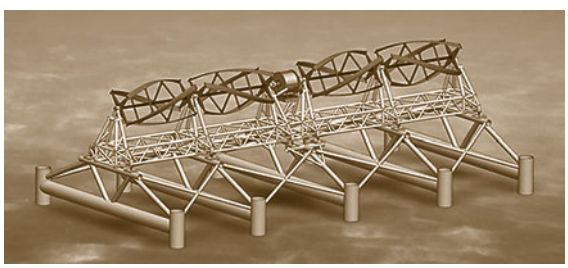


Fig. 46.12 Ocean Renewable Power Corporation's TideGen system for tidal channels. These are horizontal Darrieus rotors with helical blades (courtesy of ORPC)

46.3.3 Tidal/Open-Ocean/In-Stream Differences

Although, as shown in Fig. 46.13, axial-flow systems tend to be more efficient than cross-flow systems, the latter offer advantages in certain situations, tidal flows in particular. This is because tidal flows reverse direction twice a day, and bi-directional axial-flow rotors are unable to take advantage of the lifting characteristics of well-designed blades that makes them potentially so efficient. The alternatives, such as additional moving parts to reverse the orientation of the rotor to the flow or feather its blades through a large range, have their own practical and reliability challenges.

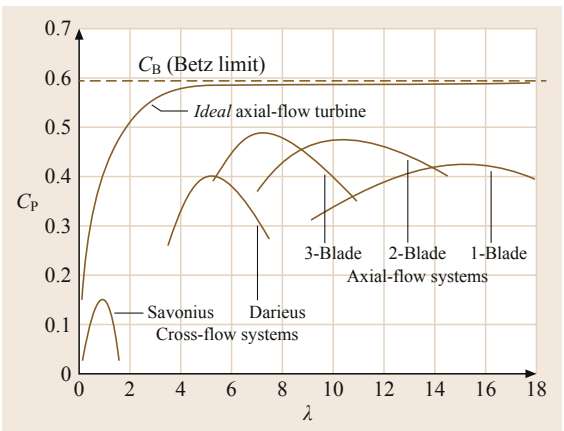


Fig. 46.13 Power coefficient as a function of TSR for several types of systems (after [46.13])

Table 46.1 Tidal/open-ocean/riverine trade-offs

Resource/issue	Current	Water depth
Tidal current	Bidirectional	Shallow: Pilings/ platforms
Open-ocean current	Unidirectional	Deep: Moorings
Riverine	Unidirectional	Shallow: Pilings/ platforms

Cross-flow systems, however, work equally well with flow in either direction, if properly designed. Thus it may be that, for certain sites, the relatively inefficient horizontal Darrieus-rotor systems would be more cost-effective. When depth restrictions of shallow tidal channels are also taken into account, this advantage multiplies.

These two factors, current and water depth, summarize the main differences between tidal, open-ocean, and in-stream river currents, as shown in Table 46.1, which notes, in the third column, the likely methods of deployment for MRE devices in each circumstance. Each type of resource has both an advantage and a disadvantage: tidal current locations offer relatively simple deployment methods but require that the bidirectional currents be managed. Open-ocean currents are sometimes unidirectional, but deployment is complex – indeed the systems must be designed around the requirement that they can be moored in a stable fashion, as discussed in the next section. River in-stream currents are both unidirectional and relatively shallow, but other considerations – the navigation hazard they present and interaction with debris, in particular – pose additional challenges.

There are, of course, other systematic differences between the two resources, such as turbulence levels and ecosystems. Turbulence levels in some tidal channels, in particular, can be daunting for designers. Careful site characterization is required to determine localized factors such as this.

46.4 Supporting Infrastructure

As suggested above, there are two main methods for deploying MRE equipment for current applications: masts (or pilings) and bases rigidly fixed to the sea bed, and moorings, where an anchor and its rode secure the equipment, which floats in the water column. Other possibilities exist for special circumstances, such as mounting equipment on bridge abutments or cables to shore securing a floating barge; the latter in use for many riverine applications.

46.4.1 Moorings

Deep water locations – and there is no strict definition of *deep* in this context – generally preclude the use of masts or bases for deployment of OCTs. For one thing, deep water makes masts impractical; for another, open-ocean currents tend to be strongest near the surface.

In deep water, OCTs will require moorings whose cables and anchors can provide sufficient holding strength and whose overall design can provide stability against variations in the current. Much of the engineering of these mooring components is discussed in the chapter on cables (Chap. 38) in this volume. However, because the currents are known to be subject to variations in both speed and direction, as functions of both time and height above the bottom (Fig. 46.14), and because most of the existing knowledge-base was developed for structures that float on the surface, it is useful to examine the stability of a simple mooring/OCT system in the presence of vertical shear. Moored OCTs will float in the water column well below the surface, for reasons noted above, and the dynamics of such systems are different from surface systems.

Consider the force diagram shown in Fig. 46.15. The current, flowing from left to right, induces a horizontal drag force $\boldsymbol{D}$ on the OCT, which is also subject

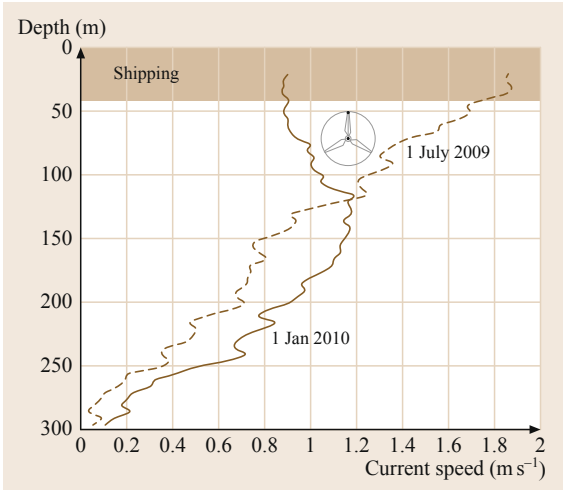


Fig. 46.14 Current speed as a function of depth in the Florida current 16 km offshore Dania beach. The stippled area at the top denotes a shipping-lane depth zone, and the symbol represents (to scale) a 40 m diameter system deployed at 70 m. Note that the sign of the vertical shear in this layer is different for the two profiles

to a net upward buoyancy force $\boldsymbol{B}$. The vector sum of these two forces is balanced, in steady state, by the tension force $\boldsymbol{M}$ on the mooring rode (the anchor line).

Figure 46.15 also shows the associated mooring geometry, in which the catenary of the mooring rode is ignored, a simplification that does not affect the analysis here. In steady state, it is clear that the scope S of the rode is R/H and that this also equals M/B (expressed as magnitudes).

A linear perturbation analysis of this system takes the various quantities to be the sum of a steady base state $[\,]_o$ and a small, time-dependent perturbation $[\,]'$. We (i) use the common notation $\frac{d[\,]}{dt} = \dot{[\,]}$, with two overdots for the second derivative, (ii) assume that the OCT has mass m, and (iii) neglect friction.

The reference frame of the OCT's mooring is properly formulated in polar coordinates, with the origin at the anchor. The two coordinates are the angle of the rode, θ, and the radial distance along the mooring rode. Projecting the drag force $\boldsymbol{D}$ onto these coordinates yields the two components D_θ and D_R shown in Fig. 46.15. As $\boldsymbol{D}$ varies with current speed, variations in D_R change the magnitude of $\boldsymbol{M}$ accordingly, while variations in D_θ induce a rotation about the anchor, changing θ. If ω is the rotation rate about the anchor, the appropriate expression of Newton's second law here is

$$mR\dot{\omega} = mR\ddot{\theta} = D_\theta = |\boldsymbol{D}|\sin(\theta)\,. \tag{46.12}$$

In order to analyze the effects of vertical shear, we set the perturbation drag $D'_\theta|\boldsymbol{D}'|$. $\sin(\theta_o) = \gamma H' \sin(\theta_o) = \gamma(H_o/R)H'$ and note that changes in $H' = R\sin(\theta')$. Finally, as shown in the upper left inset in Fig. 46.15 as a reminder, for small angles $\sin(\theta) \approx \theta$. Putting this all together gives

$$\ddot{\theta}' + \frac{\gamma}{mS_o}\theta' = 0\,, \tag{46.13}$$

where $S_o = R/H_o$ is the base-state scope of the rode.

Solutions to (46.13) take the form $\theta'\theta_1[\exp(i\omega t)]$, with θ_1 an arbitrary constant. The characteristic equation for this yields

$$\omega = \begin{cases} \varphi & \text{if } \gamma > 0 \\ i\varphi & \text{if } \gamma < 0 \end{cases}, \quad \text{where } \varphi \equiv \sqrt{\frac{|\gamma|}{mS_o}} > 0\,. \tag{46.14}$$

For cases in which the current speed, and therefore D'_θ, increase with height above the bottom (current is faster near the surface), then, small perturbations will lead to a free oscillation whose frequency depends on the base-state geometry, the mass of the OCT, and the strength of this positive shear. Of course, in the real ocean, this free oscillation would be damped by the frictional forces that are ignored here.

Cases of negative shear, $\gamma < 0$, in which the current speed decreases with height above the bottom (or, equivalently, increases with increasing depth from the surface), are of more interest. Here, the solution no longer oscillates

$$\theta' = \theta_1[\exp(-\varphi t) + \exp(\varphi t)] \text{ if } \gamma < 0\,. \tag{46.15}$$

The first term is of no concern, but the second term shows clearly that in cases of negative vertical shear, small depth perturbations will grow exponentially. Thus, in the absence of additional controls, negative shear is inherently unstable for moored OCTs.

It is worth re-emphasizing that this result depends on the assumption that the net buoyancy $\boldsymbol{B}$ is constant. It may therefore be possible to control this instability with variable (net) buoyancy, which could involve variable lift from the OCT control surfaces.

The physical basis for the instability can be stated in simple terms, because (for the assumptions here) flow perturbations affect the drag while the net buoyancy remains constant. Relative to a balanced base state, a slow-flow perturbation will create an imbalance of (positive) buoyancy, causing the OCT to rise; a fast-flow perturbation will create an imbalance of extra drag and therefore a buoyancy deficit relative to the base-state balance, causing the OCT to dive. If the OCT rises or dives into faster or slower water, respectively, then the imbalances will tend to be corrected. On the other hand, if the slow-flow, rising imbalance puts the OCT into even slower water, the imbalance will be amplified, as it will for the fast-flow imbalance that dives the

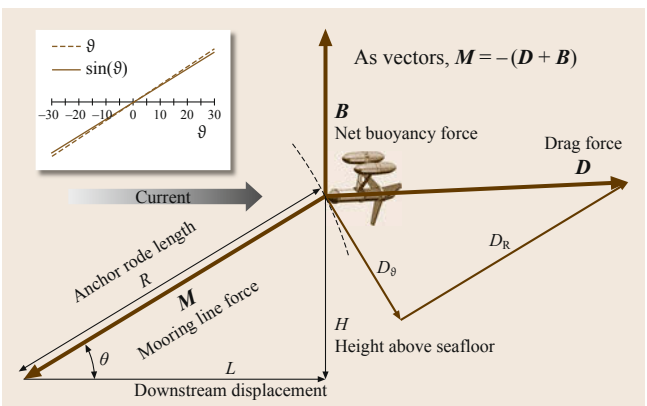


Fig. 46.15 Force diagram of a simple mooring. Current flows from left to right

OCT into faster water. In other words, the configuration of the rode is such that it pulls the OCT toward water of the same sense as the initial perturbation in the unstable case.

Note that the strength of the instability – the magnitude of the time constant φ – depends inversely on the base-state scope of the mooring: longer scope moorings, with longer time constants, are less sensitive to perturbations and therefore more amenable to control processes. Of course, these moorings are also more expensive due to the cost of the mooring rode.

In the context of this discussion of instability associated with vertical shear in the current, it is appropriate to include a qualitative look at two secondary effects on bottom-moored OCTs when one of the assumptions here is relaxed. These effects cannot be quantified without a dynamic model of a specific OCT and its mooring, but there are certain qualitative aspects of design that factor in to the stability issue.

Note that, in Fig. 46.15, the basic force balance of the system under consideration includes a *net* buoyancy force $\boldsymbol{B}$ (which, although a vector, can be assumed to have only a vertical component). This net buoyancy force is the sum of the buoyancy and whatever lift the design of the OCT generates; the previous results depend on the assumption that $\boldsymbol{B}$ is constant. In reality, the lift is likely to be dependent on both current speed and the OCT's angle of attack. Both of these dependencies factor in to the discussion of vertical shear.

Consider first the effects of vertical shear on angle of attack (Fig. 46.16). Below the lifting-surface stall limit, increases in angle of attack will tend to increase lift. To the extent that there is a cross-rotor torque introduced by vertical shear, the angle of attack becomes part of the overall effect of shear. For both leading-rotor systems and trailing-rotor systems, positive shear would tend to induce nose-up pitch, increasing the angle of attack; the opposite would happen with negative shear. Positive vertical shear, then, increases lift and negative vertical shear decreases lift. The latter would tend to exacerbate the instability associated with negative vertical shear discussed above.

The effects on lift of current speed are also relevant. In general, increased speed will increase lift (again, assuming that the system's angle of attack is below the stall limit). In the case of positive vertical shear, the overall effect here will be the same as that for the changes of angle of attack. In the case of negative vertical shear, however, they will be the opposite, tending to counterbalance the exacerbation of the fundamental instability. Quantifying the net effect of these two components of lift is not possible without a detailed model of the system under consideration. This result, however, argues for active lifting surfaces so that these effects can be controlled.

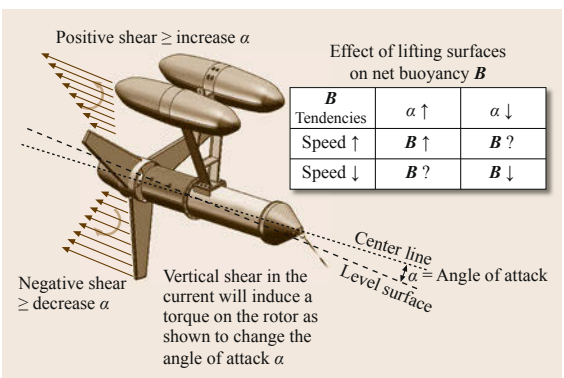


Fig. 46.16 Effects of vertical shear on lift

46.4.2 Masts, Bases, and Platforms

Shallow tidal channels, despite their potential for higher turbulence levels, eliminate the challenges associated with buoyant systems such as those discussed above. These channels allow mounting equipment directly on the sea bed in a rigid fashion, such as is implied in Figs. 46.10 and 46.12, or higher in the water column using masts, an example of which is shown in Fig. 46.17.

Thus, most tidal and river in-stream hydrokinetic devices are either rigidly attached to the channel floor or attached to a floating platform, such as a barge. Examples of barge-mounted turbines are the relatively small-scale vertical-axis devices deployed in Ruby and Eagle Alaska, which were mounted to the center of

Fig. 46.17 SeaGen tidal generation system, with two counter-rotating rotors that can be raised and lowered on the center mast (shown *raised* here) (courtesy of Marine current turbines/Siemens)

moored catamaran barges [46.14]. This mounting configuration keeps the electronics above the water line, which minimizes both equipment and maintenance costs. Drawbacks to this mounting approach include the installations being easily visible and that power cables often need to be run from the floating platform to the river bed before running to shore, leaving them exposed to possible entanglement.

Bottom-mounted devices can either be partially or completely submerged. The tidal device shown in Fig. 46.17 is attached to a mono-pile, with the control room above the water. The two horizontal axis rotors on the system can be lowered into the water during normal operation or retracted from the water for maintenance. While this approach does have a surface presence, it is much smaller than if the system were mounted from a floating platform. Alternatively, many bottom-mounted tidal devices are completely submerged. While systems such as these are attractive for many reasons, maintenance and monitoring can be challenging.

Individual device designs are typically driven by site-specific requirements, including water depth and flow characteristics. As discussed above, the flow type, tidal verses nontidal, is one important consideration. Nontidal flows are nearly unidirectional and therefore devices are optimized to take advantage of the known flow direction. In rivers, the magnitudes of these flows are driven by thawing snow or rainfall and therefore are more seasonally and event driven. Because specific events contribute to the magnitude of these flows, the power is less predictable than tidal flows. However, most large river systems that are good candidates for instream hydropower will be able to produce power year round. Conversely, devices deployed in tidal flows are typically designed to produce power from flows coming in two discrete directions. Vertical-axis systems are not typically impacted by flow direction and their performance does not degrade in these flows. Axial-flow turbines are more impacted by a changing flow direction, and therefore these devices often have rotors that have similar efficiencies regardless of the flow direction or rotate so that the rotor is always oriented in the optimal direction. Since tidal cycles are the primary drivers of these flows, the power potential for these installations are more predictable. However, systems operating in tidal flows do have a period of time between tidal cycles when no power is produced.

Design and performance of masts, bases, and platforms is a topic far beyond the scope of this chapter; some of the considerations are discussed in other parts of this volume. One factor that must be considered carefully, however, is the torque imparted to the attachment on the sea bed of masts and bases due to the forces on the OCT rotor. As noted previously, the latter are an order of magnitude larger than for wind systems, and designs must take this into account.

46.4.3 Other Components

An OCT system is, fundamentally, a machine for converting the kinetic energy of an ocean or river current into electrical energy. The rotor, turned by the current as discussed above, ultimately drives an electrical generator of some type, and the resulting voltage is either transmitted or stored. Only the most obvious of system components associated with this process are depicted in Fig. 46.5. Although most of these components, as well as those not depicted, are not unique to the world of MRE, there are challenges involved with using them in OCT systems.

As discussed previously, one advantage of OCTs is their relatively slow rotation rate, with blade-tip speeds approximately 10 times slower that wind-power systems operating at the same power densities. However, electrical generators operate most efficiently at high rotation rates. This means that the gearbox design and integrity are crucial to effective OCT operation.

Once the torque of the rotor has been transmitted through the gearbox, it drives a generator of some design. Both alternating- and direct-current generators have been proposed and designed into OCT systems, and each has its advantages and disadvantages. Alternating current (AC) generators are more efficient and, especially for high-power applications, relatively simple, but direct current (DC) transmits more efficiently through under-sea cables. Conversion between the two is always less than 100% efficient, so, as with mounting methods and overall system design, choices must be made on the basis of specific site requirements.

Similarly, cable transmission, to the on-land grid, or local use of the generated power to create a more portable fuel, is also a trade-off that depends on the nature of the resource and its location relative to load centers. The Florida current, which flows northward along the east coast of Florida approximately 25 km off shore, is relatively close to the major load center of the SE Florida metropolitan area, so direct transmission to the grid could be appropriate there. Conversely, there are significant tidal resources in many of the bays and inlets along the coast of Alaska, but there is no nearby load center to utilize generated power. Consequently, it may be more economically viable for that power to be used to create hydrogen (by electrolyzing sea water) or another portable fuel, which can be transported to where it is needed.

It is also relevant here to note that individual OCTs are unlikely to be able to capture a significant frac-

tion of the power available at a particular location, so that arrays of OCTs are likely to be deployed in commercial installations. This raises the issue of specific deployment strategies – how to design the array and the positions in it of the constituent OCTs to most efficiently capture the power of the resource. Although there is much to be learned from the experience of *wind farms* in this regard, there is also much research needed. Again, the solutions will emerge as site-specific deployment strategies are developed.

46.5 Beyond Engineering

MRE recovery requires a seamless integration of many of the ocean engineering topics discussed in this volume, especially including hydrodynamics, structures, corrosion and biofouling, acoustics and instrumentation, and underwater control systems. It also requires a clear appreciation of many topics that are normally considered to be outside engineering disciplines. Those most relevant to MRE from ocean currents are:

- Environmental impacts – from sea bed to the surface, both up- and downstream from the project
- User conflicts – including recreational, military, and commercial entities
- Public perception – because there are still so many unknowns, educating and informing the public is a necessary aspect of a project gaining access to public lands for energy generation.

All these are wrapped into the legal and regulatory processes and procedures associated with the particular location for the project. The planning process often requires a formal study of environmental impacts prior to permitting, and the various issues associated with the topics of concern need to be addressed.

In particular, the importance of public perception cannot be overestimated, as it influences the processes in place to address the technical issues associated with environmental impacts and it strongly affects issues related to user conflicts. What is critical is that initial impressions that become part of the public consciousness about a particular project tend to be extremely difficult to change. Leasing and permitting processes generally include such events as public hearings.

46.6 Summary

Equipment that can convert the kinetic energy of oceanic currents into other forms of power have much in common with wind systems, but there are critical differences between applications. In particular, while power densities in the oceans tend to be of the same order of magnitude as in the atmosphere, the forces on oceanic equipment tend to be an order of magnitude larger. On the other hand, for the rotating devices discussed here, rotation rates tend to be an order of magnitude slower. Nonetheless, animal strikes, especially when endangered or threatened species are present, are among environmental concerns.

Because oceanic systems are yet in their infancy, a general discussion has been presented here.

References

46.1 R.H. Stewart: *Introduction to Physical Oceanography* (Texas A & M University, College Station 2008), available online at http://oceanworld.tamu.edu/home/course_book.htm

46.2 C. Wunsch: What is the thermohaline circulation?, Science **298**(5596), 1179–1181 (2002)

46.3 T. Burton, N. Jenkins, D. Sharpe, E. Bossanyi: *Wind Energy Handbook* (Wiley, Chichester 2001)

46.4 A. Betz: *Wind Energie und ihre Ausnutzug durch Windmühlen* (Vandenhoeck and Ruprecht, Göttingen 1926), Wind energy and its extraction through windmills

46.5 A. Betz: *Introduction to the Theory of Flow Machines* (Pergamon, Oxford 1966), tranls. by D.G. Randall

46.6 A.S. Bahaj, A.F. Molland, J.R. Chaplin, W.M.J. Batten: Power and thrust measurements of marine current turbines under hydrodynamic flow conditions in a cavitation tunnel and a towing tank, J. Renew. Energy **32**, 407–426 (2007)

46.7 IEC: Marine Energy-Wave, Tidal and Other Current Converters. Part 2: Design Requirements for Marine Energy Systems (International Electrotechnical Commission, Geneva 2013) TC114/PT 62600-2

46.8 K.D. Leaman, R.L. Molinari, P.S. Vertes: Structure and variability of the Florida current at 27° N: April 1982–July 1984, J. Phys. Oceanogr. **17**, 565–583 (1987)
46.9 H.P. Hanson, S.H. Skemp, G.M. Alsenas, C.E. Coley: Power from the Florida current: A new perspective on an old vision, Bull. Am. Meteorol. Soc. **91**, 861–867 (2010)
46.10 D.C. Eleni, T.I. Athanasios, M.P. Dionissios: Evaluation of the turbulence models for the simulation of the flow over the national advisory council for aeronautics (NACA) 0012 airfoil, J. Mechanical Eng. Res. **4**, 100–111 (2012)
46.11 J.M. Tangler, D.M. Somers: *Airfoil Families for HAWTs* (National Renewable Energy Laboratory, Colorado 1995), NREL TP-442-7109
46.12 Office of Energy Efficiency and Renewable Energy: Water Power Program, US Department of Energy, http://www1.eere.energy.gov/water/hydrokinetic
46.13 E. Hau: *Wind Turbines: Fundamentals, Technologies, Application, Economics* (Springer, Berlin, Heidelberg 2006)
46.14 J.B. Johnson, D.J. Pride: *River, Tidal, and Ocean Current Hydrokinetic Energy Technologies: Status and Future Opportunities in Alaska* (Alaska Energy Authority Juneau, Alaska 2010)

47. Harvesting Energy by Flow Included Motions

Michael M. Bernitsas

Marine hydrokinetic (MHK) energy is clean, renewable, and available worldwide. It comes in two forms: vertical in waves and horizontal in currents, tides, and rivers. Apart from a few major ocean currents, most of the ocean currents have flow speeds less than 3 kn and most rivers have speeds less than 2 kn, making harvesting of their MHK energy by steady-lift technologies (turbines) challenging. Horizontal MHK energy can also be harnessed using alternating-lift technologies (ALTs). Fish utilize alternating lift to propel efficiently in water either as individuals or in schools. Engineered structures – bluff bodies, such as circular cylinders and prisms, or slender bodies like hydrofoils – may generate alternating lift in quasi-steady uniform flows. When these structures have scale-relevant flexibility, severe flow–structure interaction (FSI) phenomena may be induced. In typical engineering applications, FSI phenomena are destructive and, thus, avoided by design or suppressed using excessive damping or appendages. If FSI are instead enhanced, they may result in vigorous flow-induced motion (FIM) of the body, leading to the conversion of MHK energy to potential and kinetic energy in a mechanical oscillator. Hydrofoils can harvest MHK energy through flutter – a well-studied and understood form of instability. On the other hand, bluff bodies, such as circular or rectangular cross-section cylinders, may exhibit several forms of FIM, individually or in *schools* that have been studied extensively but still are not well understood for either suppression or enhancement. Those FIMs are vortex-induced vibration (VIV), galloping, buffeting, and gap flow in multibody interactions. When enhanced, they convert MHK energy to mechanical energy with high-power density (power-to-weight ratio) even from low-speed horizontal flows. This chapter presents an overview of the concepts of ALTs, the underlying physical principles, the available experimental and computational methods for studying the relevant FIM, the research challenges that have been overcome and those lying ahead, field-deployment progress, technology development, and bench marking.

47.1 **Hydrokinetic Energy in Horizontal Flow** 1166
47.1.1 Marine Hydrokinetic Energy 1167
47.1.2 Potential, Requirements, and Challenges 1168
47.1.3 Steady-Lift Technologies 1170
47.1.4 Alternating-Lift Technologies 1172
47.2 **Alternating-Lift Technologies: The VIVACE Converter as a Case Study** 1178
47.2.1 Concept, Scales, and Principles 1178
47.2.2 Scales, Models, Prototypes 1183
47.2.3 The Underlying Principles 1189
47.3 **Methodology and Tools in Support of Development** 1202
47.3.1 Experimental Facilities 1203
47.3.2 Damping Model 1209
47.3.3 Flow Visualization and Vortex Tracking 1210
47.3.4 Field Tests 1212
47.3.5 Computational Fluid Dynamics 1213
47.3.6 Mathematical Model of Harnessed and Dissipated Power 1225
47.3.7 Variable Added Mass Mathematical Model 1233
47.3.8 Physical Models and Model Equivalence 1234
47.3.9 Benchmarking 1236
47.4 **Nomenclature** 1237
References 1238

Marine hydrokinetic (MHK) energy – available in oceans and rivers – has been considered a major potential source of clean renewable energy since the Middle Ages, when tidal mills were used to grind grain in Northern Europe [47.1]. The first patent for a wave-energy converter was filed in 1799. In spite of such an early start, efforts to utilize this valuable resource have been limited all through the 1980s [47.2–9].

Through the early 1990s, the US Department of Energy (DOE) supported efforts in diverse aspects of marine energy. After a 15-year gap in funding, DOE started supporting major efforts in development of novel MHK energy technologies in 2008 [47.1, 5, 10]. The path to development of commercial devices has been very difficult due to the harshness of the marine environment, the cost of field deployment, and failures of pioneering devices leading to recent bankruptcies and project cancellations [47.11].

Several studies have estimated the available MHK energy in waves [47.3, 7, 12–16], and ocean currents [47.2, 6, 14, 16–19], rivers [47.20–23], or in all three of them [47.1, 4, 21, 24–27]. Those estimates vary widely because assumptions regarding the accessibility of energy source and technology performance, speculation, and efficiency vary widely [47.2–4, 17, 18]. Regardless of the uncertainty in estimation, comparison to past and projected worldwide production and consumption of energy in general and electricity in particular [47.4, 22, 28, 29] show that MHK energy cannot be overlooked as long as it can be generated at a competitive levelized cost of energy or LCOE [47.10, 30].

MHK energy comes in two forms, vertical in waves and horizontal in currents, tides, and rivers. In this chapter, studies in harnessing of the horizontal MHK energy from horizontal quasi-steady uniform flows are described. The majority of ocean currents are slower than 3 kn and most rivers are slower than 2 kn making harvesting of their MHK energy by turbines or watermills challenging.

Turbines and watermills [47.21, 31–33] are based on lifting surfaces in a steady flow resulting in steady lift. The term *steady-lift converter* or *steady-lift technology* (SLT) is introduced, hereafter referring to turbines and watermills. Steady lift is generated by lifting surfaces in air, such as bird wings, airplane wings, sails of sailboats, and blades of wind turbines. Horizontal MHK energy can also be harnessed using *alternating-lift converter* or *alternating-lift technology* (ALT). Fish utilize alternating lift to propel efficiently in water either as individuals or in schools [47.34–42]. SLT is more palatable to humans because we live in air. The picture in the marine environment though is totally different as fish use almost exclusively alternating lift to efficiently move through water.

Engineered structures with bluff cross-sections, such as circular cylinders and prisms, or slender structures, such as airfoils and hydrofoils, may generate alternating or oscillatory lift in uniform flows. When such structures are flexible or are rigid but mounted on flexible support, and have one or more natural frequencies of vibration in the range of frequencies of the oscillatory lift, significant flow–structure interaction (FSI) phenomena may be induced [47.43, 44]. In typical engineering applications, FSI phenomena are destructive and, thus, avoided by design or suppressed using excessive damping or appendages. If they are instead enhanced, FSI may result in vigorous flow-induced motion (FIM) of the body, leading to the conversion of MHK energy to potential and kinetic energy in a mechanical oscillator. Hydrofoils can harvest MHK energy through flutter – a well-studied and understood form of instability [47.43, 45].

On the other hand, bluff bodies such as circular or rectangular cross-section cylinders, may exhibit several forms of FIM, individually or in schools, that have been studied extensively but are still not well understood for suppression or enhancement. The five most commonly occurring FIMs are as follows:

1. Vortex-induced vibration (VIV): VIV is a nonlinear oscillation, broad-range, synchronization phenomenon between the oscillatory lift created by the alternating von Kármán vortex street in the wake of a bluff body and the body's elastic [47.46–59]. VIV is a phenomenon self-limiting in amplitude; see Sect. 47.2. Some researchers interpret it as a broad-range, lock-in phenomenon [47.43, 55] and others as a nonlinear resonance over the same broad range of flow velocities [47.60] due to variable added mass, which can be measured experimentally [47.53, 54, 61–63]. It should be noted though that the dependence of added mass on relative velocity and acceleration components is way more complex than the simple term typically used in Morison's equation [47.64].
2. Galloping: This is an instability phenomenon caused by geometric asymmetry due to the body cross-sectional geometry or the flow due to upstream turbulence [47.43, 65–67]. It results in unlimited amplitude response until the elastic structure is damaged as the flow velocity increases.
3. Coexistence of VIV and galloping: Depending on the flow speed, the body geometry, and the oscillator parameters, these two FIMs may coexist in a transition region between them, may overlap, or may be separated [47.67–73]. In the first two cases, a nonlinear oscillator can be built, which has high-amplitude response starting at low speeds using VIV and with no upper limit in flow velocity using galloping; see Sect. 47.2.
4. Buffeting: This is an instability phenomenon even more powerful than galloping due to upstream vorticity impinging on the oscillator [47.43]. Buffeting though, is random in nature and cannot be utilized to design and build a reliable energy converter.

5. Gap flow in multibody interaction: When multiple cylindrical bodies are in tandem or in close staggered formation, the interaction between them may be even more powerful than the FIMs of single cylinders [47.50, 57, 74–83]. By properly designing and positioning multiple bodies, vigorous oscillation of all elastic bodies in the array/school is induced resulting in even higher efficiency conversion of MHK energy to mechanical energy in the oscillator [47.68, 69, 84].

Elongated bodies with slender cross-section, such as hydrofoils and airfoils, in steady flow are typically subjected to steady lift perpendicular to the flow direction. Depending on the relative location of the supports and the center of pressure, the motion of such structures may become unstable resulting in flutter instability [47.43, 85]. This motion converts fluid-kinetic energy to mechanical energy and subsequently to electricity, a process that has been patented in various forms [47.86, 87].

Elongated bodies with bluff cross-section, such as circular or rectangular cylinders, in steady flow typically are subjected to alternating lift perpendicular to the flow direction. They may exhibit all five forms of FIM listed above. These FIMs convert fluid-kinetic energy to mechanical energy in an oscillator and subsequently to electricity, a process that has also been patented [47.88–90].

The VIVACE converter is a unique alternating-lift converter developed in the Marine Renewable Energy Laboratory MRELab [47.91] at the University of Michigan. Its name is the acronym for *vortex-induced vibrations for aquatic clean energy* as it started as a device to enhance VIV to convert MHK energy to mechanical in the oscillator. Since then, it has grown – as defined in the patents [47.88–90] – to include several other forms of FIM, discussed in Sect. 47.2.1, *Basic Concept*. VIVACE was introduced by *Bernitsas* et al. [47.48, 49] and developed extensively since then with model tests [47.67–69, 69, 70, 73, 77, 84, 92–110], field tests [47.111], computational fluid dynamics (CFD) [47.76, 112–118], and mathematical models [47.119–121]. VIVACE uses one [47.73, 106, 107] or multiple cylindrical oscillators [47.68–70, 77, 84, 101–103] with passive turbulence control (PTC) [47.96] or smooth cylinders to harness MHK energy and convert it to electricity, with a high-power density ratio for a renewable energy device [47.95, 100]. It utilizes several forms of FIM to generate alternating lift in a very broad range of horizontal flow velocities. It may also use more complex forms of cross-sections [47.116, 118]. The basic concepts in the VIVACE patents are presented along with the underlying physical principles implemented in its operation and development in Sect. 47.2.

Development of this novel concept involves several design parameters including one or more elongated bodies with bluff cross-sections, FIMs, multibody interactions, turbulence stimulation, wake structures, variable damping, stiffness and mass ratio, variable flow-speed placing the oscillating bodies in a variety of flow regimes, interaction of shear layers, effect of upstream vorticity wakes and shear layers on the FIM of downstream bodies, variable tandem spacing and staggering of multiple oscillators, energy conversion, and optimization of power conversion. All these make product development of VIVACE a challenge that requires pioneering research in many areas of fluid dynamics, FSI, and controls. A broad arsenal of tools is being used in the research and development of alternating-lift converters and understanding the complex relevant physics. Such tools include the low-turbulence free-surface water (LTFSW) channel of the MRELab [47.91] at the University of Michigan, two towing tanks, virtual oscillators to achieve the desired linear or nonlinear damping and stiffness models, flow visualization, and vortex tracking using lasers and aluminum oxide particles, field tests in the St. Clair River in Port Huron, Michigan, and canals in the Netherlands, dedicated CFD tools developed and validated by the MRELab, mathematical models for data post-processing and analysis, and a method developed in the MRELab to reduce data collected by different devices and experimental models and techniques to equivalent FIM data for comparison.

As an energy-generating device, designing VIVACE is a challenge more daunting than designing a propeller. The first propeller was designed by Archimedes, circa 200 BC. The first screw propeller was proposed by James Watt in the late 18th century. Optimal propeller designs were generated in the 1970s. Oscillators, even if designed to be identical, are not equivalent – as blades are in propellers – due to their relative position in a school of cylinders. Flow–structure interaction is the heart of the problem, not a higher order correction. Among the challenges encountered in the research and development of the VIVACE converter are high-damping, FIM for energy harnessing, overcoming flow transition, enhancing FIM through PTC, placing various forms of FIM back-to-back to create a broad-range, high-response oscillator, and using synergistic FIM of multiple bodies for optimal MHK energy conversion.

All FIM phenomena, which flexible cylindrical structures or rigid cylinders mounted on elastic support exhibit, are highly scalable. Cylinders, as tiny as fishnet filaments or as large as Spar offshore platforms, are subjected to vortex-induced vibrations (VIVs) for the entire range of Reynolds number with the ex-

ception of three flow-change regimes: Re$<\approx 40$; flow transition inside the vortices $\approx 150 <$ Re $<\approx 400$; transition in the boundary layer from laminar to turbulent $\approx 300\,000 <$ Re $<\approx 500\,000$. Consequently, ALTs can be developed over a very broad range of applications and configured to meet topography and flow particulars as well as power-output requirements. Most important, they can harness hydrokinetic energy even from low-velocity flows where SLTs are financially nonviable [47.103].

In Sect. 47.1, marine hydrokinetic (MHK) energy in horizontal flow is considered as a source of renewable energy. The challenges in harnessing horizontal MHK energy, converting it to mechanical energy, and meeting the requirements for a successful commercial converter are discussed. MHK energy technologies are categorized as based on steady lift or alternating lift. In Sect. 47.2, ALTs are presented using the VIVACE converter as a case study because it operates in several forms of FIMs to generate alternating lift in a very broad range of horizontal flow velocity. The basic concepts in the VIVACE patents are presented along with the underlying physical principles implemented in its operation and development. Sect. 47.3 presents the arsenal of tools being used in the research and development of alternating-lift converters and how results are used to understand the complex relevant physics. Such tools include the low-turbulence free-surface water LTFSW channel of the Marine Renewable Energy Laboratory (MRELab) of the University of Michigan, two towing tanks, virtual oscillators to achieve the desired linear or nonlinear damping and stiffness models, flow visualization and vortex tracking using lasers and aluminum oxide particles, field tests in the St. Clair River in Port Huron, Michigan, dedicated CFD tools developed and validated by the MRELab, mathematical models for data post-processing and analysis, and a method developed in the MRELab to reduce data collected by different devices and experimental models and techniques to equivalent FIM data for comparison. The challenges overcome and those lying ahead and the research approach used in addressing them are presented. Among those are high-damping FIM for energy harnessing, overcoming flow transition, enhancing FIM through PTC, placing various forms of FIM back-to-back to create a broad-range, high-response oscillator, and using synergistic FIM of multiple bodies for optimal MHK energy conversion. Targets and benchmarking are discussed as well. A summary of the state of the art, conclusions, and future development are presented at the end of this chapter.

47.1 Hydrokinetic Energy in Horizontal Flow

Hydrokinetic energy is defined as kinetic energy in moving water. The term *marine renewable energy* refers to energy generated by waves, tidal currents, open-ocean currents, river currents, ocean thermal gradients, and salinity gradients. Technologies utilizing these renewable resources are referred to as MHK energy technologies. Horizontal hydrokinetic energy usually refers to hydrokinetic energy in tidal, open-ocean, and river currents, even though there is some horizontal energy in the other sources of marine renewable energy. Typically, the term vertical hydrokinetic energy refers to hydrokinetic energy in waves, even though there is some vertical energy in the other sources of marine renewable energy.

Solar energy can be considered as the original source of all these forms of MHK energy. The sun sends 174 PW of power to earth. After reflection of energy by the atmosphere, the clouds, and the surface of the earth, 89 PW of power is absorbed by land, atmosphere, and water. That power is then radiated through the atmosphere or directly from the surface of the earth back to space. In the process, the marine environment (primarily oceans) absorbs 9 PW (285 ZJ/yr), winds capture 0.19 PW (6 ZJ/yr), and biomass captures 0.057 PW (6 ZJ/yr) [47.78]. These estimates are based on the assumptions of present technology, no loss due to mechanical efficiency, and no accessibility restrictions.

Given the vast source of renewable energy and the anticipated environmental benefit, there is pressure in the general direction of renewable energy. Thirty-seven states in the USA have introduced renewable portfolio standards (RPS) under which utility companies are mandated to generate a percentage of their total power from renewable sources in order to reduce carbon emissions. The set targets will increase in the near future as new power-generating technologies are developed [47.123]. The renewable energy use is expected to increase from 9% in 2012 to 12% in 2040 in response to RPS and Federal Tax Credits (Fig. 47.1a). Renewable energy generation is expected to grow by 1.9% per year after 2012 (Fig. 47.1b). Solar, photovoltaics, and wind dominate renewable capacity growth.

In the statistics shown in Fig. 47.1, MHK, in-steam hydroelectric, and hybrid solar–thermal combined cycle are not included. The relevant conclusion of this quick review is that the MHK-energy industry has not reached

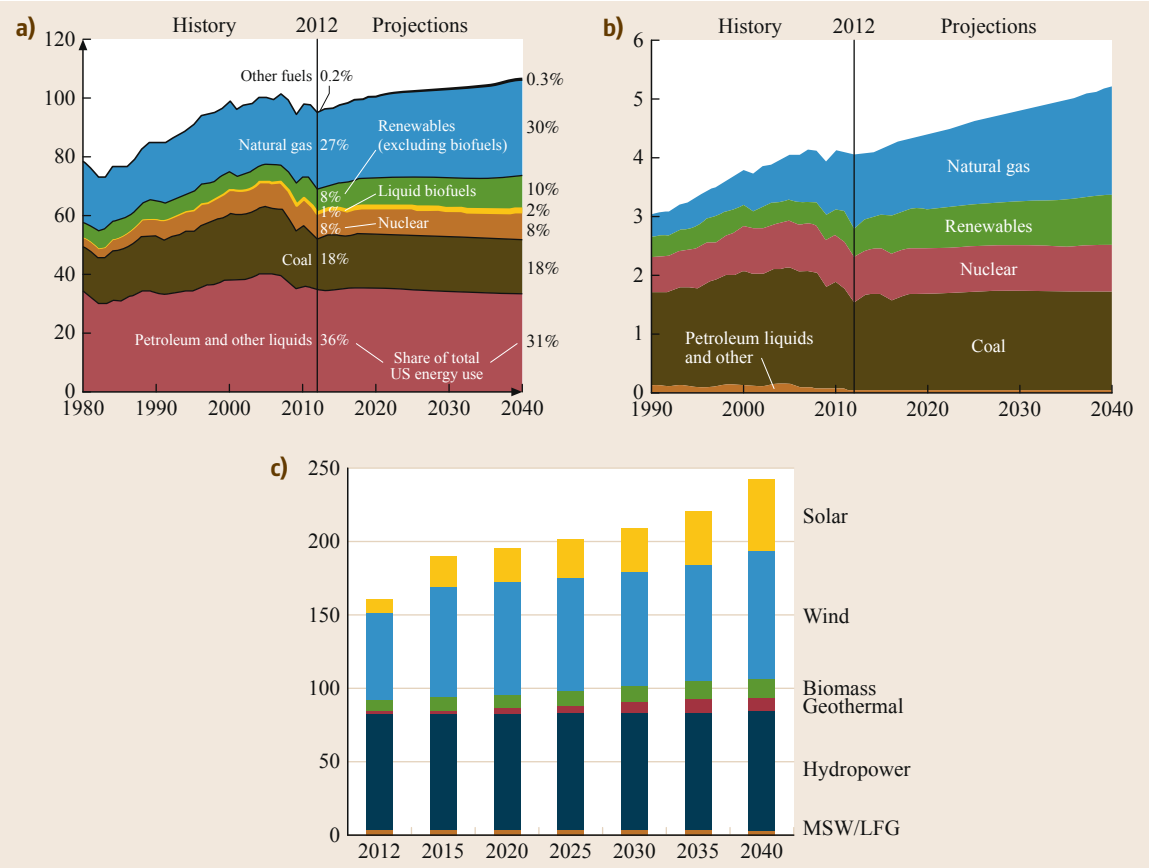


Fig. 47.1a–c Specifications of the MRELab VIVACE converter model. Energy and electricity consumption in the USA [47.122]. **(a)** US energy demand; primary energy use by fuel in 1×10^{12} BTU. **(b)** US electricity demand; generation by fuel in 1×10^{9} kWh. **(c)** US electricity generating capacity by source in GW

the maturity level where it can contribute to a national or international renewable energy portfolio.

47.1.1 Marine Hydrokinetic Energy

MHK energy, is also referred to as ocean power, marine renewable energy, and marine power. The MHK-energy industry is presently in its infancy. Many concepts have been proposed but with little technology convergence. Actually, the number of MHK-energy concepts is probably higher than the number of wind-energy concepts at the early stages of the wind-energy industry. The capacity of MHK devices installed around the world is quite small, only tens of megawatts, excluding tidal barrage plants, and these installations are generally engineering prototype test devices or small several-unit demonstration wave and tidal projects.

On the other hand, the amount of energy absorbed by the marine environment annually is vast. Ocean thermal energy is the majority with about 2 000 000 TWh, salt-gradient energy is 23 000 TWh, tidal and current energy is 22 000 TWh, and wave energy is about 18 000 TWh [47.124]. To put these numbers in perspective, the annual world electricity consumption is about 19 000 TWh and the world total energy consumption is of the order of 132 000 TWh. Thus, even though it is difficult to provide an accurate estimate of the accessible ocean power, and even more difficult to estimate the harnessable power, it is obvious that marine renewable energy is a vast resource and should not remain unexplored or underutilized. Technologies must be developed that are capable of efficiently harnessing MHK energy from areas that are easier to access in a financially viable and environmentally compatible way even if the energy source has lower power density.

The US National Renewable Energy Laboratory (NREL) has published useful reports on renewable energy. Most relevant is Chap. 9 on ocean power in reference [47.1]. It provides information on the size of the MHK energy resources for the USA. Some fig-

ures on estimate of the resources and the practicable extractable potential are summarized in the following.

Natural Wave Energy

1. The theoretical global natural wave-energy resource, including both kinetic and potential (due to the elevation of the wave) energy, is estimated between 18 000 [47.125] and 29 500 TWh/yr [47.1].
2. The total natural US wave-energy resource potential is estimated between 2100 and 2640 TWh/yr.
3. The practical extractable potential was estimated by *Bedard* et al. [47.126] at 260 TWh/yr, giving an average power output of 30 000 MW for all of the USA, including Hawaii and Alaska. They assumed a conversion of 15% wave energy to mechanical energy, a power train efficiency of 90%, and a plant availability of 90%. For a plant capacity factor of 33%, the installed capacity would be approximately 90 000 MW. The wave-generated electrical energy from this would be approximately 6.5% of the US yearly electrical energy generation in 2010 [47.1].

Natural Tidal Energy

1. The theoretical global natural tidal and current energy resources combined are estimated at 22 000 TWh/yr [47.124].
2. The total natural US tidal energy resource potential is estimated at 115 TWh/yr by *Bedard* et al. [47.17] and 111 TWh/yr by *Haas* et al. [47.19]. Both sources agree that the vast majority of the resource is in Alaska.
3. The practical extractable potential was estimated by *Bedard* et al. [47.126] only for three locations: Puget Sound in Washington, Golden Gate in California, and the Western Passage in Maine. The tidal energy available at those sources is 6 TWh/yr. The study assumed a conversion of 15% tidal kinetic energy to mechanical energy, typical power train efficiency of 90%, and a plant availability of 90% resulting in 0.73 TWh/yr of extractable power. This is equivalent to an average power of approximately 83 MW, and an installed capacity of approximately 220 MW, assuming a capacity factor of 38% [47.1].

Natural Ocean Current Energy

An ocean current is a continuous, directed movement of ocean water generated by the forces acting upon the mean flow, such as breaking waves, wind, Coriolis force, temperature and salinity differences, and tidal forces [47.1]. NREL concluded that in the United States, high-kinetic energy potential ocean current resources are found primarily in the Florida Current because of its high-core velocity of about 2 m s^{-1}. The other ocean currents have much lower flow rates and are not considered viable for energy generation. The relatively constant energy density near the surface of the Florida Current is about 1 kW m^{-2} of the flow area:

1. For the Florida Current, the natural energy of the resource was estimated at 175 TWh/yr [47.1, 126].
2. There is no established practicable extraction limit for ocean current energy. If the same engineering assumptions used for tidal energy would be applied to the Florida Current field, the extractable energy potential would be approximately 21 TWh/yr, or about 2.4 GW, of average power.

Based on the above discussion, the following conclusions are reasonable:

- The MHK energy is an abundant resource that is worth exploring by the development of new technology.
- Current technology is not capable of accessing efficiently the vast majority of the MHK energy resource at a competitive cost.
- Estimation of the available worldwide MHK-energy resource is hard.
- Estimation of the practically extractable MHK energy is approximate and depends strongly on speculation regarding the performance of technology presently under development.

47.1.2 Potential, Requirements, and Challenges

In spite of the fact that the MHK-energy industry is at its infancy, and the challenges ahead are many, MHK energy has the potential of becoming a major player in the quest for renewable and environmentally compatible energy. The potential of MHK energy, the requirements for a successful technology, and the challenges ahead are summarized below.

Potential

Besides the end product of electricity at a competitive cost, the following need to be considered in assessing the potential of MHK energy:

1. Water is the largest natural storage medium for energy and has high density; power-to-volume ratio.
2. Moving water is available worldwide and about 50% of the US population lives within 50 mile off a coast. This reduces the need for expensive storage and transportation of electricity.
3. If environmentally compatible devices that can harness MHK energy in resources of low flow speed could be developed, the environmental benefits would be substantial as shown in Table 47.1. Es-

Table 47.1 Advanced technology offers the opportunity to avoid substantial pollution

	161 MW	1700 MW	3700 MW
Offset resources			
Coal [ktons]	341	3600	7850
Natural gas [BCF]	2.3	24	52
Avoided pollution			
SO_x [tons]	7700	81 300	177 000
NO_x [tons]	4580	48 300	105 000
Particulate [tons]	3050	32 200	70 000
Greenhouse gases [equivalent ktons CO_2]	3150	33 300	72 500

Notes:
1. Offsets resources based on US average electricity generating mix
2. Avoided pollution from [47.127]

timates of the avoided pollution from installing advanced technology converters can be seen in Table 47.1. Three cases were selected to illustrate the potential savings: 161 MW (replacing the 461 MW of installed low power/low head capacity using 35% efficiency), 1700 MW (22.6% of low power/low head resources developed); and 3700 MW (includes low power/high head).

4. Horizontal MHK energy is predictable and, thus, dispatchable to an electric grid.
5. Horizontal MHK energy can be unobtrusive to people since the technology remains submerged at all times even when the resource is close to shore.

Requirements

For marine renewable energy – and particularly hydrokinetic – to contribute toward a solution of the world energy challenge, technology must be developed to generate electricity in a sustainable way. That has two attributes: First, harness energy from a clean and renewable energy source; MHK energy satisfies this requirement, second, use an environmentally compatible technology.

These two attributes can be expanded to the following list of requirements:

1. Be cost-competitive to alternative as well as conventional energy sources
2. Be unobtrusive to people; for example, remain out of sight at all times
3. Be unobtrusive to marine life; for example, by emulating fish propulsion or natural phenomena
4. Have tolerable levels of noise and electromagnetic interference
5. Have a source with high-energy density
6. Be predictable and, thus, dispatchable to an electric grid
7. Be simple to build, maintain, deploy, and retrieve
8. Be reliable and simple to maintain low cost
9. Be scalable and, thus, effective at a variety of environmental conditions
10. Be modular, reconfigurable, and adjustable to site topography
11. Be robust to environmental changes so that it functions efficiently even when the flow velocity changes
12. Be robust to floating debris and marine fouling
13. Be robust to environmental loading so that it can have a life of at least 20 years.

These requirements lead to a set of challenges as explained next.

Challenges

The challenges that the MHK-energy industry faces can be divided into research breakthroughs, technology challenges, regulatory barriers, and market issues.

Research Breakthroughs.

1. In the analysis of the horizontal MHK energy resource, the basic assumption by the MHK-energy community is that the required minimum velocity for a viable technology is $2\,m\,s^{-1}$ (≈ 4 kn). The Electrical Power Research Institute (EPRI) has concluded that for turbine technologies to be financially viable, an average flow speed of 5–7 kn (≈ 2.5–$3.5\,m\,s^{-1}$) is required [47.17]. There are only seven sites in the USA qualifying under this requirement. Furthermore, studies [47.1, 19, 126] practically limit ocean current resources to just one: the Florida Current. Since the vast majority of currents are slower than 3 kn ($\approx 1.5\,m\,s^{-1}$) and typical rivers are slower than 2 kn ($\approx 1\,m\,s^{-1}$) a breakthrough in technology is mandatory for the horizontal MHK energy to be developed. It raises the challenge of designing and building new technologies that can operate efficiently at lower speeds. The Gorlov [47.32] turbine and the ALTs discussed in Sect. 47.1.4 provide promising solutions.
2. Low-power density (power per footprint volume or power-to-weight ratio) is the Achilles heel of all renewable technologies – not just ocean-power technologies– compared to fossil fuels. It results in high cost in building, deploying, and retrieving MHK energy technologies. *Babarit* et al. [47.128] compared several wave-energy converters in operation and calculated the power density of the seven best to range from 0.065 to 0.2064 kW/Mg. Marine structures are typically large and heavy due to issues of balance between buoyancy and weight/ballast and this leads to another breakthrough requirement.

3. The challenge of low power-to-weight ratio leads to two separate challenges from the research point of view. The first one, is to increase the power-output hydrodynamically. Typical MHK-energy converters include point absorbers (buoys), line absorbers (one-dimensional (1-D) attenuators, for example, Pelamis), surface absorbers (two-dimensional (2-D), for example, oscillating water column), and area absorbers (two-dimensional, for example, turbines, propellers, watermills). Until three dimensions (3-D) are used for energy harnessing, the numerator of the power density will be proportional to two dimensions at best, while the denominator (volume or weight) will be proportional to three-dimensions. Thus, massive increase in scale will lead to lower power density for a given field of operation. A research breakthrough is mandatory in that direction. ALTs discussed in Sect. 47.1.4 provide a hydrodynamic solution to this challenge.
4. The second challenge that the low power-to-weight ratio leads to is the high weight or volume of MHK-energy converters. Classical wave-energy converters of all five concepts (point absorber, overtopping device, oscillating water column, attenuator, inverted pendulum) can be packed closer, but that would not lead to a real three-dimensional (3-D) MHK energy converter. There will be hydrodynamic interference effects that at best will result in a farm with less output than all the components harnessing power independently. That is, bringing devices closer together will be beneficial only if the support structure of the farm serves more than one device without proportional increase in mass or volume.

Technology Challenges. Several important technology development challenges have been stated by NREL [47.1]:

1. High capital cost
2. Unproven technologies that are not cost-competitive with conventional energy-generation technologies
3. Unproven functionality, performance, and reliability in the water at full scale
4. Resource quantity and variability are not well quantified
5. Undefined utility requirements.

Regulatory Barriers. NREL [47.1] has further identified the following regulatory barriers:

1. A stable, supporting policy that encourages development
2. Regulatory support to facilitate deployment and monitoring
3. Policies to minimize conflict and to align the benefits and priorities represented in environmental policy, tax policy, energy-supply policy, and energy security
4. Alignment of the regulatory process to minimize environmental impacts while facilitating responsible deployment of MHK technologies
5. Development of appropriate safety requirements and emergency procedures
6. Extensive permitting studies and lead times due to uncertain environmental impacts.

Market Issues. NREL [47.1] has further identified the following market issues:

1. New and unfamiliar technologies
2. Technologies that are not cost competitive
3. Lack of infrastructure, specialized equipment, and trained labor pool for installation, operations and maintenance.

Based on the above discussion, the following conclusions can be drawn:

- The MHK-energy industry is on a solid path to commercialization driven by the promise of the abundant MHK energy and supported by EERE (Office of Energy Efficiency and Renewable Energy) of the US-DOE since 2008.
- The MHK-energy industry has to overcome several challenges in technology development, market, and regulatory challenges before it can reach maturity.
- MHK energy needs research breakthroughs to improve its power-to-weight ratio before it can become competitive with other renewable energy technologies.
- A conceptual breakthrough is required for MHK energy technology to be able to tap into the vast majority of horizontal MHK energy, which is slower than the present technology threshold of 2 m s^{-1}.

47.1.3 Steady-Lift Technologies

The term *steady-lift converter* or *steady-lift technology* was introduced in this chapter to refer to turbines and watermills (Figs. 47.2–47.4). Steady lift is generated by lifting surfaces, such as airfoils and hydrofoils in a steady flow. Tidal turbines, ocean current turbines, watermills, propellers are SLTs. Tidal turbines are typically mounted on bottom fixed towers in shallower estuaries (Fig. 47.2a). Ocean current turbines may be mounted on a hydrofoil or a floating platform

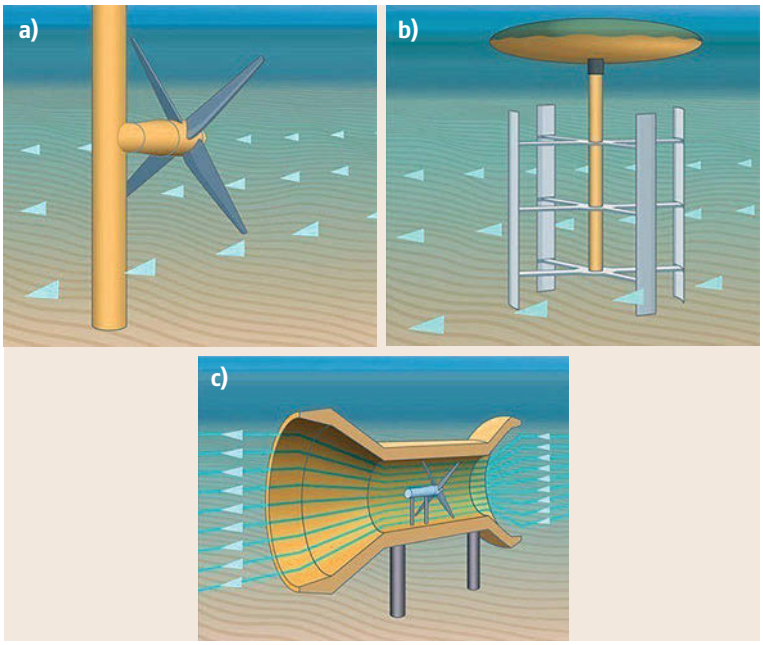


Fig. 47.2a–c Steady-lift technologies: turbines (after [47.85, 129]). **(a)** Horizontal axis turbine. **(b)** Vertical-axis turbine. **(c)** Ducted turbine

Fig. 47.3a–c Steady-lift technologies: turbines (continued). **(a)** Verdant propeller (photo by Verdant Power after [47.130]). **(b)** Sea Flow tidal stream generator (photo by Fundy, Wikicommons). **(c)** Open Hub Turbine (photo by Mike Roper)

(Fig. 47.2c) or moored in deep waters like the Florida current (Fig. 47.2c). Ocean current turbines are unrestricted in size and thus conceptually are similar to wind turbines.

Steady-lift turbines may also be classified as horizontal-axis turbines (Figs. 47.2a,c, 47.3, 47.4a) or vertical-axis turbines (Fig. 47.2b), open-hub turbines (Fig. 47.3c), and low-speed turbines (Fig. 47.4b). Ducts may be used to accelerate the flow past the turbine, thus increasing the rotor angular velocity and potentially the generator efficiency (Fig. 47.2b). Gorlov turbines may also be used horizontally (Fig. 47.4a) or vertically [47.32]. As their lifting surface is further away from the center of rotation, it results in a higher torque

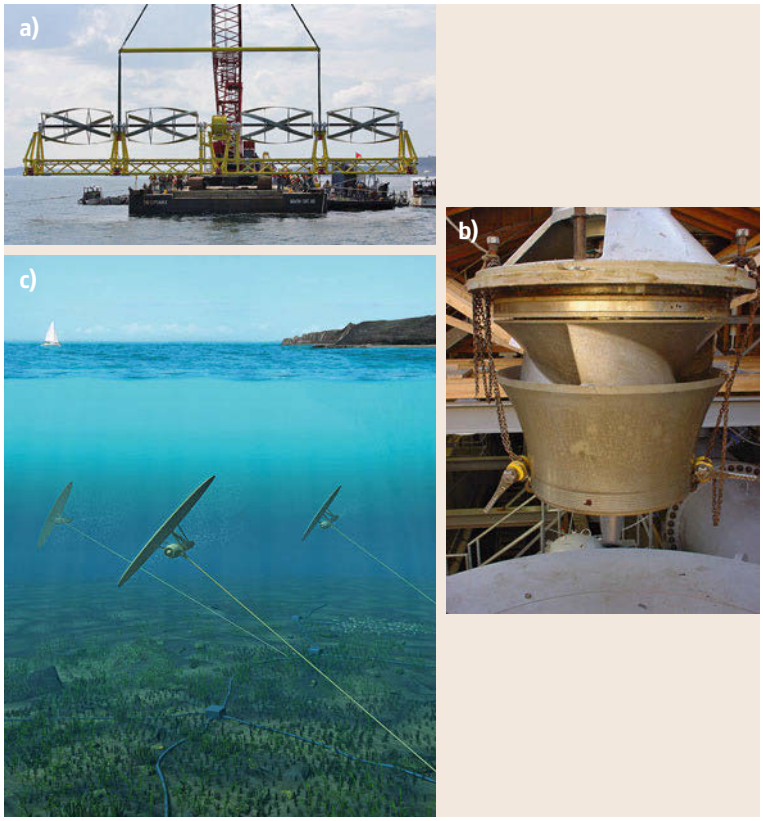


Fig. 47.4a–c Steady-lift technologies: turbines (continued). **(a)** ORPC Turbine (courtesy of Ocean Renewable Power Company after [47.131]). **(b)** Alden Turbine (courtesy of Alden Research Laboratory after [47.132]). **(c)** Underwater kite design (courtesy of Minesto)

turbine. Other turbines are designed to reduce their impact on fish, such as the Alden turbine by EPRI (Fig. 47.4b) and the open-hub turbine (Fig. 47.3c).

In the United States, several companies are developing MHK steady-lift turbines. The most notable ones are the following:

1. Verdant Power tested a 35 kW rated prototype, 5 m in diameter, 3-bladed tidal turbine in New York City's East River between 2002 and 2006. From 2006 to 2008, Verdant installed and tested six turbines in a tidal array at the same site (Fig. 47.3a) [47.122].
2. Hydro Green Energy [47.133], is developing a ducted current turbine that generates electricity from flowing water, such as river currents, tidal currents, and ocean currents.
3. In 2010, Alaska Power and Telephone installed a 25 kW, in-stream river turbine near Eagle, Alaska [47.134]. The low speed, vertical-axis turbine was mounted on a floating platform and was manufactured by New Energy Corporation in 2012.
4. Ocean Renewable Power Company is testing its new commercial TidGenTM Power System in Cobscook Bay near Eastport Maine, ORPC.

Many more technologies are undergoing testing around the world [47.135]. Numerous other technologies are at earlier stages of engineering development [47.5].

Regarding SLTs, the following conclusions/observations can be made:

- SLTs dominate the efforts to harness horizontal hydrokinetic energy.
- Turbines need a breakthrough to challenge the minimum flow-speed threshold of $2\,\mathrm{m\,s^{-1}}$.
- MHK companies, including turbine companies, face a variety of challenges and frequently go out of business, as harnessing ocean power is tougher and more expensive than expected [47.11].

47.1.4 Alternating-Lift Technologies

The term *alternating-lift converter* or *alternating-lift technology* was introduced above to refer to technologies that employ alternating lift occurring naturally in FSI phenomena. Fish utilize alternating lift to propel efficiently in water either as individuals or in schools at any speed with respect to the water flow. Lifting surfaces, such as fish fins, are used primarily for steering

rather than propulsion. Fins may contribute to propulsion as part of their entire body motion, which is alternating.

Bodies with either slender cross-sections like hydrofoils or bluff cross-sections like cylindrical pipelines are typically subjected to FSI phenomena. The scale of the fluid excitation is about the same as the size of the cross-section of the structure resulting in mutual interaction between the flow and the body motion.

Elongated structures that are flexible or are rigid on flexible support may have a natural frequency in the range of frequencies of the oscillatory lift, in which case severe FSI phenomena may be induced [47.43, 44]. FSI phenomena are destructive and, thus, avoided by design or suppressed using excessive damping or appendages. When enhanced, FSI may result in vigorous FIM of the body, thus, converting MHK energy to mechanical in an oscillator.

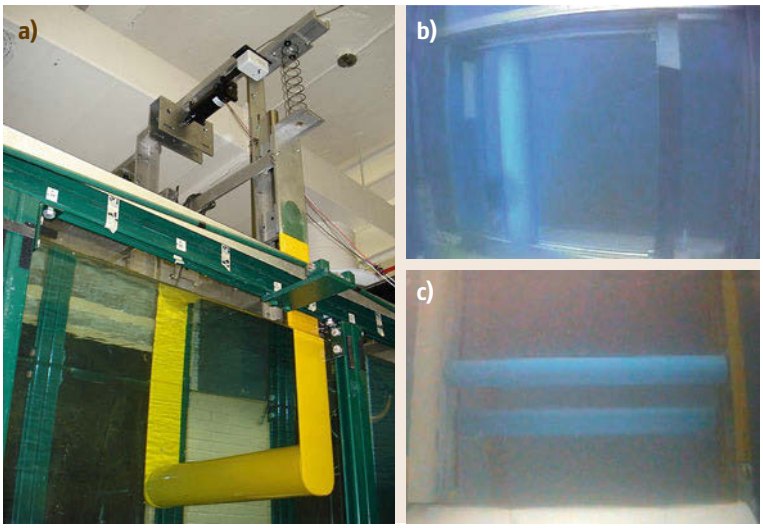


Fig. 47.5a–c Alternating-lift technologies: VIV, galloping, gap flow (see also Figs. 47.9–47.12). **(a)** MRELab model of early VIVACE based on VIV (courtesy of MRELab). **(b)** Model of VIVACE at OHMSETT (oil and hazardous materials simulated environmental test tank) towing tank based on VIV, galloping, and coexistence of VIV and galloping (courtesy of Vortex Hydro Energy). **(c)** VIVACE tests in the St. Clair river based on synergistic multibody FIMs: VIV, galloping, coexistence of VIV and galloping, and enhancement by gap flow (courtesy of Vortex Hydro Energy)

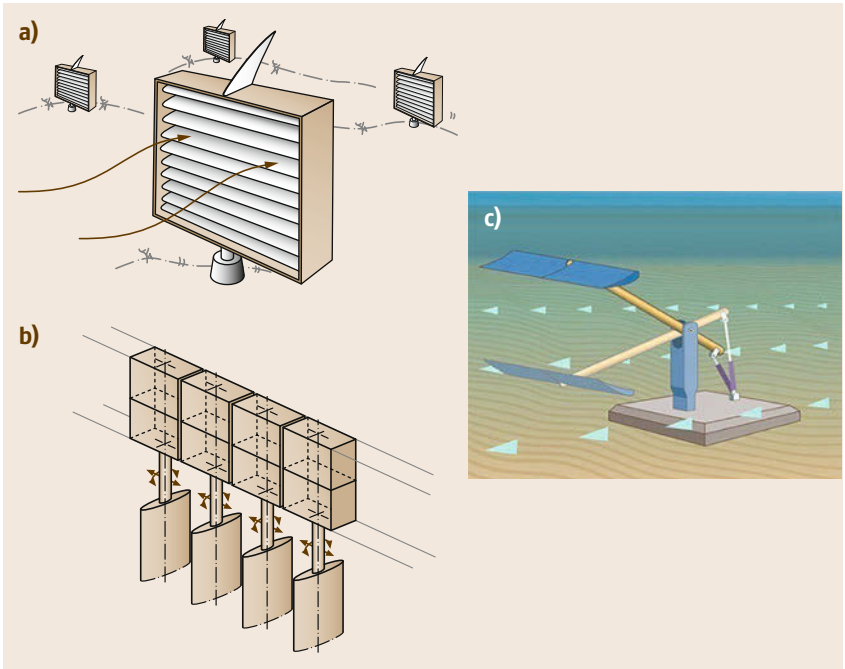


Fig. 47.6a–c Alternating-lift technologies: Flutter instability. **(a)** Flutter instability for energy harness (after [47.86]). **(b)** Flutter instability for energy harness (after [47.87]). **(c)** Hydrofoil flutter instability (after [47.85])

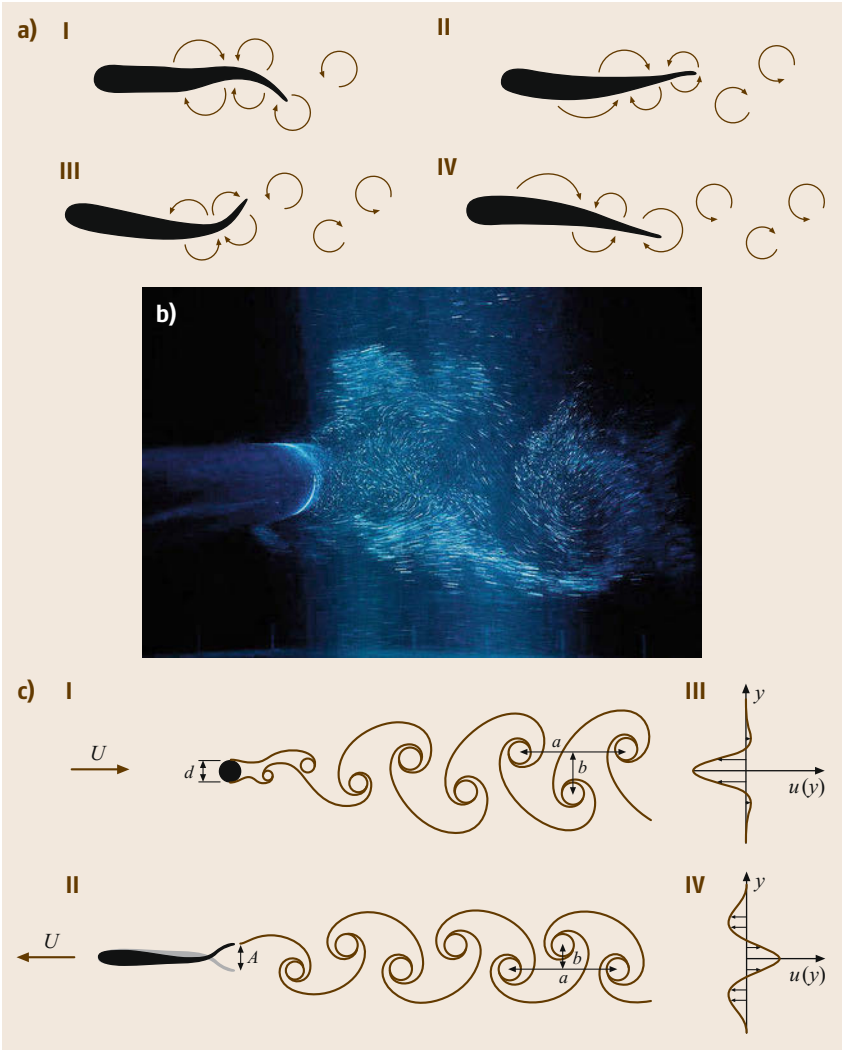


Fig. 47.7a–c Significance of alternating-lift. **(a)** Fish propulsion (after [47.136]). **(b)** Near wake of single VIVACE cylinder in VIV (courtesy of MRELab by photo by Don Boroughs). **(c)** Wake of cylinder in steady flow versus fish in forward motion (after [47.137])

Hydrofoils can harvest MHK energy through flutter – a well-studied and understood form of instability [47.43, 86, 87]. Two patents are shown in Fig. 47.5a,b and the same concept is shown for water in Fig. 47.5c.

On the other hand, bluff bodies, such as circular or rectangular cross-section cylinders, may exhibit several forms of FIM defined above. Alternating-lift MHK energy devices have been designed, individually (Fig. 47.6a,b) or in schools (Fig. 47.6c) and have been patented [47.86, 87]. VIVACE, which is a unique MHK energy device based on bluff cross-section cylinders utilizes all forms of FIM introduced above.

Significance of Alternating Lift

The details of flow mechanics of single and multiple cylinders in FIM will be discussed in Sect. 47.2. At this point, it is important to introduce the various forms of alternating lift involved in FIM of flexible cylinders with bluff cross-section. Their nature and the role each form plays in the marine environment as well as in FIMs of individual and schools of fish or elastic bodies are discussed. Figures 47.7–47.12 are used to illustrate each scenario of FIM:

1. Individual fish and vortex shedding: Fish actively curve their body, as shown in Fig. 47.7a, to collect a large vortex on the convex side. Then, actively push the vortex away forcing it to shed as they curve their bodies in the opposite direction to repeat the cycle shedding an opposite rotation vortex. Figure 47.7c II shows the alternating vortex street formed in the wake. The lift force exerted on the fish is proportional to the circulation, which is primarily

concentrated in the shed vortices. The lift force is alternating following the alternating pattern of the shed vortices. This is a thrust-wake pushing the fish forward.

2. Single cylinder in VIV and vortex-shedding frequency: In a steady flow, a street of alternating vortices forms in the wake of stationary cylinder. Figure 47.7b shows a laser-visualization picture of the near wake and Fig. 47.7c I shows schematically the alternating vortex street called von Kármán street. This is a drag wake providing zero-mean alternating lift and drag with a nonzero mean component and a zero-mean oscillatory component. Vortices shed passively downstream of the cylinder primarily following *Gerrard's* mechanism where shear layers and irrotational flow come together at a confluence point [47.138]. In the process, about 60% cancellation of vorticity occurs [47.46] resulting in 60% reduction in circulation and lift, based on the Kutta–Joukowski theorem. The wake has a dominant frequency over all Reynolds numbers with the exception of three transition regions. The vortex-shedding frequency for a stationary cylinder is called Strouhal frequency. Adding flexibility to the cylindrical body, so that it can respond to induced alternating forces, may result in nonlinear resonance with a very broad range of synchronization called VIV. This FIM phenomenon is used, as specified in Sect. 47.2, to convert MHK energy to mechanical in an oscillator.
3. Single fish in wake: As Fig. 47.7c shows, a streamlined body, such as a fish, generates a thrust wake and cylinders a drag wake. Figure 47.8a shows that fish in a school move in a diamond formation to take advantage of a street of vortices that can be used for thrust. Experiments have been conducted placing a fish in the wake of a bluff body (Fig. 47.8b) [47.36]. The fish positioned itself in such a way that the drag wake for the body becomes a thrust street for the fish. As a result, the black fish maintained its position in the flow stream with minimal effort [47.37].
4. Single cylinder in VIV and wake frequency: As velocity increases when the cylinder is in VIV, so does the number of shed vortices per cycle (Fig. 47.9) [47.76]. Thus, the vortex-shedding frequency becomes a higher multiple of the frequency of oscillation. Nevertheless, vortices shed in groups maintaining the predominant wake frequency to remain the same as the oscillation frequency. Thus, the nonlinear resonance persists and the oscillator can still be used as a powerful mechanism for converting MHK energy to mechanical.
5. Single cylinder in simultaneous VIV and galloping: Galloping is fundamentally different from VIV. The former is an instability in lift resulting from asymmetry due to geometry or impinging flow and results in much more vigorous lift. For a cylinder on elastic support or a flexible cylinder (e.g., pipeline, cable) nonlinear oscillations of amplitude even higher than VIV may occur. Typically, galloping occurs at higher velocity than VIV but depending on mass, damping ratio, absolute velocity, and cross-sectional geometry these two forms of FIM

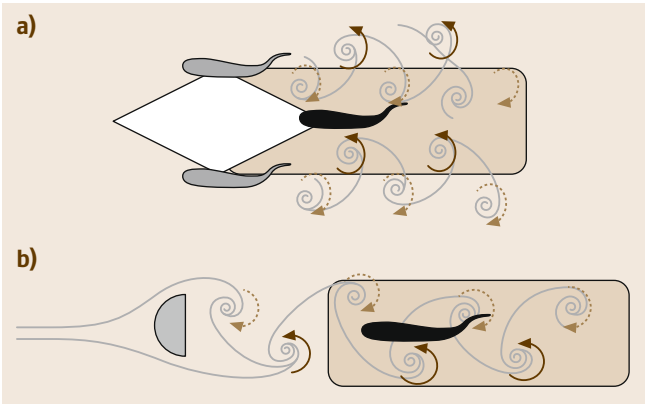


Fig. 47.8a,b Significance of alternating-lift. **(a)** Black fish utilizes reverse von Kármán street (thrust wake) generated by two grey fish. **(b)** Similar wake created by a reverse D-section cylinder. Fish exploit the energy of discrete vortices as well as the average reduced velocity in the Kármán street (after [47.36])

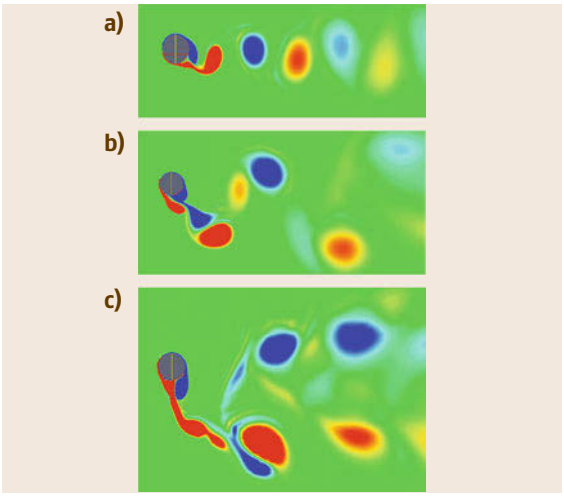


Fig. 47.9a–c Significance of alternating-lift: Single PTC cylinder in VIV (TrSL3 flow regime; $D = 3.5''$) (after [47.112, 139, 140]). **(a)** Initial-branch, 2S (S = single) vortex pattern, Re $= 35\,000$. The amplitude $A \approx 1 * D$. **(b)** Upper branch, 2P (P = pair) vortex pattern, Re $= 65\,000$. Amplitude $A \approx 1.5 * D$. **(c)** Later upper branch, PSSPSS vortex pattern, Re $= 95\,000$. Amplitude $A \approx 2 * D$

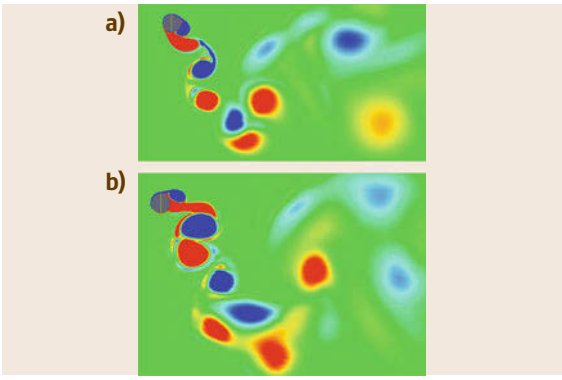


Fig. 47.10a,b Significance of alternating-lift: Single PTC cylinder in steady flow in VIV to galloping transition, and galloping (TrSL3 flow regime; $D = 3.5''$; [47.113]). **(a)** SPSSPS transition upper branch to galloping, Re = 110 000. Both the VIV and the galloping mechanisms coexist and are in phase. Amplitude $A \approx 3 * D$. **(b)** 2P + 8S galloping, Re = 130 000. The galloping mechanisms drives; vortex shedding is out of phase. Amplitude $A > 3 * D$

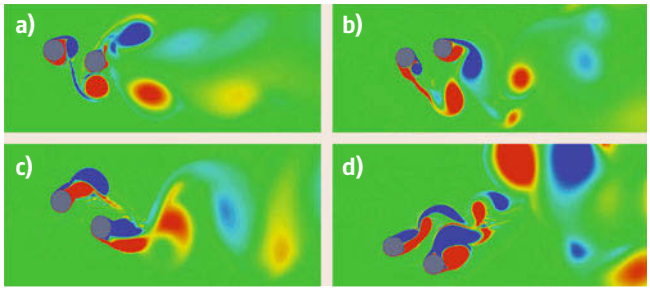


Fig. 47.11a–d Significance of alternating-lift: Two PTC cylinders in tandem in steady flow in a channel. (TrSL3 flow regime; $D = 3.5''$, 2-D center-to-center; (after [47.113]). **(a)** VIV, $R = 30\,000$. Timing of upstream vortices hitting the second cylinder is such that they grab same rotation vortices from the second cylinder leaving the second cylinder without circulation and lift. VIV of second cylinder is suppressed. **(b)** VIV, Re = 59 229. Second cylinder follows first and gap flow keeps wakes separated. Thus, each cylinder has its own wake providing alternating lift resulting in VIV. **(c)** Re = 60 000. The two cylinders are out of phase in VIV. With $m* = 1.68$ second cylinder has adequate momentum to cross the wake of the first. Partial vorticity coalescence occurs but the second cylinder maintains its wake, lift, and $A \approx 1.5 * D$. **(d)** Re = 90 000. The two cylinders are nearly in phase in VIV to galloping transition. Partial vorticity coalescence occurs increasing lift in the second cylinder as the VIV mechanism is in phase with the galloping instability. Amplitude $A > 2.5 * D$

may overlap over a transition region. Figure 47.10a shows that the vortex-shedding process is still in synchronization (Sect. 47.2.3) with the oscillations, while the galloping mechanism due to synchronization of the shear layers with the oscillations has taken effect. The asymmetry here is caused by the PTC attached to the cylinder [47.96].

6. Single cylinder in galloping: As velocity increases, the galloping mechanism becomes more powerful inducing higher amplitudes of oscillation (Fig. 47.10b). The number of vortices shed per cycle increases, but vortices fall out of synchronization with the oscillations. Shedding vortices push the cylinder away. As can be seen in Fig. 47.10b, some vortices push in the wrong direction. That is, vortex shedding continues but does not assist in increasing lift. The galloping mechanism by now is more powerful to drive the oscillations and the MHK energy conversion to higher levels [47.95, 141].
7. Multiple fish in a school: Figure 47.8a shows how fish position themselves in a school to better use the energy in the vortices. This formation fills the entire cross-section effectively making the school a three-dimensional device to convert MHK energy to mechanical energy. This is a real three-dimensional mechanism in the sense that an individual fish moves more effectively in a school than in isolation. This is not the same as packing close together converters of any kind. On the contrary, interference effects between devices may result in reduced efficiency.
8. Multiple cylinders in interference: When multiple bodies, stationary or elastic susceptible to FIM, are in close formation, interference effects may dominate their flow dynamics if they were isolated. That is, vortex shedding, wakes, and instabilities may be altered partially or completely. In that case, the basic forms of FIM discussed up to this point may not take place or may be enhanced, unless the cylinders are separated by as many as 9–10 diameters in the flow direction or 3–4 diameters perpendicular to the flow [47.57]. Figure 47.11a shows two cylinders in the early stage of the initial branch corresponding approximately to Fig. 47.9b. The configuration and dynamic particulars of the two oscillators are such that the second cylinder starts in VIV just like the first cylinder but in a few cycles its vortices coalesce with those from the upstream wake and its motion is suppressed. Figure 47.11b–d shows different kinds of interference. The timing the vortices from the upstream wake hit the second cylinder is such that the second cylinder maintains its vortex street and reaches amplitudes allowing flow between the two cylinders that further enhances FIM [47.114, 115].
9. Gap flow between cylinders: For two stationary interfering cylinders, gap flow entirely disrupts the vortex shedding and distorts the near wake of the

upstream cylinder. The region of the occurrence of the gap flow is $1 < L/D < 3.5$ and $0.2 < T/D < 0.4$, where L and T are the center-to-center in-flow and transverse spacing between the two cylinders, respectively [47.84, 113]. When two cylinders are in FIM the gap flow may enhance the FIM and result in more MHK energy converted to mechanical energy in the oscillator (Fig. 47.12).

10. Wake displacement between cylinders: For two stationary interfering cylinders, wake displacement does not disrupt the vortex shedding but only displaces the upstream wake. Wake displacement occurs for $L/D > 3.5$ and $T/D > 0$. When the two cylinders are in FIM, the wake displacement allows for flow between the two cylinders and may enhance the FIM and result in more MHK energy converted to mechanical energy in the oscillator. This scenario is shown in Fig. 47.12a between the 1st and 2nd cylinders.
11. Galloping with gap flow: At higher Reynolds number and reduced velocity – in the fully developed galloping region – flow between cylinders may enhance galloping as well, thus converting more hydrokinetic energy to mechanical energy in the oscillators. Figure 47.13 shows two cylinders in the fully developed galloping region with the 2nd cylinder consistently following the 1st. The amplitude of oscillation though changes periodically, so that sometimes the amplitude of the first cylinder (Fig. 47.13a) or the amplitude of the second cylinder (Fig. 47.13b) would be larger.

Regarding ALTs, the following conclusions/observations can be made:

1. There are two cross-sectional geometries of two-dimensional (cylindrical) bodies that when they are elastic – or rigid on elastic support – may result in alternating lift and alternating FIM:
   a) Slender cross-sections hydrofoils/airfoils that may flutter
   b) Bluff cross-sections that may be subjected to a variety of FIM due to alternating vortices, galloping instability, multibody interference, or combinations of those.
2. Flutter instability and galloping occur passively in engineered structures and are not used by marine life for propulsion.
3. Vortex streets and multibody interactions are used by fish to propel efficiently in water. Both mechanisms are highly scalable occurring practically over the entire range of Reynolds numbers.
4. By employing vortex streets, triggering galloping instability, and even superposition of the two mechanisms, nonlinear oscillators can be built using the naturally occurring alternating lift in all steady flows from very slow to very fast.
5. Interference between two-dimensional bodies with bluff cross-section in close proximity can induce much more vigorous FIM response of the same bodies when in isolation. This enables increased hy-

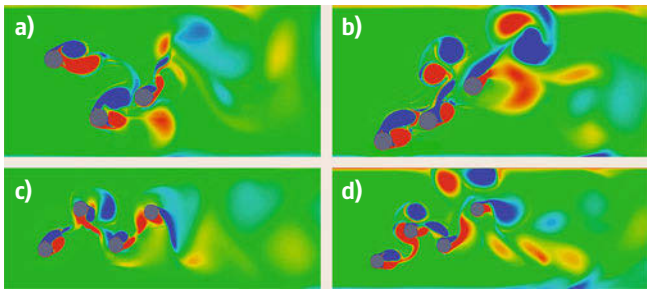


Fig. 47.12a–d Significance of alternating-lift: Three and four PTC cylinders in tandem in steady flow in a channel. (TrSL3 flow regime; $D = 3.5''$, $2.5 * D$ center-to-center spacing (after [47.114]). **(a)** Re = 62 049. 1st and 2nd cylinders are out of phase in VIV oscillating as in Fig. 47.8b. 3rd cylinder follows closely the 2nd cylinder as in Fig. 47.8a. $A \approx 1.5 * D$. **(b)** Re = 90 254. 3rd cylinder follows 2nd which follows 1st. Spacing is not optimal, thus, the 2nd cylinder does not reach as high amplitude as the 1st and 3rd. VIV to galloping transition. Partial vorticity coalescence occurs increasing lift in the 2nd and 3rd cylinders as the VIV mechanism is in phase with the galloping instability. Amplitude $A > 2.5 * D$. **(c)** Re = 62 049. Phasing between the four cylinders changes. VIV with partial vorticity coalescence occurring. Cylinders maintain their wake, lift, and $A \approx 2 * D$. **(d)** Re = 90 254. Phasing between the four cylinders changes. VIV with partial vorticity coalescence occurring. Cylinders maintain their wake, lift, and $A \approx 2 * D - 2.5 * D$ in VIV to galloping transition

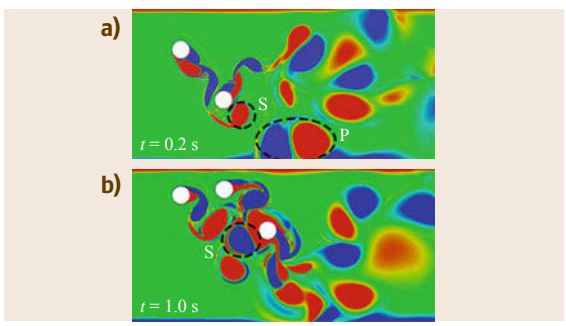


Fig. 47.13a,b Significance of alternating lift: Two PTC cylinders in tandem in steady flow in a channel. (TrSL3 flow regime; $D = 3.5''$, $2.5 * D$ center-to-center spacing (after [47.114]). **(a)** Re = 100 000. In fully developed galloping 2nd cylinder follows 1st. 1st cylinder amplitude is higher. **(b)** Re = 100 000. In fully developed galloping second cylinder follows first. 2nd cylinder amplitude is higher

drokinetic energy conversion to mechanical energy in an oscillator, while at the same time reducing the space occupied by the MHK energy conversion device.

6. The VIVACE converter is built based on conclusions 3–5 so that it has the following properties:
   a) It is based on alternating lift naturally occurring in the marine environment.
   b) The underlying physical mechanisms are highly scalable enabling the conversion of MHK energy to mechanical even at very slow flows.
   c) Multiple-body interference enables higher conversion of MHK energy to mechanical in an oscillator.
   d) Proximity of bodies reduces occupied space by VIVACE resulting in higher power density of conversion
   e) By properly configuring multiple cylindrical oscillators in close formation, a three-dimensional converter can be designed to achieve an impact on the Achilles heel of all renewable energy devices, which is the power-to-volume ratio.

47.2 Alternating-Lift Technologies: The VIVACE Converter as a Case Study

Among ALTs, the VIVACE converter is probably the closest to commercialization because it has undergone extensive laboratory testing (Figs. 47.5a,b, 47.7b) and several field deployments (Fig. 47.5c) since its introduction in 2006 [47.48, 49]. Further, several of its components and underlying principles are generic for ALTs. Accordingly, in this section, the VIVACE converter is used as a case study. It is described in Sect. 47.2.1. Its scales and applications are presented in Sect. 47.2.2, and the underlying principles are explained in Sect. 47.2.3.

47.2.1 Concept, Scales, and Principles

Physically, the VIVACE converter is a simple machine. Simplicity is a requirement (Sect. 47.1.2, *Requirements*) for reliable operation in the marine environment. The converter has few parts and they can be built to withstand the environmental loads and have low-maintenance requirements. Operating a device with moving components in the marine environment is a challenge. VIVACE prototypes and models have been built and operated in a recirculating channel, a fresh-water towing tank, a seawater-towing tank, a river, and canals. Even though each model is different, the basic concept remains unchanged and components are scaled to fit the facility and objectives of each test.

Basic Concept

Since its inception in 2005, the underlying concept for the VIVACE converter has been to enhance FIM so that more hydrokinetic energy can be converted to mechanical energy in an oscillator and subsequently to electrical energy [47.96, 100] Elastic slender bodies may exhibit FIM in response to quasi-steady flows from currents, tides, or rivers or to oscillatory flows due to waves. Given the intermittent nature of waves, quasi-steady flows present a better source of energy for eventual generation of base power and were thus selected as the source of MHK energy that VIVACE would target. Another early design decision made was to enhance, among FIMs, only VIVs, galloping, and the simultaneous presence of VIV and galloping. In multiple cylinder converters, gap flow between two cylinders in proximity and wake interference is also utilized not only to enhance FIM but, more importantly, to achieve synergistic operation resulting in energy conversion higher than that of the same number of cylinders in isolation [47.68, 69, 77, 84].

Some of these FIMs are based on synchronization and have large but bounded response and others are based on instability and may have unbounded response. Energy is harnessed by using damping, while at the same time limiting the oscillator response. A controller is required to bound the response of cylinders and harness energy through a power-take-off (PTO) system when the cylinder FIMs are enhanced.

Consequently, the following early design decisions were made:

- Buffeting, albeit being one of the most powerful forms of FIM, due to its random nature, should not be utilized in designing a commercial MHK energy converter.
- FIM due to wave excitation is a known phenomenon and can be enhanced for MHK energy conversion. The intermittent nature of waves as well as their randomness, however, makes waves a nonprimary target as a source of MHK energy for ALTs.
- VIVs, coexistence of VIV and galloping, and fully developed galloping are enhanced in single-cylinder ALTs to harness MHK.
- In multiple-cylinder ALTs, gap flow and wake-interaction phenomena are used to further enhance FIMs.

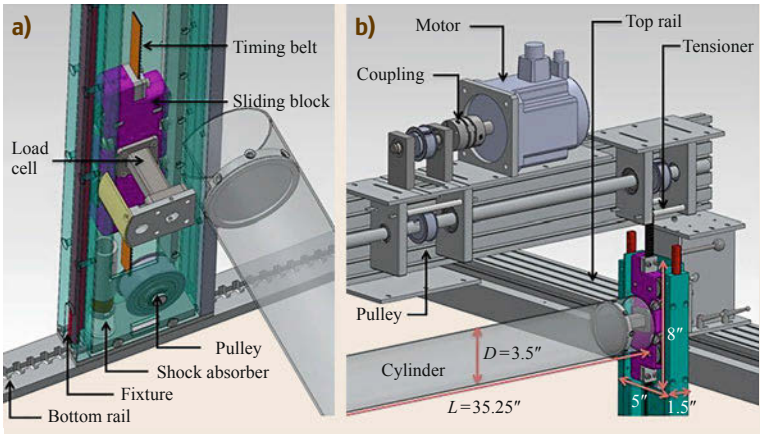


Fig. 47.14a,b Details of the VIVACE converter (after [47.142])

Description of the VIVACE Converter

The simplest form of the VIVACE converter consists of a single smooth cylinder (Fig. 47.5a) or a single cylinder with localized surface roughness (Fig. 47.5b) suspended by springs with a controller and a PTO system.

System. To describe the converter, the latest model in the MRELab is used as a generic representative of VIVACE converters. It should be pointed out that this is a model for the VIVACE converter itself not for ideal FIM tests. A separate system is used for VIV and galloping testing as explained in Sect. 47.3.8.

In early versions of the oscillator in the MRELab, the system operated on physical springs and dampers. Since 2010, *Lee* et al. [47.98, 99] introduced and implemented a virtual damping-spring device called V_{ck} system. V_{ck} replaced the physical springs and dampers in the oscillator using control feedback (Sect. 47.2.3, *Electro-Mechanical Principles*).

Between 2005 and 2012, the oscillators in galloping would reach the safety stops of the recirculating channel in the MRELab. In order to study the FIM of cylinders with fewer restrictions in amplitude, both the channel and the VIVACE converter were rebuilt. This third generation of test models was designed and built as shown in Figs. 47.14 and 47.15. In the new design, the linear motion mechanism was brought under water, which resulted in most of the structural elements being under water. The whole linear motion system, which consists of a sliding block, timing belt, and pulleys, is fitted inside of a 1.5 inch-wide slender tube. The sliding block connects the cylinder to the V_{ck} system through the timing belts and pulleys as shown in Fig. 47.14.

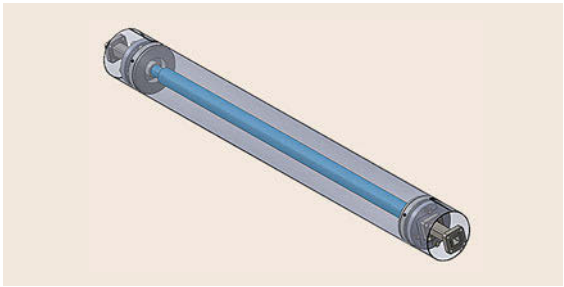


Fig. 47.15 Interchangeable cylinder with adjustable mass

The mass ratio can be changed by placing a rod of variable mass inside the cylinder, as shown in Fig. 47.15. Shock absorbers are used at both ends of the cylinder travel to protect the cylinder from hitting the plexi-glass bottom of the water channel. The center-to-center distance (d/D) between the cylinders in the flow direction can be adjusted by the rail system between 0.429″ (0.0109 m) and 5″ (0.127 m) in increments 0.5″ (0.0127 m). The features of the new VIVACE converter are summarized in Table 47.2.

The most important physical components of this alternating-lift converter are the following.

Cylinders. In tests, the cylinder diameter D has been varied between 2.0 and 6.0″ (5.08–15.24 cm) in the MRELab recirculating channel (Sect. 47.3.1), and between 6.0 and 10.0″ (15.08–25.4 cm) in towing tanks (Sect. 47.3.1) and field tests (Sect. 47.3.4). In future field-tests cylinders with $D = 36''$ (91.4 cm) have been designed by Vortex Hydro Energy (VHE) [47.111]. Up to four cylinders in tandem and staggered arrangements have been used in testing in the MRELab channel. Cylinders can be easily interchanged (Fig. 47.14) and their mass ratio $m*$ varies with insertion of mass rods inside the cylinder (Fig. 47.15). The basic properties of the cylinders used in this converter model are shown in Table 47.2.

Springs. Physical springs are required in the converter for two reasons. First, they provide a restoring force mandatory for the existence of an oscillator with a nat-

Table 47.2 Specifications of the MRELab VIVACE converter model

m_{osc} (kg) $0.634 \le m^* \le 2.00$	Equivalent mass of oscillating components including 1/3 of physical spring mass, equivalent mass of rotating pulleys, belts, and motor
m_d (kg)	Displaced fluid mass
$m^* = m_{osc}/m_d$	Mass ratio
$L = 0.89535\,\text{m} = 35.25''$	Cylinder length
$D = 0.0889\,\text{m} = 3.5''$	Cylinder diameter
$L/D = 10.7$	Aspect ratio
y_{max} (m)	Maximum allowable displacement of the upper and lower safety stops
$y_{max}/D_{3.5''} = 5.5$	Maximum allowable displacement in cylinder diameters
d	Center-to-center spacing of cylinders in flow direction
$1.429 \le d/D_{3.5''} \le 6.0$	Range of in-flow spacing of in cylinder diameters

ural frequency. Second, springs reverse the motion of the cylinder since the area spanned by the oscillating cylinder has to be finite as allowed by the size of the converter and/or the geometric restrictions imposed by the lab facility or the means of transportation in field deployments.

These two reasons would obviously collapse to just one in a regular oscillator but not in the VIVACE converter. Specifically, VIVACE is a nonlinear mechanical oscillator operating over a very wide range of velocities and, thus, spanning several FIMs as explained in Sect. 47.2.3 (Principle 9). In VIV, a spring of constant k is needed to establish a natural frequency and achieve synchronization between the frequency of oscillation and vortex shedding (Sect. 47.2.3, Principle 3). Some researchers interpret VIV as a broad range, lock-in phenomenon [47.43, 55] and others as a nonlinear resonance over the same broad range of flow velocities [47.60] with variable natural frequency in water due to variable added mass. The latter can be measured experimentally [47.53, 53, 54, 62, 63]. Both approaches are based on a simple expression of the added mass force being proportional to the relative acceleration of the fluid with respect to the body. An in-depth analysis reveals that the added mass depends on relative acceleration components as well as convective terms with relative velocities [47.123].

Galloping, however, is not a resonance phenomenon (Sect. 47.2.3, *Hydrodynamic Principles*, Principle 4). It is an instability phenomenon not driven by the oscillatory vortex structures which still exist in the wake. Springs are needed only to reverse the cylinder motion, not to establish a natural frequency and possible resonance. If the converter were not limited by size, the cylinder – once excited in galloping – would continue moving in one direction harnessing MHK energy from the stream and converting it to kinetic energy rather than alternating between kinetic and potential energy in the springs. Actually, the highest power conversion efficiency that has been achieved by the VIVACE converter was when no springs were used during most of the cylinder travel. The latter is defined by the frame holding the oscillator and is equal to twice the amplitude of oscillation. Very short springs were installed only at the two ends of the cylinder travel to reverse the motion of the cylinder in a high efficiency process. Springs have to be designed carefully to be very short to maintain the high efficiency in reversing the motion and preventing the cylinder from slamming against the converter frame, thus, wasting valuable energy.

When the converter spans over multiple FIMs, springs must be designed to accommodate the VIV oscillator and reversal of the direction of motion for galloping. These springs are highly nonlinear and their functionality is specified in synergy with the controller's function during regenerative breaking (Sect. 47.2.3, *Hydrodynamic Principle*, Principle 12).

An additional note on the springs is that in spite of the slow nature of oscillations – about a cycle a second – there are 32 million seconds in a year making the fatigue life of the springs a challenge. The design and analysis of the strongly nonlinear springs – with the dual role of serving the converter both in VIV and galloping – is difficult and is part of the optimal controller algorithm.

Finally, it should be noted that the springs are designed differently depending on the size of the converter and whether it is a model in the recirculating channel or a field prototype. In the former case, the springs are completely replaced by the virtual damping–spring system V_{ck} (Sect. 47.2.3, *Hydrodynamic Principle*, Principle 12). In field-prototypes, nonlinear springs are present and coexist with the controller for optimal regenerative breaking, motion reversal, and minimum maintenance.

Passive turbulence control (PTC) is designed and appended on the cylinder surface in the form of selective roughness. As explained in (Sect. 47.2.3, *Hydrodynamic Principle* Principle 7), PTC is applied in the form of two sandpaper strips parallel to the cylinder axis, symmetrically with respect to the flow. PTC acts as turbulence stimulation. PTC achieves the following:

1. Alters the flow around the cylinder including the boundary layer, separation point, shear layers, for-

mation length, vortex roll-up, fluid–structure interaction, stability, etc.
2. Initiates galloping at low speeds so that VIV and galloping can be reached back to back, thus, creating an open-ended high-response oscillator as explained in Sect. 47.2.3 (Principle 9).
3. Prevents suppression of FIM in the flow-transition region from the laminar to turbulent boundary layer. In this critical flow regime VIV is suppressed. With PTC properly applied, FIM persisted without any gaps in Reynolds number.

The MRELab managed to establish the *PTC-to-FIM Map* shown in Sect. 47.2.3 (Principle 7) establishing the effect that PTC can have on FIM including enhancement and suppression.

The effectiveness of the PTC depends on four parameters including the start and end points, the roughness of the grit, and the total thickness of the sandpaper in relation to the thickness of the boundary layer at the location of the PTC. Table 47.3 shows sample dimensions of PTC used routinely in the LTFSW channel of the MRELab. The boundary-layer thickness is discussed in Sect. 47.2.3, *Hydrodynamic Principle*, Principle 7.

Synergistic FIM of Multiple Cylinders. This is a critical system parameter: How many cylinders should be included in a converter; what should be the relative properties of the oscillators in terms of stiffness k, oscillating mass m_{osc}, and damping c; and in what configuration should they be placed in the school. As explained in Sect. 47.2.3 (Principle 8), gap flow and wake interference can be utilized not only to enhance FIM but, most important, to achieve synergistic operation resulting in energy conversion higher than that of the same number of cylinders in isolation [47.84]. The experimental setup in the MRELab allows for:

1. Variable center-to-center spacing between cylinders in the flow direction: $1.429 \leq d/D_{3.5''} \leq 6.0$
2. Variable center-to-center offset between cylinders transversely to the flow direction: $0.5 \leq t/D_{3.5''} \leq 1.0$
3. Reynolds number up to 130 000 for $D = 0.0889\,\mathrm{m}$ $(3.5'')$
4. Allowed amplitude of 5.5 diameters needed in galloping
5. Variable stiffness of springs $300\,\mathrm{N\,m^{-1}} < k < 5000\,\mathrm{N\,m^{-1}}$
6. Variable mass ratio
7. Variable location and properties of PTC.

All of the above parameters are important in achieving the most vigorous FIM to maximize the MHK energy converted to energy in the mechanical oscillator.

Conversion of Linear to Rotational Motion. An early design decision was made to go with off-the-shelf components in building VIVACE lab models and subsequent field prototypes. Linear generators were considered but rejected due to high cost and low efficiency even though they would be more convenient in converting the linear oscillatory motion to electricity directly. Instead, the decision was made to use high efficiency, commercially available, rotary generators at relatively low revolutions per minute (RPM), and deal with the conversion of the linear oscillatory motion to rotational oscillatory motion. The solution to this conversion is based on a pulley–belt system that has proven to be reliable both in the lab and in the field. It is described in Sect. 47.2.3, *Electro-Mechanical Principles* as part of the V_{ck} system.

Controller and Power Take Off (PTO) System. These two components are present both in models in the MRELab LTFSW channel and prototypes in towing tanks and in field tests. Their functionality and design, though, differ between models and prototypes. In field tests, the controller is used to optimize the converted power, while in the laboratory the controller is embedded in the V_{ck} system (Sect. 47.2.3, *Hydrodynamic Principles*, Principle 12) and aims to facilitate and expedite changing of testing parameters. The most important word of caution is not to include the hydrodynamics in the closed control loop, as that would introduce a phase lag, which would bias the FIM response and the converted power.

Table 47.3 Properties of roughness strips

Property	Commercial grade				
	P60	P80	P120	P150	P180
Average grit height (k/D)	3.03×10^{-3}	2.19×10^{-3}	1.42×10^{-3}	1.09×10^{-3}	0.92×10^{-3}
Thickness of the backing paper (p/D)	6.50×10^{-3}	6.24×10^{-3}	4.98×10^{-3}	5.54×10^{-3}	5.38×10^{-3}
Total height $T/D = (p+k)/D$	9.53×10^{-3}	8.43×10^{-3}	6.40×10^{-3}	6.63×10^{-3}	5.30×10^{-3}

Notes: Cylinder diameter $D = 3.5''$ (0.0889 m); T/D should be compared to the boundary layer thickness

V_{ck} System. For systematic testing in the laboratory, two issues are important:

1. Changing the physical springs and dampers requires extensive preparation for calibration and alignment.
2. The viscous damping of the system is rarely linear as modeled in classical mathematical models in vibration textbooks.

To compensate for these two major challenges, *Lee* et al. [47.98, 99] introduced and implemented the virtual damping-spring device called V_{ck} system [47.98, 142]. V_{ck} is a feedback loop that uses the servomotor encoder to keep track of the position and velocity of the cylinder, it provides the necessary torque to simulate the spring stiffness and provide the linear viscous damping required. After extensive model identification and calibration, V_{ck} is used to subtract the system damping and then impose mathematically correct linear viscous damping and spring stiffness at the press of a button. This first V_{ck} system supported the FIM experiments conducted in the MRELab between 2009 and 2011 [47.98]. A second generation of the V_{ck} system was built recently by *Sun* et al. [47.142] to support the VIVACE converter experiments being conducted presently in the MRELab in collaboration with VHE. The basic principles and mathematical models behind V_{ck} are provided in Sect. 47.2.3, *Hydrodynamic Principles*, Principle 12.

In summary, the following conclusions should be observed for the successful design of a commercial converter:

- Springs are a critical design component and should be designed carefully to meet multiple objectives:
  1. Support VIV
  2. Support only reversal of motion in galloping
  3. Work in collaboration with the controller
  4. Achieve a satisfactory fatigue life
  5. Obviously be nonlinear.
- Passive turbulence control is an essential element of the cylinders. Without PTC, only the lead cylinder would function efficiently in a school. PTC enables FIM of following cylinders required to achieve high power-density converters. In addition, only with PTC, cylinders can maintain FIM in the critical flow regime.
- Multiple cylinders properly placed in a school can achieve synergistically higher power output than the same number of isolated cylinders in FIM.

Operation

Models or prototypes of the converter are deployed in a stream with the axis of each cylinder perpendicular to the flow. Cylinders may be positioned horizontally or vertically depending on several factors that are discussed below. The process is simple in the laboratory but may take several hours in the field depending on the weather conditions. The major aspects of operation are described below.

How the VIVACE Converter Works. Following launching, the converter may be self-started in FIM or may require a push from the controller depending on various parameters, such as the uniformity of the flow, the FIM phenomenon anticipated at the flow speed, and the imposed damping. Depending on the flow speed, cylinder diameter, and system damping, the converter may be in the range of VIV, transition from VIV to galloping, or fully developed galloping. The linear oscillatory motion of the cylinder due to FIM and springs is converted to a rotational oscillatory motion of the axis of a rotary generator through belts and pulleys. The controller imposes damping through the generator in addition to the system damping in order to harness the MHK energy converted to mechanical energy in the oscillator. In laboratory tests and field tests so far, the energy generated is burnt in a heat bank. In the case of a commercial device, the power signal at the output of the generator needs to be conditioned before being dispatched to a power grid.

Operational Concerns. Operation in the marine environment may be hazardous and unforgiving to errors. Accordingly, the following words of caution are offered:

1. FIMs are powerful and destructive phenomena and in most cases self-initiated. Hard galloping is an exception requiring a threshold displacement. Thus, caution should be exercised when safety hooks are released to allow cylinders to move.
2. Flow uniformity is a strongly recommended flow characteristic. There are several situations where the flow may deviate significantly from a uniform flow and result in suboptimal conditions of operation of the converter. For example, if the cylinder is placed vertically and close to the riverbed, it will be partially immersed inside the boundary layer, which has a width of about 1 m in a 2–2.5 kn current. A turbulent boundary layer has a full profile following the $z^{\frac{1}{7}}$ rule but still creates a very strong shear flow along the cylinder length, thus, reducing the alternating lift that drives the cylinder. A second example would be that of a flow out of an orifice at the interface between flow cross-sections that differ dramatically, such as a small opening out of a big dam or wall holding back fluid. That flow would be highly nonuniform and may result in as little as

one-third of the converter cylinder to be subjected to alternating lift that would induce some form of FIM.
3. If converter cylinders are placed horizontally instead of vertically, proper positioning of the converter using a crane from a barge may take several attempts. It is very important though to position the horizontal cylinders perpendicularly to the flow.

Vertical Versus Horizontal Cylinders. Positioning of cylinders horizontally or vertically is an important issue affected by the following parameters:

1. A small misalignment of the converter with respect to the flow direction will not affect the expected FIM response of a vertical cylinder. The small error will result in asymmetric positioning of the PTC strips. A larger error can be corrected by a rotation of the cylinder with respect to its own axis. In the case of a horizontal cylinder, a misalignment of the converter will result in a less than a 90° angle between the current and the cylinder axis exposing an elliptical rather than a circular cross-section to the flow. That may result in severe implications on the cylinder FIM depending on the size of the error.
2. Buoyancy is a very large force particularly in larger field applications. In the case of a horizontal cylinder, the buoyancy must be accurately set equal to the cylinder weight to achieve a neutrally buoyant cylinder. Even a small deviation would result in significant loss of energy to compensate for the difference between weight and buoyancy.
3. Nonuniformity of the flow is expected due to the boundary layer at the bottom of a river. A horizontal cylinder will not be affected since the flow velocity may change with distance from the riverbed, but it will remain uniform along the length of the cylinder.
4. Distance from the bed will take the cylinder away from the boundary layer but will increase the overturning moment and thus the required structural strength.
5. Vertical cylinders are more likely to encounter limitation due to water depth and navigation restrictions.

Considering all five factors, a general decision on the vertical versus horizontal orientation of the cylinders in the VIVACE converter is hard to make. The design decision has to be made on a case-by-case basis.

Environmental Compatibility. ALTs, such as the VIVACE converter, are subjected to alternating lift present in various forms of FIM and cylinder interference as specified in C21 and C22. Several studies have been conducted in the past establishing that fish thrive in the wake of a cylinder in a flow. Fish relax in the oscillatory wake and, with minimal effort, stay behind the cylinder and are in general more active and spawn more. Most notable is the study by Harvard, MIT, and Oak Ridge National Laboratory [47.34–37]. More environmental studies need to be conducted for permitting in the United States, including noise levels and electromagnetic interference.

Is VIVACE exotic or natural in water? We live in air and observe around us mostly lifting surfaces – bird wings, airplane wings, sails of sailboats, windmills, etc. Lifting surfaces in air, under constant airflow, generate unidirectional lift. In water, the picture is totally different. The muscle power of a fish is smaller than its drag in forward motion. So fish move by generating alternating vortices by periodically curving their bodies (Fig. 47.7a,c) and pushing them away when those vortices have accumulated adequate level of circulation. This generates alternating lift. In motion of bodies in dense fluids like water, all alternating lift comes from bluff bodies with tail; from tiny sperm to huge whales. Vorticity, circulation, and lift generated by circular cylinders may appear exotic to humans but it is an integral part of fish kinematics in schools or as individuals. As a result, the kinematics of ALTs is compatible with the kinematics of schools of fish. In addition, fish have scales with thickness of the order of the boundary layer, which is under investigation in marine biology. Surface roughness varies with species and their speed. In cylinders, surface roughness as implemented by PTC has proven to be very effective in inducing a specific FIM.

The following design considerations should be noted:

- VIVACE can operate with horizontal cylinders when they are neutrally buoyant and can be positioned accurately perpendicular to the free stream. Vertical cylinders are recommended when the water depth is adequate and the flow velocity profile with respect to the water depth is reasonably constant above the boundary layer at the riverbed.
- As far as hydrodynamics is concerned, ALTs convert MHK energy in an environmentally compatible way whether they operate in a school or as individuals.

47.2.2 Scales, Models, Prototypes

The VIVACE converter is a highly scalable MHK energy device because the underlying FIM phenomena of VIV, galloping, gap flow, and wake interference are highly scalable. The various scales are presented in Sect. 47.2.2, *Scales* in relation to hydrodynamics and applications. The various models and prototypes

tested so far are presented in Sect. 47.2.2, *Models* and Sect. 47.2.2, *Prototypes*, respectively.

Scales

Definition of Scalability. The term *scale* has two possible interpretations. The first pertains to the wedge stabilization game [47.143] measuring the impact of pollution on the environment. The second interpretation relates to the technology or product size and the market it addresses (Table 47.4). VIVACE covers a broad range of applications from Scale-1 of just a few Watts to MW in utility Scale-4. It is expected to revolutionize some markets and have a major impact on others. In the wedge stabilization game, Scale-4 will have an impact.

Stabilization Wedge. If only 10% of the 22 000 TWh/yr horizontal MHK energy could be harnessed it would be equivalent to 0.62 of a wedge as defined by *Pacala* and *Socolow* [47.143]. The figure of 22 000 TWh/yr was calculated by *Bedard* et al. [47.18] based on present technology. VIVACE can have a major impact by harnessing also the slower flows that constitute the vast majority of water currents. For VIVACE to provide a full wedge – that is reduce CO_2 emissions by 1 Gt/yr – 16% of the 22 000 TWh/yr horizontal MHK energy or 3550 TWh/yr should be harnessed. This is well below the Betz limit. It would be an even smaller percentage when slower currents are included that can be harnessed by ALTs and not by SLTs being considered at the present time.

Returning to the more conventional definition of scale, as the size of the MHK energy device and the power it generates, the following four scales are identified. Figure 47.16 shows the hydrodynamic scales mapped on the Reynolds number and flow regimes.

Scale-1. It is the smallest scale and has very limited but unique and valuable applications. The four important aspects of this scale are:

1. Hydrodynamics: Ultraslow currents at $\approx 0.25\,\mathrm{m\,s^{-1}} \cong 0.5\,\mathrm{kn}$ abound in deep waters. This low-power scale can be used to power sensors AUVs (autonomous underwater vehicles) since battery life is too short for operations. In the MRELab, VIVACE works down to $0.4\,\mathrm{m\,s^{-1}}$ [47.67, 95] and recently to $0.31\,\mathrm{m\,s^{-1}}$, still remaining a compact device.
2. Power: About 10 W for a small converter adequate to power sensors and recharge batteries.
3. Products: The following are considered:
   a) The AUV garage shown in Table 47.4 designed by vortex hydro energy VHE with Northrop–Grumman.
   b) A self-powering AUV becoming the VIVACE cylinder itself, thus extending both the range and days of operation indefinitely. It is also shown in Table 47.4 approaching the garage.
   c) Direct powering of a spider-web of hardwired sensors.
4. Applications: Monitor coasts for defense, law-enforcement. Monitor offshore platforms, track fish populations, track spills, and other sources of pollution. Provide early alert to weather conditions.

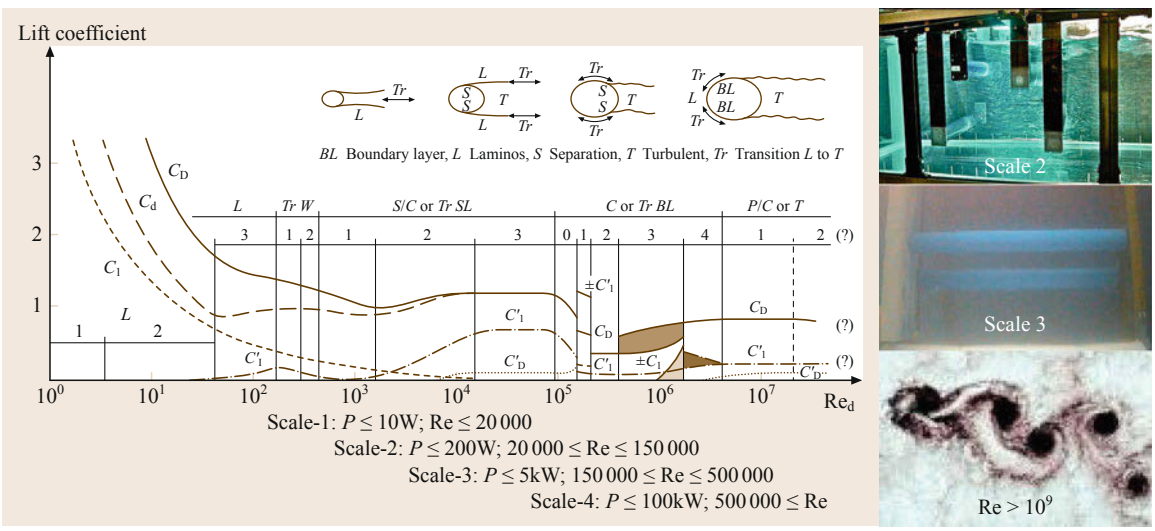
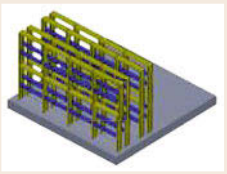


Fig. 47.16 Hydrodynamic and power scales for the VIVACE converter (after [47.56]). Scale-2: Channel MRELab test. Scale-3: Field test in St. Clair River. Large scale: NASA picture of clouds floating around an island mountain peak

Table 47.4 Hydrodynamic and power scales for the VIVACE converter (photos by Vortex Hydro Energy)

Scale	Scale-1 $1000 \leq Re \leq 20\,000$ Early laminar flow	Scale-2 $20\,000 \leq Re \leq 300\,000$ Late laminar flow	Scale-3 $300\,000 \leq Re \leq 500\,000$ Critical to turbulent	Scale-4 $500\,000 \leq Re$ Post supercritical
Unit				
Weight	≈ 500 lbs	≈ 1000 lbs	≈ 20 000 lbs	≈ 100 000 lbs
Power[a]	≈< 10 W	200 W	3500 W	> 50 kW
$/kWh	TBD	$ 2.00	$ 0.80	$ 0.08–0.12
Application	UUV, sensors, tracking, pollution, weather, fish, defence	Portable, naval, expeditions, camps	Remote communities, lighthouses, naval operations	Utility scales, coastal communities, islands

[a] Power output is tested or calculated at 3 kt

Scale-2. This is the scale that has been tested extensively as it fits in the LTFSW channel of the MRELab. The four important aspects of this scale are:

1. Hydrodynamics: Scale-2 has been demonstrated and is being developed/tested in the MRELab since 2005 (Figs. 47.5a,b and Table 47.4). It operates in the TrSL3 flow regime, where the shear layers are fully turbulent, and lift is high reaching unprecedented amplitudes even under high damping as explained in Sect. 47.2.3.
2. Power: Proven in the MRELab to have power density about 14 600 times that of wind farms and 15–100 that of other MHK energy devices [47.100]
3. Products: VHE has designed two units at this scale (Table 47.4):
   a) A lightweight compact unit that can be transported by a single soldier in the field to power electronics.
   b) A more rugged unit to mount in an offshore lab in Brazil and a canal in the Netherlands.
4. Applications: Military expeditions in remote areas where a gallon of fuel costs $ 400, small camps near rivers or areas with no resources, such as Alaska and the Caribbean.

Scale-3. This is the scale that has been tested with field tests and in the towing tank of the University of Michigan. The four important aspects of this scale are:

1. Hydrodynamics: Scale-3 was demonstrated in the St. Clair river in August 2010 (Fig. 47.5c). Scale-3 operates in the TrSL3 subcritical (laminar), critical (transition from laminar to turbulent), and early post-critical (turbulent) flow regimes. In the transition, VIV is fully suppressed. The breakthrough achieved in the MRELab was that by applying PTC properly. FIM was sustained through the critical flow regime into the post-critical without losing amplitude.
2. Power: Of the order of 5 kW for a VIVACE unit of $\approx 10\,m^3$ in 3 kn current.
3. Product: A modular product that can build farms is VHE's target in this scale.
4. Applications: Powering of river and coastal camps and small communities.

Scale-4. This is the scale that has not been tested with field tests but is currently under funding by DOE for development and deployment in the summer of 2016. The four important aspects of this scale are:

1. Hydrodynamics: Scale-4 does not represent any challenge given that there is no change in hydrodynamics/FIM between Scales 3 and 4 that could affect operation. It operates in the fully turbulent flow.
2. Power: About 33 kW for a $80\,m^3$ VIVACE module at 3.2 kn.
3. Product: A modular product that can be repeated to build farms to achieve the next scale of 500 kW. There is no upper limit to this scale, as flow properties related to VIV and galloping are not known to change within the super-post-critical flow.
4. Applications: Powering of river/coastal communities at utility scale. Offshore farms can generate power of the order of MW. Ultimately, large-scale VIVACE converters can be installed along ocean currents for multi-MW power generation

(e.g., the Florida Current). According to the Energy Power Research Institute, the annual hydrokinetic tidal power output potential for North America is 115 000 GWh [47.18].

The following design considerations should be noted:

- Alternating-lift technologies, such as VIVACE, are highly scalable because the underlying principles are highly scalable. In this particular case study, all FIMs, as well as gap flow and wake interference are highly scalable.
- Four scales have been identified in this case study, primarily based on the consistency of hydrodynamics in each scale when using PTC. Scale-1, for low Reynolds numbers including the TrSL2 (Transition Shear Layer 2 – *Zdravkovich* [47.56], Fig. 47.16) flow regime. Scale-2, for moderate Reynolds numbers covering the first half of the TrSL3 flow regime. Scale-3, covering the second half of the TrSL3 flow regime and the critical flow. Scale-4, covering the post-critical flow regime.
- Commercial products with unique applications can be developed in each one of the four scales.

Models

Models are typically defined as scaled-down geometrically similar (geosims) devices built for the sole purpose of testing them in experimental facilities to extrapolate measurements to enable analysis and design of prototypes. Often, several geosims of different size are built and tested prior to building a prototype. A five-step protocol for the development and evaluation of tidal-current energy devices was published by *Bahaj* et al. [47.144]. They also showed the correspondence between the five steps in the proposed protocol to the nine levels of the NASA TRL (technology readiness level) process [47.145].

If the underlying hydrodynamic principles, however, hold over a very broad range of Reynolds numbers (Sect. 47.2.3, *Mechanical Principles*) as in the present case study, models could be considered and serve as smaller scale prototypes. Those scales were defined in Sect. 47.2.1, *Basic Concepts*. Hereafter, to minimize confusion, devices of Scale-1 and Scale-2 tested in the LTFSW channel or towing tanks are referred to as models while devices tested in field tests in Scale-2 and Scale-3 are referred to as prototypes. Several prototypes have been tested at full size in towing tanks before used in field tests.

Figure 47.17a shows a four, horizontal, PTC-cylinder model of a VIVACE converter tested in the old (prior to 2012) LTFSW channel of the MRELab. All FIMs used by VIVACE were observed and measured: VIV, galloping, coexistence of VIV and galloping, and enhancing gap flow [47.84]. Response in galloping reached three diameters hitting the safety stops, which led to rebuilding of the channel to accommodate larger amplitudes up to 5.5 times the cylinder diameter of $D = 3.5''$ (0.0889 m), as explained in Sect. 47.3.1.

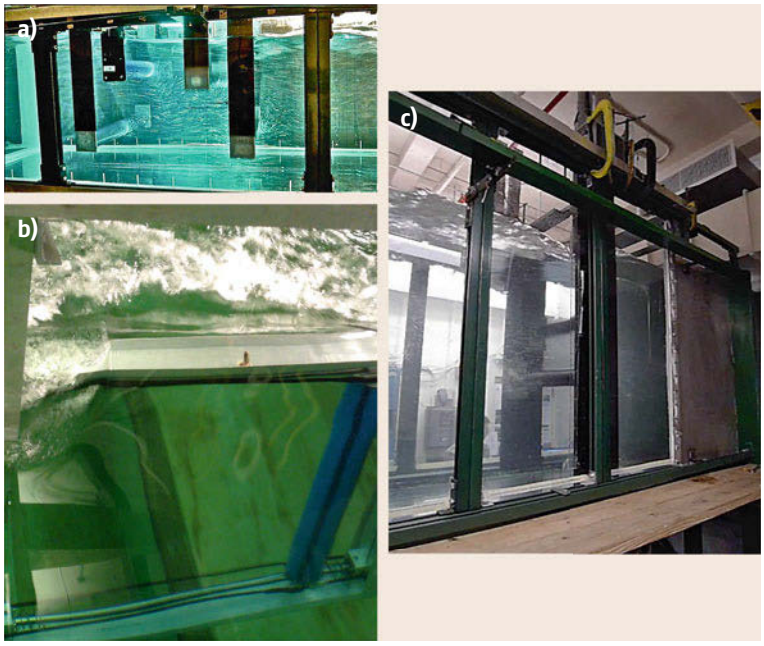


Fig. 47.17a–c Laboratory models of VIVACE: Scale-2 (photos courtesy of MRELab) **(a)** $D = 3.5''$ (0.0889 m), $L = 36''$ (0.914 m). MRELab tests observed synergistic multibody FIM: VIV, galloping, coexistence of VIV and galloping, and enhancing gap flow, **(b)** $D = 8''$ (0.203 m), $L = 60''$ (1.524 m). Towing-tank tests at the University of Michigan observed VIV, galloping, and coexistence of VIV and galloping, **(c)** $D = 3.5''$ (0.0889 m), $L = 35.25''$ (0.895 m). MRELab tests of 2 PTC cylinders in galloping with enhancing gap flow

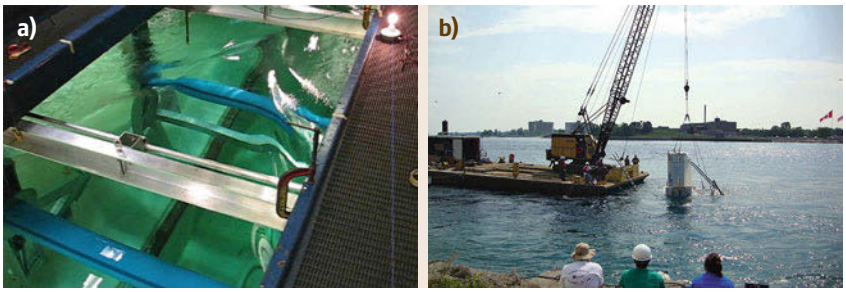


Fig. 47.18a,b Prototypes of VIVACE in field tests: Sacle-3 (photos courtesy of Vortex Hydro Energy). **(a)** VHE with MRELab tests of two PTC cylinders in galloping with enhancing gap flow. $D = 10''$ (0.254 m), $L = 105''$ (2.70 m), **(b)** St. Clair river tests based on synergistic multibody flow induced motions: VIV, galloping, coexistence of VIV and galloping, and enhancing gap flow

Figure 47.17b shows a single, vertical, PTC-cylinder model of a VIVACE converter tested in the towing tank of the Marine Hydrodynamics Laboratory at the University of Michigan. VIV, galloping, and coexistence of VIV and galloping used by VIVACE were observed and measured. Response in galloping reached four diameters for $D = 8''$ when hitting the safety stops.

Figure 47.17c shows a two, horizontal, PTC-cylinder model of a VIVACE converter tested in the new (since 2013) LTFSW channel of the MRELab at the University of Michigan. All FIMs used by VIVACE were observed and measured: VIV, galloping, coexistence of VIV and galloping, and enhancing gap flow [47.84, 142]. Safety stops were not hit as the new channel allows for amplitudes of 5.5 cylinder diameters for $D = 8''$ (0.203 m). In the new V_{ck} system (Sect. 47.2.3, *Electro-Mechanical Principles*, Principle 12), programmable stoppers are set by the controller at both ends of the cylinder travel, in addition to the physical bumpers, in order to protect the channel.

Prototypes

In contrast to models, which are tested only in experimental facilities, prototypes are field deployed. Since the VIVACE converter is highly scalable, devices that have been tested in the field are referred to as prototypes of various scales even if they have been tested in experimental facilities prior to their field deployment. VHE in collaboration with the MRELab has deployed two different prototypes in the St. Clair river and two in canals in the Netherlands in collaboration with TAUW.

River Prototypes. Two different river prototypes were deployed and tested at the same location in the St. Clair river, at Dunn's paper, near the Blue Water Bridge. The flow was steady with velocity in the range of 2.1–2.5 kn and water depth varying from 20 to 60 ft.

The first prototype was deployed in August 2010 after extensive testing in the towing tank at the University of Michigan as shown in Fig. 47.18. The synergistic kinematics of two PTC cylinders worked very well. Each horizontal cylinder was neutrally buoyant with $D = 10''$ (0.254 m) and $L = 105''$ (2.70 m), and weighed 150 pounds in air without ballast. FIM was self-initiated.

The second river prototype is shown in Fig. 47.19 during testing at the towing tank of OHMSETT

Fig. 47.19a,b Prototypes of VIVACE in field tests: Sacle-2; $D = 8''$, $L = 60''$ (photos courtesy of Vortex Hydro Energy). **(a)** VHE tests of VIVACE prototype at OHMSETT towing tank based on VIV, galloping, and coexistence of VIV and galloping, Pictures courtesy of VHE. **(b)** Deployment in St. Clair River in September 2012. $D = 8''$ (0.203 m), $L = 60''$ (1.524 m)

Fig. 47.20a,b VIVACE Prototype in field tests in the Netherlands: $D = 10''$ (0254 m), $L = 60''$ (1.524 m). **(a)** Model of VIVACE at MHLab towing tank. Shorter models require bigger diameter to overcome system friction (photo courtesy of MRELab). **(b)** Shorter models require uniform flow and bigger diameter to ensure adequate lift (after [47.146])

[47.147], in Leonardo, New Jersey, deployment in the St. Clair River, and in operation in the field viewed through an underwater camera. It was field tested in September to November 2012. It was smaller than the previous prototype, with $D = 10''$ (0.254 m) and $L = 105''$ (2.70 m). One cylinder was used and positioned vertically. Upon installation FIM was not self-initiated. It was quickly realized that the cylinder was placed too close to the riverbed with about 45% of its length inside the boundary layer of the river flow. The cylinder was raised by about 1 m and FIM was self-initiated immediately.

Canal Prototypes. Two deployments of the same, small, Scale-2 prototype took place in collaboration with TAUW, in canals in the Netherlands, in January 2013. The device was tested in the University of Michigan towing tank (Fig. 47.20a) and functioned well. It was not self-initiated but went into FIM when it was given an initial displacement. This is typical of hard-galloping [47.43]. This device was the smallest possible as it was constrained by the size of the opening in the gate shown in Figs. 47.20b and 47.21. Thus, overcoming the system friction required a bigger diameter cylinder reducing the allowable maximum amplitude of oscillation. The increased diameter provided the required force. The damping of the system associated with harnessing the energy was increased to achieve higher MHK energy conversion. That also reduced the amplitude response to avoid collision of the cylinder in FIM with the frame, which would have wasted valuable energy. When the water passed through the orifice in Fig. 47.20b, the flow was uniform only in about one-third of the length around the center of the cylinder. That is, two-thirds of the cylinder length was providing drag and one-third was providing lift. The lift was not sufficient to overcome the drag from the two-thirds of the cylinder length, plus the system damping. FIM was not sustained even with initial displacement. Side cuts were opened around the gate. This resulted in more uniform flow and initiation of FIM. The electrical output was small as expected under the adverse circumstances.

Scale-4 Farm. Vortex hydro energy (VHE) is developing a large-scale device called *Oscylator-4*. It is designed to produce 4 kW in a flow of 3.2 kn. It has four large PTC cylinders in close proximity that allows for high power-to-weight/volume density, which is the Achilles heel of all renewable energy devices. Several *Oscylator-4s* will form a small farm for commercial applications.

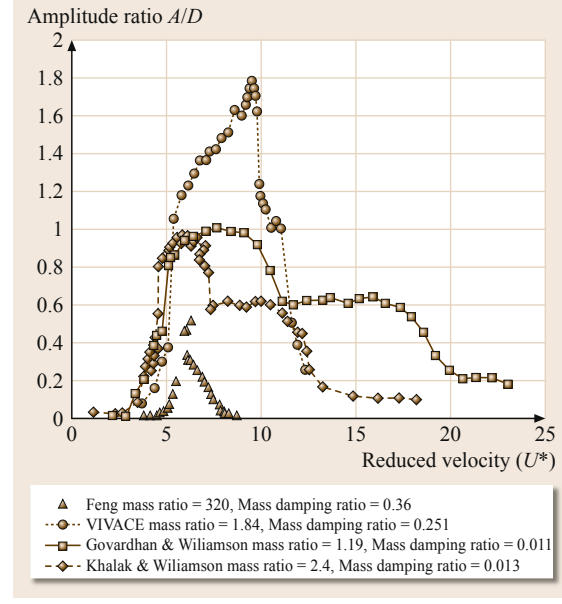


Fig. 47.21 In spite of high damping for energy harnessing, VIVACE amplitude in the TrSL3 flow regime is significantly higher than VIV response in the TrSL2 flow regime. In addition, the upper branch overtakes the lower branch and with amplitude increasing with velocity (after [47.107])

On the basis of the extensive laboratory experiments and four field deployments of models and prototypes of various scales, the following conclusions are important to record for the specific case study considered in Sect. 47.2.2:

1. Development of a MHK energy device – or any ocean structure for that matter – requires several stages of development described in detail in the Southampton Protocol [47.144]. It takes into consideration the special requirements of scaling, testing, building, and launching of ocean structures while observing the nine levels of Technology Readiness Levels developed by NASA [47.145].
2. ALTs may have an advantage – as in the case study of the VIVACE converter – that the hydrodynamic phenomena on which they are based are highly scalable. Accordingly, a small-scale prototype may serve as a larger scale model.
3. Small-scale models face additional challenges related to end flow conditions, which typically reduce the two dimensionality of the flow. Nonuniform flows result in the reduction of the length of the cylinder, which provides lift. Instead, they are subject to opposing forces in the lift direction in addition to the system damping. This double jeopardy may result in failure of an ALT to function at satisfactory efficiency.

47.2.3 The Underlying Principles

Following the analysis of the basic concepts behind ALTs, the discussion of scales, and the case study and its operational conditions considered in Sect. 47.2, it is appropriate to understand the underlying principles of ALTs. Twelve physical principles are implemented in the discovery and development of the case study of the VIVACE converter. Most of these would apply to a certain extent to all forms of ALT.

For easier presentation and understanding, they are divided into three categories: hydrodynamics, mechanical, and electromechanical.

Hydrodynamic Principles

Principle 1 – Fluid–Structure Interaction (FSI). When a flexible body or structural system is exposed to a fluid flow and the hydrodynamics, and the structural dynamics scales are comparable, the presence of each (fluid, structure) affects the other. In the case study of the VIVACE converter, FSI results in dynamic bifurcations and instabilities, such as various forms of FIM.

Principle 2 – Flow-Induced Motions (FIM). A buoy floating on waves is not in FIM as the buoy is too small to affect the frequency of the wave field. The diffracted waves have the same frequency as the incident wave field. On the contrary, a flexible cylinder or a rigid cylinder on elastic support has a diameter close in size to that of the von Kármán vortices generated in the cylinder wake. The frequency of the vortex street, which induces oscillatory lift (and drag) on the body, is affected drastically by the interaction with the body motion. A rigid cylinder on springs is excited nonlinearly in FIM in transverse current. The presence of the body affects the flow and vice versa. VIVs, galloping, buffeting, fluttering, and gap flows in schools of bodies are FIM phenomena.

Principle 3 – Vortex-Induced Vibrations (VIV). A cylinder with its axis perpendicular to a steady flow generates the von Kármán vortex street where the vortex size is about equal to the cylinder diameter (Figs. 47.7b, 47.16). When the cylinder is flexible it may oscillate driven by the alternating vortex street at a frequency of oscillation f_{osc}. The wake frequency f_{wake} is the same as f_{osc}. The vortex shedding frequency f_{vs}, though, is a multiple of the wake frequency f_{wake} depending on the vortex structure being 2S (two single vortices) or 2P (two pairs of vortices), etc. The vortex shedding frequency f_{vs} when the cylinder is in FIM, should not be confused with the vortex-shedding frequency f_s when the cylinder is stationary. f_s is also called the Stouhal frequency defined as

$$\mathrm{St} = \frac{f_s D}{U} \,. \tag{47.1}$$

This synchronization between f_{osc} and f_{wake} persists over a very broad range of stream velocities resulting in large amplitude oscillations. The frequency of oscillation f_{osc} changes with flow speed as the vortex structures change. Some researchers interpret this synchronization phenomenon between excitation and response as a lock-in phenomenon and others as a nonlinear resonance over the same broad range of flow velocities [47.60] with variable natural frequency in water ($f_{n,water}$) due to variable added mass. The latter can be measured experimentally. The mathematical modeling of VIV as lock-in is presented in Sect. 47.3.6. The mathematical model of VIV as a variable natural frequency – and thus added mass – phenomenon is presented in Sect. 47.3.7. $f_{n,water}$ is defined as

$$f_{n,water} = \frac{1}{2\pi}\sqrt{\frac{k}{m_{osc}+m_a}} \,. \tag{47.2}$$

This broad-band synchronization phenomenon between excitation and cylinder vibration is called VIV.

Leonardo da Vinci first observed VIV of a taut wire exposed to wind-flow transverse to the wire axis, in 1504. He called the emitted sound *Aeolian Tone*. Typically, VIV occurs in the range

$$\approx 5 < U^* = \frac{U}{f_{\mathrm{n,water}} D} <\approx 10\,, \tag{47.3}$$

where U = stream velocity, D = cylinder diameter, U^* = reduced velocity,

$$f_{\mathrm{n,water}} = \frac{1}{2\pi}\sqrt{\frac{k}{m_{\mathrm{osc}} + m_{\mathrm{a}}}}\,, \tag{47.4}$$

and the ideal flow added mass coefficient $C_{\mathrm{A}} = 1$ is used resulting in $m_{\mathrm{a}} = m_{\mathrm{d}}$ where m_{d} is the displaced fluid mass

$$m_{\mathrm{d}} = \rho\pi\frac{D^2}{4}L\,. \tag{47.5}$$

This is purely for data presentation reasons and does not support or negate any of the two interpretations of VIV explained in the previous paragraph.

As a synchronization phenomenon, VIV is destructive like linear resonance. Two differences are important:

1. VIV is self-limiting and in that sense not as destructive as liner resonance
2. VIV is broad-banded while linear resonance is very narrow-banded.

In that respect, VIV is much more difficult to prevent in designing structures in dynamic response. On the other hand, the broad range of synchronization makes VIV a much more suitable principle to base an MHK energy converter than the principle of a linear oscillator.

Another very important characteristic of VIV is its strong dependence on Reynolds number [47.107, 148] as shown in Fig. 47.21 and explained further (Figs. 47.42– 47.45) in Sect. 47.3.1, *First Low-Turbulence Free-Surface Water (LTFSW) Channel*. Dependence on Reynolds number brings the VIV response to new heights in the TrSL3 flow regime – around two diameters amplitude – and reveals a very different response operator as a function of reduced velocity+U^* than the VIV response operator in the TrSL2 flow regime. That is, the TrSL3 flow regime is much more efficient for VIVACE to convert MHK energy. This is confirmed by the high-lift coefficient C_{L} as shown in Fig. 47.16.

Principle 4 – Galloping. It is a FIM more vigorous than VIV. It is an instability phenomenon caused by geometric/hydrodynamic asymmetry and once initiated has no end with increasing velocity until structural failure [47.43]. A smooth circular cylinder in a uniform stream could not be destabilized to gallop.

A simple mathematical model of an oscillator leading to galloping is shown in Fig. 47.22

$$(m_{\mathrm{osc}} + m_{\mathrm{a}})(\ddot{y} + 2\zeta_{\mathrm{water}}\omega_{\mathrm{n,water}}\dot{y} + \omega^2_{\mathrm{n,water}}y) = F_{\mathrm{Y}}(t)\,, \tag{47.6}$$

where the damping ratio is defined as

$$\zeta_{\mathrm{water}} = \frac{c}{2\sqrt{k(m_{\mathrm{osc}} + m_{\mathrm{a}})}}\,, \tag{47.7}$$

the natural frequency in water is

$$\omega_{\mathrm{n,water}} = \sqrt{\frac{k}{m_{\mathrm{osc}} + m_{\mathrm{a}}}}\,, \tag{47.8}$$

and

$$\begin{aligned} F_{\mathrm{Y}}(t) &= -F_{\mathrm{L}}\cos\alpha - F_{\mathrm{D}}\sin\alpha \\ &= -\frac{1}{2}\rho DLU^2 C_L\cos\alpha - F\frac{1}{2}\rho DLU^2 C_{\mathrm{D}}\sin\alpha\,, \end{aligned} \tag{47.9}$$

and, thus, C_{Y} is given by

$$C_{\mathrm{Y}} = -C_{\mathrm{L}}\cos\alpha - C_{\mathrm{D}}\sin\alpha\,. \tag{47.10}$$

For a small angle of attack, C_{Y} can be expanded in Taylor series to yield

$$C_{\mathrm{Y}} = -\left[\left(\frac{\partial C_{\mathrm{L}}}{\partial\alpha}\right)_{\alpha=0} + C_{\mathrm{D}}(\alpha = 0)\right]\frac{\dot{y}}{U}\,, \tag{47.11}$$

resulting in the following equation of motion

$$\ddot{y} + 2\zeta_{\mathrm{total}}\omega_{\mathrm{n,water}}\dot{y} + \omega^2_{\mathrm{n,water}}y = 0\,, \tag{47.12}$$

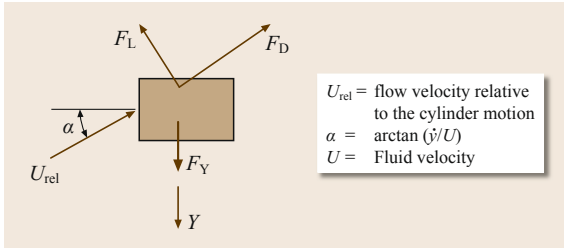


Fig. 47.22 Kinematics of a square prism leading to galloping instability (after [47.43])

where

$$\zeta_{\text{total}} = \zeta_{\text{system}} - \frac{1}{4}\frac{\rho D U}{(m_{\text{osc}} + m_{\text{a}})\omega_{\text{n,water}}}\left(\frac{\partial C_L}{\partial \alpha}\right)_{\alpha=0} . \tag{47.13}$$

Galloping initiates when

$$\zeta_{\text{total}} < 0 . \tag{47.14}$$

There are several forms of galloping, which are well described in the literature:

1. Soft or regular galloping which is self-initiated and thus does not require a push on the VIVACE cylinders by the controller. This is a typical Hopf bifurcation [47.45].
2. Hard or regular galloping which is not self-initiated and thus does require a push on the VIVACE cylinders by the controller. This is a subcritical Hopf bifurcation [47.45].
3. Interference galloping, which is further divided as: proximity (P), wake interference (W), and combination of both proximity plus wake interference ($P + W$) depending on the relative position of the two bodies [47.81].

Differences Between VIV and Galloping. VIV and galloping are fundamentally different phenomena. It is important to understand these differences as ALTs would be designed differently to take advantage of the underlying FIM (VIV verses galloping) for MHK-energy harnessing. Most important differences are the following:

1. The hydrodynamic driving mechanisms are different. VIV is driven by synchronization of the wake frequency with the frequency of oscillation, that is

$$f_{\text{wake}} = f_{\text{osc}} . \tag{47.15}$$

   In galloping, even though vortex streets are generated in the cylinder's wake, these vortices are not the driving FIM mechanism in galloping. This is obvious in Fig. 47.10b where galloping is fully developed at Re = 130 000. Some vortices are out of phase with the cylinder motion. Vortices push the cylinder away from the side they shed.
2. The galloping instability is due to the negative lift created by the geometric asymmetry. Thus, the cylinder is being pushed in the direction it moves. If it was not for the limited width of the converter's frame and the restoring springs, the cylinder's motion would not be reversed.
3. VIV is self-limiting in amplitude. When the relative flow velocity exceeds a certain limit, either due to increased stream velocity or due to increased amplitude of oscillation (in forced oscillations), vortices mature and shed prior to the end of the cylinder travel. At that point, the wake frequency from vortex shedding f_{wake} falls out of synchronization with the frequency of oscillation f_{osc} and results in the lift opposing the transverse motion and terminating VIV.
4. In galloping, f_{osc} drops below $f_{\text{n,water}}$. In VIV, f_{osc} increases above $f_{\text{n,water}}$.
5. The VIV synchronization range hardly changes in terms of reduced velocity U^*. As stated in the double inequality (47.3), the range of synchronization for VIV is dictated by the range $\approx 5 < U^* = U/(f_{\text{n,water}}D) <\approx 10$. Accordingly, VIV can be moved into the range of stream velocity for energy harnessing by changing the cylinder diameter D, or its $f_{\text{n,water}}$ by changing its mass m_{osc} or its spring constant K. For example, if we assume that the end of the VIV synchronization range occurs approximately at $U^* = 10.0$, we have

$$U^* = \frac{U}{f_{\text{n,w}}D} = 10 \Leftrightarrow U = 10D\frac{1}{2\pi}\sqrt{\frac{K}{(m + m_{\text{a}})}} . \tag{47.16}$$

That is, the end of the VIV synchronization range increases proportionally with the square root of the spring stiffness K. Obviously, the onset of synchronization will also follow the same rule.

On the other hand, galloping initiates at an absolute value of velocity $U = U_{\text{g}}$, which is not affected by the spring constant K [47.43].

$$U_{\text{g}} = \frac{2cDm^*}{a} , \tag{47.17}$$

where c is the linear viscous damping coefficient, a is the geometry-dependent constant, and

$$m^* = \frac{m_{\text{osc}}}{m_{\text{d}}} . \tag{47.18}$$

That is, initiation of galloping does not depend on K but only on the absolute flow velocity and geometric and dynamic particulars of the oscillator. Higher damping would require a higher velocity to initiate galloping.

Principle 5 – Transition from VIV to Galloping. Typically, VIV occurs before galloping in the U^* scale, where galloping starts at $U^* \approx 20$. Then, a gap will

exist between the ending of VIV and the onset of galloping. This is discussed further in Principle 9 and in Fig. 47.29a below. In that gap, the VIVACE converter will remain idle or its parameters would require adjustment to either shift the VIV range to higher speeds, or initiate galloping earlier. By comparing (47.16) and (47.17), it is easy to see that it is feasible to bridge the gap and make VIV and galloping overlap. What happens during such transition given that VIV and galloping differ at least in the five points (1–5) stated earlier in Sect. 47.2.3, *The Underlying Principles*:

1. The amplitude of oscillation increases sharply over a very short range of U^*.
2. The two hydrodynamic driving mechanisms of VIV and galloping coexist. This is shown in Fig. 47.10a, where vortices still push the cylinder in the direction of its motion.
3. The amplitude of oscillation increases to the point that the VIV mechanism of wake synchronization cannot follow and fully developed galloping takes over. Figure 47.10b shows that vortices shedding out of synchronization and pushing the cylinder against its motion.

Principle 6 – Enhancement of FIM. ALTs, such as the VIVACE converter, take advantage of catastrophic phenomena such as FIMs, control them, and in the process harness energy from fluid flows using an appropriate damping mechanism. Enhancement of FIM is desirable in order to convert more MHK energy to mechanical in the oscillator; as long as the enhanced response of the oscillator can be controlled.

There are several flow regimes and configurations where FIM enhancement is desirable. The following have been pursued and successfully enhanced in VIVACE converter models and prototypes using properly designed turbulence stimulation – called PTC – which is discussed under Principle 7:

1. In the critical flow regime, at the transition from the laminar boundary layer to turbulent boundary layer, VIV is fully suppressed, the lift coefficient is nearly zero as shown in Fig. 47.16. In that flow regime there is no predominant Strouhal frequency f_s. The Strouhal number is defined as in (47.1), where f_s is the vortex-shedding frequency for a stationary circular cylinder. This regime is important for Scale-2 and Scale-3 because a cylinder with $D = 0.30\,\text{m}$ in a flow speed of about $U = 2.5\,\text{kn}$ results in Reynolds number $\text{Re} = 327\,000$ placing it in the critical flow regime.
2. The lift coefficient, and consequently the amplitude in VIV, is relatively low in the TrSL2 flow regime (Fig. 47.16) [47.56]. Using properly designed turbulence stimulation, the effective Re can be pushed to higher values to move the cylinder flow into the TrSL3 regime where the lift and VIV responses are significantly higher.
3. VIV is a broad range, high-response phenomenon but has a finite range as defined by the double inequality (47.3). Past the end of the VIV synchronization range the response of ALT would be zero leaving a range of velocities with no response from the ALT. Using relations (47.16) and (47.17), in terms of absolute flow velocity U, the VIVACE converter can be redesigned to close the gap between VIV and galloping. This is also achieved by using PTC defined in Principle 7.
4. In case of multiple cylinders in FIM, it is possible to adjust the in-flow and transverse spacing between PTC cylinders to achieve gap flow that can enhance the interaction between cylinders in a school. This is further discussed in Principle 9.

Enhancement of MHK energy production can be achieved by several different means. PTC though, can be applied in the four flow regimes and configurations of FIM enhancement listed above. In all of those the passive enhancement was achieved by affecting the hydrodynamics directly. In addition though, there are ways to enhance the final output of an ALT by affecting the interaction between fluid and structure. Means that are still under development in the MRELab and VHE include:

1. Most of the information on FIM available in the literature uses linear restoring springs. In field tests, VHE has achieved maximum power output using strongly nonlinear restoring force as discussed in Sect. 47.2.1, *Description of the VIVACE Converter* (springs).
2. In the traditional linear time invariant (LTI) mathematical model of oscillators (Sect. 47.3.6), the term that generates power is the damping term. Typical LTI models in textbooks and papers assume that the linear viscous model is representative of one degree-of-freedom LTI oscillators. Papers by *Lee* et al. [47.98] and *Sun*et al. [47.142], prove that damping in real mechanical oscillators is strongly nonlinear and may exhibit memory and hysteresis. In the process of system identifying damping, in VIVACE laboratory models, nonlinear forms of damping have proven to be more effective in affecting the Fluid-Structure Interaction and generating MHK power.
3. The circular cross-section of cylinders in an oscillator provides many advantages from the hydrodynamic and fabrication points of view. Other cross-sectional shapes have been studied experi-

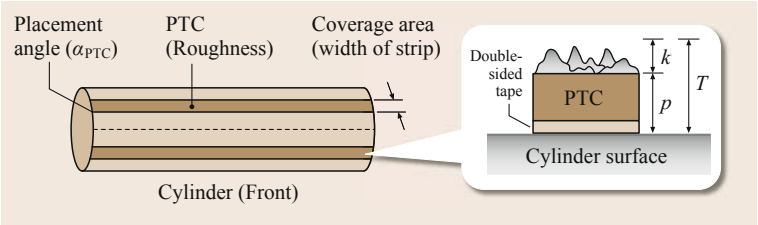


Fig. 47.23 Passive turbulence control (PTC) parameters

mentally and computationally as discussed under Principle 11.

Principle 7 – Passive Turbulence Control (PTC). In the MRELab, a method was developed and patented to induce the FIM of a circular cylinder by designing and applying turbulence stimulation in a passive way [47.67, 70, 96]. This method, named PTC, is based on the selective distribution of surface roughness on a circular cylinder. PTC induces several different FIMs.

The design parameters of PTC are listed below and shown in Fig. 47.23:

1. α_{PTC}, placement angle
2. Area coverage in degrees
3. k, roughness grit size
4. $T = k + p$, PTC total height.

The mechanics of PTC affect the flow around the cylinder and consequently its FIM in the following ways:

1. It trips the boundary layer
2. It may or may not allow reattachment
3. It introduces turbulence into the flow
4. It reorganizes the flow in downstream cylinders and affects the correlation length
5. It introduces rotational asymmetry which may initiate galloping instability.

This four-parameter problem was solved over a period of two years. The results are shown for the TrSL3 flow regime in the PTC-to-FIM map in Fig. 47.24 [47.70]. The following FIM zones were established as a function of the starting point of the PTC location angle in the clock-wise direction from the forward stagnation point:

1. Weak suppression zone 1, WS1
2. Hard galloping zone 1, HG1
3. Soft or regular galloping zone, SG
4. Hard galloping zone 2, HG2
5. Weak suppression zone 2, WS2
6. Strong suppression zone, SS.

Obviously, PTC can act as an FIM amplifier (HG1, HG2 and SG) and an FIM suppressor (WS1, WS2, and SS).

Additional information that is useful in terms of the PTC and its impact on the developed *PTC-to-FIM map* are the following:

1. Several types of commercial sandpaper were used in the experimental investigation. Their properties are shown in Table 47.3.
2. The two maps developed, for two commercial grade sandpapers P60 and P180, exhibit minor variations in each zonal range. Response features within each zone remain nearly independent of the height of the PTC.
3. The thickness of the PTC should be compared to the thickness of the boundary layer where the effect of the PTC is initiated (Fig. 47.25 and Table 47.5).

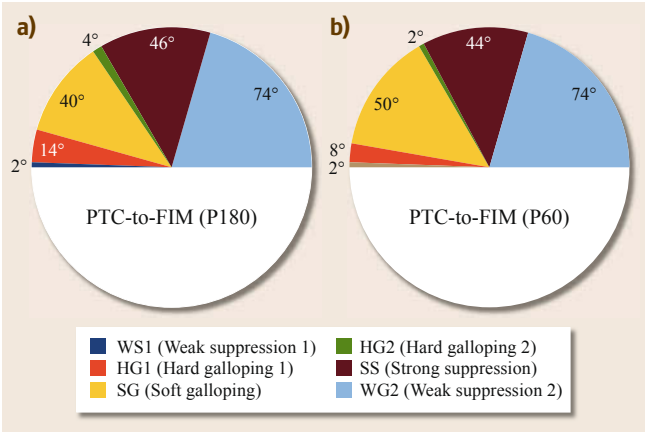


Fig. 47.24a,b PTC-to-FIM map for two different roughness values: **(a)** P180, **(b)** P60. Roughness strip width covers 16° on each side of the cylinder as shown in Fig. 47.23 (after [47.149])

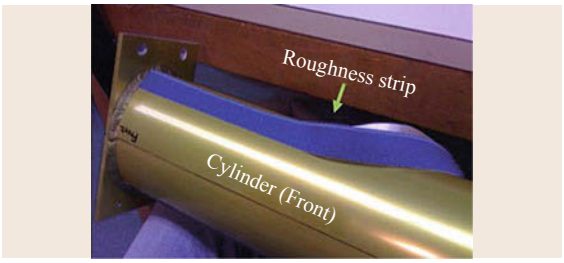


Fig. 47.25 PTC application on cylinder (after [47.101])

Table 47.5 Boundary-layer thickness δ(mm) for various Re and α_{TC}. Calculation of boundary layer thickness for smooth cylinder by the approximate method of Pohlhausen (after [47.150]); $D = 3.5''$ (0.0889 m)

α_{PTC}	Re 10 000	40 000	70 000	80 000	90 000	100 000	118 000
20	1.283	0.6416	0.48500	0.4537	0.4277	0.4058	0.3735
25	1.247	0.6234	0.4712	0.4408	0.4156	0.3943	0.3629
30	1.219	0.6092	0.4605	0.4308	0.4062	0.3853	0.3547
35	1.249	0.6246	0.4721	0.4417	0.4164	0.3950	0.3637

With the application of PTC, the following results have been achieved:

1. Even with a smooth PTC strip of total thickness $T = 0.192$ mm on a $D = 0.0889$ m cylinder, amplitude of three diameter was reached. That is, such a small geometric asymmetry-induced fully developed galloping.
2. By introducing PTC with total thickness of the order of magnitude of the boundary layer, and positioning it properly along the boundary layer (before the separation point) several different forms of FIM were achieved.
3. By placing the PTC in the SS zone, the separation point is covered, which oscillates around 81° in laminar flow for a smooth stationary cylinder.
4. With PTC, galloping was initiated earlier which is an important step toward achieving back-to-back VIV and galloping on a single-circular cylinder [47.70]. See also Principle 9.
5. FIM of multiple cylinders in tandem was studied by *Kim* et al. [47.68, 69, 76, 84, 113, 114, 117] using PTC. Smooth cylinders down stream from the lead cylinder would not move without PTC. The oscillatory amplitudes of all cylinders (up to four tested) reached the limitation of the facility at a flow speed

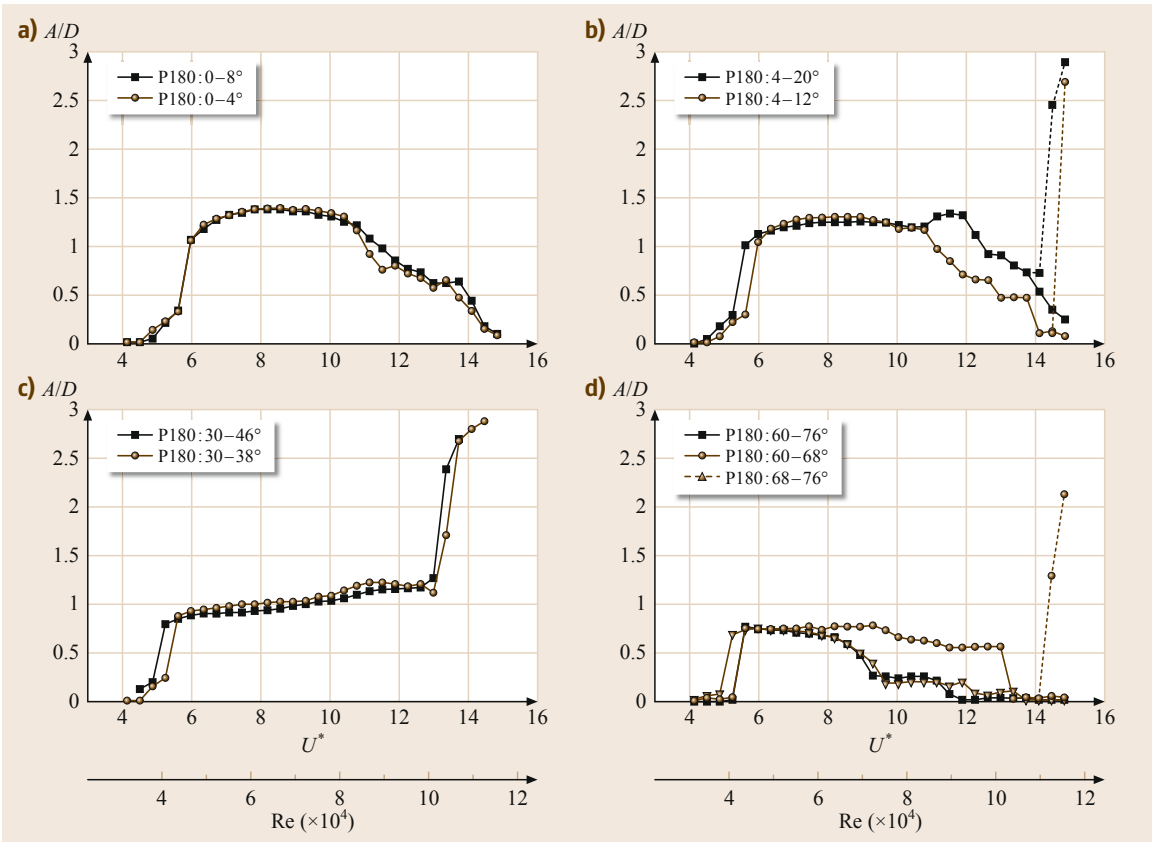


Fig. 47.26a–d Amplitude response A/D for cylinder, with $D = 3.5''$ (0.0889 m), with half-width strip P180. Zones: **(a)** WS1, **(b)** HG1, **(c)** SG, **(d)** SS (after [47.70])

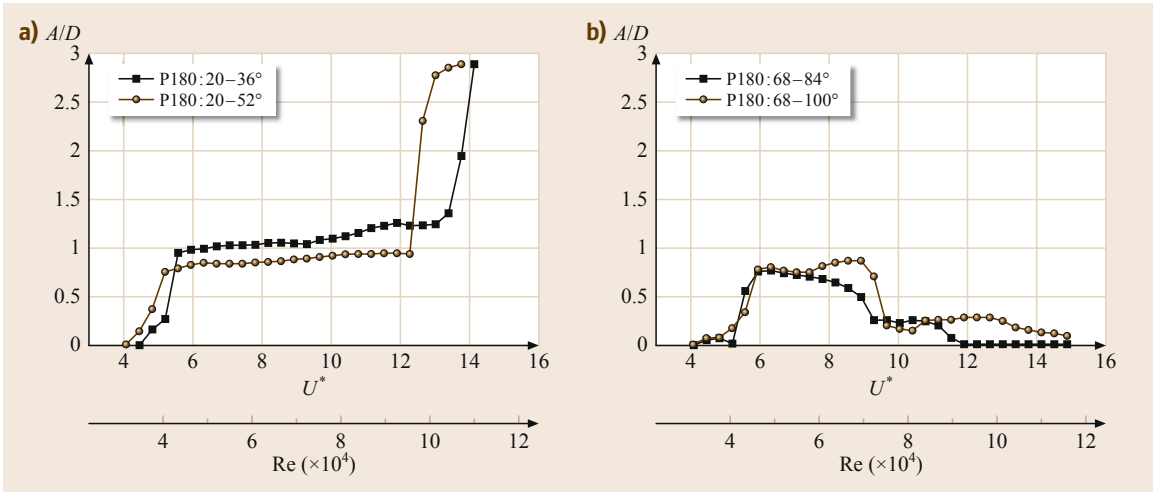


Fig. 47.27a,b Amplitude response A/D for cylinder, with $D = 3.5''$ (0.0889 m), with double-width strip P180. Zones: **(a)** SG, **(b)** SS (after [47.70])

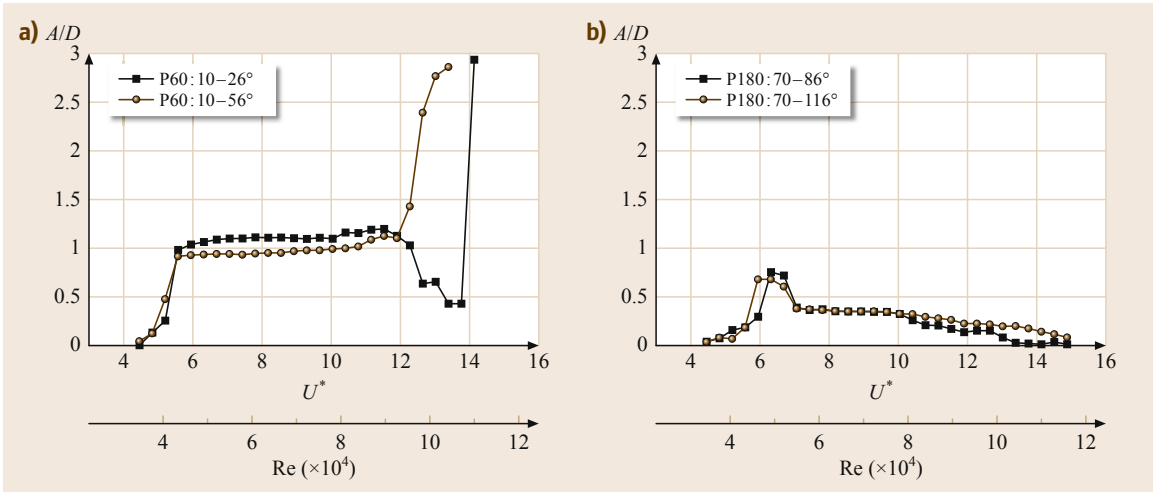


Fig. 47.28a,b Amplitude response A/D for cylinder, with $D = 3.5''$ (0.0889 m), with staggered strip pattern P180 (after [47.70])

of $1.5\,\mathrm{m\,s^{-1}}$, which is below the maximum flow speed of the water channel.

The final issue raised about the effectiveness of the *PTC-to-FIM map* relates to its robustness. The conclusions of this investigation follow supported by Figs. 47.26–47.28 [47.70]:

1. The zones specified in the *PTC-to-FIM map* are robust (insensitive) to width and configuration of the PTC. That is, if the roughness strips are located in their entirety inside a single zone, the cylinder response will be the same regardless of covering (8, 16, 32°) or design.
2. The SS zone is the most dominant zone in the *PTC-to-FIM map*. When the SS zone is covered by PTC along with other adjacent zone(s), the amplitude response of the PTC cylinder corresponds always to strong suppression.
3. The soft (regular) galloping (SG) zone is the second most dominant zone in the map. It has stronger effects than the WS1, HG1, and HG2 zones. When the four zones WS1, HG1, HG2, and SG are covered by roughness, the resulting FIM was soft galloping as long as the SS zone was not covered.
4. A staggered configuration covering all zones is an effective means of suppression. The range of the

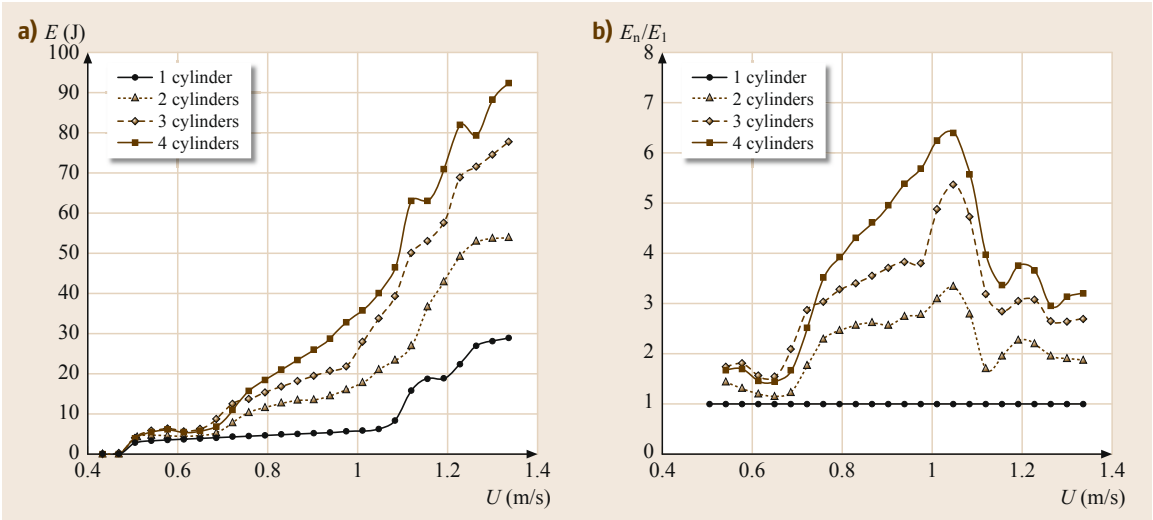


Fig. 47.29a,b Synergistic operation of 2, 3, 4 PTC cylinders for $d/D = 2.5$ (after [47.68, 84]). Smallest spacing is $d/D = 2.5$ for 3 or 4 cylinder tests. For $0.8\,\mathrm{m\,s^{-1}} \le U \le 1.1\,\mathrm{m\,s^{-1}}$, multiple cylinders in synergistic operation harness 60–80% more energy than acting in isolation. **(a)** Total converted energy. **(b)** Normalized converted energy

upper branch is about half that of the smooth cylinder and its maximum amplitude is about 56% of that of the smooth cylinder. Also, the desycnhronization region starts at much lower U^* compared to the smooth cylinder.

Principle 8 – Multi-Body Interaction. A single-cylinder converter cannot achieve high enough power-to-volume/weight ratio. That would correspond to a single-blade propeller or turbine. Placing cylinders far apart, either in tandem or side-by-side, would require significant spacing to minimize interference. To achieve high enough power-to-volume/weight ratio to overcome this Achilles heal of renewable energy technologies, multiple cylinders should be placed in close proximity and made to work efficiently. That would be the equivalent to a multiblade propeller/turbine. The additional challenge in an ALT is that unlike for a propeller, for which all blades have equivalent position as they span the same space and are hydrodynamically equivalent, in a school of cylinders:

1. The lead cylinder is not in the wake of any other cylinder.
2. Each cylinder following the lead cylinder is in the wake of a different number of upstream cylinders.
3. The flow velocity may be reduced due to power harnessing by upstream cylinders.
4. In an unconfined space, the flow may be diverted by the reaction of the oscillating cylinders on the fluid.

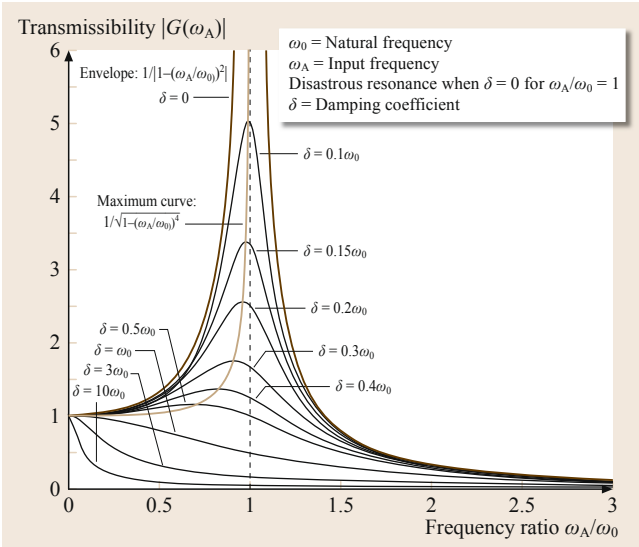


Fig. 47.30 Linear oscillator

Actually, the average drag coefficient in the direction of the flow may increase from $C_D = 1.05$ for a stationary cylinder in steady uniform flow in the laminar regime (Fig. 47.16) to for a cylinder oscillating transversely to a uniform flow for the same Reynolds number [47.151].

The only solution to this problem is that cylinders are placed in close proximity in a school and work synergistically. This has been achieved in the MRELab with up to four cylinders by taking advantage of PTC and interference phenomena such as gap flow. This is

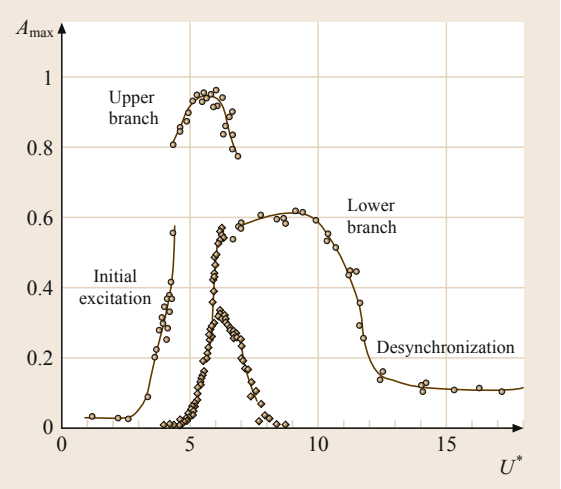


Fig. 47.31 Nonlinear oscillator: low-Re (< 10 000) VIV (after [47.154])

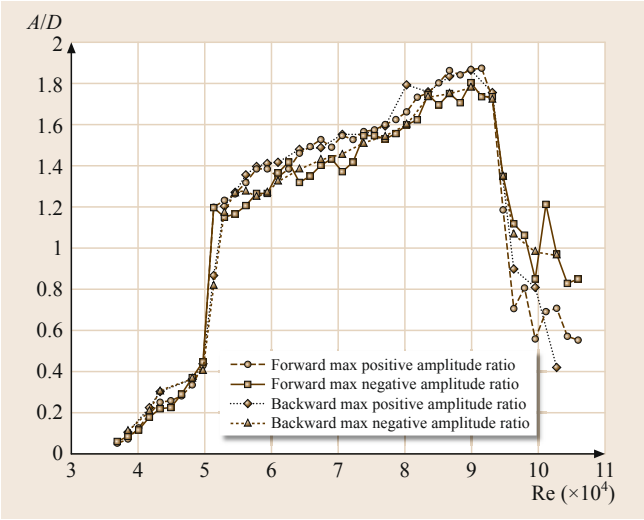


Fig. 47.32 Nonlinear oscillator: higher Re (≈ 30 000–150 000) VIV (after [47.107])

further discussed in the research (Sect. 47.3.5). At this point, it is sufficient to introduce two results:

1. A four-cylinder VIVACE converter has been designed and built in the MRELab, where the cylinders work synergistically and are all in vigorous FIM as shown in the two 30 s videos [47.152, 153]. The converter operates like a 4-cylinder reciprocating engine driven by water. There are no mechanical connections between the cylinders.
2. Figures 47.28 and 47.29 show that two, three, four cylinders placed in a school properly can harness up to 60–80% more MHK energy that the same number of cylinders operating in isolation.

Mechanical Principles

Principles involved in an FSI system cannot rationally be separated into fluid and mechanical principles. Nevertheless, they are separated in Sect. 47.2 for more efficient presentation.

Principle 9 – Nonlinear Oscillator. A one degree-of-freedom linear oscillator has a response amplitude operator like the one shown in Fig. 47.30. It has high response, at low damping, with small bandwidth near the natural frequency f_n. Figure 47.31 shows the response of a cylinder in VIV in the TrSL2 flow regime. It is broad banded but self-limiting with a maximum amplitude of one diameter. Figure 47.32 shows the response of a cylinder in VIV in the TrSL3 flow regime. It is broad banded but self-limiting with a maximum amplitude $A = 1.8D$. The VIVACE oscillator in Fig. 47.33 has open-ended bandwidth to the right, with high response even at high damping, only limited by the experimental facility limitations.

As explained in Principle 5, with PTC (Principle 7) back-to-back VIV and galloping have been achieved resulting in an open-ended range of high-amplitude response. By increasing damping to harness more energy, the gap between VIV and galloping may reappear. This gap can be bridged again by increasing the spring constant K as shown by *Chang* et al. [47.67].

Thus, the VIVACE converter is a nonlinear oscillator, which initiates at low velocities with VIV and continues with galloping. The latter – being an instability – increases in amplitude with increasing velocity unless post-instability nonlinearities introduce additional strength. That makes the VIVACE converter operate in FIM at a very broad range of synchronization and therefore broad range of velocities.

Principle 10 – Viscous Damping for Energy Harnessing. High-amplitude response of the nonlinear VIVACE oscillator is maintained even at high damping as shown in Fig. 47.33. High damping is used to harness energy and lower the oscillator response tending to separate VIV from galloping. High-damping FIM is very different from low mass and damping VIV [47.155], With increased spring stiffness, the back-to-back VIV and galloping feature of the VIVACE oscillator is restored. Nonlinear forms of damping have proven to be more effective in affecting the FSI and generating more MHK power. This concept is still at the research stage in the MRELab and VHE (Vortex Hydro Energy).

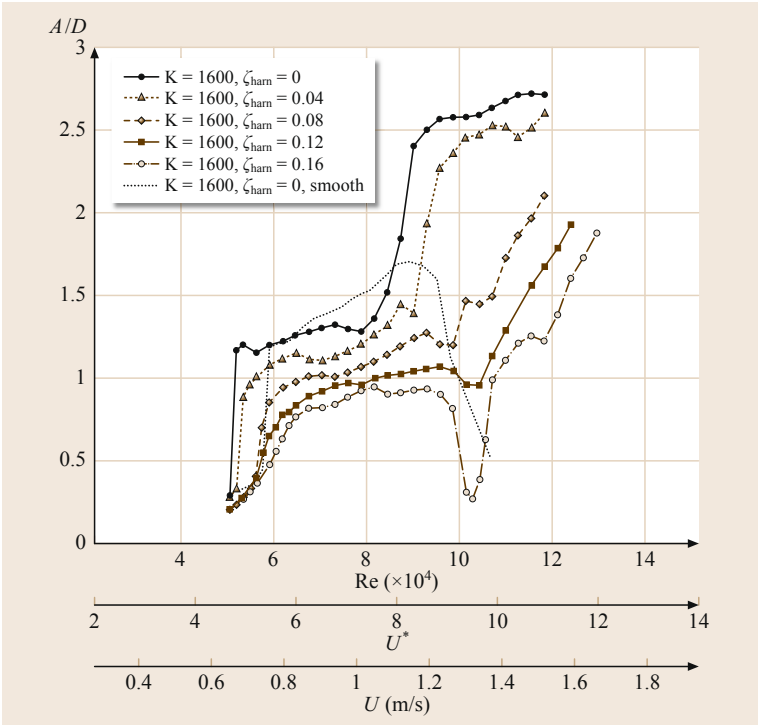


Fig. 47.33 Nonlinear oscillator: higher Re($\approx$ 30 000–150 000) with back-to-back VIV and galloping and variable damping (after [47.96])

Principle 11 – Shape Optimization. Fish in schools glide between vortices of preceding fish to move effortlessly through water. They take advantage of the thrust wake shown in Fig. 47.7c. Testing has shown that fish placed in the wake of a stationary cylinder in a steady flow can maintain their position by gliding between the cylinder alternating wake vortices. That leads to the conclusion that a cross-section of different shape may have some advantages over the simple-to-fabricate, circular cylinder. Figures 47.34 and 47.35 show parallel efforts to optimize the cylinder cross-section for the maximum oscillatory lift-to-drag ratio [47.118]. The early conclusion of this research is that there are shapes, which in some reduced velocity range may perform better than a circular cylinder but not over the entire range of velocities of interest. An important factor in this outcome is the difference in the location of vortex shedding as well as the shape of the vortex in the wake of circular and noncircular cylinders as can be seen in comparison of Figs. 47.34–47.36.

Electro-Mechanical Principles

A converter, regardless of its scale and whether used as a model or as a prototype, would not be complete without an integrated controller with the PTO system. As part of this case study, only the advanced features of the system developed in the MRELab are discussed. The PTO is conventional and will not be presented here. Suffice to say that the power generated is either burned in a heat bank or used to light a device for demonstration purposes. The advancement in the MRELab converter is in the V_{ck} (Virtual damping–spring) system,

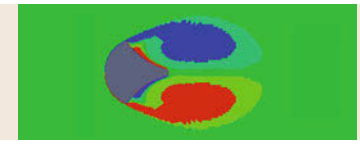


Fig. 47.34 CFD for fish-shaped after-body

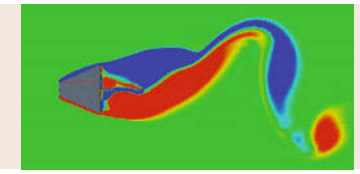


Fig. 47.35 CFD for wedge-shaped fore-body

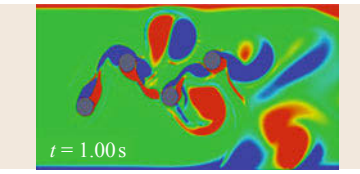


Fig. 47.36 Channel CFD for four circular cylinders in FIM

which enables efficient and fast changing of parameter values with little to no downtime between experiments.

Principle 12 – Virtual c-k Control-Law (V_{ck}). As mentioned in Sect. 47.2.1, *Description of the VIVACE Converter* in spite of the slow nature of oscillations – about a cycle a second – there are 32 million seconds in a year making the fatigue life of springs a challenge. In a laboratory setting, changing physical springs and dampers in order to change values of experimental parameters is time consuming requiring ordering of new physical components, calibrating them, mounting them, and retesting them periodically to ensure that their values have not changed with time. In the MRELab, to overcome this problem, a real-time controller with feedback force $-ky - c\mathrm{d}y/\mathrm{d}t$ emulating spring and damper has been developed [47.156]. It does not include the hydrodynamic forces in the closed loop. That is, it simulates only the mechanical part of the oscillator. The V_{ck} system requires extensive calibration and system identification as well as comparison with real springs and damping devices as shown in Fig. 47.39. V_{ck}, though, enables performing hundreds of tests in a few days and generating power envelopes (Fig. 47.38) by just adjusting the parameters c and k in the feedback force. Figures 47.37 and 47.38 show the results from the development of the first generation of the V_{ck} system [47.97, 98].

The mathematical model of the V_{ck} system and its practical implementation in the MRELab converter are presented next.

Oscillator simulated by the V_{ck} system: V_{ck} is a feedback loop that uses the servomotor encoder to keep track of the position and velocity of the cylinder and provide the necessary torque to simulate the spring stiffness and provide the linear viscous damping required. This logic keeps the hydrodynamics out of the control loop making it possible to simulate the mechanical part of the oscillator only without affecting the hydrodynamic excitation. The V_{ck} end product provides to the experimental setup the flexibility to input the values of spring stiffness and linear viscous damping into the controller and continue the experiments without interruption. This huge advantage comes at the expense of extensive time in V_{ck} system development.

It is critical to compare the accuracy of the V_{ck} system to implement a mechanical oscillator with the damping and spring constants desired. That requires systematic calibration followed by comparison of V_{ck} to an oscillator with real springs and dampers. The process requires performing repeated experiments in free air and in the LTFSW channel of the MRELab for system identification (SI). A nonlinear model for damping has been defined after many trials. The resulting damping model is subtracted by the controller in the V_{ck} system. That results in a zero damping system. Then, a linear viscous damping model is imposed by the controller, thus, producing a perfectly linear viscous damping matching the textbook oscillator models.

The first generation of V_{ck} [47.97] used the National Instruments data acquisition system for reading the position and completing the force feedback. This system was a success and showed improvements in the opera-

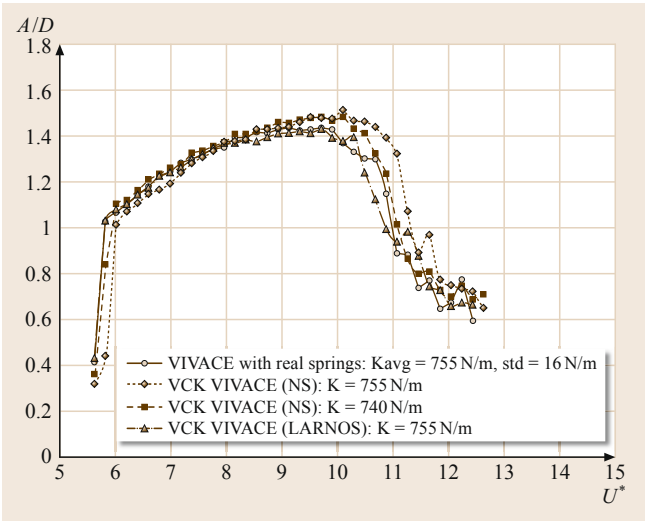


Fig. 47.37 Comparison between real spring damper and V_{ck} experiments for a single cylinder in VIV (after [47.100])

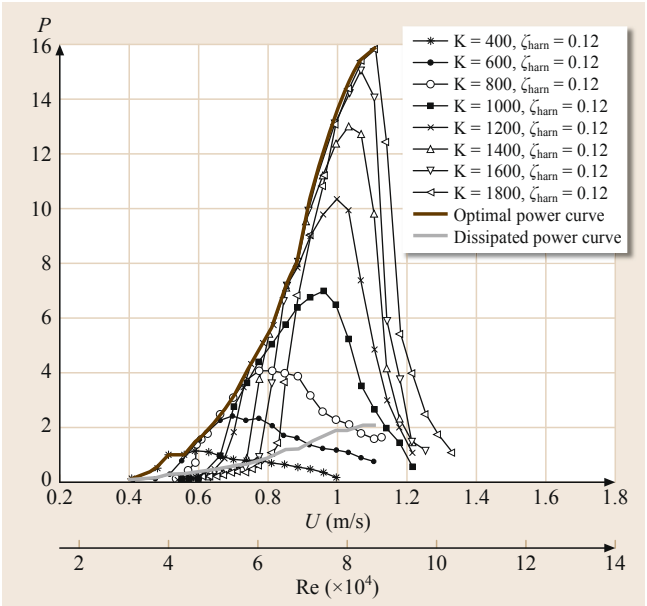


Fig. 47.38 Power envelope of converter with a single smooth cylinder (after [47.100])

tional efficiency but it had its own disadvantages as it introduced a significant amount of phase lag between force feedback and displacement into the system. The phase lag was due to the data acquisition system where the analog-to-digital conversion (ADC) and digital-to-analog conversion (DAC) ran through a digital filter. The phase lag was clocked between 30 and 50 ms, which was not desirable for FIM measurements or energy harnessing. The phase lag was compensated by introducing four dynamic terms with memory to model the hysteresis effects [47.98].

In the second generation of V_{ck}, a controller-embedded virtual spring-damping system (V_{ck})) for FIM experiments was developed for each oscillator [47.142]. The controller simulates the mechanical part of the oscillator by providing the feedback through belts and pulleys (Figs. 47.39 and 47.40). An Arduino-embedded board controls the servomotor, which has an optical encoder. Digital signal is used, making the system more responsive, theoretically with no lag except for that due to the CPU operation clocked to 10 µs. This approach enables real-time position/speed measurement.

Physical V_{ck} VIVACE System. Extensive experimentation has been completed on FIM with PTC with all the bearings and shafts outside the water (Figs. 47.5a, 47.17a, and 47.18a). The end conditions of the experimental oscillator reduced the two-dimensional flow along the cylinder length only by 3%. That is 97% of the length of the cylinder provided lift resulting in FIM. In 2011, in order to investigate further the FIM of cylinders with fewer restrictions, both the channel and the VIVACE converter were rebuilt. This allowed for amplitudes of oscillation reaching 5.5 diameters for $D = 3.5''$ (0.0889 m). The new test models were de-

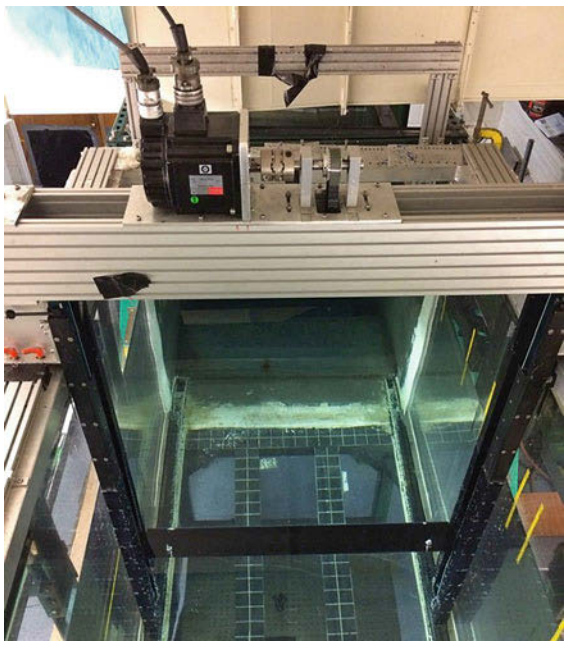


Fig. 47.39 One VIVACE converter with V_{ck} mounted in the LTFSW channel (photo courtesy MRELab)

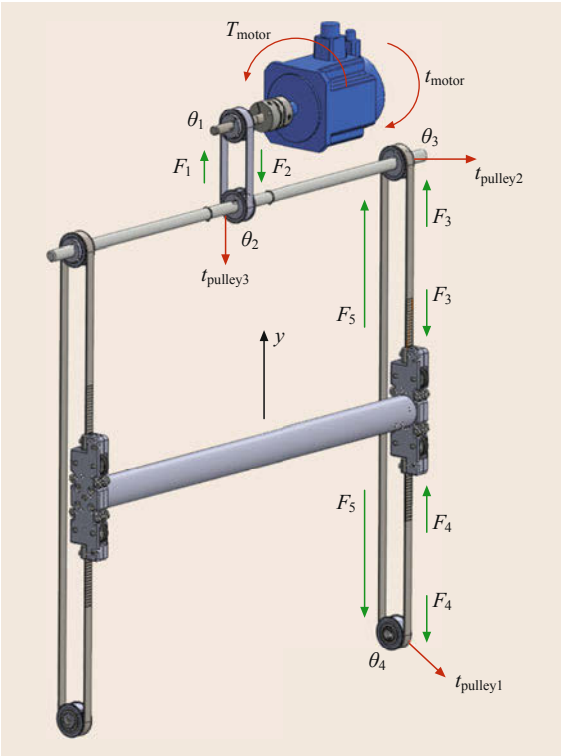


Fig. 47.40 Symmetric V_{ck} model with notation

Table 47.6 Components of the embedded V_{ck} system

θ_i, $i = 1, 2, 3, 4$	Angle of motor coupling and pulley $i = 2$: angle of the pulley top-center $i = 3$: angle of the pulley top-right $i = 4$: angle of the pulley lower right
y	Displacement of the cylinder
M_{osc} [kg]	Equivalent mass of oscillating components
J_{motor} [kg m^2]	Mass moment of inertia of the motor
t_{motor} [N]	Nonlinear damping torque of motor
T_{motor} [Nm]	Torque generated by the motor
J_{pulley} [kg m^2]	Mass moment of inertia of the pulley
$t_{pulley,i}$, $i = 1, 2, 3, 4$	Nonlinear damping torque of i-th pulley
r_{pulley} [m]	Radius of all pulleys
$f_{bearing}$ [N]	Nonlinear damping force of all bearings
F_i [N]	i-th tension of the time belt

signed and built as shown in Figs. 47.14 and 47.15 and were described in Sect. 47.2.1, *Description of the VIVACE Converter*.

Math Model for a V_{ck} VIVACE System. The physical model (Fig. 47.39) is assumed to have structural symmetry and the gravitational force is ignored assuming that the cylinder oscillates around its equilibrium position after the virtual springs and damper have been implemented.

The equations of motion of the new VIVACE system with the embedded V_{ck} are (47.19)–(47.23) where the symbols are defined in Table 47.6 and Fig. 47.40

$$(J_{motor} + J_{pulley})\ddot{\theta}_1 = T_{motor} - t_{motor} + r_{pulley}(F_1 - F_2)\,, \tag{47.19}$$

$$J_{pulley}\ddot{\theta}_2 = -t_{pulley\,3} + r_{pulley}(F_1 - F_2) - 2t_{pulley\,2}\,, \tag{47.20}$$

$$J_{pulley}\ddot{\theta}_3 = 2t_{pulley\,2} + r_{pulley}(F_3 - F_5) - t_{pulley\,3}\,, \tag{47.21}$$

$$J_{pulley}\ddot{\theta}_4 = -t_{pulley\,1} + r_{pulley}(F_5 - F_4)\,, \tag{47.22}$$

$$M_{OSC}\ddot{y} = -f_{bearing} + 2(F_4 - F_3)\,. \tag{47.23}$$

Assuming that the timing belts are inelastic, the kinematic relations among θ_i and y are

$$\theta_1 = \theta_2 = \theta_3 = \theta\,, \tag{47.24}$$

$$y = r_{pulley}\theta\,. \tag{47.25}$$

Thus, (47.19)–(47.23) are simplified to (47.26), by utilizing (47.24) and (47.25)

$$(J_{motor} + 6J_{pulley} + M_{osc}r_{pulley}^2)\ddot{\theta} = T_{motor} - t_{motor} - 6t_{pulley2} - 3t_{pulley3} - 2t_{pulley1} - f_{bearing}r_{pulley}\,. \tag{47.26}$$

$$\begin{aligned}(J_{motor} + 6J_{pulley} + M_{osc}r_{pulley}^2)\ddot{\theta} &= T_{motor} - (t_{motor} - 2t_{pulley} + 6t_{pulley2} - 3t_{pulley3} + f_{bearing}r_{pulley})\,.\\ \left(\frac{J_{motor}}{r_{pulley}} + \frac{6J_{pulley}}{r_{pulley}^2} + M_{OSC}\right)\ddot{y} &= \frac{T_{motor}}{r_{pulley}} - \left(\frac{t_{motor}}{r_{pulley}} + \frac{2t_{pulley}}{r_{pulley}} + \frac{6t_{pulley2}}{r_{pulley}} + \frac{3t_{pulley3}}{r_{pulley}} + f_{bearing}\right)\,.\end{aligned}$$

Assuming that $F_{motor} = T_{motor}/r_{pulley}$, the above equation can be simplified for convenience as

$$m\ddot{y} = F_{motor} - f\,, \tag{47.27}$$

where

$$m = \frac{J_{motor}}{r_{pulley}^2} + \frac{6J_{pulley}}{r_{pulley}^2} + M_{OSC}\,, \tag{47.28}$$

$$f = \frac{t_{motor}}{r_{pulley}} + \frac{2t_{pulley}}{r_{pulley}} + \frac{6t_{pulley\,2}}{r_{pulley}} + \frac{3t_{pulley3}}{r_{pulley}} + f_{bearing}\,. \tag{47.29}$$

Virtual, nonlinear, viscous damping model: First, the system calibration for stiffness k is performed. The results are shown in Fig. 47.40. Next the system identification process can be performed. One of the functions of the V_{ck} system is adding linear adjustable damping to the system to harness MHK energy. To develop the sought power envelope, systematic tests need to be conducted. Prior to adding prescribed damping, the physical system damping needs to be identified and subtracted. The latter is due to all moving parts, especially the timing belts and pulleys, which bring undesirable/excessive amount of nonlinear damping to V_{ck}. The mechanical oscillator equation of the V_{ck} system is

$$m\ddot{y} + c(\dot{y}) + ky = F_{total}\,, \tag{47.30}$$

where F_{total} is the total hydrodynamic transverse force acting on the cylinder. Thus, the damping model can be obtained through

$$c(\dot{y}) = F_{total} - m\ddot{y} - ky\,. \tag{47.31}$$

In order to find the nonlinear viscous damping and static friction, the total damping force f, which contains motor, pulleys, and bearings, can be expressed as

$$f = f_{motor} + f_{pulley}f_{bearing} \tag{47.32}$$

which can also be expressed as a function of dy/dt. Here, all the damping force is assumed to be be modeled adequately by a third-order polynomial as

$$c(\dot{y}) = c_3\dot{y}^3 + c_2\dot{y}^2 + c_1\dot{y}^1 + c_c\,\mathrm{sign}(\dot{y})\,. \tag{47.33}$$

c_1, c_2, and c_3 are the first, second, and third-order coefficients, respectively, and c_c, is the Coulomb friction coefficient.

The first generation of V_{ck} system [47.98] had a closed loop of the vibration components and the cylinder could be detached from the belt. Constant forces were applied continuously to the timing belt to achieve sufficient speed. The current V_{ck} system has sliding blocks and timing belts inside the structure and, thus, they cannot be disconnected (Fig. 47.14a). To achieve sufficient speed and acceleration, various single direction forces are fed to the system by the mo-

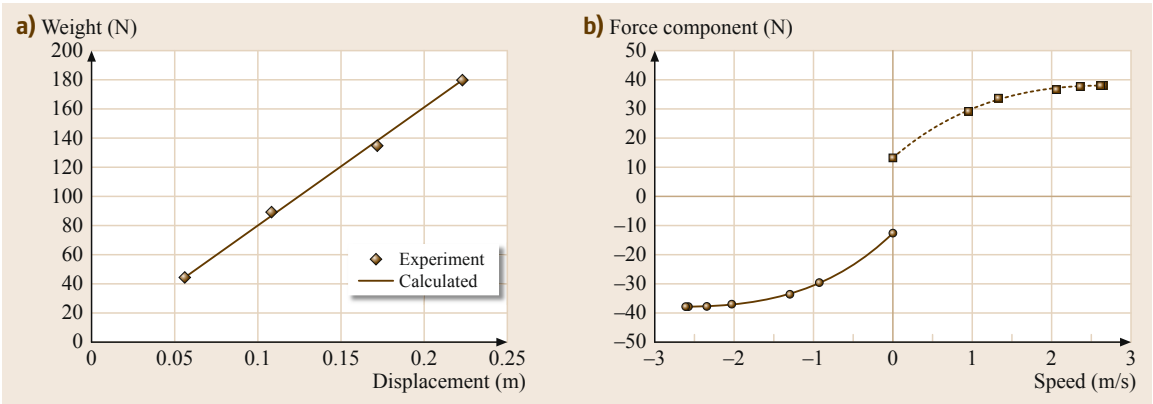


Fig. 47.41a,b System identification of second generation V_{ck} system (after [47.142]). **(a)** Virtual spring test curve fitting. **(b)** Nonlinear static damping

tor during the system-identification process [47.142]. Programmed safety stoppers are activated when the cylinder reaches the upper and lower ends of the travel to prevent collision with the plexiglass bottom of the LTFSW channel. The regression equations derived from the system identification process are (Fig. 47.41b)

$$c(\dot{y}) = 0.8793\dot{y}^3 - 7.7678\dot{y}^2 + 23.736\dot{y} + 13.439, \quad \text{for } \dot{y} > 0\,, \tag{47.34}$$

$$c(\dot{y}) = 1.1544\dot{y}^3 + 9.2064\dot{y}^2 + 25.992\dot{y} - 12.484, \quad \text{for } \dot{y} < 0\,. \tag{47.35}$$

Controller design: Based on this nonlinear viscous damping model, the controller force can be expressed as

$$F_{\text{controller}} = F_{\text{nonlinear damping}} - c_{\text{virtual}}\dot{y} - k_{\text{virtual}}y\,, \tag{47.36}$$

where, $F_{\text{controller}}$ is the force applied by the virtual system, F_{fluid} is the total hydrodynamic force acting on the cylinder, k_{virtual} is the virtual spring constant, c_{virtual} is the virtual linear viscous damping. c_{virtual} is added after subtracting the nonlinear viscous damping defined by (47.34) and (47.35). That results in a textbook case of linear viscous damping.

It should be noted that following this procedure, the V_{ck} system can be readily used to apply nonlinear springs functions or nonlinear damping.

The major conclusions that can be drawn from the analysis of underlying principles of the case study of the VIVACE converter are the following:

- Enhancing naturally occurring FIMs and altering their onset through PTC have made it possible to develop an ALT that has a high-response amplitude operator starting at low flow speeds and having no upper limit.

47.3 Methodology and Tools in Support of Development

The MRELab has pioneered several areas of research in support of the development of ALT and particularly the VIVACE converter. The outcome is implemented immediately in the design and field tests conducted by VHE for the product development. Research conducted in several labs worldwide citing the pioneering work of the MRELab and further developing these areas has been most valuable in this pursuit; *Vandiver* [47.157], *Konstantinidis* [47.158], *Vinod* and *Banerjee* [47.159], *Huynh* et al. [47.160]. These research areas include:

1. Energy harnessing by enhancing FIMs using alternating lift like fish, rather than steady lift like birds [47.48, 49].
2. Use of cylinder interference to improve synergistic operation of multicylinder arrays. It has been shown in the lab, for up to two/three/four cylinders, that an array of cylinders can actually harness more energy when cylinders work synergistically rather than in isolation (Figs. 47.28 and 47.29a) [47.68, 69, 77].
3. Cylinder cross-section shape optimization to maximize their oscillatory lift-to-drag ratio. This is the counterpart research conducted for decades on steady lift-to-drag ratio for airfoils (Figs. 47.33 and 47.34) [47.118].
4. Achieve three-dimensional distribution of cylinders in a flow so that the VIVACE converter may harnesses energy from the entire flow domain rather

than being a point absorber like buoys, line absorber like Pelamis, surface absorber like oscillating water columns, or area absorber like turbines and watermills.
5. Establish nonlinear spring constant for cylinder motion reversal and energy harnessing (Figs. 47.17b and 47.19a).
6. Build a virtual damping–spring system, V_{ck}, that does not include the hydrodynamic force in the loop to enable fast experimental change of parameters for testing [47.98, 142].
7. Implement different nonlinear models of damping and spring stiffness using the virtual system V_{ck} (Figs. 47.40 and 47.41).

To be effective in all these areas needed in support of ALT development, several research tools are used and/or being developed by MRELab and VHE. These include experimental facilities, flow visualization, field tests, mathematical models, and CFD codes. These are described briefly in the following sections.

47.3.1 Experimental Facilities

The experimental facilities used in the development of the VIVACE converter include two towing tanks and two recirculating channels providing complementary information through testing and data processing.

First Low-Turbulence Free-Surface Water (LTFSW) Channel

FIM experiments were conducted in the first LTFSW channel between 2005 and 2012 in support of the VIVACE converter development. Some of the outcomes and limitations of research and testing in this first channel are summarized below.

The first LTFSW channel was built in 1992 that proved to be a very valuable facility till its decommission in 2012. The purpose of the LTFSW channel was to facilitate the study of the fundamental structural aspects of turbulent flows near boundaries (solid walls, free surfaces, or both). The two-story high water channel was recirculating approximately 8000 gal of water. Its maximum flow speed was 2 ms^{-1}. The test section was 2.44 m long, 1 m across and 0.8 m deep. All walls of the test section were acrylic to allow for flow visualization experiments and to facilitate measurements with optical instrumentation. The measured background turbulence level at the test section was less than 0.1% of the free-stream velocity. Velocity and turbulence measurements are accomplished by utilizing a three-component, fiber optic, laser-Doppler velocimetry (LDV) system. The LDV system was specifically designed to allow for simultaneous laser-induced fluorescence concentration measurements. A three-axis traverse system allows movement of the LDV crossing point across the test section [47.161].

The tests conducted in this first LTFSW channel were FIM tests, not VIVACE model tests. The qualifying difference is very important. Specifically, in the FIM tests, only the cylinder and the supporting struts are submerged making it possible to bring the cylinder very close to the channel walls, thus, limiting the tip flow (three dimensionality) effects to 3% of the lift force (Fig. 47.17a). In VIVACE model tests, conducted in the second LTFSW channel (Sect. 47.3.1, *Second Low Turbulence Free Surface Water (LTFSW) Channel*) roller carts, bearings, belts, springs, and safety stops are placed in the channel creating a strong tip-flow effect of 43% (Fig. 47.17c). This is further discussed in Sect. 47.3.8. As a result, neither FIM test results can be used for VIVACE predictions nor VIVACE model-test measurements can be used as pure FIM experimental results. The mathematical post-processing for the reduction of data to a common equivalent has been developed in the MRELab and is introduced in Sect. 47.3.8. It will appear in a journal publication shortly after this work is published.

The first contribution of FIM testing by MRELab in the LTFSW channel has been tested in the TrSL3 flow regime [47.48, 49, 67, 107]. Typical VIV tests until then were conducted in the TrSL2 flow regime [47.55, 162–164]. The resulting response amplitude operators differ dramatically as shown in Fig. 47.30 for TrSL2 and Fig. 47.31 for TrSL3.

There are several fundamental differences in the flow dynamics past a circular cylinder between the TrSL3 ($1000 < Re < 300\,000$) and TrSL2 ($1000 < Re < 20\,000$) flow regimes as defined by *Zdravkovich* [47.56]. Most notably, the shear layers on either side of the cylinder are saturated in TrSL3 carrying more vorticity to the von Kármán vortices. As a result, the shear layers have a stronger roll-up much closer to the cylinder. The circulation in the von Kármán vortices reaches shedding strength faster based on the fact that the Strouhal number (47.1) is about constant ($St = 0.20 - 0.21$) practically over the entire laminar flow regime for $400 < Re < 300\,000$. Thus, the frequency of vortex shedding increases proportionally to the velocity. Further, the formation length is shorter [47.165]. At the moment of shedding, the combination of a shorter formation length with strong roll-up and higher circulation closer to the afterbody of the cylinder cross-section induces higher lift force. Thus, as shown in Fig. 47.16, the lift coefficient in TrSL3 is several times higher than that in the TrSL2 regime.

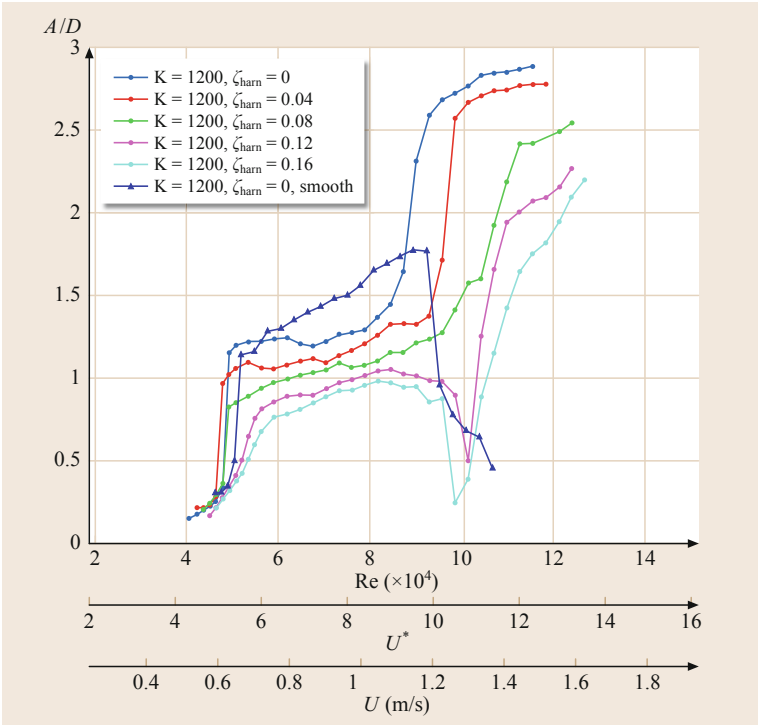


Fig. 47.42 The effect of Reynolds number is strong in the TrSL3 flow regime (after [47.105])

The following observations can be made comparing the response amplitude operators in TrSL2 (Fig. 47.30) and TrSL3 (Fig. 47.31):

1. Both exhibit the initial branch, the upper branch, the lower branch, and desynchronization.
2. In TrSL2, the upper branch shows constant response around $A/D = 1$, which led for years to the incorrect conclusion that VIV response depends only on the reduced velocity U^* (47.3) and not on Reynolds number.
3. In TrSL3, the upper branch constantly increases and overtakes the lower branch. The response A/D starts around 1.2 and increases up to 1.9 depending on the Reynolds number as shown in Figs. 47.41b, 47.42, 47.43, and 47.85 [47.48, 49, 73, 105]. This high response and Reynolds dependence were confirmed by a single data point at even higher Reynolds number, around 10^6, by *Ding* et al. [47.148].
4. Placing the MRELab measurements [47.48, 49] on the modified *Griffin* [47.166] plot shows that the extrapolation performed by *Govardhan* and *Williamson* [47.167] does not capture correctly the dependence of VIV on Reynolds number.
5. The measurements performed by *Bernitsas* et al. [47.48, 49, 67, 70, 105] show that VIV can be maintained in the TrSL3 flow regime even when values of damping are high for energy harnessing (Fig. 47.44).

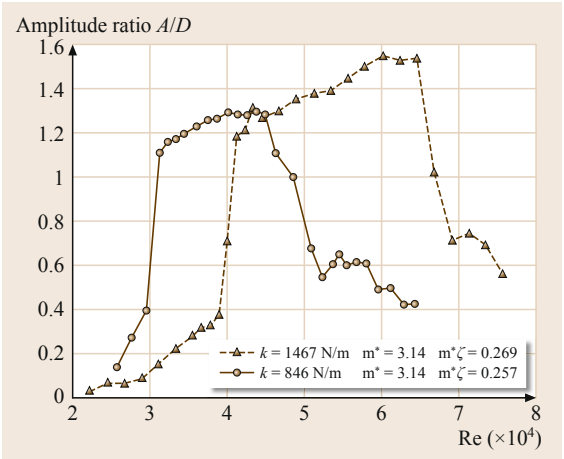


Fig. 47.43 Re has a stronger influence than m^* and $m^*\zeta$ on A/D (after [47.105])

Enhancing FIM through PTC: This was the second major contribution of the MRELab tests in the first LTFSW channel. Enhancement of FIM was introduced in Principle 6 and PTC in Principle 7 in Sect. 47.2. The PTC-to-FIM map was presented in Fig. 47.24 along with its robustness and showed how the motion of a cylinder can be enhanced or partially suppressed us-

ing PTC. Suffice to mention here that the ways in which our understanding of PTC have been implemented to improve the performance of this ALT converter:

1. Figures 47.45 and 47.46, when combined with the boundary-layer thickness in Fig. 47.25 provide guidance on the size of the PTC to be used for a specific size cylinder.
2. PTC has the following effects on VIV: It starts the TrSL3 flow regime earlier with the turbulators; it maintains the TrSL3 regime with the turbulators longer; it eliminates the transition region, thus, maintaining VIV without allowing its suppression in transition from laminar to turbulent boundary layer as is the case with smooth cylinders.
3. PTC increases the synchronization range in VIV.
4. PTC partially suppresses VIV amplitude. No significant influence from PTC roughness was found. For larger PTC coverage area, more amplitude reduction is observed.
5. PTC causes higher oscillation frequency in the VIV synchronization range as it induces more vorticity.
6. PTC induces galloping at smaller velocities, thus, enabling back-to-back VIV and galloping as shown in Fig. 47.32, which is the major advantage of the VIVACE converter, that is, an open-ended large amplitude response amplitude operator.
7. Galloping instability depends on PTC location mostly. PTC is more effective in galloping when placed at α_{PTC} of 20–64° (Fig. 47.23).
8. Effective PTC thickness for galloping is $T \geq 0.19$ mm.
9. Rougher PTC surface and larger PTC coverage promote early fully developed galloping. Surface roughness enhances maximum amplitude in galloping branch.

Blockage effects are a concern with all the experiments and thus the following assessment is offered. In both LTFSW channels, most experiments were conducted with a smooth circular cylinder or a PTC circular cylinder made of anodized aluminum with diameter $D = 3.5''$ (0.0889 m) and length $L = 36''$ (0.914 m). The corresponding aspect ratio (L/D) is 10.29. Tests are subject to blockage, free surface, and bottom effects. The ratio of cross-section of cylinder over water depth was about 12% ($D/d = 3.5/29$) for the first channel and about 7% for the second channel (Sect. 47.3.1, *Second Low-Turbulence Free-Surface Water (LTFSW) Channel*. The cylinder was suspended by two tension/compression coil springs and attached to the supporting struts. The test cylinder was constrained to move in the vertical direction (perpendicular

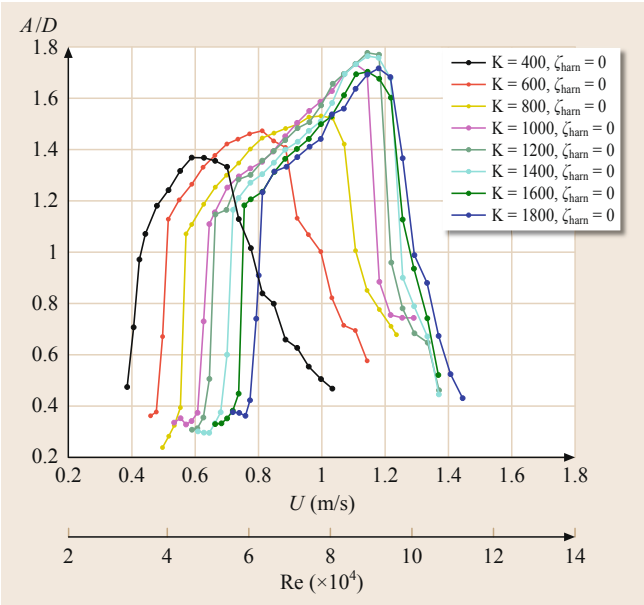


Fig. 47.44 VIV synchronization vs Reynolds number shows an increase of A/D with Reynolds number in the TrSL3 and TrBL regimes (after [47.100])

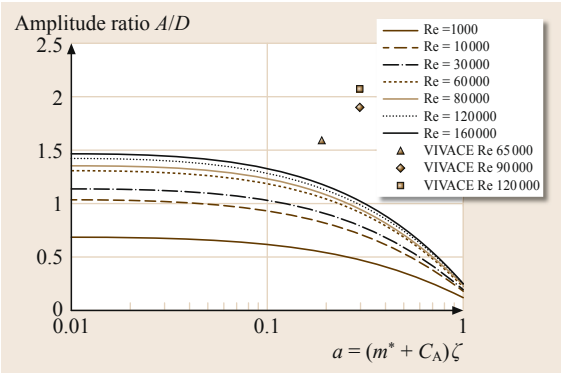


Fig. 47.45 Modified Griffin plot (after [47.167]) with VIVACE tests (after [47.49, 107]). TrSL3 tests revealed much stronger Re number dependence than predictions by extrapolation

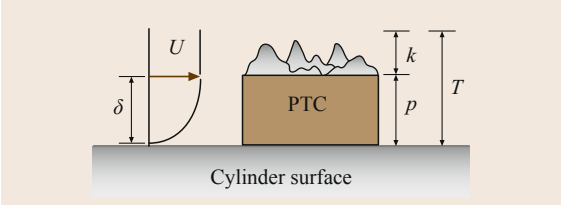


Fig. 47.46 Definition of PTC roughness strip parameters

to the flow direction) as shown in Fig. 47.17a. The ratio of cylinder length L over channel width w is nearly 1 ($L/w = 36/38$). All the experiments were conducted in the Reynolds number (Re=UD/ v) range of 10 000–118 000.

Four different boundary effects need to be discussed:

1. Side-to-side blockage
2. Top-to-bottom blockage
3. Free-surface effect
4. Bottom-boundary effect.

Boundary effects are expected to influence amplitude response particularly when the cylinder approaches the free surface or bottom boundary. There is limited information in the literature. For a stationary circular cylinder, with increase in blockage ratio but in the absence of free surface, vortices possibly become stronger [47.56] and span-wise correlation increases [47.168]. Both these factors, that is, increase in vortex strength and span-wise correlation of vortex shedding, are conducive for enhancement of lift force and consequent higher amplitudes of oscillation. In the FIM channel experiments, the testing arrangement was very different. Side-to-side blockage effects were eliminated by taking special care to maintain a two-dimensional flow side-to-side (Sect. 47.3.8, *Tip-Flow Effect*). The cylinder along with its end brackets occupied practically the entire width of the channel. Thus, there was no room for flow acceleration due to side-to-side restrictions. Hence, in the present case, boundary effects from flow restrictions were limited only to the vertical direction.

For nonstationary, smooth cylinders, free-surface effects have been studied by *Raghavan* and *Bernitsas* [47.73] and bottom-boundary effects by *Roshko* and *Chattoorgoon* [47.169] and *Raghavan* et al. [47.108–110]. Typically, free surface and bottom boundary partially suppress VIV. For the cylinders with PTC used, these effects can be surmised only by comparison to the smooth cylinder effects as presented in these references and by observing change in behavior as the cylinder approaches the boundaries. Corrections for free surface and bottom boundary were not carried out.

When the cylinder comes close to a wall, significant lift force is produced repelling the cylinder away from the wall according to the results by *Rosko* and *Chattoorgoon* [47.169]. Hence, applying that conclusion to this study, this implies a negative hydrodynamic force tending to reduce the amplitude of oscillation of the cylinder. This agrees with the results by *Raghavan* et al. [47.109, 110] confirming the effect of bottom boundary suppressing the oscillatory amplitudes.

In galloping, there will always be blockage in a channel since there is no limit to galloping amplitude. Even when there is no hydrodynamic blockage in very deep channels, there will always be mechanical constraints since galloping stops only with structural failure. Galloping is not self-limiting like VIV. This blockage is observed in Fig. 47.32, where the amplitude of oscillation reaches a plateau; actually at a different U^* depending on the system damping. Until that point, the results are not affected by blockage and, in an unbounded flow, galloping amplitude will increase until structural failure.

The following observations are useful. In the tests by *Chang* et al. [47.67], using smooth cylinder and thus being subjected only to VIV – not galloping $-A^* = A/D$ reached 1.65, which allows a gap ratio (ratio of gap between the top of the cylinder surface and the free surface) of about 2.0. For this gap ratio, results by *Raghavan* and *Bernitsas* [47.73] show that free-surface effects could bring an amplitude reduction for smooth cylinder in the range of 8–11%. It should be specifically noted that, for the corresponding stationary smooth cylinder, located midway between the free surface and the bottom boundary when mounted, the gap ratio comes to about 3.5. At that value, no free-surface effects are expected [47.170]. Furthermore, for the same gap ratio (2.0) with respect to the bottom boundary of the channel, there is an amplitude reduction of about 20% [47.108–110]. So, effects due to bottom boundary and free surface may have jointly contributed to an amplitude reduction of up to 30% for the smooth cylinder. In other words, without these effects, amplitude would have been up to 30% higher for smooth cylinders.

Second Low-Turbulence Free-Surface Water (LTFSW) Channel

After several years of testing in the first LTFSW channel, researching a variety of new phenomena regarding enhancement of FIM and synergistic operation between cylinders in arrays of up to four cylinders in tandem, it became clear that there were limitations in amplitude of oscillation in galloping. Thus, the channel was decommissioned and rebuilt out of stainless steel with bigger depth allowing for A/D of up to 5–7. At the same time, MRELab turned its attention to VIVACE tests as opposed to FIM tests. The relation between these two types of tests is discussed in Sect. 47.3.8.

The second LTFSW channel was built in 2012–2013. All stainless steel sections of the first two-story channel [47.161] were maintained and adapted to the new design. The section of the channel above the working level (second story) made originally by PVC was replaced by stainless steel and was made deeper. The depth increased from 80 to 140 cm allowing up

Table 47.7 Dimensions of PTC roughness strips. For $\mathrm{Re} > 7\times10^4$, $T/\delta > 1$ (after [47.102])

	P60	P80	P120	P150
K: Avg. grit size [mm]	0.26	0.195	0.127	0.097
$K_{max} \approx K_{min}$ [mm]	$0.416 \approx 0.175$	$0.302 \approx 0.112$	$0.175 \approx 0.60$	$0.150 \approx 0.48$
p: thickness of sieve [mm]	0.587	0.554	0.454	0.492
T: total thickness [mm]	0.847	0.749	0.581	0.589
k/D (10^{-3})	2.92	2.19	1.42	1.09
T/D (10^{-3})	9.53	8.43	6.40	6.63

to 7 diameters amplitude, depending on the design of the oscillator model. The recirculating water increased from 8000 to 10 000 gal. The impeller and motor were maintained resulting in a reduced maximum velocity (from 2 to about 1.5 m s^{-1}) for the same level of turbulence intensity. Additional vanes were installed to reduce air entrapment at high speeds. On the other hand, increased speed is used and air entrapment is allowed when realistic conditions are tested for converter models. Figure. 47.17c and Table 47.7 show pictures of the test section of the new channel. Systematic tests are being conducted with up to four oscillators in synergistic operation. The objective of the MRELab is to optimize the multicylinder design of converters with systematic tests enabled by the V_{ck} systems [47.142].

The VIVACE models for the new channel were designed and built in 2012. They are described in detail as shown in Sect. 47.2.1, *Description of the VIVACE Converter* (Figs. 47.17c, 47.38, 47.39, 47.47).

Towing Tank at the University of Michigan

The need to test bigger models used in field tests, measure FIM response with no blockage effects, conduct measurements at larger Reynolds numbers and particularly in flow transition which suppresses VIV, led to experimentation in two towing tanks starting with the one at the University of Michigan, Ann Arbor.

The tank is a carriage-drive system with power provided by four 5 kW brushless servo motors using computer control for optimum speed regulation. Tank length is 109.7 m (360.0 ft), width is 6.7 m (22 ft), and depth (to edge of trough) is 3.05 m (10.0 ft). Typical water depth is 3.2 m (10.5 ft). Carriage type is a manned bridge with an unmanned trailer. Carriage speed is 0.08–6.10 m s^{-1} (0.25–20.0 ft s^{-1}).

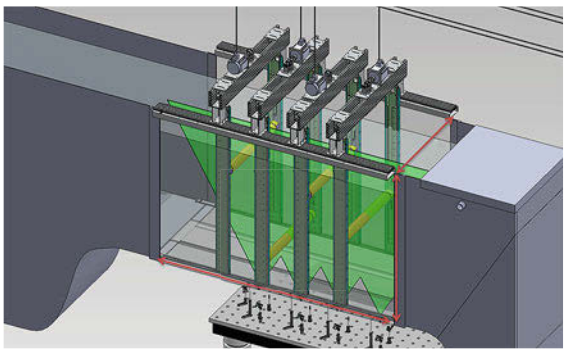


Fig. 47.47 Schematic of the mew LTFSW channel with four oscillators and laser visualization

Three different prototypes were tested in towing tank:

1. The prototype deployed in the St. Clair river in August 2010 consisting of two, horizontal, PTC cylinders with $D = 10''$ (0.254 m) and $L = 106''$ (2.70 m) shown during testing in the towing tank in Fig. 47.18a, in field deployment in Fig. 47.18b, and in river operation.
2. The prototype deployed in the St. Clair river in September–November of 2012 consisting of one, vertical, cylinder with $D = 8''$ (0.203 m) and $L = 60''$ (1.52 m) shown during testing in the towing tank in Fig. 47.17b, at the OHMSETT towing tank in Fig. 47.19a, in field deployment in Fig. 47.19b, and in river operation.
3. The small prototype built for canals in the Netherlands consisting of one, vertical, short cylinder with $D = 10''$ (0.254 m) and $L = 39.4''$ (1.00 m) shown during model tests at the University of Michigan in Fig. 47.20a, and deployment in Fig. 47.20b.

Flow Transition. Testing of larger models and prototypes is required in the development of any marine structure. For MHK energy devices, a procedure of gradually increasing size models and testing components is described by *Bahaj*et al. [47.144]. Figure 47.48 shows results of four towing tank tests at the University of Michigan with two different prototypes both of length $L = 106''$ (2.70 m) and with diameters $D = 10''$ (0.254 m) and $D = 6''$ (0.152 m). The objective of those tests performed over a period of 8 months (2009–2010) was to advance the design of the converter to maintain FIM though flow transition from laminar to turbulent. The amplitude gradually increased with each test. The first test was performed with the $D = 10''$ (0.254 m) smooth cylinder and VIV shut down at the onset of transition. In the second test, PTC was added to the same cylinder initiating VIV earlier, inducing higher response in VIV, and triggering galloping before the onset

Fig. 47.48 Two oscillators in the new LTFSW channel (photo courtesy of MRELab)

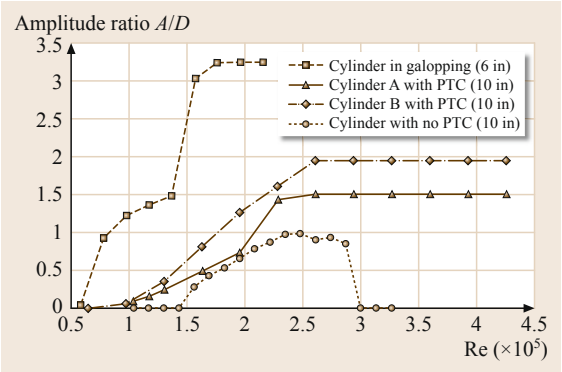


Fig. 47.49 Transition from laminar to turbulent boundary layer shuts down VIV at $\mathrm{Re} = 3 \times 10^5$. With turbulence stimulation placed as per Fig. 47.24, galloping was initiated prior to VIV desynchronization (courtesy of VHE)

of flow transition. The safety stops set at $A/D = 1.5''$ were hit through $\mathrm{Re} \cong 500\,000$. In the third test, the converter frame was redesigned to allow $A/D = 1.9$ for the $D = 10''$ (0.254 m) PTC cylinder. Due to towing tank limitations, the frame could not be redesigned to allow for higher A/D for the $D = 10''$ (0.254 m) cylinder. Thus, in the fourth test, the $D = 6''$ (0.152 m) cylinder was used resulting in the earlier onset of galloping in the middle of the upper branch and amplitudes limited again by the safety stops at $A/D = 3.25$. A similar effect is shown in Fig. 47.49 for a $3.5''$ diameter cylinder in the first LTFSW channel.

Amplitude Limits. Figure 47.48 made it clear that it would be important to conduct tests allowing for yet higher amplitude of oscillation. This led to testing of a smaller converter model in the towing tank of the University of Michigan (Fig. 47.17b). Amplitudes exceeded 4.5 diameters.

Towing Tank at OHMSETT

Limitation in amplitudes as well as concerns about blockage effects led VHE and MRELab to conduct testing in the towing tank of OHMSETT [47.147], which is a much larger, sea-water facility.

OHMSETT's above ground concrete test tank is one of the largest of its kind, measuring 203 m long by 20 m wide by 3.4 m deep. The tank is filled with 2.6 million gallons of crystal-clear saltwater. The OHMSETT test tank allows testing of full-scale equipment. The tank's wave generator creates realistic sea environments, while state-of-the-art data collection and video systems record test results.

In this tank, testing of vertical cylinders without horizontal or vertical blockage effects was made possible. Measurements conducted in the LTFSW channel and the towing tank of the University of Michigan were validated by comparison.

Linear or Nonlinear Springs or no Springs at all? The ever-increasing amplitude of oscillation recorded in FIM tests by MRELab and VHE resulted in continuously higher energy generation. This observation naturally led to the conclusion that a cylinder in galloping could convert more MHK energy to mechanical if it were allowed to move in one direction without frame and/or facility limitations. Several nonlinear springs were tested in 2009–2010, leading to the surprising conclusion that the optimal arrangement was for no springs at all. Short and stiff jounce bumpers were introduced at the end of the frame to reverse the direction of oscillation.

The major conclusions to be drawn from this discussion on the tests conducted in two channels and two towing tanks with various size models and prototypes of the VIVACE converter are the following:

- In the development of novel MHK-energy converters, as with all marine structures, model tests need to be conducted at several scales providing complementary information on the system performance.
- Among MHK-energy converters, an advantage that ALTs may have is that the underlying phenomena may be widely scalable and, thus, smaller models are prototypes of a smaller scale. This makes it possible to design multiple-scale converters in the process of developing a larger scale commercial converter.

- A V_{ck} (virtual damping–spring) system, which does not involve the hydrodynamic force in the loop, enables systematic change of the spring not only in value but also in functional form. The optimization process complexity increases dramatically as it becomes the counterpart of a variational calculus problem rather than a size optimization problem.
- In the design of the VIVACE converter, it has already been established that the optimal energy harnessing ALT requires strongly nonlinear springs.

47.3.2 Damping Model

The system identification process developed by *Lee* et al. [47.98] and *Sun* et al. [47.53] has established that the classical, linear viscous damping model does not represent reality in actual real-life oscillators. Consequently, comparison between experiments, even at the same scale, is of questionable validity. The V_{ck} simulator of the mechanical oscillator, which keeps the hydrodynamic force out of the controller loop, enables performing experiments even with high damping with mathematically accurate damping functional form and value. As explained in Sect. 47.2.3, *Electro-Mechanical Principals* under Principle 12, the following elaborate system identification, the system-damping model is removed from the oscillator making the system-damping zero before damping of specific form and value is added. That process, albeit tedious, enables accurate modeling of damping, systematic variation of damping values for a given damping model, and comparison of experimental results with different oscillators.

It is the author's opinion that the reduction of experimental results to a common equivalent even at the same scale is important for comparison of experiments from different laboratories. The process is not completed by the damping modeling and implementation established by the V_{ck} system by the MRELab. It is further discussed in Sect. 47.3.8.

The damping model of the VIVACE converter models established in Sect. 47.2.3, *Electro-Mechanical Principals* is specific to the oscillators used presently in the LTFSW channel experiments. Every oscillator tested in the MRELab undergoes this procedure. Then, all system damping is removed, before a specific damping model is implemented by the controller. The system-damping model identified by *Sun* et al. [47.142], given by (47.33) is

$$c(\dot{y}) = c_3\dot{y}^3 + c_2\dot{y}^2 + c_1\dot{y}^1 + c_c\text{sign}(\dot{y}) ,$$

where c_1, c_2, and c_3 are the first-, second-, and third-order coefficients, respectively, and c_c, is the Coulomb

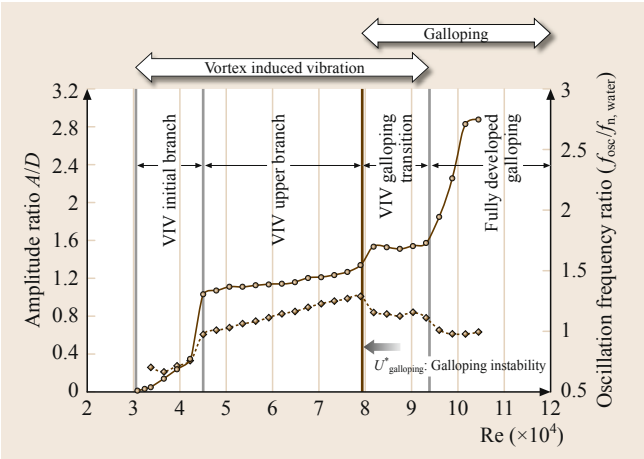


Fig. 47.50 Effect of turbulence stimulation on VIV and galloping using roughness strips (after [47.141])

friction coefficient. This model is generic enough but does not cover all systems encountered in testing in the MRELab. *Lee* et al. [47.98] identified a more complex damping system with hysteresis modeled with four dynamic terms; that is a total of eight coefficients. In both models, the values of the coefficients were different in the two halves of the periodic cycle; that is, when the oscillator was moving in one direction or the other. This is expressed by (47.34) and (47.35).

The following conclusions are worth noting from the preceding discussion on damping:

- A V_{ck} (virtual damping–spring) system, which does not involve the hydrodynamic force in the loop,

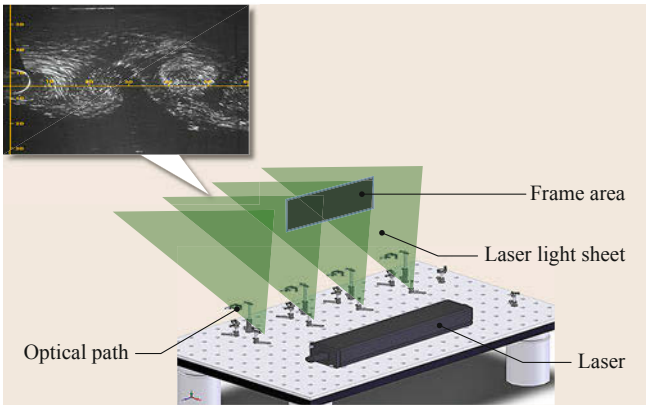


Fig. 47.51 Laser visualization arrangement to span the wake of all four oscillators. Lighting source: argon Laser, continuous laser. Maximum power: 5 W. Particles: AlO_2, ($D_{avg} = 100$ nm), fluorescent particles. Digital camera: Kodak 1000×1050, video; and Imprex 1500×3000, video. Nikon: 1800×2400, snap-shot image (after [47.104])

enables elimination of the system damping and implementation of a mathematically accurate damping model. The actual process has shown that damping varies not only from model to model even in the same laboratory, but also in a given period depending on the location of the oscillator. This is neglected in the MRELab system identification process [47.142].

- Implementation of a specific damping model with mathematical/controller accuracy is a solid step toward reduction of experimental data to a common equivalent model that would enable data comparison across models and labs at the same hydrodynamic scale.

47.3.3 Flow Visualization and Vortex Tracking

Amplitude and frequency response are not the only important information in FIM testing. Vortex shedding, timing and phase of vortex shedding, vortex structures, vortex strength, shear layer motion, and vortex formation length are also important at the scale of the cylinder diameter.

Broad-field visualization of the FIM properties listed above, at the cylinder diameter scale, is achieved as explained in Fig. 47.51. The result is vortex tracking images like those shown in Figs. 47.51–47.54, which show clearly the information needed at this scale in

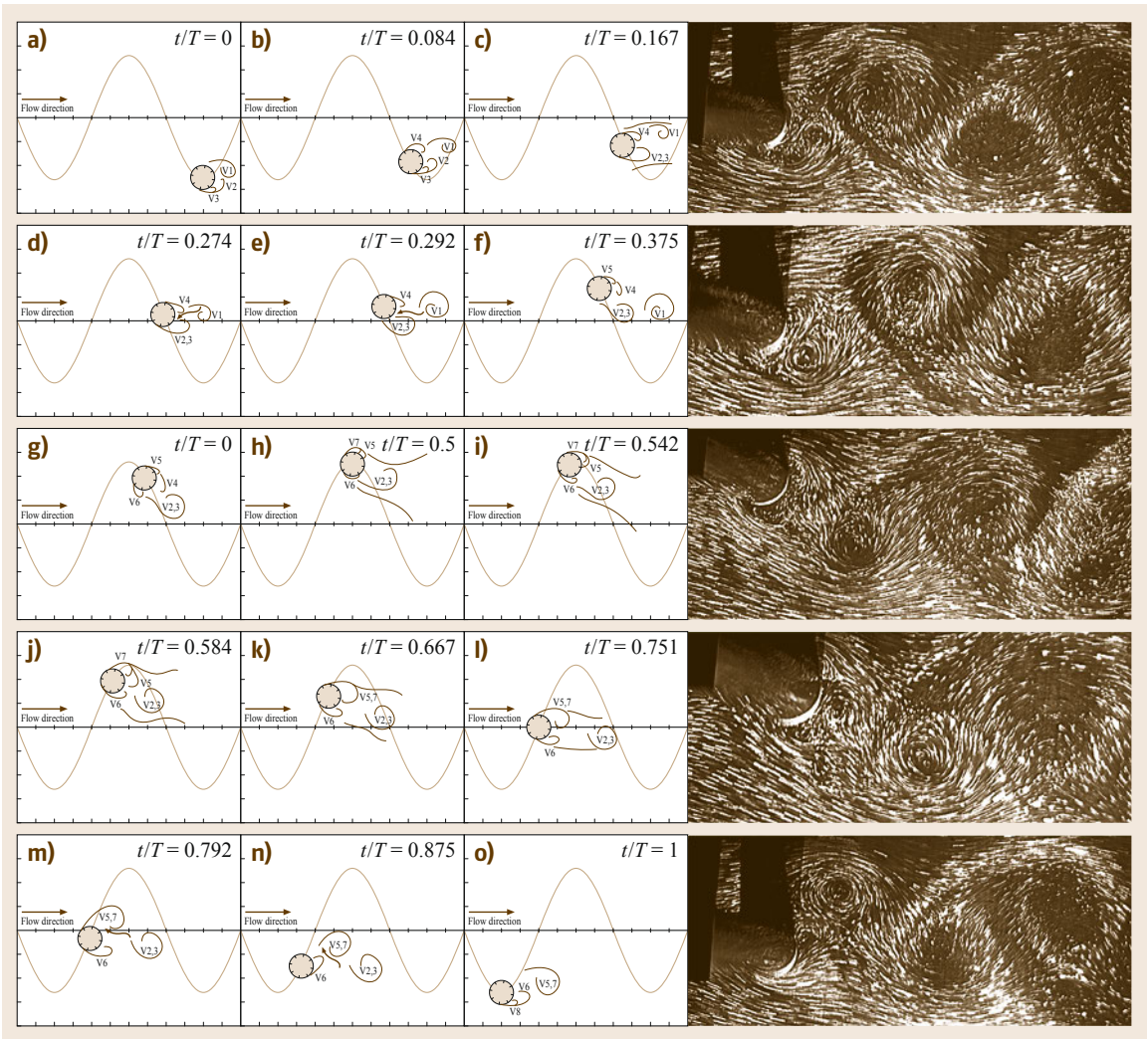


Fig. 47.52a–o Wake vortex structure of smooth cylinder over a cycle: $U^* = 5.58$ (upper end of initial branch), Re$= 4.33 \times 10^4$, $A^* = 0.404$, $f_{osc}/f_s = 0.64$. (after [47.104])

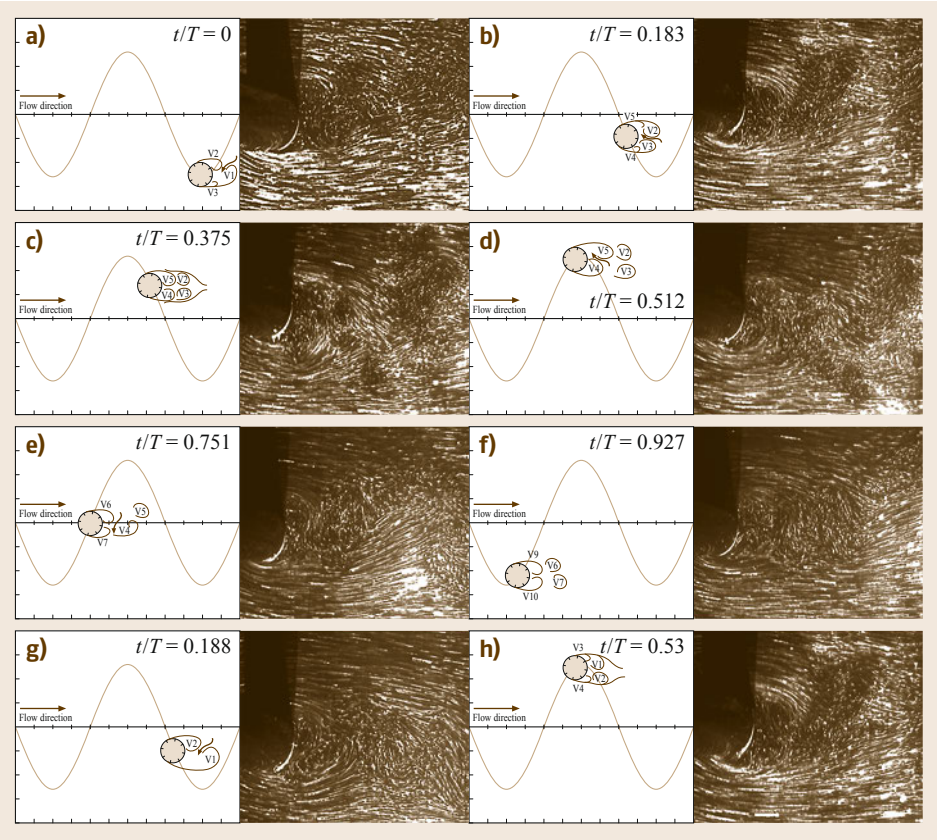


Fig. 47.53a–h Wake vortex structure of smooth cylinder in lower branch, $U^* = 12.28$, $\mathrm{Re} = 9.53 \times 10^4$, $A^* = 0.76$. Sketches **(a–f)** show the cycle with seven shed vortices. Variations exist. Sketch **(g)** is a variation of **(b)** and sketch **(h)** is a variation of **(d)** (after [47.104])

three different tests. Nevertheless, tedious post processing is needed to track vortices and reveal properties of vortex structures in the cylinder wake.

Figure 47.52 shows vortex tracking over a complete period for $U^* = 5.58$, which is close to the end of the upper branch. Figure 47.53 shows vortex tracking over more than a cycle for $U^* = 12.28$ which is in the lower branch. Figure 47.54 shows typical wake structures in desynchronization, for $U^* = 12.28$, which does not form von Kármán vortices. Finally, Fig. 47.55 shows how PTC can suppress VIV creating a narrow escape path for vortices. The cylinder plus its immediate wake form a hydrofoil shape which does not generate a von Kármán street and therefore it does not induce VIV.

Boundary-layer scale information is also important in FIM, when trying to understand the nature of the phenomenon under study, but that is retrieved in the MRELab using validated and verified CFD codes. The alternative of a high-speed camera tracking the motion of the cylinder at high amplitudes or using PIV (particle image velocimetry) at high Reynolds numbers would be desirable but impractical. As explained in Sect. 47.3.5, due to CFD limitations the diameter-scale information collected in visualization and vortex tracking is most important in verification of CFD codes before trusting CFD results at the boundary-layer scale.

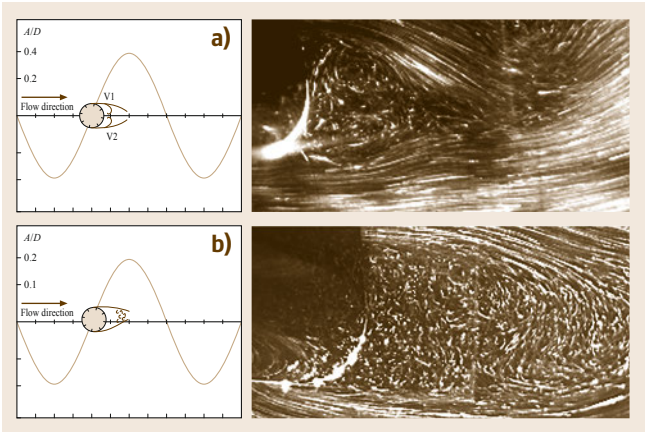


Fig. 47.54a,b Typical wake vortex structure of smooth cylinder at $U^* = 14.51$ (desynchronization), $\mathrm{Re} = 1.12 \times 10^5$, $A^* = 0.06$. **(a)** V1 and V2 shed nearly simultaneously, **(b)** irregularly shed-turbulent eddies (after [47.104])

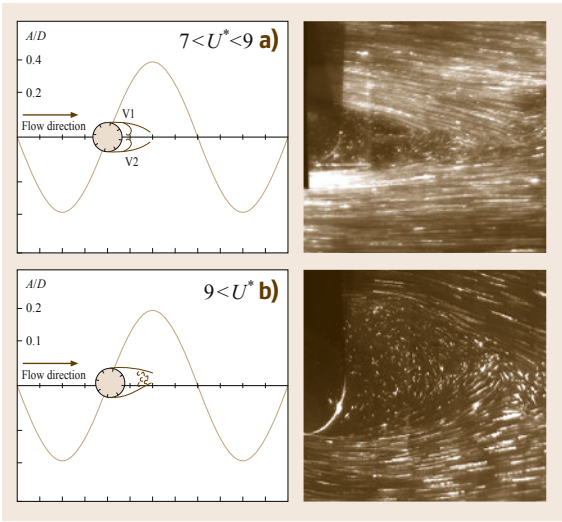


Fig. 47.55a,b Flow structures around a circular PTC cylinder in FIM suppression. Near-wake demonstrates absence of stable vortex shedding when PTC is placed in the SS zone. **(a)** V1 and V2 shed nearly simultaneously, **(b)** irregularly shed-turbulent eddies (after [47.104])

In conclusion it should be noted that:

- Broad-field visualization of FIM and vortex tracking is needed to understand the diameter scale for FIM dynamics as well as interaction of wakes with additional oscillators operating in the wake of other cylinders.
- Broad-field visualization of FIM is also needed for verification of CFD codes so that information at the boundary-layer scale can be trusted.

47.3.4 Field Tests

Field tests, regardless of the model or prototype scale, provide the ultimate environment for testing a technology in real life. Among the conditions that can be tested in field tests but not in a laboratory are the following:

1. Realistic flow conditions, including nonuniformity of flow caused by shear flows, boundary layers from the cite bed or walls, or variation along the cylinder length; variation of flow velocity and direction in time.
2. Data and power transmission to shore.
3. Hazards caused by other uses of the cite for navigation, recreation, fishing, etc.
4. Extended exposure to the marine environment including floating debris.
5. Recording of data for possible impact on the environment.
6. Deployment, retrieval, and decommissioning.

In the process of developing the ALT for the VIVACE converter, two river and two canal deployments have been conducted briefly described below along with some lessons learnt.

River Tests

The first field test was conducted in the St. Clair River in August 2010 with the two horizontal-cylinder converter shown in Fig. 47.18 in lab tests, during the deployment, and in underwater operation, respectively. In spite of the variation of flow from one end of the cylinder to the other by about 15% both cylinders operated in FIM and video and response were recorded for a short period of time.

The second field test tool place in September 2012 at the same location with a smaller, one-cylinder, horizontal-cylinder converter shown in Fig. 47.19 in towing tank testing, deployment, and underwater operation. The converter stayed in place for three months.

Canal Tests

In January 2013, two tests were conducted in canals in the Netherlands under the lead of *TAUW* [47.146]. Figure 47.20 shows the small model tested in the University of Michigan towing tank, and at location, respectively. The modelas small to fit in the available opening allowing for no more than 1.5 diameters amplitude. The testing took place under the most adverse conditions and had limited success. Specifically, the flow was nonuniform because it was coming into a small flat opening from a wider area (Fig. 47.20b). As a result, just as CFD conducted by TAUW showed, there was uniform flow only over one-third of the length of the cylinder while the rest was providing drag instead of lift. That means, that the generated lift power was only $(\frac{1}{3})^2$ of what that VIVACE model produced in the towing tank. In addition, the drag power in the same (transverse=lift) direction due to the two-thirds of the cylinder length was overcome by the lift. That is, the power capacity of the VIVACE canal model tested by TAUW was well over 10 times higher that what was actually produced in the field tests in the canals.

The conclusion to be drawn from field tests in the river and in the canals is:

- Flow uniformity is mandatory for success of the converter. At a 15% velocity variation from one end of the cylinder to the other, the oscillators performed very well. At the other extreme of field conditions, lift was generated by only one-third of the cylinder length. The converter still generated power, albeit about one-tenth of its capacity.

47.3.5 Computational Fluid Dynamics

Design of an ALT by CFD simulations can be a very time consuming and expensive process particularly as the geometry of the oscillator in FIM becomes more complex. The circular cross-section of a PTC cylinder or even a smooth cylinder is deceivingly simple. The complexity lies in the fact that CFD codes have difficulty predicting accurately the flow separation points one on each side of the cylinder. In simulation of flat plates or slender bodies, the location of the separation point is easily identifiable as it is expected to be at $\mathrm{Re} \cong 300\,000$ and can be set at that point in a CFD code. In flows past a stationary circular cylinder, the flow separation point is known experimentally to oscillate $\pm 5{-}10^\circ$ due to the von Kármán vortices. Finding the exact location of the separation point is critical in cylinder fluid dynamics as it affects all the important properties of vortex shedding, such as the beginning of the shear layer, the formation length, the circulation of the von Kármán vortices, and thus the alternating lift on the cylinder, the vortex structures, the width of the immediate wake, and thus, the interference with following cylinders in tandem. The problem becomes even more challenging when the cylinder is in FIM with large amplitude.

Another consideration in design by CFD is the need for a computational tool that can perform simulations fast. For four cylinders in FIM, simulation of 15 s real time may require 100 h of computational time converted to equivalent single-processor time. Most of this time is consumed in dynamic remeshing due to the finite fluid domain in the channel and the interface between the computational subdomains in the vicinity of each cylinder. DNS, DES, DNS three-dimensional computations would be too time consuming and do not resolve the separation point to a satisfactory degree.

In the MRELab, 2-D-URANS (two-dimensional Unsteady, Reynolds-averaged, Navier–Stokes) solvers in commercial codes are used for simulation. Specifically, OpenFOAM and Fluent/ANSYS are used. For a smooth cylinder, for $\mathrm{Re} > 10\,000$, CFD codes do not work well because of their inability to identify the exact location of the separation point even for a stationary cylinder. The separation point is calculated to be stationary at 90° which is incorrect and leads to poor predictions [47.139, 140]. Up to $\mathrm{Re} \cong 10\,000$, agreement between experiments and CFD for a smooth cylinder in VIV is acceptable (Fig. 47.56). *Wanderley* et al. [47.172, 173], achieved agreement between CFD and smooth cylinder VIV for Re up to 12 000.

On the other hand, when the separation point is forced to occur at the correct point on the cylinder, CFD predictions match remarkably well with experimental results. This is the case of PTC cylinders where separation occurs at the leading edge of the PTC both experimentally and computationally [47.139, 140]. Following that discovery, CFD tools have become very useful in the research of FIM of multicylinder VIVACE converters.

Next CFD modeling and simulation results are presented for two cylinders in tandem in FIM. Results are compared to experimental measurements. Few ad-

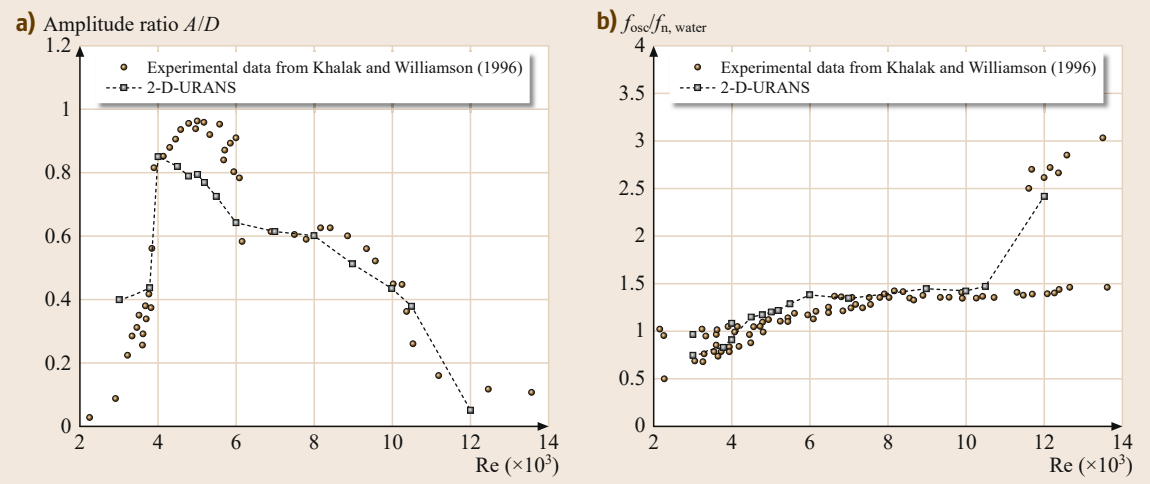


Fig. 47.56a,b For $\mathrm{Re} > 10\,000$, 2-D-URANS fails to predict cylinder forces and motion due to poor prediction of the oscillation of the separation point. For PTC cylinders, separation point is forced and agreement of CFD with experiments is excellent (after [47.112, 139, 140, 171]). **(a)** Smooth-cylinder comparison of A/D between experiments and VIV. **(b)** Smooth-cylinder comparison of $f_{\mathrm{osc}}/f_{\mathrm{n}}$, water between experiments and VIV

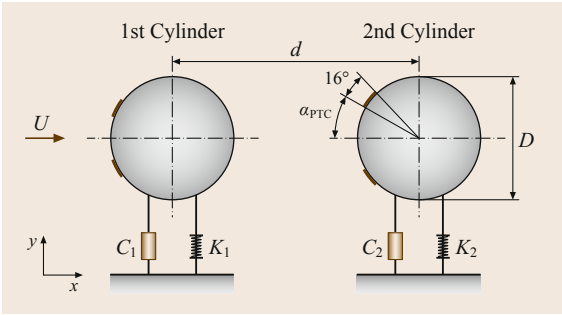


Fig. 47.57 Schematic of two PTC-cylinder physical model

ditional results are presented for three and four cylinder oscillators.

Physical Model

The first physical model consists of two oscillatory systems as shown in Fig. 47.57. Their properties are listed in Table 47.8. The other physical systems consist of two, three, or four oscillators shown in Fig. 47.58 with their properties summarized in Table 47.9. The elements of each oscillatory system are a rigid circular cylinder with diameter D and length L, two end-supporting linear-springs of stiffness K, and the system linear viscous damping C due to friction. Cylinders are arranged in tandem and constrained to oscillate in the y-direction, which is perpendicular to the flow velocity direction. The center-to-center distance d, between two cylinders is varied from $2D$ to $6D$. Two straight roughness strips are attached to the surface of each cylinder symmetrically, one on each side of the cylinder. The angle α_{ptc}, is measured in degrees from the forward stagnation point in the corresponding ideal flow (Fig. 47.23). The type and positions of the roughness strips are determined based on the PTC-to-FIM map (Fig. 47.24). The coverage provided by each sand strip is 16°. For the two cylinders in tandem configuration, $\alpha_{ptc} = \pm 20°$ is applied for the 1st cylinder and $\alpha_{ptc} = \pm 30°$ for the 2nd cylinder.

Simulations are always verified by comparison to experimental measurements on FIM of two PTC cylinders in tandem. The system parameters in the numerical simulations are the same as those used in the corresponding experiments.

Mathematical Models

Flow simulations are performed by a code developed in the MRELab based on the open source CFD tool OpenFOAM, which is composed of C++ libraries solving continuum mechanics problems with a finite-volume discretization method. In addition, code was developed:

1. For the dynamics of the multiple oscillators
2. For dynamic adjustment of the grid near a cylinder as it approaches a boundary.

Table 47.8 Physical model parameters in two-cylinder tests and CFD

Item	Symbol	1st cylinder	2nd cylinder
Diameter [m]	D	0.0889	0.0889
Length [m]	L	0.9144	0.9144
System mass [kg]	m	9.5121	9.5756
Spring constant [N m^{-1}]	K	758.11	726.84
Damping ratio	Z	0.0161	0.017
Damping [Ns m^{-1}]	C	2.7274	2.8434
Natural frequency [Hz]	$F_{n,water}$	1.1246	1.0989
Mass ratio	m^*	1.6774	1.6886
Strip placement angle [degree]	α_{ptc} (Fig. 47.23)	20°	30°
Thickness of strip [mm]	T	0.847	0.847

Table 47.9 Physical model parameters in three and four cylinder tests and CFD

Parameters	Symbol	Isolated cylinder	1st cylinder	2nd cylinder	3rd cylinder	4th cylinder
Diameter [m]	D	0.0889	0.0889	0.0889	0.0889	0.0889
Length [m]	l	0.9144	0.9144	0.9144	0.9144	0.9144
Oscillating mass [kg]	m_{osc}	10.75	9.53	9.59	9.51	9.58
Spring stiffness [N m^{-1}]	K	1600	744.29	757.41	737.24	747.39
Damping ratio	η	0.0122	0.0206	0.0198	0.0172	0.0168
Damping [N s m^{-1}]	C	3.2	3.464	3.382	2.882	2.834
Naturalfrequency [Hz]	$f_{n,water}$	1.57	1.114	1.121	1.109	1.114
Mass ratio	m^*	1.896	1.681	1.690	1.677	1.689
Added mass [kg]	m_a	5.671	5.671	5.671	5.671	5.671

Fig. 47.58 Four cylinders with PTC in tandem in the first LTFSW channel (photo courtesy of MRELab)

The time-dependent viscous flow solutions are obtained by numerical approximation of the incompressible 2-D-URANS equations in conjunction with the one-equation *Spalart–Allmaras* [47.174] turbulence model. The model is as follows

$$\frac{\partial \boldsymbol{U}_j}{\partial x_i} = 0\,, \tag{47.37}$$

$$\frac{\partial \boldsymbol{U}_i}{\partial t} + \boldsymbol{U}_j \frac{\partial \boldsymbol{U}_i}{\partial x_j} = -\frac{1}{\rho}\frac{\partial P}{\partial x_i} + \frac{\partial}{\partial x_j}\left(2\upsilon \mathbf{S}_{ij} - \overline{u_j' u_i'}\right), \tag{47.38}$$

where υ is the molecular kinematic viscosity and S_{ij} is the mean strain-rate tensor

$$\mathbf{S}_{ij} = \frac{1}{2}\left(\frac{\partial \boldsymbol{U}_i}{\partial x_j} + \frac{\partial \boldsymbol{U}_j}{\partial x_i}\right), \tag{47.39}$$

where $\boldsymbol{U}_i$ is the mean flow velocity vector. The quantity $\boldsymbol{\tau}_{ij} = -\overline{u_i' u_j'}$ is known as the Reynolds stress tensor, which is symmetric. In the Spalart–Allmaras model, a common method employs the Boussinesq approximation to relate the Reynolds stresses to the mean velocity gradient as

$$\boldsymbol{\tau}_{ij} = 2\nu_\mathrm{T}\mathbf{S}_{ij}\,, \tag{47.40}$$

where υ_T is the kinetic eddy viscosity and the related defining equations are as follows

$$\mu_t = \rho\tilde{\nu} f_{\nu 1},\; f_{\nu 1} = \frac{\chi^3}{\chi^3 + c_{\nu 1}^3},\; \chi = \frac{\tilde{\nu}}{\nu}\chi = \frac{\tilde{\nu}}{\nu}\,.$$

$\tilde{\nu}$ is an intermediate working variable of the turbulence model and obeys the following transport equation

$$\frac{\partial \tilde{\nu}}{\partial t} + u_j \frac{\partial \tilde{\nu}}{\partial x_j} = c_{\mathrm{b}1}\tilde{S}\tilde{\nu} - c_{\mathrm{w}1} f_\mathrm{w}\left(\frac{\tilde{\nu}}{d}\right)^2 + \frac{1}{\sigma}\left\{\frac{\partial}{\partial x_j}\left[(\nu + \tilde{\nu})\frac{\partial \tilde{\nu}}{\partial x_j}\right] + c_{\mathrm{b}2}\frac{\partial \tilde{\nu}}{\partial x_i}\frac{\partial \tilde{\nu}}{\partial x_i}\right\}. \tag{47.41}$$

Additional definitions of functions and constants are given by *Spalart* and *Allmaras* [47.174]. In this study, the trip terms $f_{\mathrm{t}1}$, $f_{\mathrm{t}1}$, and $f_{\mathrm{t}2}$ are turned off, and the *trip-less* initial condition for ν is used following *Travin* et al. [47.175].

The classical mass-damper-spring oscillator model is used to describe the cylinder dynamics in FIM. The one-degree of freedom equation of motion can be expressed as

$$m\ddot{y} + C\dot{y} + Ky = f(t)\,, \tag{47.42}$$

where m is the total oscillating mass of the cylinder and attachments including one-third of the spring mass and $f(t)$ is the total hydrodynamic excitation.

In the numerical simulations, a second-order Gauss integration scheme with a linear interpolation for the face-centered value of the unknown are used for the divergence, gradient, and Laplacian terms in the governing equations. The second-order backward Euler method is adopted for time integration. Thus, the numerical discretization scheme gives second-order accuracy in space and time. A pressure implicit with splitting of operators (PISO) algorithm is employed to solve the momentum and continuity equations together in a segregated way. The equation of motion for the cylinder in FIM is solved using a second-order mixed implicit and explicit time integration method.

Numerical Model

The computational domain size is different in the stream direction depending on the tandem spacing between the two cylinders. The domain length in the y-direction is $9D$ for all cases. As shown in Figs. 47.59 and 47.60, the entire domain includes five boundaries: inflow, outflow, top, bottom, and the cylinder walls. The distance between the inlet boundary and the center of the 1st cylinder, l_up, is set at $25D$. The downstream length of the domain, l_down, is also set at $25D$. The spacing d is varied from $2d$ to $6d$. The inflow velocity is considered as uniform and constant. At the outflow boundary, a zero gradient condition is specified for velocity. The bottom condition is defined as a wall boundary to match the experimental conditions. In the present numerical study, the free surface is simplified

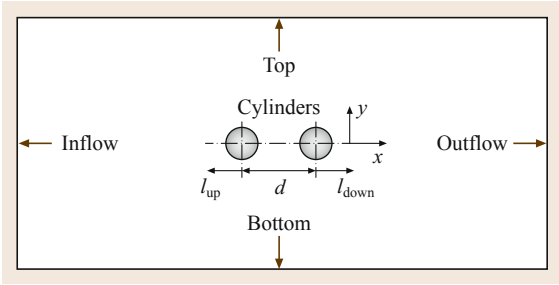


Fig. 47.59 Computational domain (after [47.113])

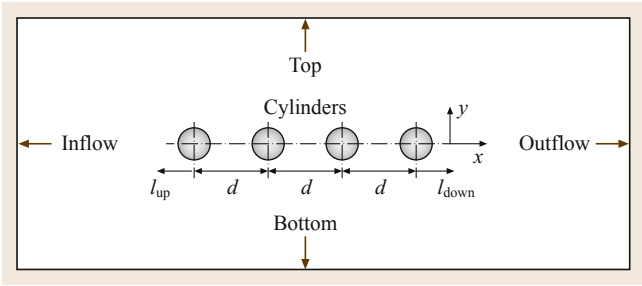


Fig. 47.60 Computational domain for four cylinders (after [47.114])

by modeling it as a wall. A moving wall boundary condition is applied for the cylinder surface when the cylinders are in FIM.

Two-dimensional, structured, computational grids were generated for all cases. A dynamic mesh technique of topological change is used to minimize the mesh deformation. A $2D \times 2D$ square subdomain for the near-wall region of each cylinder moves up and down when the cylinder is in motion. The cell layers at the top and bottom of the subdomain are removed or generated when the mesh is compressed or expanded during the simulation. Due to the specifically modified surface geometry of a PTC cylinder, a wall-function-type boundary condition is implemented to account for the effect of surface roughness. The near-wall grid spacing was selected to produce a $y+$ between 30 and 70, depending on Reynolds number. The medium grid resolution for the two PTC cylinders was selected and is identical to the density used in earlier work for CFD of one cylinder in FIM [47.139, 140] or two cylinders in FIM [47.113]. A close-up of the medium grid for the two PTC cylinders when $d = 2.5D$ is shown in Fig. 47.61. Figure 47.62 shows the grid configuration for four cylinders.

FIM of Two PTC Cylinders in Tandem

A series of simulations are conducted for investigation of the effect of tandem spacing on FIM of two cylinders with PTC. The Reynolds number range is $30\,000 < \mathrm{Re} < 100\,000$ which is in the high lift TrSL3

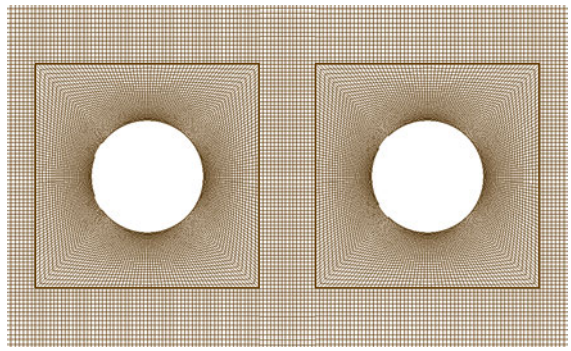


Fig. 47.61 Close-up of medium resolution grid (after [47.113])

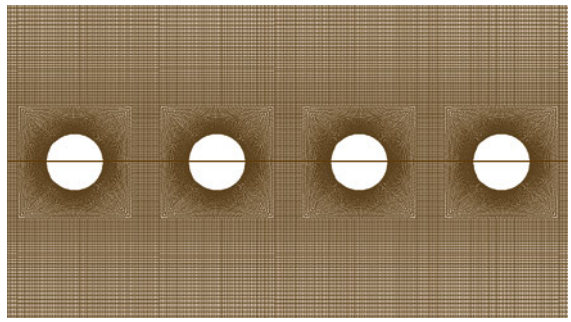


Fig. 47.62 Close-up of the medium grid for four PTC cylinders (after [47.114])

regime (Fig. 47.16). The corresponding reduced velocity ranges are $3.84 < U^*_{\mathrm{water}} < 12.81$ for the 1st cylinder and $3.93 < U^*_{\mathrm{water}} < 13.11$ for the 2nd cylinder. The simulation results are compared with the experimental data derived in the LTFSW channel of the MRELab. The two PTC cylinders start from the neutral position with zero initial velocity and displacement. In this study, the amplitude A_{peaks} of each PTC cylinder is calculated by averaging the absolute values of the 60 largest amplitude (negative and positive). The oscillation frequency f_{osc} of each cylinder is calculated by fast Fourier transform (FFT) over the recorded, steady-state, displacement history of the cylinder.

FIM of First Cylinder. Figure 47.63 shows the variation of amplitude and frequency response of the 1st cylinder with respect to Reynolds number (Re), free stream velocity (U), and reduced velocity (U^*_{water}) defined as in (47.3) $U^*_{\mathrm{water}} = U/(f_{\mathrm{n,water}}D)$. Within the test range of simulations and experiments, four regions can be clearly observed, including the VIV initial branch ($30\,000 < \mathrm{Re} < 40\,000$, $3.84 < U^*_{\mathrm{water}} < 5.12$), the VIV upper branch ($40\,000 < \mathrm{Re} < 80\,000$, $5.12 < U^*_{\mathrm{water}} 10.25$), transition from VIV to galloping ($80\,000 < \mathrm{Re} < 95\,000$, $10.25 < U^*\mathrm{water} <$

12.17), and galloping (Re > 95 000, U^*_{water}12.17). The experimental tests show that the oscillatory trends in A_{peaks}/D and f_{osc}/f_n, water of the 1st cylinder are not significantly influenced by the value of the center-to-center distance between the two PTC cylinders. For the simulation results, however, the VIV initial branch is initiated numerically earlier than in the experiments for all cases. This is attributed to the fact that the classical linear-viscous damping model used in the numerical simulations is different from the physical damping model in the experimental apparatus, which is more accurate but extremely complex [47.98, 100, 142]. When Re> 40 000 (U^*water > 5.12), the amplitude ratio follows closely the experimental data except for the case of $d = 2.5D$. Specifically, the amplitude of the 1st cylinder is slightly lower than other cases at Re = 50 000 ($U^*_{water} = 6.40$), and higher at Re = 80 000 ($U^*_{water} = 10.25$) when $d = 2.5D$. In all cases, Fig. 47.63 shows the rapid rise in amplitude, which is accompanied by a decrease in frequency for Re > 80 000 ($U^*_{water} > 10.25$). This phenomenon indicates that transition from the VIV upper branch to galloping is successfully initiated at $U^*_{water} = 10.25$ (Re = 80 000) instead of the typical $U^*_{water} \approx 20$, resulting in back-to-back VIV and galloping [47.67]. It should be noted that, for the smallest spacing ($d = 2.0D$), both the amplitude and frequency ratios of the 1st cylinder are lower than that in the other cases in the galloping region when Re > 95 000 ($U^*_{water} > 12.17$).

FIM of Second Cylinder. The amplitude and frequency ratios of the 2nd cylinder are plotted in Fig. 47.64. The motion of the 2nd cylinder is affected by the 1st cylinder and the response trend is quite different from that of the 1st cylinder. As shown in Fig. 47.64, during the VIV initial branch (30 000 < Re < 40 000, 3.93 < U^*_{water} < 5.24), the motion of the 2nd cylinder is almost suppressed in all the experimental tests. Similar phenomenon can be observed for the cases of $d < 3D$ in the numerical results. The major harmonic frequency is higher in the numerical simulations than the experiments in the initial-branch due to the difference in the viscous damping models between CFD and experiments. With increasing Re, the experimental results show that the amplitude bumps appear before the occurrence of the upshot of the initial-branch at Re ≈ 55 000 ($U^*_{water} \approx 7.21$) for all cases. The frequency trends also show a drop around Re ≈ 55 000 ($U^*_{water} \approx 7.21$) for the 2nd cylinder. Simulation cannot predict the phenomenon of decreasing in amplitude response at the start point of the upper branch. In the range of 60 000 < Re < 80 000 (7.87 < U^*_{water}10.49) in the VIV upper branch, the amplitude of the second cylinder shows a rising trend for all cases, and the oscillatory frequencies are not affected by the tandem spacing.

For $d = 2D$, the amplitude ratio is higher than other cases in the VIV upper branch. Similar bump in FIM response can also be observed in the transition region from the VIV upper branch to galloping at Re ≈ 95 000 ($U^*_{water} \approx 12.46$) when the tandem spacing d is smaller than $3D$. Thus, the galloping region is distinctly separated from the VIV upper branch for $d \le 3D$. However, with the increase in d, the VIV upper branch merges with the galloping branch. For the cases of $d > 3D$, it

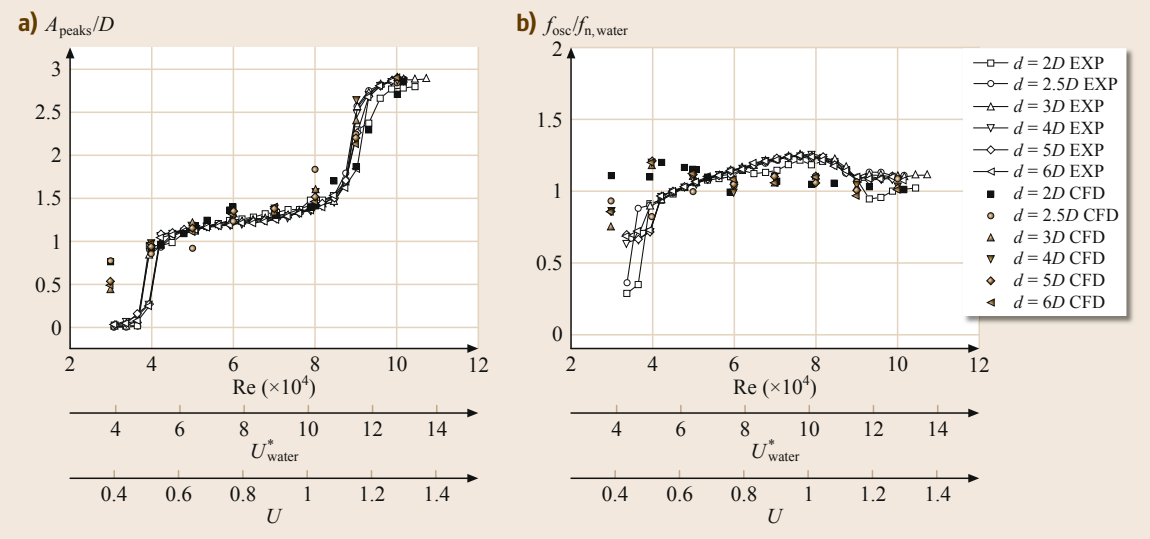


Fig. 47.63a,b Amplitude and frequency ratios of the 1st cylinder at different spacing values (after [47.113]) **(a)** A_{peaks}/D **(b)** $f_{osc}/f_{n,water}$

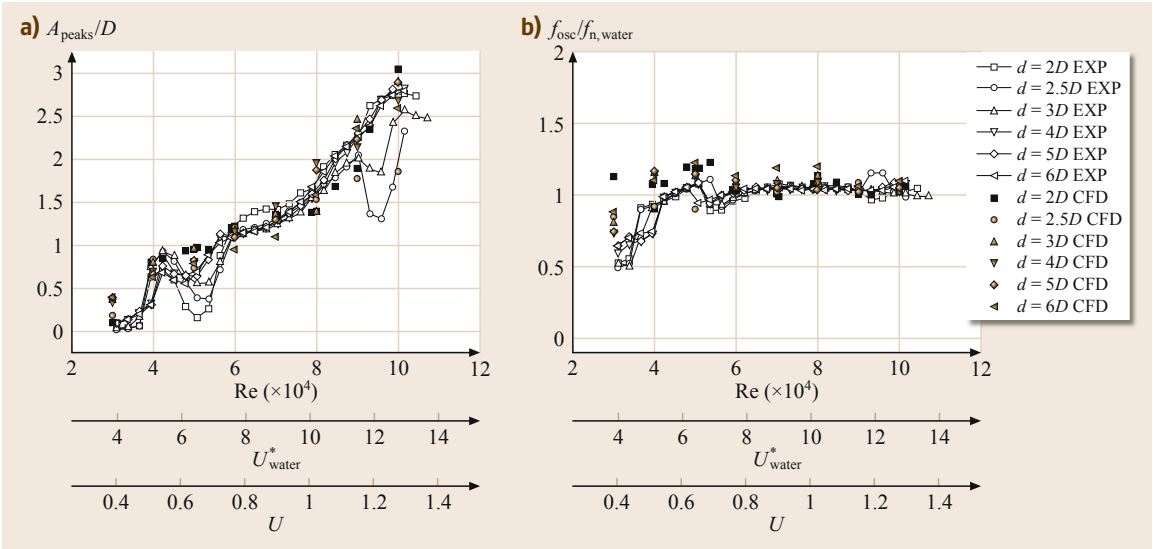


Fig. 47.64a,b Amplitude and frequency ratios of the 2nd cylinder at different spacing values **(a)** A_{peaks}/D **(b)** $f_{osc}/f_{n,water}$ (after [47.113])

is difficult to separate them. In addition, as shown in Fig. 47.63 102, the frequency of the 1st cylinder is not affected by the tandem spacing over the majority of the U^*_{water} range except for $d = 2D$. Similar is the case of the 2nd cylinder as well, but with minor deviations observed at very low and very high Re values.

Near-Wake Structures. For the two PTC cylinders in tandem, the 1st cylinder has great influence on the motion and vortex formation of the 2nd cylinder. To investigate the effect of tandem spacing on the FIM response of the two PTC cylinders, the vortex structures are discussed at four typical Reynolds numbers, as shown in Figs. 47.65 and 47.66 Re = 30 000, Re = 59 229, Re = 93 074, and Re = 10 000:

1. Re = 30 000. The wake vortex structures of two PTC cylinders with tandem spacing $d = 2D$ Re = 30 000 are illustrated in Fig. 47.65. The 2S mode of the vortex shedding can be clearly observed for the 1st cylinder. Two vortices are shed from the 1st cylinder per cycle of oscillation.
However, the vortex formation of the 2nd cylinder is affected by the vortices from the 1st cylinder, especially for the case of $d = 2D$. The motion of the 2nd cylinder is suppressed as the vortices from the 1st cylinder grab away the same rotation vortices of the second cylinder depriving it of circulation and consequently lift. For $d = 2.5D$, the shed vortices of the 1st cylinder allow only generation of weak vortices in the 2nd cylinder resulting in very small displacement of the 2nd cylinder (Fig. 47.64). When $d = 2.5D$, despite the formation of vortices behind the 2nd cylinder is affected by the 1st cylinder, the shed vortices can be clearly identified as 2S for the 2nd cylinder.

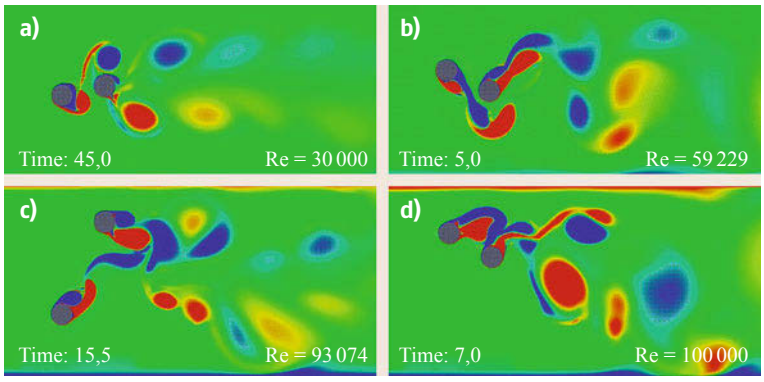


Fig. 47.65 Vorticity showing vortex structures of two PTC cylinders at various Re numbers at specific times; $d = 2D$ (after [47.113])

2. Re = 59 229. The two PTC cylinders oscillate in the VIV upper branch. Figure 47.65b shows the vortex structures of two PTC cylinders with $d = 2D$ center-to-center spacing. For the 1st cylinder, four vortices are shed per cycle of oscillation and the vortex patterns are 2P. When $d < 3D$, the upstream vortices directly and closely interact with the downstream cylinder. The shed vortices of the 2nd cylinder merge with the vortices from upstream as shown in Fig. 47.65b. The vortex pattern of the 2nd cylinder is hard to identify. It should be noted that the FIM of the 2nd cylinder is stimulated by the alternating vortices shed from the 1st cylinder at Re = 59 229. This means that the effect of the 1st cylinder on the motion of the 2nd cylinder in the VIV upper branch is different from the suppression results of the VIV initial-branch when $d < 3D$ for Re = 30 000. For $d \geq 3D$, 2P mode of vortex shedding can be observed for the 2nd cylinder.
3. Re = 93 074. As Re exceeds 90 000 the transition to galloping is initiated for the cylinders. There is variation in the vortex-shedding structure for the 1st cylinder, but the 2P+2S mode is observed in most cycles, such as the cases shown in Fig. 47.65c.

Figure 47.66 is very useful because it shows the pressure distribution on the cylinders at the same time as the corresponding picture in Fig. 47.65. This confirms the understanding that a shed vortex pushes the cylinder away.

The effect of spacing is shown in Fig. 47.67 for six values of center-to-center spacing for Re = 90 000. As expected, increased spacing reduces the impact on the second cylinder.

Several observations were made based on the numerical simulations and the experimental results regarding FIM of two PTC cylinders in tandem. The major conclusions are summarized below:

1. Simulation results are similar to experiments in VIV and galloping for PTC cylinders, capturing the most important features of amplitude, frequency, and vortex structures. Thus, the developed numerical tool can be powerful in understanding the FIM of two PTC cylinders in tandem.
2. For the upstream cylinder, the amplitude and frequency responses are not significantly influenced by the downstream cylinder when the spacing is greater than $2D$.
3. For the downstream cylinder, a rising amplitude trend in the VIV upper branch can be observed in all cases. The galloping region merges with the VIV upper branch for spacing greater than $3D$.

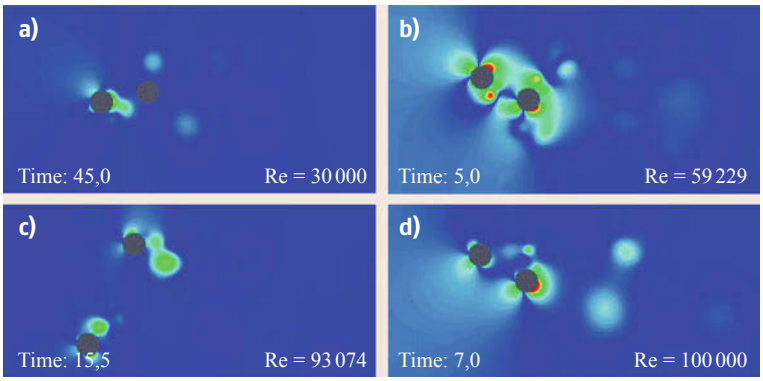


Fig. 47.66 Corresponding pressure distribution at the same Reynolds number and timing as in Fig. 47.65

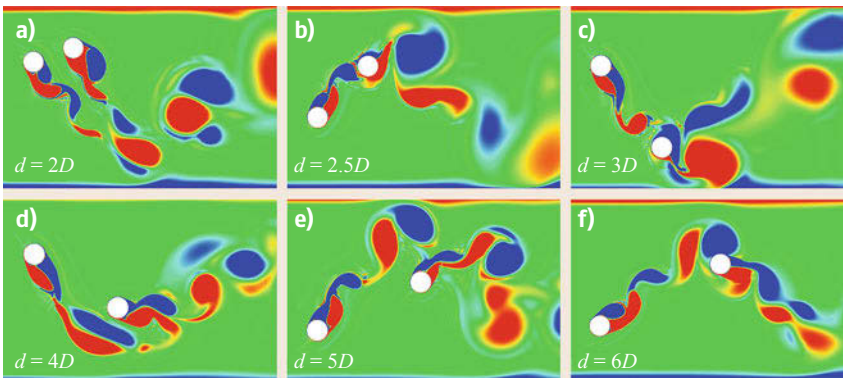


Fig. 47.67 Vortex structures of two PTC cylinders with different spacing at Re = 90 000 (after [47.113])

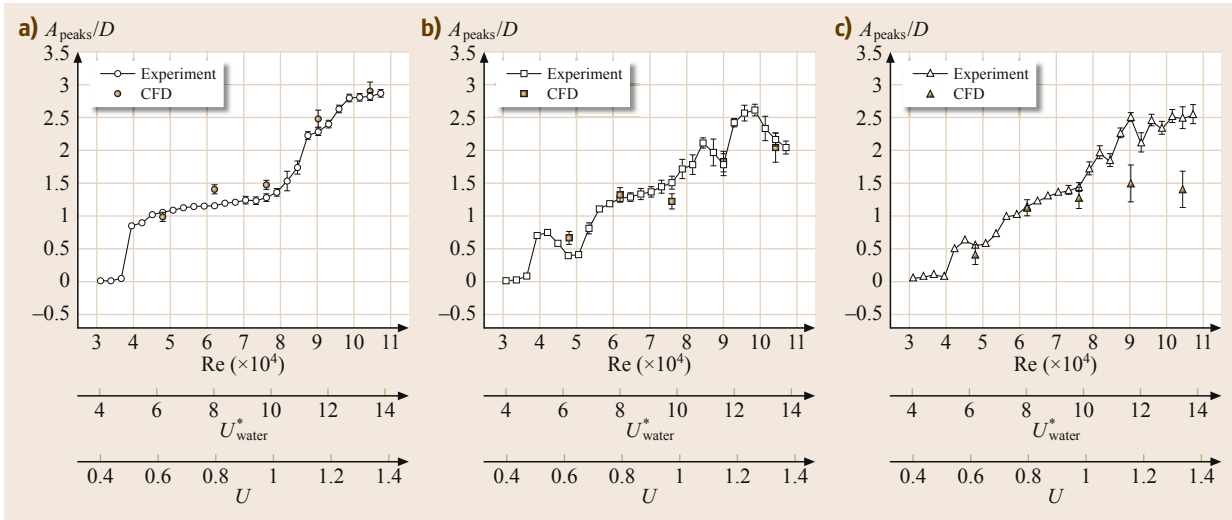


Fig. 47.68a–c Amplitude ratios comparison for three PTC cylinders: **(a)** 1st cylinder, **(b)** 2nd cylinder, **(c)** 3rd cylinder $d = 2.5D$ (after [47.113])

4. For Re = 30 000, the 2S mode of the vortex shedding can be clearly observed for the 1st cylinder in all cases. For the 2nd cylinder, vortex shedding is suppressed for $d = 2D$.
5. For Re = 60 000, the vortex patterns of the 1st cylinder are 2P for all cases. The vortex pattern of the 2nd cylinder is hard to identify for $d < 3D$. The 2P mode can be observed for the 2nd cylinder for $d \geq 3D$.
6. For Re = 90 000, the $2P + 2S$ mode is observed in most cycles for the 1st cylinder.
7. For Re > 90 000, the cycle-to-cycle variation phenomenon in vortex shedding occurs.

FIM of Three PTC Cylinders in Tandem

A few cases are presented here where the in-flow center-to-center distances between cylinders are kept equal to $d = 2.5D$. The angles of the PTC are $\alpha_{PTC} = \pm 30°$ for all the three cylinders. Results of three PTC cylinders are shown in Figs. 47.68–47.70. The different oscillating branches can be clearly observed in the curves of amplitude and frequency ratios. The dynamic response of the cylinders placed downstream of the 1st cylinder is affected by the 1st one. Comparing the results of the single PTC cylinder and two PTC cylinders, we observe that the downstream cylinders have an influence on the FIM for the 1st cylinder as well. As shown in Fig. 47.68, the amplitude ratios of the 1st cylinder rise from the VIV upper branch to galloping by an easy gradient instead of a sharp jump. For the 2nd and 3rd cylinders, similar to the behavior of the downstream cylinder of the 2-PTC cylinder case, the results show drops in the amplitude and frequency response around Re = 50 000 ($U^*_{water} = 6.47$). When galloping is fully developed, (Re > 95 000, $U^*_{water} > 12.29$), the response trends of the two downstream cylinders are profoundly affected by the upstream cylinder. Furthermore, the amplitude ratio for the 2nd and 3rd cylinders is lower than that of the 1st cylinder. Notably, $2.9D$ is obtained for the 1st cylinder in the simulation results.

For the three-cylinder cases, the motion of the downstream cylinders is strongly influenced by the vortices shed from the 1st cylinder. When Re =

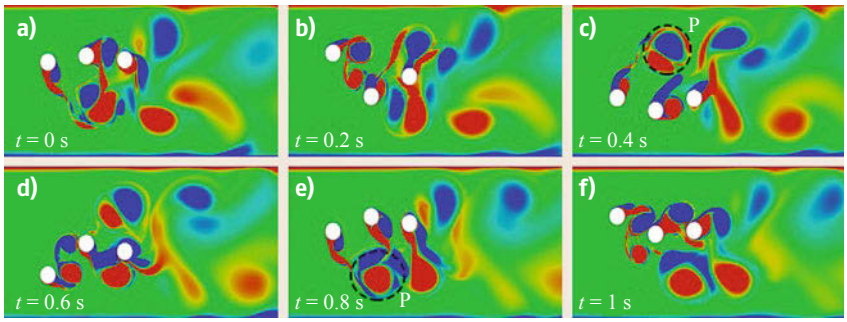


Fig. 47.69 Vortex pattern for three PTC cylinders at Re = 62 049; $d = 2.5D$ (after [47.113])

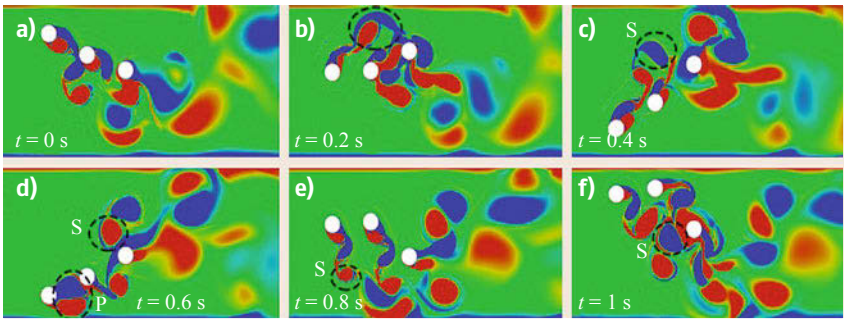


Fig. 47.70a–f Vortex pattern for three PTC cylinders at Re = 90 254; $d = 2.5D$ (after [47.113])

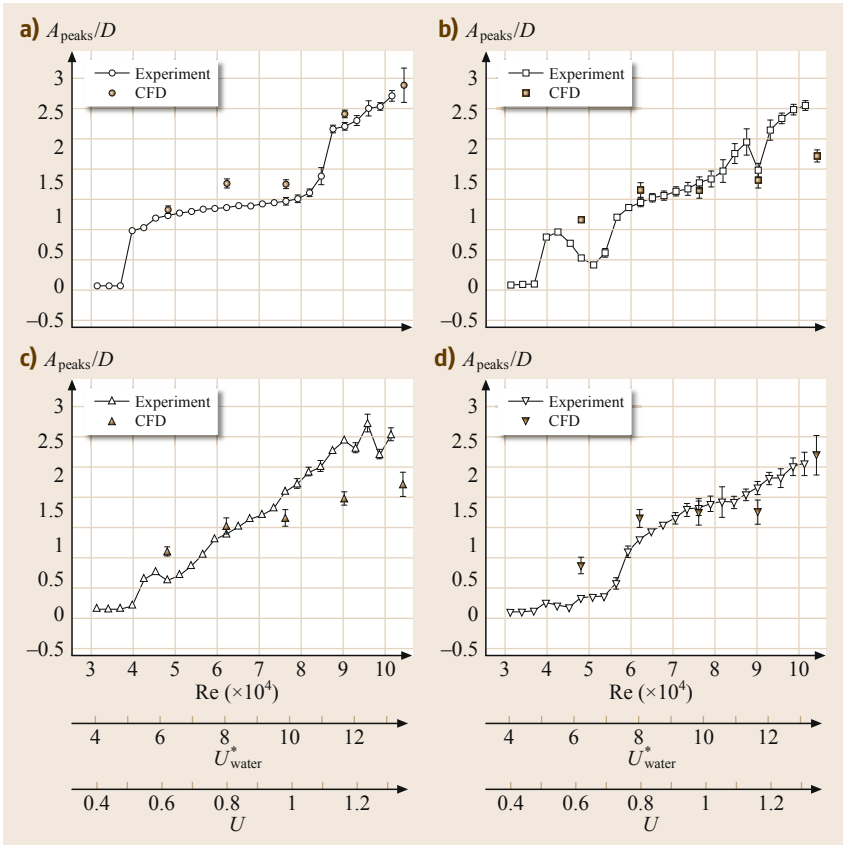


Fig. 47.71a–d Four PTC cylinders: **(a)** 1st cylinder, **(b)** 2nd cylinder, **(c)** 3rd cylinder, **(d)** 4th cylinder; A/D comparison between experiments and CFD in VIV, transition, and galloping; $d = 2.5D$ (after [47.113])

62 094($U^*_{water} = 8.03$), as presented in Fig. 47.69, the vortex pattern of the 1st cylinder is 2P and it is the same as that of the 1st cylinder for the two-cylinder case. The vortices developed by the 2nd and 3rd cylinders are taken away and absorbed by the upstream vortices. It is difficult to identify the vortex pattern for the 2nd and 3rd cylinders in animations due to the presence of the wake of the 1st cylinder. Moreover, the amplitude is higher for the upstream cylinder as shown in Fig. 47.68. Due to the specific spacing between the 1st and 2nd cylinder, the vortex pair from the 1st cylinder does not collide into the body of the 2nd one. Instead, the vortex pair from the first cylinder triggers the vortex shedding of the 2nd cylinder and can change the direction of movement. Therefore, the oscillatory frequency of the 2nd cylinder is approximately equal to the wake pattern (2P) frequency of the 1st cylinder. The 3rd cylinder follows the 2nd one. Thus, the frequencies of the three cylinders are close to each other.

In Fig. 47.70, the vortex pattern of three PTC cylinders for Re = 90 254 ($U^*_{water} = 11.67$) is plotted. Eight vortices are shed by the 1st cylinder in a 2P+4S vortex structure. The vortex patterns are not clear for the 2nd and 3rd cylinders due to the presence of the 1st

cylinder's wake. It should be noted that the 2nd cylinder follows the 1st cylinder and its vortex shedding is disturbed by upstream vortices. Thus, the displacement of the 2nd cylinder is influenced and small amplitude is obtained, as shown in Fig. 47.68. The amplitude ratio of $A_{peaks}/D = 2.49$ is reached for the 1st cylinder.

FIM of Four PTC Cylinders in Tandem

In all cases, the in-flow center-to-center distance between consecutive cylinders is fixed at $2.5D$ and PTC is placed symmetrically on cylinders at $\pm 30°$ from the front stagnation point of the corresponding ideal flow. The amplitude and frequency ratios are presented in Figs. 47.71 and 47.72, respectively. The maximum amplitude of $2.9D$ is achieved by the 1st cylinder in galloping. Compared with the results of the three PTC-cylinder cases, the amplitude and frequency responses are very similar for the three upstream cylinders. That means the 4th cylinder hardly affects the three upstream cylinders. But the FIM of the 4th cylinder is strongly disrupted by the vortices shed from the upstream cylinders. As shown in Fig. 47.71, four oscillating branches can be clearly observed and the VIV upper branch starts at Re$\approx 40\,000$ ($U^*_{water} \approx 5.17$) for the 1st cylinder. However, due to the presence of the upstream cylinders, the velocity corresponding to the start point of the VIV upper branch moves to higher velocities. Especially for the 4th cylinder, the VIV upper branch appears at Re $\approx 60\,000$ ($U^*_{water} \approx 7.76$).

For Re $= 62\,049$ ($U^*_{water} = 8.03$), the presence of the other three cylinders practically does not affect the oscillations of the 1st cylinder. In this case, the 1st cylinder experiences VIV and the 2P vortex pattern is observed in Fig. 47.73. It is worthwhile to note that significant phase difference exists between two adjacent cylinders. As shown in Fig. 47.73, there is a nearly 180° phase lag between the 3rd and 4th cylinders. The shear layers of the downstream cylinders are significantly affected by the wake of the upstream cylinders. The oscillating frequencies are very close for all four cylinders as can be seen in Fig. 47.71. The amplitude of the 1st cylinder is higher than the downstream cylinders and reaches $1.48D$.

As shown in Fig. 47.9b for Re $= 90\,254$ ($U^*_{water} = 11.67$), when the 1st cylinder moves down from the point of maximum displacement, a pair (P) of vortices sheds from the cylinder, followed closely by a single vortex (S). The same is observed in the upward journey of the cylinder. Thus, the vortex model is 2P+2S for the 1st cylinder. For the same Re as shown in Fig. 47.74, the downstream cylinders, the vortex pattern is hard to identify as the shed vortices are affected and modified by the vortices from the upstream cylinder wakes. In addition, the motion of the 1st cylinder has a regular form, which can be observed in the displacement ratio curve in Fig. 47.9c. The amplitude ratio $A_{peaks}/D = 2.49$ is achieved by the 1st cylinder which is much higher than those of the downstream cylinders.

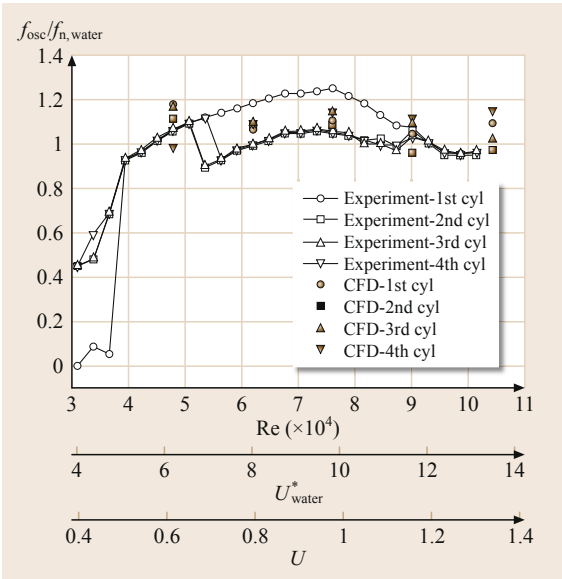


Fig. 47.72 Four PTC cylinders; comparison f_{osc}/f_n, water between experiments and CFD in VIV, transition, and galloping (after [47.113])

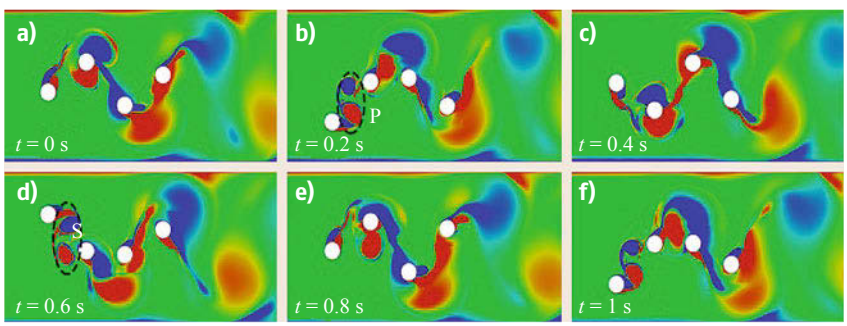


Fig. 47.73a–f Vortex pattern for four PTC cylinders at Re $= 62\,049$, $d = 2.5D$ (after [47.113])

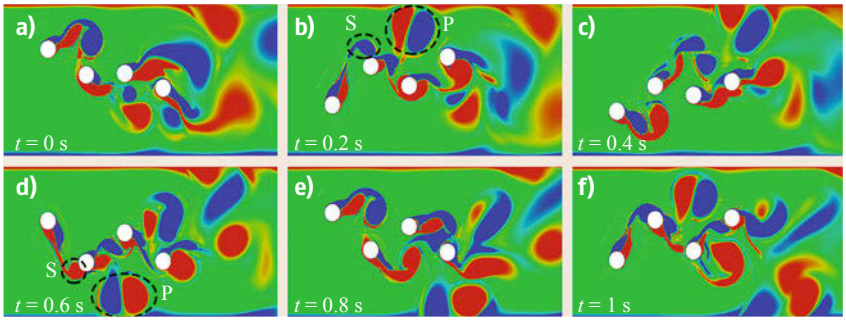


Fig. 47.74a–f Vortex pattern for four cylinders with PTC at Re = 90 254, $d = 2.5D$ (after [47.113])

Computational Visualization

Upon validation and experimental verification of the CFD code, it can become a very useful tool in investigating FIM of multiple PTC cylinders in tandem. To gain a greater insight into the flow details, the flow has to be visualized. It is difficult to visualize the flow experimentally at high Reynolds numbers using PIV. CFD can provide a unique tool to assist in understanding the underlying fluid dynamics. Most importantly, once the CFD code has been proven to predict correctly integral flow properties, such as amplitude and frequency responses, and nonintegral large-scale behavior – that is, von Kármán-scale vortex structures, it is expected to be able to predict with acceptable accuracy other properties difficult to measure experimentally and even smaller scale properties. Examples are the pressure distribution as in Fig. 47.66a–d, the separation points, and the interaction between shear layers and upstream wakes as in the significant discovery in Fig. 47.65a, which explained the experimentally observed suppression of the second cylinder's FIM.

A good case study is provided by *Kinaci* et al. [47.176] on computational and experimental assessment of turbulence stimulation on FIM of PTC cylinders. Different grades of commercial sandpaper were tested experimentally (P60 and P180 – Table 47.3). Figure 47.75 show vorticity contours to visualize wake structures. Pressure distributions, as in Fig. 47.76, are used to integrate pressure and calculate force. Velocity vector fields in the body vicinity in Fig. 47.77 are used to visualize separation of the flow. The following observations can be made based on Figs. 47.75–47.77:

1. The vorticity contours given in Fig. 47.75a,b are shown at the lowest position of the cylinder in oscillation. Both images are at the reduced velocity of $U^* = 7$ for the two PTC thicknesses defined in Table 47.3. The vorticity contours for the two roughness strips explain the difference in amplitude response as the thicker PTC induces higher amplitude when compared to the thinner PTC. The reason

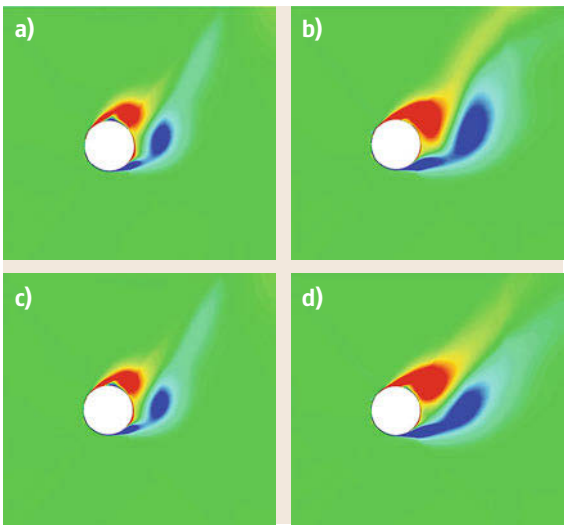


Fig. 47.75a–d Vorticity contours at the lowest position in vertical oscillation (after [47.176]). **(a)** PTC P60 at $U^* = 7$. **(b)** PTC P60 at $U^* = 14$. **(c)** PTC P180 at $U^* = 7$. **(d)** PTC P180 at $U^* = 14$

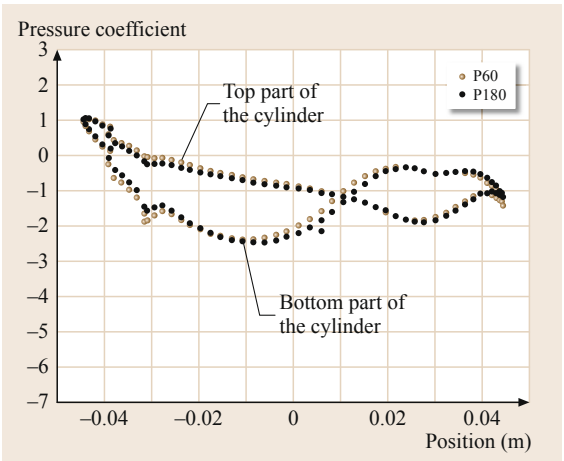


Fig. 47.76 Pressure coefficient around the cylinder with PTC P60 and P180 (after [47.176])

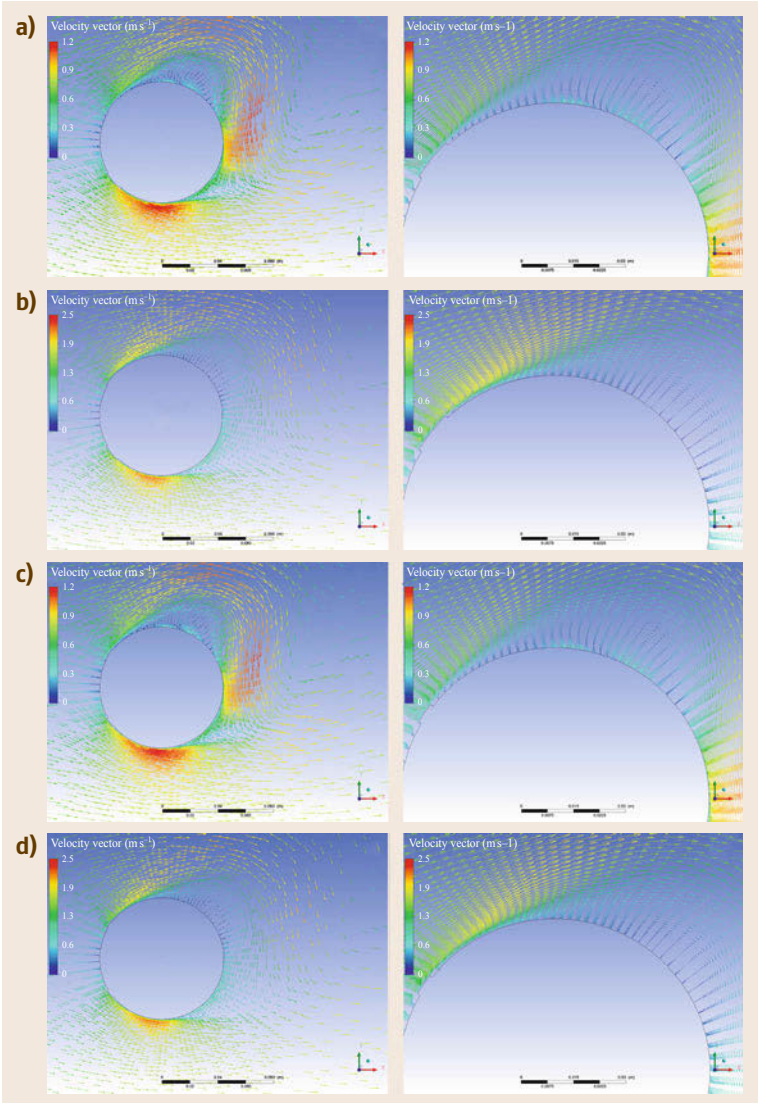


Fig. 47.77a–d Velocity vector round a stationary cylinder with PTC P60 and P180 (after [47.176]). **(a)** Velocity vectors at $U^* = 7$ for PTC P60 of 16° coverage. **(b)** Velocity vectors at $U^* = 14$ for PTC P60 of 16° coverage. **(c)** Velocity vectors at $U^* = 7$ for PTC P60 of 8° coverage. **(d)** Velocity vectors at $U^* = 14$ for PTC P60 of 8° coverage

for this is related to the vortex strength, which is proportional to the circulation and the resulting lift force per the Kutta–Joukowski theorem. Furthermore, shorter formation-length brings the shed vortex closer to the after body, where the pressure variation applies when the vortex sheds and thus induces lift.

2. The difference in amplitudes may be shown by the pressure differences at the upper and lower parts of the cylinders. The pressure coefficients at the lowest position of the cylinder for reduced velocity $U^* = 7$ for the cylinders with PTC P60 and P180 are given in Fig. 47.76. The pressure difference of the cylinder with PTC P60 is higher than that of the cylinder with PTC P180. It is clear that in Fig. 47.76 the difference in pressure coefficients is greater, which in return, creates higher amplitude.
3. Figure 47.77 presents the velocity vectors around the cylinder for 16° and 8° coverage at $U^* = 7$ and $U^* = 14$. These figures capture the flow-speed vector field at the lowest position of the cylinder oscillation. The boundary layer separation on the lower side of the cylinder occurs earlier at $U^* = 7$ when compared to the separation at $U^* = 14$.
4. There is a reattachment of the flow at the upper part of the cylinder which is not examined at the lower part independent of the flow speed. The formation of the flow characteristics given in Fig. 47.77 is periodic and at the upper end of the oscillation, they will approximately be reversed. The boundary layer

separation and flow reattachment at the upper part of the cylinder occurs later in galloping compared to the VIV range.
5. The trapped dead fluid in the VIV region ($U^* = 7$) is greater when compared to the galloping region ($U^* = 14$). The higher flow momentum in galloping, compared to VIV, pushes the trapped dead fluid zone to a later part of the cylinder.
6. In the VIV region, the trapped dead fluid restricts the amplitude of the cylinder when compared to the cylinder in the galloping region. In the galloping region, the dead-fluid zone is not carried with the cylinder in oscillation.
7. At $U^* = 7$, which is in the VIV region, the vortex that will be shed once the cylinder reaches its highest position, is built at the upper part of the cylinder and waits to be shed. At $U^* = 14$, which is in the galloping region, the vortex formed at the upper part of the cylinder is about to be shed. In the galloping region, the number of shed vortices per half cycle is higher when compared to the VIV region.
8. The trapped dead fluid in Fig. 47.77a,c increases the added mass at $U^* = 7$. This CFD result is consistent with the added mass curve by *Vikestad* et al. [47.60], showing that the added mass decreases with respect to increasing fluid velocity.

Numerous specific conclusions regarding CFD use in FIM of multiple cylinders were presented in this section. The conclusions to be retained are the following:

- 2-D-URANS CFD codes can be used to accurately model FIM of a single or multiple cylinders in tandem as long as turbulence stimulation is properly imposed both in CFD and experiments as specified in Fig. 47.24 and following the relevant robustness rules.
- CFD codes verified by experiments not only in predicting integral properties, such as amplitude and frequency, but also nonintegral properties at the von Kármán scale can provide valuable information at the small-scale level. Thus, separation points, pressure distribution, and wake interactions can be studied using such CFD codes.

47.3.6 Mathematical Model of Harnessed and Dissipated Power

FIM data collected experimentally in a lab, or in the field, or numerically can be post-processed to yield useful information. A mathematical model and a procedure are required and are presented in this section.

The motion of the cylinder in the y-direction (Fig. 47.78) is modeled by a second-order linear differential equation as

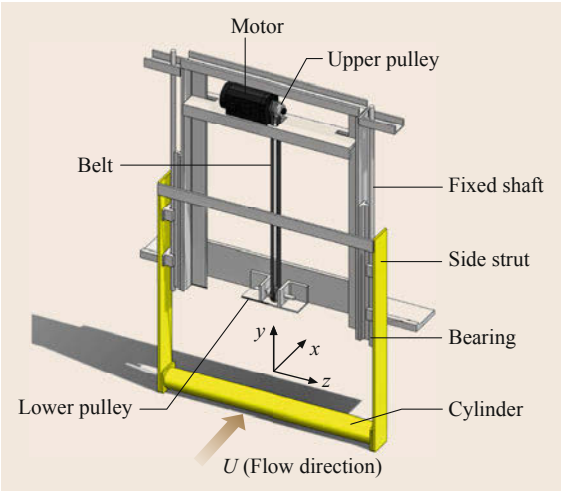


Fig. 47.78 Simple schematic of a VIVACE module with coordinate system (after [47.98])

$$m_{\text{osc}}\ddot{y} + C_{\text{total}}\dot{y} + Ky = F_{\text{fluid}}\,, \tag{47.43}$$

where y is the direction perpendicular to the flow and the cylinder axis, m_{osc} is the total oscillating system mass, which includes one-third of the spring mass, K is the spring stiffness, C_{total} is the total damping coefficient, and F_{fluid} is the force exerted by the fluid on the body in the y direction. This model differs from (47.6) in the sense that the added mass has not yet been separated from the total hydrodynamic force.

There are two approaches to the added mass issue:

1. The first approach separates the inviscid added mass from the total hydrodynamic force and is presented in this section. This assumes that VIV is a lock-in phenomenon resulting in synchronization between vortex shedding (excitation) and cylinder motion, which has a similar effect as resonance, albeit of self-limiting amplitude.
2. The second approach considers the added mass as part of the total lift force in the y direction and calculates it from experimental or numerical data from the force component in phase with the acceleration $\ddot{y}$ [47.61]. This is presented in Sect. 47.3.7 and leads to the conclusion that VIV is a resonance phenomenon with variable natural frequency in water due to variable added mass [47.60].

It is the author's opinion that this controversy is unnecessary as both are simple modeling methods that cannot account for the range of synchronization or the self-limiting nature of the amplitude of oscillation. Both just model the response of the cylinder in

FIM for a specific flow velocity and a given experiment. On the other hand, both approaches are useful in post-processing data and understanding this complex phenomenon better. For better understanding of the complexity of the added mass expression in forces and moments on a small body moving in a 3-D unsteady flow, the reader is referred to the detailed analysis by *Foulhoux* and *Bernitsas* [47.64]. The dependence of the inertia terms on absolute and/or relative velocity as well as the form of convective terms is explained thoroughly putting into perspective the simplicity of the inertia term used in Morison's equation.

Basic Mathematical Equations

The fluid force is separated into viscous and inviscid components as follows

$$m_{\text{osc}}\ddot{y} + C_{\text{total}}\dot{y} + Ky = F_{\text{viscous}} + F_{\text{inviscid}} \,. \tag{47.44}$$

The inviscid force may be defined in terms of the inviscid added mass m_{a} (this model) or may be measured experimentally or numerically (Sect. 47.3.7). Further,

$$F_{\text{inviscid}} = -m_{\text{a}}\ddot{y} \,, \tag{47.45}$$

$$F_{\text{viscous}} = \frac{1}{2}c_{\text{y}}(t)\rho U^2 DL \,. \tag{47.46}$$

The displaced fluid mass m_{d} for a circular cylinder is $\rho\frac{\pi}{4}D^2L$. This reduces the equation of motion to

$$(m_{\text{osc}} + m_{\text{a}})\ddot{y} + C_{\text{total}}\dot{y} + Ky = \frac{2}{\pi D}c_{\text{y}}(t)m_{\text{d}}U^2 \,. \tag{47.47}$$

Fluid Power Conversion

The power in the flowing fluid over the projected area of the cylinder in the direction perpendicular to the flow direction can be calculated as follows. The kinetic pressure head in the fluid form Bernoulli's equation is $\frac{1}{2}\rho U^2$. The force acting over the projected area DL of the cylinder in the direction y of the flow is $\frac{1}{2}\rho U^2 DL$. Then, the power in the fluid, being the product of the acting force times the velocity in that same direction x is

$$\text{Power in the fluid } = \frac{1}{2}\rho U^3 DL \,. \tag{47.48}$$

The work done by the fluid force acting on a VIVACE cylinder during a vibration cycle is the inner vector product of the force by the displacement vector $\mathrm{d}y$ integrated over one cycle of oscillation of the cylinder

$$W_{\text{VIVACE-Fluid}} = \int_0^{T_{\text{cyl}}} F_{\text{fluid}}\dot{y}\mathrm{d}t \,. \tag{47.49}$$

Then, the fluid power in VIVACE is

$$P_{\text{VIVACE-Fluid}} = \frac{W_{\text{VIVACE-Fluid}}}{T_{\text{cyl}}} \,. \tag{47.50}$$

The viscous force exerted by the fluid on the cylinder is given by the right-hand side of (47.47). Assuming a sinusoidal response which is accurate enough in all but the desynchronization zone of VIV, we multiply the force by the instantaneous velocity, integrate the right-hand side, and average it over the cycle period T_{cyl} to yield

$$P_{\text{VIVACE-Fluid}} = \frac{1}{T_{\text{cyl}}}\int_0^{T_{\text{cyl}}} \frac{2}{\pi D}c_{\text{y}}(t)m_{\text{d}} \times U^2 2\pi f_{\text{fluid}} y_{\max}\cos(2\pi f_{\text{fluid}})\mathrm{d}t \,. \tag{47.51}$$

In synchronization, we have $f_{\text{fluid}} = f_{\text{cyl}} = \frac{1}{T_{\text{cyl}}}$.

Therefore,

$$\begin{aligned} P_{\text{VIVACE-Fluid}} &= f_{\text{cyl}}\int_0^{\frac{1}{f_{\text{cyl}}}} \frac{2}{\pi D}C_{\text{y}}m_{\text{d}}U^2 2\pi f_{\text{cyl}} y_{\max} \\ &\quad \times \sin(2\pi f_{\text{cyl}}t + \emptyset)\cos(2\pi f_{\text{cyl}}t)\mathrm{d}t \\ &= \frac{4\pi}{D}C_{\text{y}}m_{\text{d}}U^2 f_{\text{cyl}} y_{\max}\sin\emptyset \,. \end{aligned} \tag{47.52}$$

Inserting the definition of $m_{\text{d}} = \frac{\pi}{4}\rho D^2 L$, yields

$$P_{\text{VIVACE-Fluid}} = \frac{1}{2}\rho\pi C_{\text{y}}U^2 f_{\text{cyl}} y_{\max} DL\sin\emptyset \,. \tag{47.53}$$

Mechanical Power

From (47.47), we integrate the left-hand side after multiplying it by the instantaneous velocity and we average it over the cycle period T_{cyl}. That yields the mechanical power in VIVACE [47.48, 49].

$$P_{\text{VIVACE-Mech}} = \frac{1}{T_{\text{cyl}}}\int_0^{T_{\text{cyl}}} [(m_{\text{osc}} + m_{\text{a}})\ddot{y} + C_{\text{total}}\dot{y} + Ky]\dot{y}\mathrm{d}t \,. \tag{47.54}$$

Only the term in phase with the velocity generates a nonzero energy term in sinusoidal response. In addition, we have

$$\zeta_{\text{total}} = \frac{C_{\text{total}}}{2\sqrt{(m_{\text{osc}} + m_{\text{a}}K)}} \,, \tag{47.55}$$

$$f_{\mathrm{n,water}} = \frac{1}{2\pi}\sqrt{\frac{K}{(m_{\mathrm{osc}} + m_{\mathrm{a}})}}\,. \tag{47.56}$$

Thus,

$$P_{\mathrm{VIVACE\text{-}Mech}} = \frac{1}{T_{\mathrm{cyl}}}\int_0^{T_{\mathrm{cyl}}} 4\pi(m_{\mathrm{osc}} + m_{\mathrm{a}})\zeta_{\mathrm{total}}\dot{y}^2 f_{\mathrm{n,water}}\mathrm{d}t\,. \tag{47.57}$$

Using the sinusoidal expression for

$$y = y_{\max}\sin(2\pi f_{\mathrm{cyl}}t)$$

in the above integral, we have

$$\begin{aligned}
&\frac{1}{T_{\mathrm{cyl}}}\int_0^{T_{\mathrm{cyl}}} 4\pi(m_{\mathrm{osc}} + m_{\mathrm{a}})\zeta_{\mathrm{total}}y_{\max}^2 \\
&\quad [\cos(2\pi f_{\mathrm{cylt}})]^2(2\pi f_{\mathrm{cyl}})^2 f_{\mathrm{n,water}}\mathrm{d}t \\
&= \frac{1}{T_{\mathrm{cyl}}}\int_0^{T_{\mathrm{cyl}}} 4\pi(m_{\mathrm{osc}} + m_{\mathrm{a}})\zeta_{\mathrm{total}}y_{\max}^2 \\
&\quad \frac{1+\cos(4\pi f_{\mathrm{cyl}}t)}{2}(2\pi f_{\mathrm{cyl}})^2 f_{\mathrm{n,water}}\mathrm{d}t \\
&= \frac{1}{T_{\mathrm{cyl}}}\int_0^{T_{\mathrm{cyl}}} 4\pi(m_{\mathrm{osc}} + m_{\mathrm{a}})\zeta_{\mathrm{total}}y_{\max}^2 \\
&\quad (2\pi f_{\mathrm{cyl}})^2 f_{\mathrm{n,water}}\mathrm{d}t \\
&= \left[\frac{T_{\mathrm{cyl}} + \frac{\sin(4\pi f_{\mathrm{cyl}}T_{\mathrm{cyl}})}{4\pi f_{\mathrm{cyl}}}}{2}\right] - \left[\frac{0 + \frac{\sin(4\pi f_{\mathrm{cyl}}0)}{4\pi f_{\mathrm{cyl}}}}{2}\right] \\
&= 2\pi(m_{\mathrm{osc}} + m_{\mathrm{a}})\zeta_{\mathrm{total}}y_{\max}^2(2\pi f_{\mathrm{cyl}})^2 f_{\mathrm{n,water}} \\
&= 8\pi^3(m_{\mathrm{osc}} + m_{\mathrm{a}})\zeta_{\mathrm{total}}(y_{\max}f_{\mathrm{cyl}})^2 f_{\mathrm{n,water}}\,.
\end{aligned} \tag{47.58}$$

Harnessed Power and Dissipated Power

In order to convert MHK energy to mechanical, additional damping is introduced into the system. The total damping (C_{total}) is defined as

$$C_{\mathrm{total}} = C_{\mathrm{stucture}} + C_{\mathrm{harness}}\,, \tag{47.59}$$

where $C_{\mathrm{structure}}$ is the existing damping due to frictional losses in the transmission system, C_{harness} is the damping added through the generator to convert the mechanical energy in the oscillating cylinder to electrical energy an in field prototypes.

From (47.55), we know that $C_{\mathrm{structure}}$ and C_{harness} can be expressed in dimensionless form using the damping ratios $\zeta_{\mathrm{structure}}$ and ζ_{harness}

$$\zeta_{\mathrm{structure}} = \frac{C_{\mathrm{structure}}}{2\sqrt{(m_{\mathrm{osc}} + m_{\mathrm{a}})K}}\,, \tag{47.60}$$

$$\zeta_{\mathrm{harness}} = \frac{C_{\mathrm{harness}}}{2\sqrt{(m_{\mathrm{osc}} + m_{\mathrm{a}})K}}\,. \tag{47.61}$$

Using (47.58), we obtain the harnessed power and the dissipated power as

$$P_{\mathrm{VIVACE\text{-}Dissipated}} = 8\pi^3(m_{\mathrm{osc}} + m_{\mathrm{a}})\zeta_{\mathrm{structure}}(y_{\max}f_{\mathrm{cyl}})^2 f_{\mathrm{n,water}}\,, \tag{47.62}$$

$$P_{\mathrm{VIVACE\text{-}Harness}} = 8\pi^3(m_{\mathrm{osc}} + m_{\mathrm{a}})\zeta_{\mathrm{harness}}(y_{\max}f_{\mathrm{cyl}})^2 f_{\mathrm{n,water}}\,. \tag{47.63}$$

Harnessed Power Calculation Using Experimental and Numerical Simulation Results

To calculate the harnessed power using (47.63), several parameters should be determined. Oscillating system mass m_{osc} and spring constant K can be measured directly. The ideal added mass m_{a} is equal to the displaced fluid mass. Using spring constant K, oscillating system mass m_{osc} and added mass m_{a}, the natural frequency of cylinder in water can be calculated using (47.11). Structural damping $C_{\mathrm{structure}}$ can be measured during the experiment and added damping C_{harness} is set by the V_{ck} system (virtual damping–spring system) developed by the MRELab (Sect. 47.2.3, *Electro-Mechanical Principles*) [47.98].

The remaining two parameters, $y_{\max}$ and f_{cyl}, are obtained through measured displacement history recorded in experiments or calculated displacement history based on numerical simulations. Next, the harnessed-power envelope for a single cylinder VIVACE converter is calculated using CFD simulations and compared to experimental results. *Lee* and *Bernitsas* [47.100] derived experimentally the power envelope in Fig. 47.38 for a smooth cylinder. *Chang* and *Bernitsas* [47.96] derived the power envelope in Fig. 47.33 for a PTC cylinder shown.

Liu and *Bernitsas* [47.177] derived numerically the power envelope for the same PTC cylinder and the results are presented in Table 47.10 and later in this section.

Amplitude Response. The amplitude ratio A/D graph obtained by numerical simulation is shown in Fig. 47.79 based on the optimal values in Table 47.10. The amplitude A of the PTC cylinder is calculated by averaging

Table 47.10 Power generated by single PTC cylinder VIVACE (after [47.177])

Case	U [m s^{-1}]	Re 1×10^3	U^*	K [N m^{-1}]	c [N s m^{-1}]	ζ	Power CFD [w]
1	0.38	30	5.50	400	18.28	0.12	1.27
2	0.51	40	7.34	400	18.28	0.12	1.31
3	0.64	50	7.49	600	21.85	0.12	3.72
4	0.77	60	7.78	800	24.86	0.12	3.68
5	0.83	65	7.54	1000	27.51	0.12	5.79
6	0.90	70	8.12	1000	27.51	0.12	7.13
7	0.96	75	7.94	1200	29.91	0.12	10.49
8	1.02	80	7.84	1400	32.12	0.12	14.87
9	1.09	85	8.34	1400	32.12	0.12	19.39
10	1.15	90	8.26	1600	34.17	0.12	25.19
11	1.18	92	8.44	1600	34.17	0.12	26.66
12	1.22	90	8.71	1600	34.17	0.12	30.09
13	1.28	100	8.65	1800	36.09	0.12	40.33
14	1.31	102	8.82	1800	36.09	0.12	41.22
15	1.35	105	9.08	1800	36.09	0.12	35.53
16	1.41	110	9.02	2000	37.92	0.12	49.38
17	1.54	120	9.85	2000	37.92	0.12	55.53

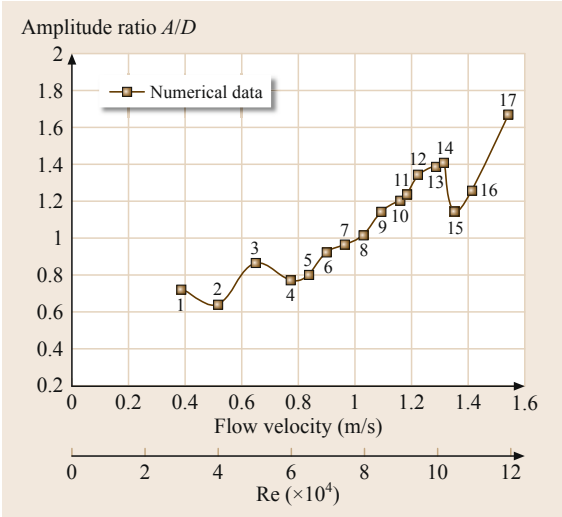


Fig. 47.79 Amplitude ratio of one cylinder with PTC for the power envelope points in Table 47.10 (after [47.177])

the absolute values of the 40 highest values negative and positive peaks in 30 s of simulation time. The maximum displacement of the oscillating cylinder reached in galloping is 1.67D with high damping as shown in Table 47.10 for each case. Figure 47.79 is not a typical VIV-galloping response graph as K and ζ are different for each point on the power envelope. Due to the variation of K and ζ on the envelope, it is not easy to identify the various branches of FIM from Fig. 47.79.

The various branches are easily identified for this application in Figs. 47.80 and 47.81:

1. Re $<$ 30 000: No FIM takes place in this range experimentally or numerically.
2. 30 000 $\le$ Re $\le$ 35 000: This is the VIV initial branch for $K = 600\,\mathrm{N\,m^{-1}}$ (Fig. 47.4c). The upper end of the range changes dramatically depending on K (Fig. 47.4b). Under the condition of constant spring stiffness K and damping ratio ζ, the amplitude ratio increases rapidly with increasing Re [47.88, 113, 114]. For optimal power though, *Liu* and *Bernitsas* [47.177] showed using 2-D-URANS that the amplitude ratio changes from 0.65–0.87. This is due to the effect of spring stiffness going up from $K = 400\,\mathrm{N\,m^{-1}}$ at Re = 30 000 ($U^* = 5.50$) to $K = 800\,\mathrm{N\,m^{-1}}$ at Re = 60 000 ($U^* = 7.78$) and system damping C rising from $18.28\,\mathrm{N\,s\,m^{-1}}$ at Re = 30 000 to $24.86\,\mathrm{N\,s\,m^{-1}}$ at Re = 60 000 used in the numerical simulation.
3. 35 000–55 000 $\le$ Re $\le$ 85 000–95 000: This is the upper branch in VIV for $K = 600\,\mathrm{N\,m^{-1}}$ (Fig. 47.81). But as Fig. 47.79 shows, the upper and lower ends change widely depending on K 35 000–55 000 $\le$ Re $\le$ 85 000–95 000. As far as power harnessing is concerned, the numerical results in Fig. 47.77d shows that the amplitude ratio increases approximately linearly from the starting point 0.78 to the end point 1.41 of the upper branch with the rising of spring stiffness from $K = 800\,\mathrm{N\,m^{-1}}$ at Re = 60 000 ($U^* = 7.78$) to $K = 1800\,\mathrm{N\,m^{-1}}$ at Re = 100 000 ($U^* = 8.65$). System damping rises from $\zeta = 24.86\,\mathrm{N\,s\,m^{-1}}$ at Re = 60 000 to $\zeta = 36.09\,\mathrm{N\,s\,m^{-1}}$ at Re = 100 000 per Table 47.10. This increase is a characteristic of VIV in the TrSL3 regime not in TrSL2 [47.107].

4. $85\,000–95\,000 \le \mathrm{Re} \le 95\,000–105\,000$: This is the transition region from the VIV upper branch to galloping. The amplitude of a single cylinder with PTC decreases dramatically while the spring stiffness and system damping are held at the level of $K = 1800\,\mathrm{N\,m^{-1}}$ and $36.09\,\mathrm{N\,s\,m^{-1}}$, respectively, with the Reynolds number increasing form $\mathrm{Re} = 100\,000$ ($U^* = 8.65$) to $\mathrm{Re} = 105\,000$ ($U^* = 9.08$). Here the transition region from the VIV upper branch to galloping is different from that in Ding's studies, which is $80\,000 \le \mathrm{Re} \le 95\,000$ [47.113, 114]. The reason is that the velocity range between VIV and galloping requires additional resolution to find out if the lower branch in VIV has been overtaken by the instigation of galloping as explained thoroughly by *Park* et al. [47.70–72, 149]. See also Sect. 47.2.3, *Hydrodynamic Principles* under Principle 4. In this application, of all the parameters that can affect the separation between VIV and galloping, D, a, and m^* in (47.17) are constant. Thus, the separation between VIV and galloping is influenced by the change in K, ζ, and flow velocity (Figs. 47.80 and 47.81).
5. $\mathrm{Re} \ge 105\,000$: This is the galloping region. Galloping is known to be a high-amplitude and low-frequency FIM. The maximum amplitude of the single cylinder with PTC reaches $1.67 * D$ at $\mathrm{Re} = 120\,000$ ($U^* = 9.85$) in the CFD numerical simulation results.

Frequency Response. In Fig. 47.82, the frequency ratio $f_{\mathrm{osc}}/f_{\mathrm{n}}$, water of the PTC cylinder is plotted versus Reynolds number Re and flow velocity U. The oscillation frequency f_{osc} is calculated by FFT of the displacement of the cylinder over the recorded period.

The following observations can be made based on Figs. 47.82 and 47.83:

1. Frequency increases from point 1 to 2, which have the same stiffness and system damping as in experiments in Fig. 47.83. The decreasing part – points 2,3,4 – is mainly caused by the difference in spring stiffness and system damping.
2. At $\mathrm{Re} = 60\,000$ ($U^* = 7.78$), the VIV upper branch begins. With increasing spring stiffness K, the system's natural frequency in water increases from 1.11–1.67 and the frequency of oscillation rises from 1.18–1.81. But the frequency ratio remains nearly constant around 1.07, which is typical in VIV. As Re approaches 100 000, the frequency drop indicates transition to galloping. The frequency drop is accompanied by a rapid fall in the amplitude of the cylinder in agreement with the experimental results in Fig. 47.83.

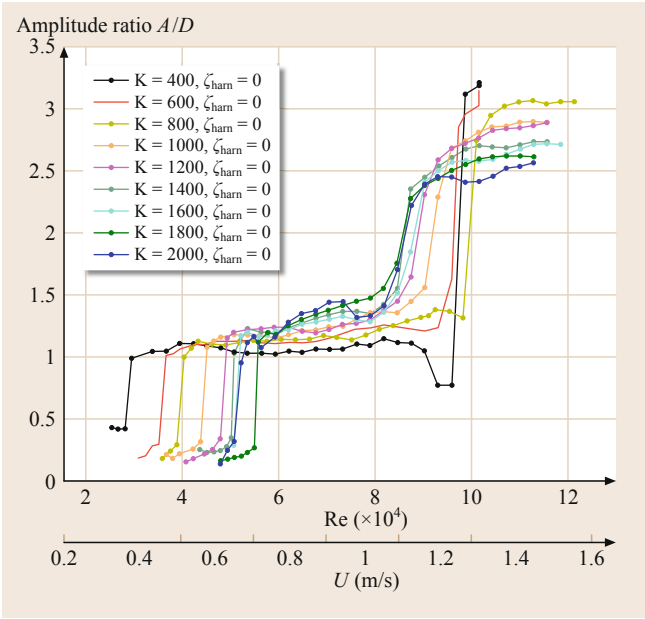


Fig. 47.80 Rough cylinder oscillatory responses at different stiffness values at no external damping (after [47.95])

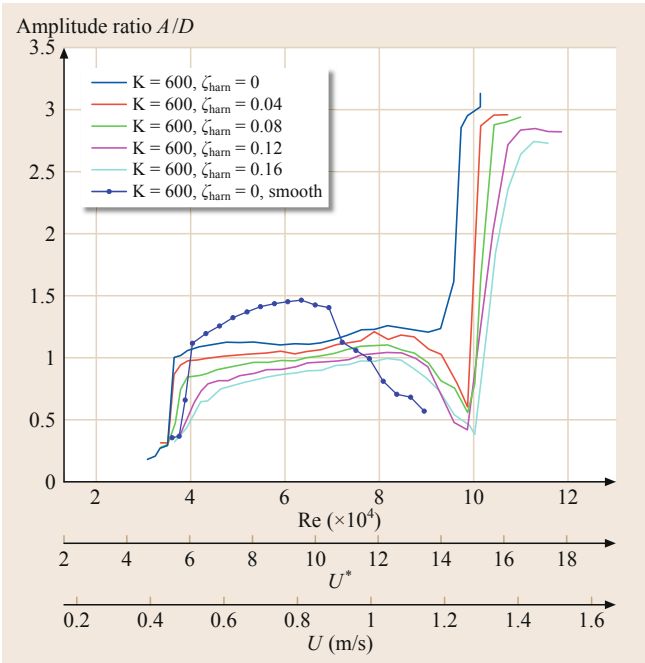


Fig. 47.81 Rough cylinder amplitude response at various damping values at $K = 600\,\mathrm{N\,m^{-1}}$ (after [47.95])

3. In the galloping branch for $\mathrm{Re} \ge 105\,000$ ($U^* = 9.08$), the frequency ratio reduction – characteristic of galloping – appears as a drop from 1.08–1.04. For the case of $\mathrm{Re} = 120\,000$ ($U^* =$

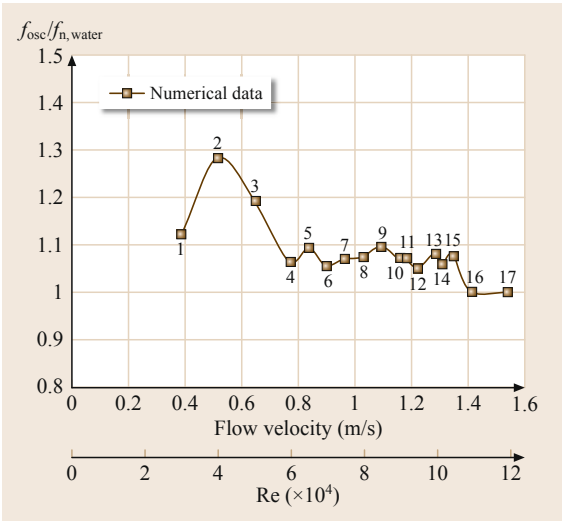


Fig. 47.82 Frequency ratio of one cylinder with PTC for the power envelope points in Table 47.10 (after [47.177])

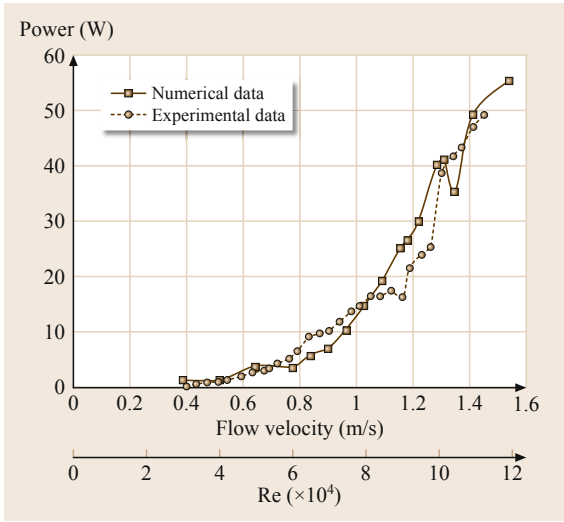


Fig. 47.84 Comparison of CFD optimal harnessed power envelope with experiments for single PTC-cylinder VIVACE (after [47.177])

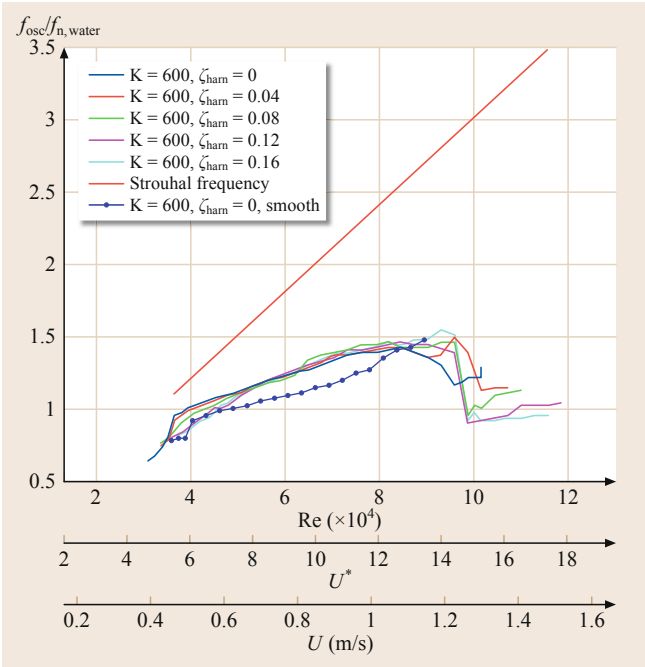


Fig. 47.83 Rough cylinder amplitude response at various damping values at $K = 600\,\mathrm{N\,m^{-1}}$ (after [47.95])

9.85) in CFD simulation and in the experiments the frequency of oscillation remains close to the system natural frequency; also a characteristic of galloping.

CFD versus Experimental Harnessed Power-Envelope. The power that can be harnessed by a single PTC-cylinder VIVACE converter is the power extracted by the generator, which is equal to the power that VIVACE extracts from the fluid minus the power dissipated by structural, transmission, and the internal generator losses. Based on the numerical results of amplitude and frequency response, the harnessed power P_w by the single PTC cylinder VIVACE system was calculated by *Liu* and *Bernitsas* [47.177] using the mathematical model presented above. The harnessed power envelope versus Reynolds number Re and flow speed (U), calculated by CFD, are depicted in Fig. 47.84, along with the experimental results by *Chang* [47.95]. The corresponding CFD results are presented in Table 47.10.

Power can be harnessed over a broad range of flow velocity from 0.38 to $1.54\,\mathrm{m\,s^{1-}}$. The upper limit is only due to limitations of the LTFSW channel. In the initial branch of VIV, the harnessed power of the single PTC cylinder VIVACE is very small and only 1.27 W is achieved at Re = 30 000 ($U^* = 5.50$). This is due not only to the fact that the available power in the fluid is small at such low speeds but also due to the low amplitude and frequency in the initial branch. With the increase in flow velocity, the power in the fluid increases by the cube of the velocity – as per (47.48) and at the same time the VIV upper branch begins, resulting in a steep increase in the energy conversion. This can be observed in the CFD calculations and the experimental measurements in Fig. 47.84. A second steep increase in converted power occurs at the beginning of galloping. In the present CFD results, it climbs from 35.53 W (Re = 105 000, $U^* = 9.08$) to 55.53 W (Re = 120 000, $U^* = 9.85$). In the experimen-

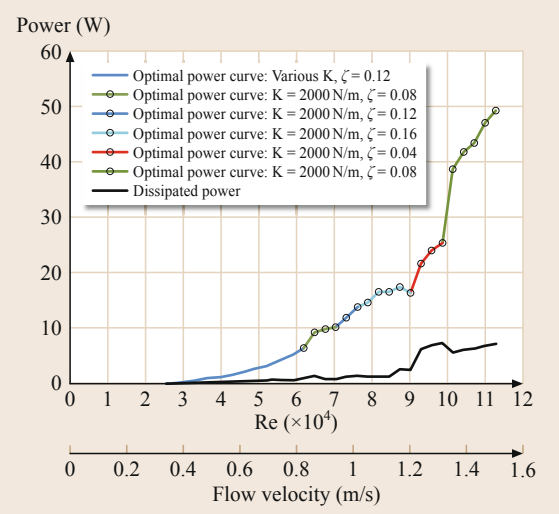


Fig. 47.85 Experimentally harnessed power envelope and corresponding dissipated power with information of system stiffness and damping (after [47.95])

tal measurements, converted power rises from 16.37 W (Re = 90 667, $U^* = 7.43$) to 49.35 W Re = 113 076, $U^* = 9.29$). The harnessed power envelope calculated by 2-D-URANS simulation is close to that obtained by experiments. The small discrepancy is mainly due to two reasons. First, the difference between the damping models used in the numerical simulation and the experimental apparatus (Sect. 47.3.5, *FIM of Two PTC Cylinders in Tandem*). Second, in the experiments conducted in the LTFSW channel in the MRELab, the cylinder would hit the safety stops limiting its travel in galloping, which leads to the smaller amplitude in the tests than in the simulations.

Even when the values differ a little, the slope of the two graphs in Fig. 47.84 are similar rendering the graphs to have the same trend and thus capturing the important variations in power conversion versus flow speed.

Based on the comparison of experimental and numerical results in Fig. 47.84, we can draw two conclusions:

1. The experimental measurements were conducted for variable spring stiffness K. The optimal points on the power envelope, though, were reached for $K = 2000\,\mathrm{N\,m^{-1}}$ and variable harnessed damping ratio from 0.04 to 0.14 corresponding to different Reynolds numbers or flow velocity (Fig. 47.85). The numerical results were acquired based on fixed harnessed damping ratio $\zeta = 0.12$ by changing spring stiffness K from 400 to $2000\,\mathrm{N\,m^{-1}}$ corresponding to different Reynolds numbers or flow velocity. The simulation results of the harnessed power envelope match very well with the experimental measurements. This demonstrates that optimal points – lying on the power envelope of single PTC-cylinder VIVACE converter – can be achieved in two different ways, for two different pairs of (K, ζ). For example, at Re = 75 000 ($U^* = 7.94$), the harnessed power is 10.49 W for ($K = 1200\,\mathrm{N\,m^{-1}}$, $C = 29.91\,\mathrm{N\,s\,m^{-1}}$), and 10.43 W for ($K = 1036\,\mathrm{N\,m^{-1}}$). The difference in harnessed power between these two pairs is only 0.57%.
2. The experimental results exhibit a local reduction at Re = 90 667 ($U^* = 7.43$). The numerical results exhibit a deep at Re = 105 000 ($U^* = 9.08$). This phenomenon indicates that transition between VIV and galloping can be changed by adjusting the spring stiffness and harnessed damping ratio of the PTC cylinder.

Post-Processing of Experimental or Numerical Simulation Results

In the MRELab, data collected through experimental measurements or CFD simulations are post-processed to extract useful information. Actually, five different mathematical models are used providing complementary information. All five are based on (47.43) and are differentiated based on the simplifying assumptions. They are defined as follows:

1. Math Model #1a [47.119]:
   - The ideal added mass cannot be separated from the hydrodynamic force
   - The response of the oscillator in FIM can be approximated by a sinusoidal function.
2. Math Model #1b [47.119]:
   - The ideal added mass can be separated from the hydrodynamic force
   - The response of the oscillator in FIM can be approximated by a sinusoidal function.
3. Math Model #2a [47.120]:
   - The ideal added mass cannot be separated from the hydrodynamic force
   - The response of the oscillator in FIM cannot be approximated by a sinusoidal function.
4. Math Model #2b [47.120]:
   - The ideal added mass can be separated from the hydrodynamic force
   - The response of the oscillator in FIM cannot be approximated by a sinusoidal function.
5. Math Model #3 [47.121]:
   - It is based on the approach by *Vikestad* et al. [47.60], where the added mass is calculated from the available data. It is further described in Sect. 47.3.7.

Reconstruction of the Lift Force. A key element in the mathematical post-processing of data with any of

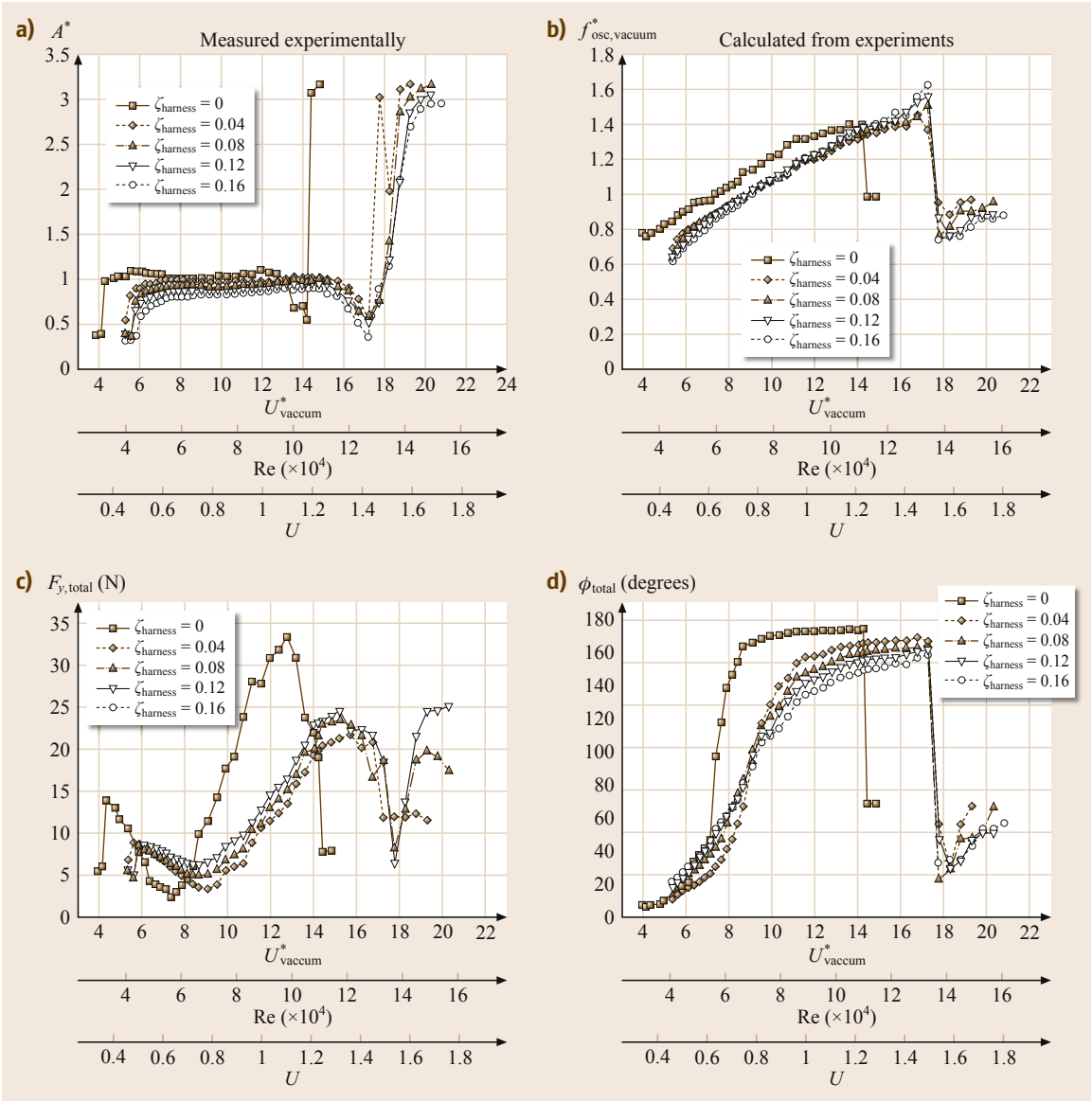


Fig. 47.86a–d Example of post-processing of experimental or CFD data for a given set of runs versus flow speed (after [47.119, 120]). **(a)** Amplitude ratio versus U^*, Re, U. **(b)** Frequency ratio versus U^*, Re, U. **(c)** Total reconstructed transverse force versus U^*, Re, U. **(d)** Total force phase lead over the displacement versus U^*, Re, U

the five mathematical models is the availability of the time history of the lift force. That can be obtained by a load cell or by reconstruction of the force from the displacement time history. Specifically the following steps are used at the MRELab to reconstruct the lift force time history:

1. The displacement is measured through the encoder of the motor of the V_{ck} controller.
2. The time derivative of the displacement is taken using the backward difference.
3. Low-pass filtering is used in MATLAB to eliminate the noise that typically comes with digital differentiation.
4. The phase shift induced by the filtering is eliminated by a MATLAB code.
5. The process is repeated for the second derivative. The accuracy of the methodology has been tested

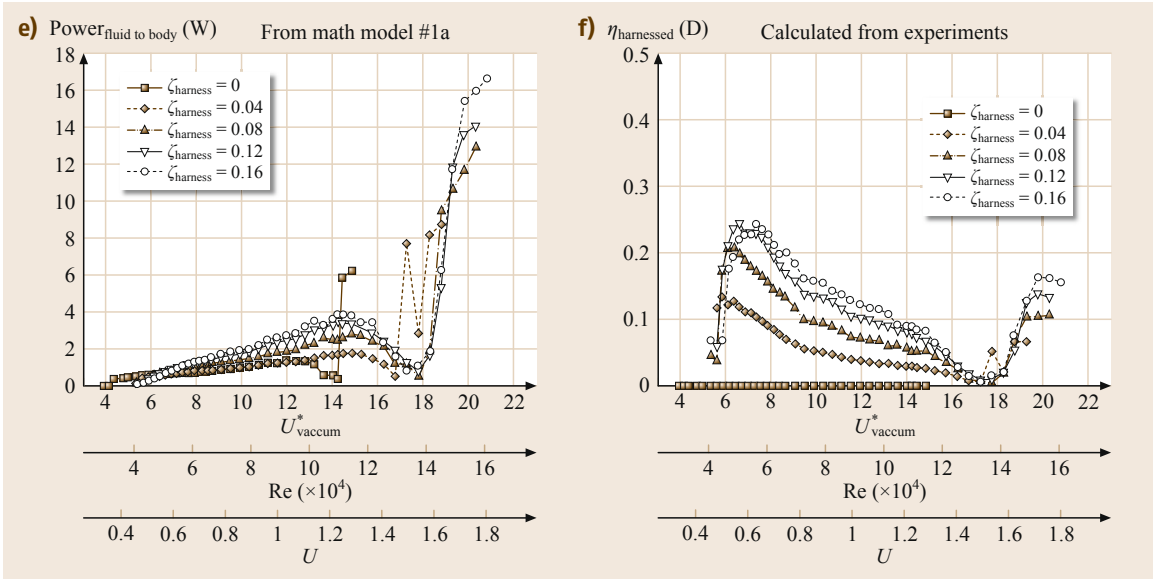


Fig. 47.86e,f **(e)** Power extracted from fluid to body versus U^*, Re, U. **(f)** Total hydrodynamic force versus U^*, Re, U

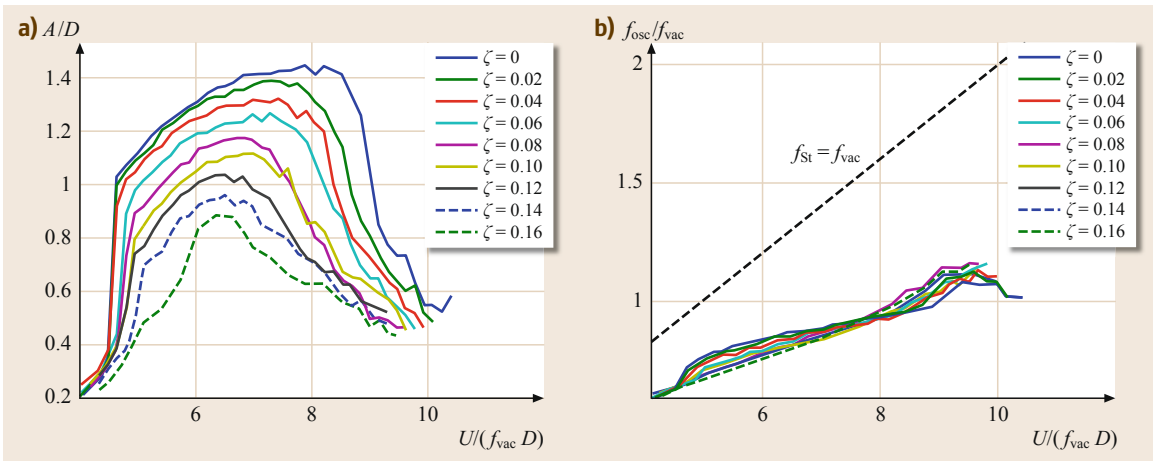


Fig. 47.87a,b Post-processing data by *Lee* and *Bernitsas* [47.100] using the variable added mass mathematical model [47.121]. **(a)** A/D for smooth cylinder in VIV versus **(b)** U^*. f_{osc}/f_{vacuum} for smooth cylinder in VIV versus U^*

repeatedly and the signal reproduction has been excellent.

6. The calculated derivatives along with the values of the coefficients in (47.43) are used to reconstruct the lift force at each time step.

Example of Post-Processing. Some of the results of this post-processing procedure are presented in Fig. 47.86. Specifically, in Fig. 47.86a, the peak amplitude ratio is calculated from the time history, for each velocity, by averaging the highest 40–60 peaks over a time interval of 60 s. Using FFT, the frequency is calculated in Fig. 47.86b. The lift force is reconstructed in Fig. 47.86c following the procedure outlined above. The force phase lead with respect to the displacement is shown in Fig. 47.86d. Generated power and efficiency are shown in Fig. 47.86e,f respectively.

47.3.7 Variable Added Mass Mathematical Model

Based on the procedure presented in Sect. 47.3.6, *Post-Processing of Experimental or Numerical Simulation Results* the lift-force time history can be reconstructed.

A time-averaged variable added mass is then calculated by averaging all periods of oscillation for a given data set. Starting with (47.43) for a cylinder oscillatory system in FIM, VIV or galloping, the following procedure can be used to post-process numerical or experimental data in yet another way.

If the lift force F_{fluid} in (47.43) is assumed to be sinusoidal, shifted in phase from the response of the body in VIV, the following substitutions can be made for $F_{\text{fluid}}(t)$ and $y(t)$

$$F_{\text{fluid}}(t) = F_{\text{o}} \sin(\omega t + \phi) \,, \tag{47.64}$$

$$y(t) = y_{\text{o}} \sin(\omega t) \,. \tag{47.65}$$

This allows (47.43) to be written as

$$\left(m_{\text{osc}} + \frac{F_{\text{o}} \cos\phi}{\omega^2 y_{\text{o}}}\right) \ddot{y} + \left(C_{\text{total}} - \frac{F_{\text{o}} \sin\phi}{\omega y_{\text{o}}}\right) \dot{y} + ky = 0 \,. \tag{47.66}$$

The added mass coefficient C_{a} can be calculated then based on the forcing function as

$$C_{\text{a}} = -\frac{8}{nTm_{\text{d}}(\omega^2 y_{\text{o}})^2} \int_{t}^{t+nt} F_{\text{fluid}}(t)\ddot{y}\mathrm{d}t \,. \tag{47.67}$$

The corresponding natural frequency in water can be calculated, then, as

$$f_{\text{n,water}} = \frac{1}{2\pi} \sqrt{\frac{K}{m_{\text{osc}} + m_{\text{d}} C_{\text{a}}}} \,. \tag{47.68}$$

This analysis is straightforward and was presented by *Vikestad* et al. [47.60].

In order to test the idea that a time-averaged variable added mass coefficient better captures VIV with the resonance model, previously collected VIV data by *Lee* et al. [47.98] are post-processed below and plotted in Figs. 47.87–47.89. The particulars of the VIVACE system used for analysis are given in Table 47.11. Figure 47.88 shows that the added mass coefficient may take negative values. On the other hand, the frequency of oscillation becomes equal to the calculated natural frequency in water over the entire range of VIV synchronization.

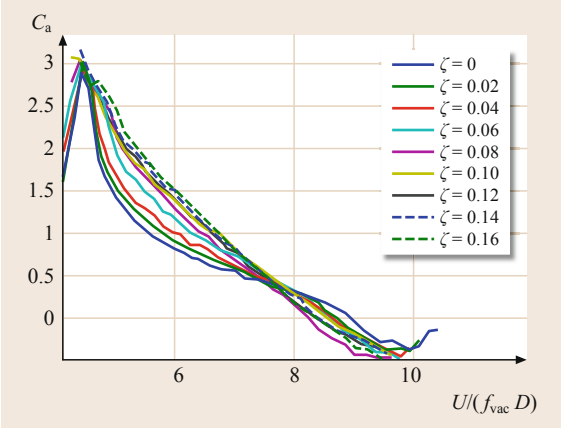


Fig. 47.88 Calculated added mass coefficient for smooth cylinder in VIV versus U^*

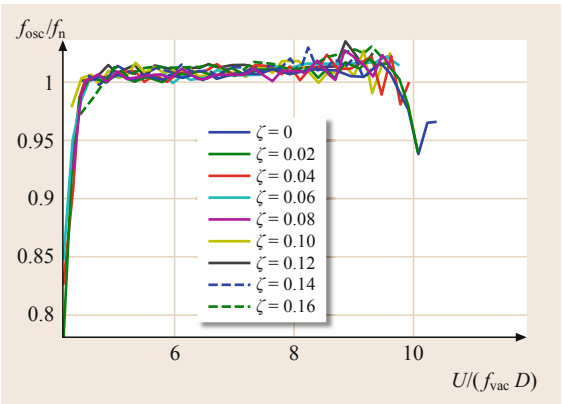


Fig. 47.89 Calculated $f_{\text{osc}}/f_{\text{n}}$, water for smooth cylinder in VIV versus U^*

Table 47.11 Properties of oscillator

Parameter	Value
m_{osc} [kg]	10.94
D [m]	0.0889
L [m]	0.9144
ρ [kg m^{-3}]	999.103
K [N m^{-1}]	400–1800
ζ	0–0.16

47.3.8 Physical Models and Model Equivalence

A major challenge that the MRELab has been addressing in its research on MHK energy conversion using FIM is the reduction of data and models to a common equivalence. That is a challenge within the models in the same laboratory and of course it becomes even more of a challenge when processing data from different laboratories. This challenge has multiple facets:

1. Springs are not necessarily linear even when they are used for the first time
2. Viscous damping is by no means linear
3. Flows perpendicular to the cylinder may not be fully two-dimensional even in VIV due to tip-flow effects

Table 47.12 Experimental parameters of two models

Device parameter	VIV oscillator	VIVACE converter
$m*$	1.725	1.67
$f_{n,water}$	1.1183	1.0141
ζ	0.0158	0.02

4. FIM models are designed totally different from MHK-energy converters and thus results would not be comparable.

The first two cases have been handled in the MRELab using the V_{ck} controller, which simulates a mathematically accurate oscillator with whatever linear or nonlinear function for damping and spring stiffness desired. The third challenge is assessed next in Sect. 47.3.8, *Tip-Flow Effect*. The last challenge is major and is only stated in Sect. 47.3.8, *Experimental Model Equivalence*.

Tip-Flow Effect

Tip flow limits the cylinder length over which positive lift is exerted in the transverse direction when the cylinder is in FIM. Negative lift is typically exerted on the rest of the cylinder length, which is affected by the tip flow. Figure 47.90 shows schematically the first LTFSW channel setup with the VIV oscillator and the VIVACE converter as set in the new LTFSW channel. The parameters of the oscillators used in the experiments in both setups are summarized in Table 47.12. Although the parameters are not exactly equal, it is believed that they are sufficiently close to illustrate quantitatively the effect of the tip flow.

Figure 47.91 compares the results of tests of the VIV oscillator versus the VIVACE converter. The maximum amplitude and the range of synchronization obtained with the converter setup are significantly smaller than those obtained by *Park* et al. [47.70, 71, 149] for the VIV oscillator. This is attributed to the tip-flow effects evaluated by *Kinaci* et al. [47.178].

This effect is different in each oscillator and has to be evaluated before FIM data can be considered comparable. The tip-flow effects may become dramatic as in the field tests in canals in the Netherlands with the converter shown in Fig. 47.20.

The VIVACE converter model for energy extraction includes all the conversion equipment and the supporting struts underwater resulting in high tip-flow effects. It allows, though, for up to seven-diameter amplitude due to the bigger depth of the test section of the new channel (Fig. 47.92). Measurements of the setup are provided in the same figure. The struts are inside the water and take a considerable amount of space of the test cross-section. The width of the channel is $w_c =$ 100.8 cm with 3.8 cm wide struts at each side. The struts

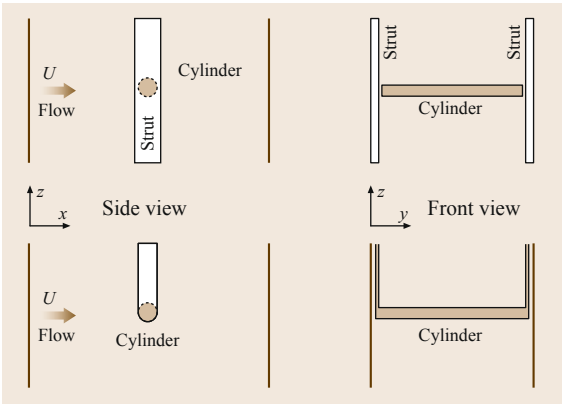


Fig. 47.90 Schematic view of the VIVACE converter used in the new channel and the VIV oscillator used in the first LTFSW channel (after [47.178])

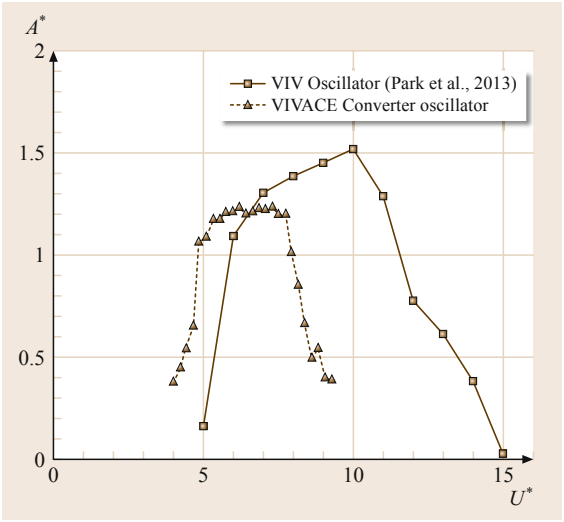


Fig. 47.91 Comparison of VIV response of VIV oscillator and converter oscillator with parameters shown in Table 47.1 (after [47.71, 72, 178])

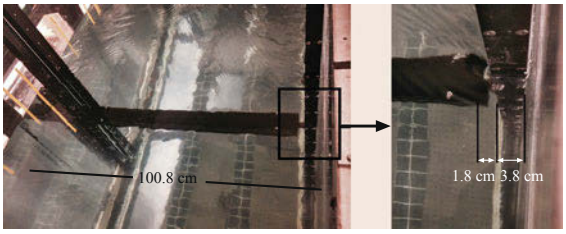


Fig. 47.92 VIVACE converter model in the new LTFSW channel (photo courtesy of MRELab)

are attached to the plexiglass side-walls of the channel. There is 1.8 cm of free space between the strut and the cylinder at each side. The length of the cylinder is

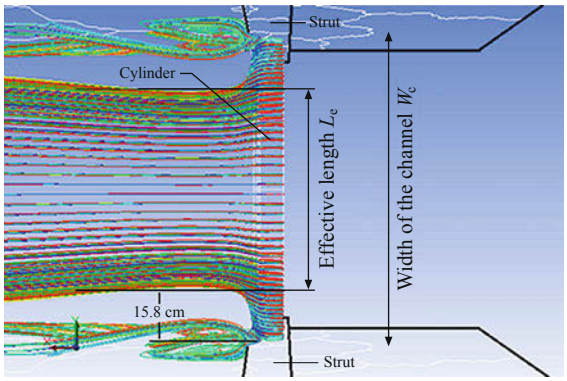


Fig. 47.93 Effective cylinder length from 3-D-URANS simulations for the new LTFSW channel at Re = 75 000 (after [47.176])

89.6 cm. That is, the converter oscillators serve a different purpose than the VIV oscillators in the first channel. Comparison between these two extreme cases of tip-flow effect provides valuable information on tip flow.

The effective length of the cylinder, which is generating lift, can be detected by studying the streamlines produced by the 3-D steady CFD analysis (Fig. 47.93). It is concluded that 15.8 cm of the cylinder at each end do not contribute to generating lift, reducing the effective length of the cylinder. So the effective length, L_e becomes

$$L_e = 100.8\,\text{cm} - 2(3.8\,\text{cm} + 1.8\,\text{cm} + 15.8\,\text{cm}) = 58\,\text{cm}\,. \quad (47.69)$$

Substituting the effective length of the cylinder L_e and the channel width w_c into (47.17), a tip-flow effect factor α, for this experimental setup, can be calculated as

$$\alpha = 0.5752\,. \quad (47.70)$$

Experimental Model Equivalence

The five blockage effects discussed in Sect. 47.3.1, *First Low-Turbulence Free-Surface Wake (LTFSW) Channel* and the four factors listed at the beginning of Sect. 47.3.8 result in sources of variability in the data collected about FIM. For proper comparison of data collected from different labs, different experimental devices, and particularly from diverse sources, such as FIM oscillators, ALT converters, or flexible risers, a major research effort would be required, which is beyond the scope of this review study.

47.3.9 Benchmarking

Benchmarking is difficult at best given that different technologies, and particularly MHK energy devices, have different ways of assessing their performance. More recently instead of efficiency data, comparisons have been based on power density ratios, such as power-to-weight or power-to-volume data. Table 47.13 shows data from wave-energy devices and data from recent VIVACE converter calculations. Conditions of course are very different between wave fields and ocean currents. In addition, it is debatable whether ballast should be part of the weight considered or volume includes space occupied by the structure only or its anchoring or mooring system as well since those do not allow alternative use of their space.

The advantage that the VIVACE converter has besides being able to operate efficiently even in low speed flows is the ability of cylinders to work synergistically generating more power in a properly deigned compact array rather than in isolation (Fig. 47.29). The gain from the hydrodynamic interaction is significant but the weight reduction due to the compact size of the converter has a bigger positive effect on the power density of VIVACE.

Table 47.13 Power density for wave-energy devices and the VIVACE converter (after [47.128])

Device	kW	m^2	Mg	kW/m^3	kW/Mg
F-30F	108.8	2160	1622	0.0059	0.0671
B-HBA	3090.4	4350	1600	0.0055	0.1933
Bref-SHB	19.86	220	200	0.0045	0.0993
F-OWC	371.6	6500	1800	0.0044	0.2064
Bref-HB	3.24	42	31	0.002	0.1045
B-OF	498.6	2020	3800	0.0019	0.1312
F-2HB	338.6	4750	5233	0.0008	0.0647
# Devices	VIVACE [Mg]	Concrete [Mg]	kW	kW/Mg VIVACE + base	kW/Mg VIVACE no-base
1	0.2267	8.165	1.2	0.143	5.29
2	0.4535	8.165	2.9	0.336	639
3	0.6803	8.165	4.2	0.475	6.17
4	0.9072	8.165	6.0	0.661	6.61

Closing Remarks

Horizontal MHK energy, encountered in rivers, ocean currents, and tides, is an abundant resource. Current technology is not capable of accessing efficiently and at a competitive cost the vast majority of the horizontal MHK energy because ocean currents are typically slower than 3 kn and rivers are slower than 2 kn. The MHK-energy industry has to overcome several challenges in technology development, market, and regulatory challenges before it can reach maturity. MHK energy harnessing technologies need research breakthroughs to improve their power-to-weight ratio before they can become competitive with other renewable energy technologies. Steady-lift technologies (SLT) are turbines and dominate the efforts to harness horizontal hydrokinetic energy. They are in need of a breakthrough to reduce their minimum flow speed threshold of $2\,\mathrm{m\,s^{-1}}$.

Alternating-lift technologies (ALTs) provide an alternative means to harness horizontal MHK energy. ALTs are in the form of an oscillator taking advantage of fluid–structure interaction instabilities. They come in two forms: hydrofoils in flutter instability or bluff bodies in vortex-induced oscillations, galloping, or coexistence of the two. In multiple body ALTs, gap flow may be used to further enhance the oscillator dynamics. Vortex streets and multibody interactions provide alternating lift and are used by fish to propel efficiently in water. Thus, ALT dynamics is natural in the marine environment. Both mechanisms are highly scalable occurring practically over the entire range of Reynolds numbers. By employing vortex streets, triggering galloping instability, and even superposition of VIV and galloping, nonlinear oscillators can be built using the naturally occurring alternating lift in all steady flows from very slow to very fast. Interference between two-dimensional bodies with bluff cross-section in close proximity can induce much more vigorous FIM response of the same bodies in isolation. That enables increased hydrokinetic energy conversion to mechanical energy in an oscillator, while at the same time reducing the space occupied by the MHK energy conversion device.

The VIVACE converter is probably closest to commercialization among ALTs and has the following properties:

1. It is based on alternating lift naturally occurring in the marine environment
2. The underlying physical mechanisms are highly scalable enabling the conversion of MHK energy to mechanical even at very slow flows
3. Multiple body interference enables higher conversion of MHK energy to mechanical in an oscillator
4. Proximity of bodies reduces occupied space by VIVACE resulting in higher power density of conversion
5. By properly configuring multiple cylindrical oscillators in close formation, a three-dimensional converter can be designed to have a positive impact on the power-to-volume ratio, which is the Achilles heel of all renewable energy devices.

Development of new MHK energy technologies requires extensive basic research using large experimental facilities like towing tanks, recirculating channels, control mechanisms to ensure a mathematically accurate oscillator model without including the hydrodynamic force in the closed loop, flow visualization, and computational fluid dynamics.

The MHK-energy industry has not reached the maturity level, where it can contribute to a national or international renewable energy portfolio. The next 5–10 years will be a critical period for development of new MHK energy technologies. If the industry does not succeed in demonstrating a robust and reliable SLT or ALT, which can compete with other renewable and conventional energy technologies on the basis of LCOE, governments and private investors alike may lose interest in this vast source of energy.

47.4 Nomenclature

Latin symbols:

- A: Amplitude value of the transverse displacement
- A^*: Amplitude ratio A/D, average of the 60 highest peaks
- ALT: alternating-lift technology
- c, C_{total}: Total damping
- c_{bearing}: Friction damping
- c_{harness}: Harnessing damping
- C_{a}: Potential flow added-mass coefficient of a circular cylinder ($C_{\mathrm{a}} = 1.0$)
- $C_{\mathrm{Y,total}}(t)$: Instantaneous lift force coefficient
- $C_{\mathrm{y,total}}$: Lift force coefficient
- $C_{\mathrm{Y,vortex}}(t)$: Instantaneous vortex force coefficient
- $C_{\mathrm{y,vortex}}$: Vortex force coefficient
- d: Spacing between cylinders in tandem
- D: Diameter of cylinder
- $f_{\mathrm{n,vacuum}}$: Natural frequency in vacuum $= f_{\mathrm{n,vacuum}} = 1/(2\pi)\sqrt{k/m_{\mathrm{osc}}}$
- $f_{\mathrm{n,water}}$: Natural frequency in water $= f_{\mathrm{n,water}} = 1/(2\pi)\sqrt{k/(m_{\mathrm{osc}} + m_{\mathrm{a}})}$
- f_{osc}: Body oscillation frequency

- $f^*_{\mathrm{osc,vacuum}}$: Vacuum oscillation frequency ratio $= f^*_{\mathrm{osc,vacuum}} = f_{\mathrm{osc}}/f_{\mathrm{n,vacuum}}$
- $f^*_{\mathrm{osc,water}}$: Water oscillation frequency ratio $= f^*_{\mathrm{osc,water}} = f_{\mathrm{osc}}/f_{\mathrm{n,water}}$
- $F_{\mathrm{Y,potential}}$: Instantaneous potential force in transverse direction
- $F_{\mathrm{Y,total}}$: Instantaneous total force in transverse direction
- $F_{\mathrm{Y,vortex}}$: Instantaneous vortex force in transverse direction
- k: Average grit height
- k, K, K_{virtual}: Spring constant
- L: Cylinder length
- m_{osc}: Oscillating mass
- m_{a}: Added mass
- m_{d}: Displaced fluid mass
- m^*: Mass ratio $m^* = m_{\mathrm{osc}} m_{\mathrm{d}}$
- P: Thickness of sand paper
- $P_{\mathrm{F-B}}$: Hydrokinetic power converted to body mechanical power
- $P_{\mathrm{F(D)}}$: Hydrokinetic power in fluid through height D
- $P_{\mathrm{F(2A+D)}}$: Hydrokinetic power in fluid through height $2A + D$
- $P_{\mathrm{dissipated}}$: Dissipated power P_{fluid} power in the fluid
- $P_{\mathrm{harnessed}}$: Harnessed hydrokinetic power converted to body mechanical power
- PTC: passive turbulence control
- Re: $= UD/\nu$ Reynolds number
- S_{ij}: Mean strain-rate tensor
- St: Strouhal number
- t: Time
- T: Total thickness of strip
- T_{osc} : A period of cylinder oscillation
- U: Flow velocity
- U^*_{vacuum} : Reduced velocity $= U/(f_{\mathrm{n,vacuum}} D)$
- U^*_{water}: Reduced velocity $= U/(f_{\mathrm{n,water}} D)$
- $y(t)$: Instantaneous cylinder motion in transvers direction.

Greek symbols:

- ζ_{vacuum}: $c/(2\sqrt{m_{\mathrm{osc}} k})$ Total damping ratio including friction and power harnessing
- ζ_{water}: $c/(2\sqrt{k(m + m_{\mathrm{a}})})$ Total damping ratio including friction and power harnessing considering added mass
- $\eta_{(\mathrm{D})}$: $\eta_{\mathrm{D}} = P_{\mathrm{F-B}}/P_{\mathrm{F(D)}}$ Efficiency with respect to fluid of height D
- $\eta_{(2A+D)}$: $\eta(2A + D) = P_{\mathrm{F-B}}/P_{\mathrm{F(2A+D)}}$ Efficiency with respect to fluid of height $2A + D$
- $\eta_{\mathrm{harnessed}(D)}$: $\eta_{\mathrm{harnessed}(D)} = P_{\mathrm{harnessed}}/P_{\mathrm{F(D)}}$ efficiency with respect to fluid of height D
- $\eta_{\mathrm{harnessed}(2A+D)}$: $\eta_{\mathrm{harnessed}(2A+D)} = P_{\mathrm{harnessed}}/P_{\mathrm{F(2A+D)}}$ efficiency with respect to fluid of height $2A + D$
- μ: Dynamic viscosity of water
- μ_{t}: Turbulence eddy viscosity
- ν: Fluid kinematic viscosity
- ρ: Water density
- τ_{ij}: Reynolds-stress tensor
- ϕ_{total}: Phase angle between total lift force and cylinder motion
- ϕ_{vortex}: Phase angle between vortex lift force and cylinder motion
- $\omega_{\mathrm{n,vacuum}}$: Undamped natural frequency of the system in vacuum
- $\omega_{\mathrm{n,water}}$: Undamped natural frequency of the system in water
- ω_{osc}: Angular frequency of oscillation.

References

47.1 NREL: Renewable electricity futures study, Vol. 2, Renewable electricity generation and storage technologies, http://nrel.gov/analysis/re-futures/ (2012)

47.2 EERE-DOE: Assessment of energy production potential from tidal streams in the United States, Final project report, http://energy.gov/eere/water/downloads/assessment-energy-production-potential-tidal-streams-united-states (2011)

47.3 EERE-DOE: Mapping and assessment of the United States ocean wave energy resource, http://energy.gov/eere/water/downloads/mapping-and-assessment-united-states-ocean-wave-energy-resource (2011)

47.4 EERE-DOE: *An Evaluation of the US Department of Energy's Marine and Hydrokinetic Resource Assessments* (Marine and Hydrokinetic Energy Technology Assessment Committee, Board on Energy and Environmental Systems, National Research Council, Washington 2013)

47.5 EERE-DOE: *Marine and Hydrokinetic Energy Projects* (US Department of Energy, Wind and Water Power Technologies Office, Washington 2013), DOE/EE-0924

47.6 EERE-DOE: Ocean tidal power, http://www.eere.energy.gov/consumer/renewable_energy/ocean/index.cfm/mytopic=50008 (2015)

47.7 EERE-DOE: Ocean wave power, http://www.eere.energy.gov/consumer/renewable_energy/ocean/index.cfm/mytopic=50009 (2015)

47.8 Renewables 2015 Global Status Report, (REN21 Secretariat, Paris 2007)

47.9 REN21: Global Status Report 2014, http://www.ren21.net/states-of-renewables/global-stocking-report (2014)

47.10 NREL: Levelized cost of energy calculator, http://www.nrel.gov/analysis/tech_lcoe.html (2015)

47.11 C. Matlack: Hopes dim for renewable power from ocean waves and tides, Bloomberg Businessweek, http://www.businessweek.com/articles/2014-12-09/hopes-dim-for-renewable-power-from-ocean-waves-and-tides (2014)

47.12 J. Falnes: *Ocean Waves and Oscillating Systems* (Cambridge Univ. Press, Cambridge 2002)

47.13 J. Brooke: *Wave Energy Conversion, Elsevier Ocean Engineering Book Series*, Vol. 6 (Elsevier, Amsterdam 2003)

47.14 Commission of the European Committee: Wave energy project results: The exploitation of tidal marine currents, http://www.eere.energy.gov/consumer/renewable_energy/ocean/index.cfm/mytopic=50010 (1996) Report EUR16683EN

47.15 European Ocean Energy Association: Wave energy technology http://www.eu-oea/index.asp?bid=232 (2015)

47.16 R. Boud: *Status and Research and Development Priorities 2003: Wave and Marine Current Energy*, DTI Rep. FES-R-132 (UK Department of Trade and Industry, London 2003)

47.17 R. Bedard, O. Siddiqui, M. Previsic, B. Polagye: *Economic Assessment Methodology for Tidal In-Stream Plants*, Tech. Rep. EPRI-TP-002NA (Electric Power Research Institute, Palo Alto 2006)

47.18 R. Bedard, M. Previsic, G. Hagerman, B. Polagye, W. Musial, J. Klure, A. von Jouanne, U. Mathur, J. Partin, C. Collar, C. Hopper, S. Amsden: North American ocean energy status, Proc. 7th Eur. Wave Tidal Energy Conf.'07 (2007)

47.19 K.A. Haas, H. Fritz, S. French, B. Smith, V. Neary: *The Assessment of Energy Production Potential From Tidal Streams in the United States*, Tech. Rep. (Georgia Tech Research Corporation, Atlanta 2011)

47.20 M.J. Khan, M.T. Iqbal, J.E. Quaicoe: River current energy conversion systems: Progress, prospects and challenges, Renew. Sustain. Energy Rev. **12**, 2177–2193 (2008)

47.21 M.J. Khan, G. Bhuyan, M.T. Iqbal, J.E. Quaicoe: Hydrokinetic energy conversion systems and assessment of horizontal and vertical axis turbines for river and tidal applications: A technology status review, Appl. Energy **86**, 1823–1835 (2009)

47.22 D.G. Hall, K.S. Reeves, J. Brizzee, R.D. Lee, G.R. Carroll, G.L. Sommers: *Feasibility Assessment of the Water Energy Resources of the United States for New Low Power and Small Hydro classes of Hydroelectric Plants*, Tech. Rep. DOE-ID-11263 (Idaho National Lab, Idaho Falls 2006)

47.23 P. Jacobson: *Assessment and Mapping of Riverine Hydrokinetic Resource in Continental United States* (Electric Power Research Institute, Palo 2012)

47.24 A.S. Bahaj: Generating electricity from the oceans, Renew. Sustain. Energy Rev. **15**, 3399–3416 (2011)

47.25 A.S. Bahaj: Marine current energy conversion: The dawn of a new era in electricity production, Phil. Trans. R. Soc. **371**, 20120500 (2012)

47.26 R. Bedard: *Prioritized Research, Development, Deployment and Demonstration (RDD and D) Needs: Marine and Other Hydrokinetic Renewable Energy* (Electric Power Research Institute, Palo Alto 2008)

47.27 EPRI: http://www.epri.com/Our-Work/Pages/Renewables.aspx (2015)

47.28 E. Bachynski, M.M. Bernitsas, M.M. Brown, S. Coblitz, S. Dole, H.Y. Kim, T. Witkin, P.E. Bernitsas, E. Fassezke: *Marine Renewable Energy: Potential Contribution to the World's Energy Production and Consumption* (MERLab Dept. of NA&ME, University of Michigan, Ann Arbor 2013)

47.29 International Energy Agency: World energy outlook 2012, http://www.iea.org (2013)

47.30 International Renewable Energy Agency: Renewable power generation costs 2012, http://www.irena.org (2013)

47.31 A.S. Bahaj, L.E. Myers: Fundamentals applicable to the utilization of marine current turbines for energy production, Renew. Energy **28**, 2205–2211 (2003)

47.32 Gorlov: Helical turbine http://en.wikipedia.org/wiki/Gorlov_helical_turbine (2015)

47.33 Underwater Ocean Trbines: http://inhabitat.com/underwater-power-generating-ocean-turbines/ (2015)

47.34 R.W.J. Lacey, V.S. Neary, J.C. Liao, E.C. Enders, H.M. Tritico: The IPOS framework: Linking fish swimming performance in altered flows from laboratory experiments to rivers, River Res. **28**, 429–443 (2011)

47.35 J.C. Liao: A review of fish swimming mechanics and behaviour in altered flows, Phil. Trans. R. Soc. B. **362**, 1973–1993 (2007)

47.36 J.C. Liao, D.N. Beal, G.V. Lauder, M.S. Triantafyllou: Fish exploiting vortices decrease muscle activity, Science **302**, 1566–1569 (2003)

47.37 J.C. Liao, D.N. Beal, G.V. Lauder, M.S. Triantafyllou: The Ka'rma'n gait: Novel kinematics of rainbow trout swimming in a vortex street, J. Exp. Bio. **206**, 1059–1073 (2003)

47.38 A. Ojima, K. Kamemoto: Numerical simulation of unsteady flows around a fish, Proc. 3rd Int. Conf. Vortex Flows Vortex Models, Yokohama'05 (2005)

47.39 M.S. Triantafyllou, G.S. Triantafyllou, D.K.P. Yue: Hydrodynamics of fishlike swimming, Annu. Rev. Fluid Mech. **32**, 33–53 (2000)

47.40 G.S. Triantafyllou, A.H. Techet, Q. Zhu, D.N. Beal, F.S. Hover, D.K.P. Yue: Vorticity control in fish like propulsion and maneuvering, Integr. Comp. Biol. **42**, 1026–1031 (2002)

47.41 G.S. Triantafyllou, M.S. Triantafyllou: An efficient swimming machine, Sci. Am. **272**(3), 64–70 (1995)

47.42 M.J. Wolfgang, J.M. Anderson, M.A. Grosenbaugh, D.K.P. Yue, M.S. Triantafyllou: Near-body flow

dynamics in swimming fish, J. Exp. Biol. **202**, 2303–2327 (1999)
47.43 R.D. Blevins: *Flow-Induced Vibration* (Krieger, Florida 1990)
47.44 M.P. Paidoussis, S.J. Price, E. de Langre: *Fluid-Structure Interactions* (Cambridge Univ. Press, Cambridge 2011)
47.45 J. Guckenheimer, P. Holmes: *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, Applied Mathematical Sciences, Vol. 42 (Springer, Berlin, Heidelberg 2002)
47.46 P.W. Bearman: Vortex shedding from oscillating bluff bodies, Annu. Rev. Fluid Mech. **16**, 195–222 (1984)
47.47 P.W. Bearman: Circular cylinder wakes and vortex-induced vibrations, J. Fluid Mech. **27**, 648–658 (2011)
47.48 M.M. Bernitsas, K. Raghavan, Y. Ben-Simon, G.M.H. Garcia: VIVACE (vortex induced vibration aquatic clean energy): A new concept in generation of clean and renewable energy from fluid flow, J. Offshore Mech. Arct. Eng. **130**(4) (2006)
47.49 M.M. Bernitsas, K. Raghavan, Y. Ben-Simon, G.M.H. Garcia: The VIVACE converter: Model tests at Reynolds numbers around 10^5, J. Offshore Mech. Arct. Eng. **131**(1), 1–13 (2006)
47.50 S.S. Chen: A review of flow induced vibration of two circular cylinders in crossflow, J. Press. Vessel Techno. **108**, 382–393 (1986)
47.51 R.D. Gabbai, H. Benaroya: An overview of modeling and experiments of vortex-induced vibration of circular cylinders, J. Sound Vib. **282**, 575–616 (2005)
47.52 R.A. Kumar, C.H. Sohn, B.H.L. Gowda: Passive control of vortex-induced vibrations: An overview Recent Patents on Mech, Eng. **1**, 1–11 (2008)
47.53 T. Sarpkaya: Vortex-induced oscillations: A selective review, J. Appl. Mech. **46**, 241–257 (1979)
47.54 T. Sarpkaya: A critical review of the intrinsic nature of vortex induced vibrations, J. Fluids Struct. **19**(4), 389–447 (2004)
47.55 C.H.K. Williamson, R. Govardhan: Vortex-induced vibrations, Annu. Rev. Fluid Mech. **36**, 413–455 (2004)
47.56 M.M. Zdravkovich: *Flow Around Circular Cylinders*, Vol. 1 (Oxford Univ. Press, Oxford 1997)
47.57 M.M. Zdravkovich: *Flow Around Circular Cylinders*, Vol. 2 (Oxford Univ. Press, Oxford 2002)
47.58 Y. Constantinides, O.H. Oakley Jr., S. Holmes: Analysis of turbulent flows and VIV of truss spar risers, Proc. 25th Int. Conf. Offshore Mech. Arct. Eng., Hamburg'06 (2006)
47.59 Y. Constantinides, O.H. Oakley Jr.: Numerical prediction of VIV and comparison with field experiments, Proc. 27th Int. Conf. Offshore Mech.Arct. Eng., Estoril'08 (2008)
47.60 K. Vikestad, J.K. Vandiver, C.M. Larsen: Added mass and oscillation frequency for a circular cylinder subjected to vortex-induced vibrations and external disturbance, J. Fluids Struct. **14**, 1071–1088 (2000)
47.61 T. Sarpkaya: In-line and transverse forces on cylinders in oscillatory flow at high Reynolds numbers, J. Ship Res. **21**, 200–216 (1977)
47.62 T. Sarpkaya: Hydrodynamic resistance of roughened cylinders in harmonic flow, J. R. Inst. Nav. Archit. **120**, 41–55 (1978)
47.63 T. Sarpkaya: Fluid forces on oscillating cylinders, J. Waterw. Port Coast. Ocean Div. **104**(3), 275–290 (1978)
47.64 L. Foulhoux, M.M. Bernitsas: Forces and moments on a small body moving in a 3-D unsteady flow, J. Offshore Mech. Arct. Eng. **115**(2), 91–104 (1993)
47.65 A.S. Richardson, J.R. Martucelli, W.S. Price: Research study on galloping of electric power transmission lines, Proc. 1st Int. Conf. Wind Eff. Build. Struct.'65 (1965) pp. 612–686
47.66 A. Bokaian: Galloping of a circular cylinder in the wake of another, J. Sound Vib. **128**, 71–85 (1989)
47.67 C.C. Chang, R.A. Kumar, M.M. Bernitsas: VIV and galloping of single circular cylinder with surface roughness at $3.0\times10^4 \le Re \le 1.2\times10^5$, Ocean Eng. **38**(16), 1713–1732 (2011)
47.68 E.S. Kim, M.M. Bernitsas, A.R. Kumar: Multi-cylinder flow induced motions: Enhancement by passive turbulence control at $28000 < Re < 120000$, Proc. 30th Offshore Mech. Arct. Eng. Conf., Rotterdam'11 (2011), Paper 49405
47.69 E.S. Kim, M.M. Bernitsas, A.R. Kumar: Multi-cylinder flow induced motions: Enhancement by passive turbulence control at $28000 < Re < 120000$, J. Offshore Mech. Arc. Eng. ASME Trans. **135**(1), Paper OMAE-11-1069 (2013)
47.70 H.R. Park, M.M. Bernitsas, C.C. Chang: Robustness of the map of passive turbulence control to flow-induced motions for a circular cylinder at $30000 < Re < 120000$, Proc. 31st Offshore Mech. Arct. Eng. Conf., Nantes'13 (2013), Paper 10123
47.71 H.R. Park, M.M. Bernitsas, E.S. Kim: Selective surface roughness to suppress flow-induced motions of two circular cylinders at $30,000 < Re < 120000$, Proc. 31st Offshore Mech. Arct. Eng. 2013 Conf., Nantes'13 (2013)
47.72 H.R. Park, M.M. Bernitsas, E.S. Kim: Selective surface roughness to suppress flow-induced motions of two circular cylinders at $30,000 < Re < 120000$, J. Offshore Mech. Arct. Eng. ASME Trans. **136**(4), Paper OMAE-13-1067 (2014)
47.73 K. Raghavan, M.M. Bernitsas: Effect of free surface on VIV for energy harnessing at $8\times10^3 < Re < 1.5\times10^5$, Proc. 26th Int. Offshore Mech. Arct. Eng. Conf., San Diego'07 (2007), Paper 29726
47.74 G.R.S. Assi, J.R. Meneghini, J.A.P. Aranha, P.W. Bearman, E. Casaprima: Experimental investigation of flow-induced vibration interference between two circular cylinders, J. Fluids Struct. **22**, 819–827 (2006)
47.75 G.R.S. Assi, P.W. Bearman, J.R. Meneghini: On the wake-induced vibration of tandem circular cylinders: The vortex interaction excitation mechanism, J. Fluid Mech. **661**, 365–401 (2010)
47.76 L. Ding, L. Zhang, E.S. Kim, M.M. Bernitsas: 2-URANS versus experiments of flow induced mo-

tions of multiple circular cylinders with passive turbulence control, J. Fluids and Structures **54**, 612–628 (2015)
47.77 E.S. Kim, A. Bas, B. Francis, N.A. Melliti, P.E. Bernitsas, A. Vahid, A. Kana, H.R. Park, M.M. Bernitsas: Two-cylinder flow-induced motions at $28000 < Re < 120000$: Enhancement in ultra-low speeds, Proc. 31st Offshore Mech. Arct. Eng. Conf., Nantes'13 (2013), Paper 10870
47.78 H.P. Ruscheweyh: Aeroelastic interference effects between slender structures, J. Wind Eng. Ind. Aerodyn. **14**, 129–140 (1983)
47.79 M.M. Zdravkovich: Flow induced oscillations of two interfering circular cylinders, J. Sound Vib. **101**(4), 511–521 (1985)
47.80 M.M. Zdravkovich: The effects of interference between circular cylinders in cross flow, J. Fluids Struct. **1**, 239–261 (1985)
47.81 M.M. Zdravkovich: Review of interference induced oscillations in flow past two parallel circular cylinders in various arrangements, J. Wind Eng. Ind. Aerodyn. **28**, 183–200 (1988)
47.82 M.M. Zdravkovich, E.B. Medeiros: Effect of damping on interference-induced oscillations of two identical circular cylinders, J. Wind Eng. Ind. Aerodyn. **38**, 197–211 (1991)
47.83 M.M. Zdravkovich, D.L. Priden: Interference between two circular cylinders; series of unexpected discontinuities, J. Ind. Aerodyn. **3**, 255–270 (1997)
47.84 E.S. Kim: Synergy of Multiple Cylinders in Flow Induced Motion for Hydrokinetic Energy Harnessing, Ph.D. Thesis (University of Michigan, Ann Arbor 2013)
47.85 Decarboni.se: Flutter instability, http://decarboni.se/publications/renewable-electricity-futures-study-volume-2-renewable-electricity-generation-and-storage-technologies/95-technology-characterization (2015)
47.86 L. Arnold: Fluid energy converting method and apparatus, US Patent 4 184 805 B2 (1980)
47.87 C.C. Kerr: Transfer of kinetic energy to and from fluids, US Patent 8 469 663 B2 (2013)
47.88 M.M. Bernitsas, K. Raghavan: Converter of current, tide, or wave energy, US Patent 7 493 759 B2 (2009)
47.89 M.M. Bernitsas, K. Raghavan: Enhancement of vortex induced forces and motion through surface roughness control, US Patent 8 047 232 B2 (2011)
47.90 M.M. Bernitsas, K. Raghavan: Reduction of vortex induced forces and motion through surface roughness control, US Patent 8 684 040 B2 (2014)
47.91 Marine Renewable Energy Laboratory, University of Michigan, http://name.engin.umich.edu/facilities/mhl/facilities/basin/ (2015)
47.92 M.M. Bernitsas, K. Raghavan, G. Duchene: Induced separation and vorticity using roughness in VIV of circular cylinders at $8 \times 10^3 < Re < 1.5 \times 10^5$, Proc. 27th Offshore Mech. Arct. Eng. Conf., Estoril '08 (2008)
47.93 M. Bernitsas, K. Raghavan: Reduction/suppression of VIV of circular cylinders through roughness distribution at $8 \times 10^3 < Re < 1.5 \times 10^5$, Proc. of the 27th Offshore Mech. Arct. Eng. Conf., Estoril'08 (2008) pp. 15–20
47.94 E.M.H. Garcia: Prediction by Energy Phenomenology for Harnessing Hydrokinetic Energy Using Vortex-Induced Vibrations, Ph.D. Thesis (University of Michigan, Ann Arbor 2008)
47.95 C.C. Chang: Passive Turbulence Control for VIV Enhancement for Hydrokinetic Energy Harnessing Using Vortex Induced Vibrations, Ph.D. Thesis (University of Michigan, Ann Arbor 2011)
47.96 C.C. Chang, M.M. Bernitsas: Hydrokinetic energy harnessing using the VIVACE converter with passive turbulence control, Proc. 30th Offshore Mech. Arct. Eng. Conf., Rotterdam'11 (2011), Paper 50290
47.97 J.H. Lee: Hydrokinetic Power Harnessing Utilizing Vortex Induced Vibrations Through a Virtual c-k VIVACE Model, Ph.D. Thesis (University of Michigan, Ann Arbor 2010)
47.98 J.H. Lee, N. Xiros, M.M. Bernitsas: Virtual damper-spring system for VIV experiments and hydrokinetic energy conversion, Ocean Eng. **5–6**, 732–747 (2011)
47.99 J.H. Lee, C.C. Chang, N. Xiros, M.M. Bernitsas: Integrated power take off and virtual oscillator system for the VIVACE converter: VCK system identification, Proc. ASME Int. Eng. Congr. Expo. Lake Buena Vista'09 (2009), Paper IMECE2009-11430
47.100 J.H. Lee, M.M. Bernitsas: High-damping, high-Reynolds VIV tests for energy harnessing using the VIVACE converter, Ocean Eng. **38**(16), 1697–1712 (2011)
47.101 H.R. Park: Expansion of the High-Lift Regime in Vortex Induced Vibration to Enhance VIV for Harnessing Hydrokinetic Energy of Even Slow Water Currents, Ph.D. Thesis (University of Michigan, Ann Arbor 2011)
47.102 H.R. Park, M.M. Bernitsas, A.R. Kumar: Selective roughness in the boundary layer to suppress flow induced motions of circular cylinder at $30000 < Re < 120000$, Proc. 30th Offshore Mech. Arct. Eng. Conf., Rotterdam'11 (2011)
47.103 H.R. Park, M.M. Bernitsas, A.R. Kumar: Selective roughness in the boundary layer to suppress flow induced motions of circular cylinder at $30000 < Re < 120000$, J. Offshore Mech. Arct. Eng. ASME Trans. **134**(4) (2012)
47.104 H.R. Park, R.A. Kumar, M.M. Bernitsas: Suppression of flow induced motions of rigid circular cylinder on springs by localized surface roughness at $3 \times 10^4 \leq Re \leq 1.2 \times 10^5$, Ocean Eng. **111**, 218–233 (2015)
47.105 K. Raghavan: Energy Extraction from a Steady Flow Using Vortex Induced Vibration, Ph.D. Thesis (University of Michigan, Ann Arbor 2007)
47.106 K. Raghavan, M.M. Bernitsas: Enhancement of high damping VIV through roughness distribution for energy harnessing at $8 \times 10^3 < Re < 1.5 \times 10^5$, Proc. 27th Offshore Mech. Arct. Eng. Conf., Estoril'08 (2008), Paper 58042

47.107 K. Raghavan, M.M. Bernitsas: Experimental investigation of Reynolds number effect on vortex induced vibration of rigid cylinder on elastic supports, Ocean Eng. **38**, 719–731 (2010)

47.108 K. Raghavan, M.M. Bernitsas, D. Maroulis: Effect of Reynolds number on vortex induced vibrations, Proc. IUTAM Symp., Hamburg'07 (2007) pp. 23–26

47.109 K. Raghavan, M.M. Bernitsas, D. Maroulis: Effect of bottom boundary on VIV for energy harnessing at $8\times10^3 < Re < 1.5\times10^5$, Proc. Offshore Mech. Arct. Eng.'07 (2007)

47.110 K. Raghavan, M.M. Bernitsas, D. Maroulis: Effect of bottom boundary on VIV for energy harnessing at $8\times10^3 < Re < 1.5\times10^5$, J. Offshore Mech. Arct. Eng. **131**(3), 031102 (2009)

47.111 Vortex Hydro Energy: http://www.vortexhydroenergy.com/ (2014)

47.112 W. Wu: CFD Modeling and Model-Test Calibration of VIV of Cylinders with Sroughness, Ph.D. Theory (University of Michigan, Ann Arbor 2011)

47.113 L. Ding, M.M. Bernitsas, E.S. Kim: 2-D URANS vs. experiments of flow induced motions of two circular cylinders in tandem with passive turbulence control for $30000 < Re < 105000$, Ocean Eng. **72**, 429–440 (2013)

47.114 L. Ding, Y. Chen, E.S. Kim, M.M. Bernitsas: 2-D URANS vs. experiments of flow induced motions of multiple circular cylinders with passive turbulence control, Proc. 32nd Offshore Mech. Arct. Eng. Conf., Nantes'13 (2013), Paper 10911

47.115 L. Ding, L. Zhang, C.C. Chang, M.M. Bernitsas: Numerical simulation and experimental validation for energy harvesting of single-cylinder VIVACE converter with passive turbulence control, J. Renew. Energy **85**, 1246–1259 (2016)

47.116 L. Ding, L. Zhang, C. Wu, X. Mao, D. Jian: Flow induced motion and energy harvesting of bluff bodies with different cross sections, Energy Convers. Manag. **91**, 416–426 (2015)

47.117 L. Ding, L. Zhang, C.M. Wu, E.S. Kim, M.M. Bernitsas: Numerical study on the effect of tandem spacing on flow induced motions of two cylinders with passive turbulence control, Proc. 34th Offshore Mech. Arct. Eng. Conf., Newfoundland'15 (2015), Paper 42301

47.118 A. Kana, R. Alter, M.M. Bernitsas: *Two-dimensional RANS Simulations of a Series of Fish-shaped Geometries for* $10000 < \mathrm{Re} < 150000$ (Marine Renewable Energy Laboratory and Vortex Hydro Energy, Dept. of NA&ME, University of Michigan, Ann Arbor 2013)

47.119 J. Wang, H.R. Park, M.M. Bernitsas: Mathematical processing of experimental data for VIV and galloping of rigid circular cylinders on linear springs: Mathematical model #1 (Dept. of NA&ME, University of Michigan, Anne Arbor 2013) (MRELab #007 2013)

47.120 B. Yucekaya, H.R. Park, M.M. Bernitsas: Mathematical processing of experimental data for VIV and galloping of rigid circular cylinders on linear springs: Mathematical model #2, MRELab #008 (2013) (Dept. of NA&ME, University of Michigan, Anne Arbor 2013)

47.121 E.M.H. Garcia, C.C. Chang, H.R. Park, M.M. Bernitsas: Effect of damping on variable added mass and lift of circular cylinders in vortex-induced vibrations, Proc. MTS/IEEE Ocean'14 (2014)

47.122 US Energy Information Administration: Annual Energy Outlook 2014 with Projections to 2040, DOE/EIA-0383 Tech. Rep. (DOE, Washington 2014)

47.123 R. Ferry, E. Monoian: *A Field Guide to Renewable Energy Technologies* (Land Art Generator Initiative, Pittsburgh 2012)

47.124 World Energy Council: Survey of Energy Resources 2004, http://www.worldenergy.org/documents/ser2004.pdf (2004)

47.125 S.S. Schwartz (Ed.): *Proceedings of the Hydrokinetic and Wave Energy Technologies Technical and Environmental Issues Workshop* (Revolve Inc., Washington 2006)

47.126 R. Bedard, M. Previsic, B. Polagye, G. Hagerman, A. Casavant: North America Tidal In-Stream Energy Conversion Technology Feasilbility Study, EPRI Tidal In Stream Energy Conversion (TISEC) Project (2006)

47.127 Franklin and Assoc.: *Avoided pollution report*, appendix A (1992)

47.128 A. Babarit, J. Hals, M.J. Muliawan, A. Kurniawan, T. Moan, J. Krokstad: Numerical benchmarking study of a selection of wave energy converters, Renew. Energy **41**, 44–63 (2012)

47.129 R. Bedard, M. Previsic, O. Siddiqui, G. Hagerman, M. Robinson: *Tidal in Stream Energy Conversion (TISEC) Devices*, Tech. Rep. EPRI-TP-004NA (Eletric Power Research Institute, Palo Alto 2005)

47.130 Verdant Power: http://www.verdantpower.com/technology.html (2014)

47.131 Ocean Renewable Power Company: http://www.orpc.co (2015)

47.132 Alden Turbine: http://energy.gov/articles/fish-friendly-turbine-making-splash-water-power (2015)

47.133 Hydro Green Energy Technology: http://hgenergy.com/index.php (2015)

47.134 J.B. Johnson, D.J. Pride: River, Tidal, and Ocean Current Hydrokinetic Energy Technologies: Status and Future Opportunities in Alaska, Report prepared for the Alaska Energy Authority, by the (Alaska Center for Energy and Power, 2010)

47.135 Marine Current Turbines: https://www.google.com/search?q=Images+of+marine+hydrokinetic+energy+technology&client=safari&rls=en&tbm=isch&tbo=u&source=univ&sa=X&ei=LkNmVLC0JouUyQTI3YKwDw&ved=0CB8QsAQ

47.136 M.J. Wolfgang, J.M. Anderson, M.A. Grosenbaugh, D.P.K. Yue, M.S. Triantaphyllou: Near body flow dynamics in swimming fish, J. Exp. Biol. **202**, 2303–2327 (1999)

47.137 C. Eloy: Optimal Strouhal number for swimming animals, J. Fluids Struct. **30**, 205–218 (2012)

47.138 J.H. Gerrard: The mechanics of the formation region of vortices behind bluff bodies, J. Fluid Mech. **25**, 401–413 (1966)

47.139 W. Wu, M.M. Bernitsas, K.J. Maki: 2D-URANS simulation vs. experimental measurements of flow induced motion of circular cylinder with passive turbulence control at $30000 < Re < 120000$, Proc. 30th Offshore Mech. Arct. Eng. Conf., Rotterdam'11 (2011), Paper 50293

47.140 W. Wu, M.M. Bernitsas, K.J. Maki: 2D-URANS simulation versus experimental measurements of flow induced motion of circular cylinder with passive turbulence control at $30000 < Re < 120000$, J. Offshore Mech. Arct. Eng. **136**(4) (2014)

47.141 A. Barrero-Gil, G. Alonso, A. Sanz-Andres: Energy harvesting from transverse galloping, J. Sound Vib. **329**(14), 2873–2883 (2010)

47.142 H. Sun, M.P. Bernitsas, E.S. Kim, M.M. Bernitsas: Virtual spring-damping system for fluid induced motion experiments, Proc. 34th Offshore Mech. Arct. Eng. Conf., Newfoundland'15 (2015), Paper 42179

47.143 S. Pacala, R. Socolow: Stabilization wedges: Solving the climate problem for the next 50 years with current technologies, Science **305**, 968–971 (2004)

47.144 A.S. Bahaj, L. Blunden, A. Anwar: *Tidal-Current Energy Device Development and Evaluation Protocol*, Tech. Rep. URN 08/1317 (University of Southampton, Southampton 2008)

47.145 J.C. Mankins: *Technology Readiness Levels: A White Paper* (Advanced Concepts Office, Office of Space Access and Technology, NASA, Washington 1995)

47.146 TAUW: http://www.tauw.nl/ (2013)

47.147 OHMSETT: http://ohmsett.com/facility.html (2015)

47.148 J. Ding, S. Balasubramanian, R. Lokken, T. Yung: Lift and damping characteristics of bare and straked cylinders at riser scale Reynolds numbers, Proc. Offshore Technol. Conf.'04 (2004), Paper 16341

47.149 H.R. Park, R.A. Kumar, M.M. Bernitsas: Enhancement of flow induced motions of rigid circular cylinder on springs by localized surface roughness at $3\times10^4 \le Re \le 1.2\times10^4 \le Re \le 1.2\times10^5$, Ocean Eng. **72**, 403–415 (2013)

47.150 H. Schlichting: *Boundary Layer Theory*, 7th edn. (Springer, Berlin, Heidelberg 1955)

47.151 J. Mercier: Large Amplitude Oscillations of a Circular Cylinder in a Low Speed Stream, Ph.D. Thesis (Stevens Institute of Technology, Hoboken 1973)

47.152 Marine Renewable Energy Laboratory: Four-Cylinder School in Synergistic FIM, Video I, https://www.youtube.com/watch?v=IcR8HszacOE (2010)

47.153 Marine Renewable Energy Laboratory: Four-Cylinder School in Synergistic FIM, Video II, https://www.youtube.com/watch?v=Z6q0Nd4hsjo (2010)

47.154 A. Khalak, C.H.K. Williamson: Investigation of the relative effects of mass and damping in vortex-induced vibration of a circular cylinder, J. Wind Energy Ind, Aerodyn **69–71**, 341–350 (1997)

47.155 A. Khalak, C.H.K. Williamson: Motions, forces and mode transitions in vortex-induced vibrations at low mass-damping, J. Fluids Struct. **13**, 813–851 (1999)

47.156 A. Isidori: *Nonlinear Control Systems*, Communications and Control Engineering (Springer, Berlin, Heidelberg 1995)

47.157 J.K. Vandiver: Damping parameters for flow-induced vibration, J. Fluids Struct. **35**, 105–119 (2012)

47.158 E. Konstantinidis: Apparent and effective drag for circular cylinders oscillating transverse to a free stream, J. Fluids Struct. **39**, 418–426 (2013)

47.159 A. Vinod, A. Banerjee: Surface protrusion based mechanisms of augmenting energy extraction from vibrating cylinders at Reynolds number 3×10^3–3×10^4, J. Renew. Sustain. Energy **6**, 063106 (2014)

47.160 B.H. Huynh, T. Tjahjowidodo, Z.W. Zhong, Y. Wang, N. Srikanth: Nonlinearly enhanced vortex induced vibrations for energy harvesting, Proc. IEEE Int. Conf. on Adv. Intell. Mechatron. (AIM), Busan'15 (2015) pp. 91–96

47.161 D.T. Walker, D.R. Lyzenga, E.A. Ericson, D.E. Lund: Radar backscatter and surface roughness measurements for stationary breaking waves, Proc. R. Soc. Lond. '96, Vol. A452 (1996) pp. 1953–1984

47.162 R. Govardhan, C.H.K. Williamson: Modes of vortex formation and frequency response for a freely-vibrating cylinder, J. Fluid Mech. **420**, 85–130 (2000)

47.163 R. Govardhan, C.H.K. Williamson: Resonance forever: Existence of a critical mass and an infinite regime of resonance in vortex-induced vibration, J. Fluid Mech. **473**, 147–166 (2002)

47.164 C.H.K. Williamson: Vortex dynamics in the cylinder wake, Annu. Rev. Fluid Mech. **28**, 477–539 (1996)

47.165 M.F. Unal, D. Rockwell: On vortex formation from a cylinder, Part 1: The initial instability, J. Fluid Mech. **190**, 491–512 (1988)

47.166 O.M. Griffin: Vortex shedding from bluff bodies in a shear flow: A review, J. Fluid Eng. **107**(3), 298–306 (1985)

47.167 R. Govardhan, C.H.K. Williamson: Defining the modified Griffin plot in vortex-induced vibration: Revealing the effect of Reynolds number using controlled damping, J. Fluid Mech. **561**, 147–180 (2006)

47.168 H.M. Blackburn: Effect of blockage on spanwise correlation in a circular cylinder wake, Exp. in Fluids **13**, 134–136 (1994)

47.169 A. Roshko, V. Chattoorgoon: Flow force on a cylinder near a wall or another cylinder, J. Fluid Mech. **15**, 1–3 (1975)

47.170 P. Reichl, K. Hourigan, M.C. Thompson: Flow past a cylinder close to a free surface, J Fluid Mech. **533**, 269–296 (2005)

47.171 A. Khalak, C.H.K. Williamson: Dynamics of a hydroelastic cylinder with very low mass and damping, J. Fluids Struct. **10**(5), 455–472 (1996)

47.172 J.B.V. Wanderley, S.H. Sphaier, C. Levi: A numerical investigation of vortex induced vibration on an elastically mounted rigid cylinder, Proc. 27th Offshore Mech. Arct. Eng. Conf.'08 (2008), Paper 57344

47.173 J.B.V. Wanderley, S.H. Sphaier, C. Levi: A numerical investigation of vortex induced vibration on

an elastically mounted rigid cylinder, Proc. 28th Offshore Mech. Arct. Eng. Conf.'09 (2009), Paper 79725

47.174 P.R. Spalart, S.R. Allmaras: A one-equation turbulence model for aerodynamic flows, Rech. Aerosp. **1**, 5–21 (1994)

47.175 A. Travin, M. Shur, M. Strelets, P. Spalart: Detached-eddy simulations past a circular cylinder, Flow Turbul. Combust. **63**, 293–313 (2000)

47.176 O.K. Kinaci, S. Lakka, H. Sun, M.M. Bernitsas: Computational and experimental assessment of turbulence stimulation on flow induced motion of circular cylinder, Proc. 34th Offshore Mech. Arct. Eng. Conf., Newfoundland'15 (2015), Paper 42357

47.177 J.H. Liu, M.M. Bernitsas: Envelope of power harvested by a single-cylinder VIVACE converter with passive turbulence control, Proc. 34th Offshore Mech. Arct. Eng. Conf., Newfoundland'15 (2015), Paper 42333

47.178 O.K. Kinaci, M.M. Bernitsas, S. Lakka: Effect of tip-flow on vortex induced vibration of circular cylinders for $Re < 1.2 \times 10^5$, Ocean Eng. **117**, 130–142 (2015)

48. Ocean Thermal Energy Conversion

Muthukamatchi Ravindran, Raju Abraham

Ocean thermal energy conversion (OTEC) utilizes the thermal gradient available in the ocean to operate a heat engine to produce work output. Even though the concept is simple and old for almost one century, during last three decades, it has gained momentum due to worldwide search for a clean, continuous energy source to replace the fossil fuels. There are technological hurdles to overcome to tap the immense potential of OTEC. But still the technology is mature enough to establish commercial power plants. This chapter presents the technological and economical aspects of the OTEC with an overview of the current status. It also suggests some further research and technological guidance based on the Indian OTEC experiences for this promising technology for future.

48.1 **OTEC Principles and Systems** 1245
48.1.1 Spin-Off Technologies 1246
48.1.2 Types of OTEC Plants 1247
48.2 **History of OTEC Installations Worldwide** 1250
48.2.1 Indian Experiences 1253
48.3 **Current Status of OTEC Technologies** 1254
48.3.1 Developing Technologies on OTEC 1254
48.3.2 Research on OTEC in International Scenario 1254
48.3.3 Commercial Designs for OTEC Plants 1254
48.3.4 Industrial Proposals for OTEC Commercialization 1254
48.4 **Design Considerations for Future OTEC Plants** 1257
48.4.1 Choice of Working Fluids 1257
48.4.2 Choice of Materials 1258
48.4.3 Selection of Equipment 1259
48.4.4 Sea Water Intake System and Deployments 1261
48.4.5 Bio-Fouling and Control 1263
48.4.6 Power Transmission from Offshore OTEC Plants 1264
48.5 **Conclusion** 1265
References 1265

Sun warms the oceans at the surface and wave motion mixes the warmed water downward to depths of about 100 m. This mixed layer is separated from the deep cold water, formed at high latitudes, by a thermocline. This boundary is sometimes marked by an abrupt change in temperature, but more often the change is gradual. The resulting vertical temperature distribution, therefore, consists of two layers separated by an interface with temperature differences between them ranging from 10–25 °C. Higher values are found in equatorial waters. This simplistic description implies that there are two enormous reservoirs, in some oceanic regions, providing the heat source and sink required to operate a heat engine. The concept of utilizing the ocean temperature difference is referred to as OTEC or ocean thermal energy conversion.

48.1 OTEC Principles and Systems

OTEC is an energy technology that converts solar radiation to electric power. As long as the temperature between the warm surface water and the cold deep water differs by about 20 °C, an OTEC system can produce a significant amount of power which is economical. This potential is estimated to be about 10^{13} W of base load power generation. The cold deep seawater used in the OTEC process is also rich in nutrients, and it can be used to culture both marine organisms and plant life near the shore or on land.

The annual incidence of solar energy, in the form of short wavelength radiation, on the Earth is about 1.52×10^{18} kWh (1.73×10^{14} kW in units of power) of which as much as 70% is absorbed as heat by our ocean–land–atmosphere system to eventually be radiated back to space as long-wave radiation. This is

equivalent to an incidence of 240 W/m^2. The actual amount absorbed by the oceans corresponds to an incidence of 95 W/m^2 or 37×10^{12} kW over the surface of the oceans [48.1]. This incidence value represents an enormous amount of power compared to our human consumption. The amount of solar power absorbed by our oceans is equivalent to at least 4000 times the amount presently consumed by humans [48.2]. For a conversion efficiency of 3% from ocean energy to electricity as in the case of OTEC, we would need 1% of this renewable energy to meet the total demand of present requirement. The removal of such a relatively small value of ocean solar energy does not create an adverse environmental impact and technologies must be developed to transform it to a useful form and to transport it to the user.

In an OTEC plant, large quantity of water is to be transported to the surface from 800 to 1000 m depths, with temperatures ranging from 8 to 4 °C. Utilizing the warm surface water at 27–29 °C and the deep-sea cold water, energy is extracted. This concept is based in the thermodynamic Rankine cycle used in steam power plants. Figure 48.1 shows the distribution in the temperature difference between the surface seawater and

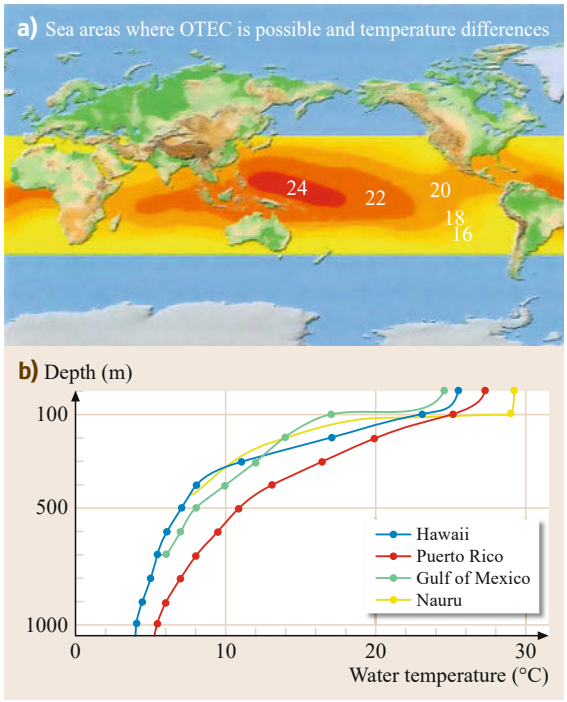


Fig. 48.1a,b Temperature variation with depth of the sea. **(a)** Ocean surface with the availability of different temperature differences in the tropical region; **(b)** vertical temperature distribution at four different locations (courtesy of Saga University, Japan)

1000 m depth across the world. The performance of OTEC power generating cycles is assessed with the same elementary concepts of thermodynamics used for conventional steam power plants. The major difference arises from the large quantities of warm and cold seawater required for heat transfer processes, resulting in the consumption of 20–40% of the power generated by the turbine-generator in the operation of pumps.

OTEC is capital intensive; however, improvements in technology and higher rated plants can bring down the unit cost considerably. For the mainland, the cost of power generation for plants upward of 25 MW is expected to be comparable to fossil fuel units. But for islands, an OTEC power plant of any size is cheaper than the conventional generation units. Countries like India are geographically well placed as far as the OTEC potential is concerned. Upwelling of deep sea cold water can be an answer to issues such as global warming and food scarcity.

48.1.1 Spin-Off Technologies

The primary benefit of OTEC system is to produce electricity. The turbine–generator system of the OTEC power cycle can generate electricity in a continuous mode compared to other renewable energy systems which are intermittent. There are spin-off technologies associated with OTEC which is of commercial interest. Larger OTEC plants can be used for producing electricity, water, and other spin-off benefits such as air conditioning and aquaculture (Fig. 48.2).

Desalination

A two-stage OTEC cycle, wherein electricity is produced in the first stage followed by water production in the second-stage, has also been proposed to maximize the use of the thermal resource available to produce both water and electricity [48.3]. In the second stage, the temperature difference available in the seawater effluents from an OTEC plant (i. e., about 10–12 °C) is used to produce desalinated water through a system consisting of a flash evaporator and a surface condenser. The simultaneous production of desalinated water and power has been considered to be economically attractive for the Island market. For example, a 50 MW hybrid cycle plant producing as much as 62 000 m^3 of water per day could be tailored to support a developing community of approximately 300 000 people or as many as 100 000 people in an industrialized nation.

Air-Conditioning

The use of the cold deep water as the chiller fluid in air-conditioning systems has a tremendous economic potential by providing significant energy conservation, in-

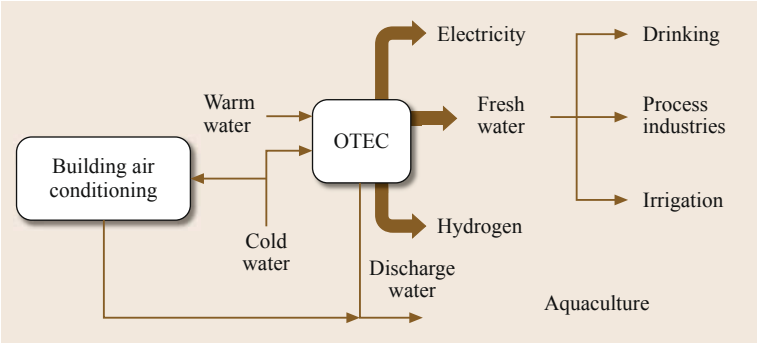


Fig. 48.2 OTEC spin-off technologies

dependent of OTEC. For example, only 1 m^3/s of 7 °C deep ocean water is required to produce 5800 t (roughly capable of cooling 5800 rooms) of air conditioning, consuming 360 kW, mostly in pumping seawater, while saving the equivalent of 5000 kWe with an investment payback period estimated at 3–4 years [48.4]. The same amount of seawater could only support a 500 kWe OTEC plant (1/10 of the power saved with the AC (air-conditioning) system).

Aqua-Culture

The deep-sea cold water is very rich (nearly 28 times) in inorganic nutrients such as nitrates, phosphates, silicates, and it can be commercially used for sea farming. There are several significant deep sea-water aquaculture research facilities in many countries such as Japan (Toyama Bay and Kochi) and US (Hawaii). The commercial utilization of deep sea water is already initiated in Hawaii. It has been estimated that the annual revenue from fish culture from a 10 MW OTEC mari-culture facility is $3.3 million/year [48.5]. The aquaculture of abalone, kelp, nori, micro algae, lobster, salmon, oysters sea urchins, etc. were successfully demonstrated at Hawaii. As the deep-sea cold water is comparatively pure, sensitive larval stages of species can thus be grown without expensive purification and filtration needed when surface water is used.

Production of Metals

Sea water contains metals such as uranium, lithium, etc. Large quantity of sea water is pumped through the OTEC system and discharged back into the sea. Deep sea water is having higher concentration of many of these metals. Extractions of these minerals are an additional benefit of the OTEC system and attempts are made in many research laboratories [48.6].

48.1.2 Types of OTEC Plants

OTEC plants are classified based on the location of the plant established as well the thermodynamic cycle used. The cost and size of the plant is determined by the thermodynamic cycle and the working fluid used.

OTEC Plants Based on Location

Depending upon the availability of deep sea near the coast for a depth of 600–1000 m, the OTEC system could be installed in three possible types of locations.

Floating Plants. When the distance of deep sea from coast is large, that is, 10–15 km or more from the shoreline the OTEC plant could be placed on a floating platform with a cold water pipe suspended vertically down. An underwater cable is needed for power transmission to shore. Alternately, the energy generated may be utilized to produce energy intensive materials like ammonia or hydrogen from the seawater. The products have to be transported to the main land by ships. Keeping the position of the floating platform under the action of waves, wind, and currents needs careful considerations during the design.

Shelf-Mounted Plants. The sea bed is shallow for some distance and then moves steep and the distance of 1000 m contour is around 10 km, the OTEC plant could be mounted on a platform at the end of shallow water area and the generated power can be transmitted to the mainland by underwater cables. A tower is built in a water depth at around 50–100 m where the bed starts sloping down.

Land Based Plants. If the deep-water conditions are available within 2–3 km of the coast, the entire plant could be situated on land with the cold-water pipeline running along the ocean bed to a depth of 800–1000 m. This eliminates the costly structures for floating platform or shelf-mounted platform.

OTEC Plants Based on Power Cycles

There are basically three types of OTEC processes: closed cycle, open cycle, and hybrid cycle based on the thermodynamic process. Much efforts were made in the

past for optimizing the cost of OTEC using power cycle analysis [48.7, 8].

Closed Cycle Plants. In the closed-cycle system, heat transferred from the warm surface seawater causes the low boiling point working fluid (such as ammonia), to evaporate. The expanding vapor drives a turbine attached to a generator, which produces electricity. Cold seawater passing through a condenser containing the vaporized working fluid turns the vapor back into liquid, which is then recycled through the system (Fig. 48.3). Technology for the manufacture of ammonia turbines with a low pressure ratio required for the closed-cycle plant is available. The heat exchangers having close temperature approach for OTEC applications can be of either titanium or aluminum plate heat exchangers. It is easy to scale up a closed cycle plant to 5, 25, or even 100 MW OTEC plant.

Open Cycle Plants. Open-cycle OTEC plant uses the warm surface water itself as the working fluid. The water vaporizes in a near vacuum at surface water temperatures. The expanding vapor drives a low-pressure turbine attached to a generator, which produces electricity. The vapor, which has lost its salt and is almost pure fresh water, is condensed back into a liquid by exposure to cold temperatures from deep ocean water. If the condenser keeps the vapor from direct contact with seawater, the condensed water can be used for drinking water, irrigation, or aquaculture. A direct contact condenser (DCC) instead of surface condenser produces more electricity, but the vapor is mixed with cold seawater and that mixture is returned to the ocean. The process is repeated with a continuous supply of warm surface seawater (Fig. 48.4).

Since the pressure drop across the turbine is the difference between the pressure at which the water vaporizes and the lower pressure remaining after condensation, open-cycle systems require very large turbines to capture relatively small amounts of energy. George Claude, the inventor of the open-cycle process, calculated that a 6 MW turbine would need to be about 10 m in diameter, and he could not design a realistic turbine larger than this. Recent re-evaluation of Claude's work indicates that modern technology cannot improve significantly on his design, so it appears that the open-cycle turbines are limited to about 6 MW. The multiple turbines required for a commercial-sized OTEC plant will significantly increase its complexity and reduce its efficiency. Also, a complicated vacuum system is required to create and maintain the vacuum up to 25 m bar.

In an ideal open-cycle plant, the vacuum pumps could be shut down after start-up, since all the water vaporized in the evaporator would be condensed in the condenser, leaving behind a vacuum. However, vacuum leaks and noncondensable gases dissolved in the surface and deep seawater necessitate continuous operation of the vacuum pumps. The overall thermal to electrical efficiency of these traditional open- and closed-cycle OTEC plants is very similar, approaching 2.5%. The following are the advantages of using an open cycle:

- Conventional surface heat exchanger is eliminated.
- Production of potable water is an additional benefit.
- Water itself is the working fluid and hence no environmental pollution.

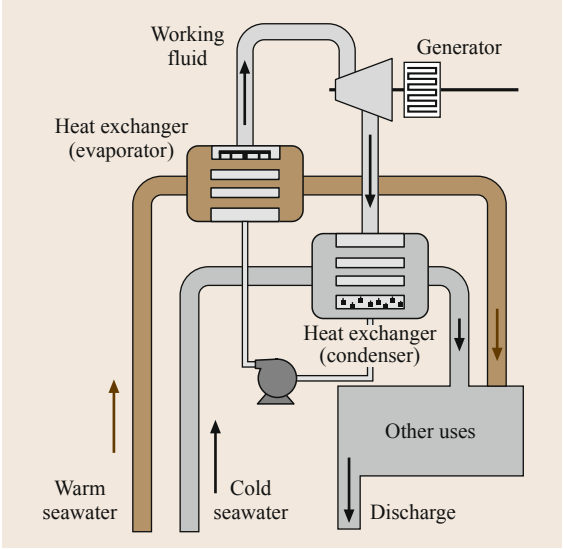


Fig. 48.3 Closed cycle OTEC plant

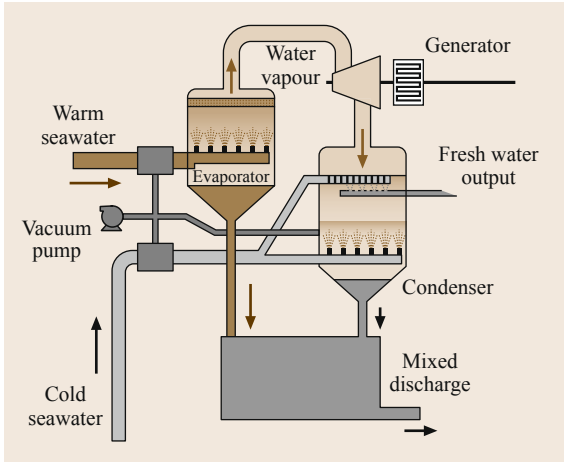


Fig. 48.4 Open cycle OTEC plant

Panchal–Bell Cycle. Panchal–Bell cycle is a combination of the closed and the open cycle. In the Panchal–Bell cycle, steam from flash evaporation of warm water is condensed on the surface of a heat exchanger containing liquid ammonia, which vaporizes to form the working fluid for the closed cycle. Condensed steam is discharged to the ocean or may be saved as a source of fresh water. The Panchal–Bell cycle has a number of potential advantages, compared to the closed and open cycles such as the elimination of low-pressure turbine, reduced heat exchanger area.

Hybrid Cycle Plants. Hybrid systems is similar to Panchal–Bell cycle in which parts of both open- and closed-cycle systems to optimize production of electricity and fresh water. The warm water discharge from the evaporator is sent to a flash chamber where it is flash evaporated. The vapor thus produced is condensed back to fresh water in a desalination condenser using the cold-water discharge from the closed-cycle condenser. The warm water enters a flash chamber under vacuum condition where it is flash-evaporated into steam, in a similar fashion to the open-cycle evaporation process. The vaporized working fluid in the evaporator would then drive a turbine to produce electricity and is condensed back to liquid in the condenser. A vacuum pump is connected to the desalination condenser to maintain the vacuum in the flash chamber and desalination condenser (Fig. 48.5). The simultaneous production of power and water is the great advantage and the quantity of power and water can be varied depending on the requirements.

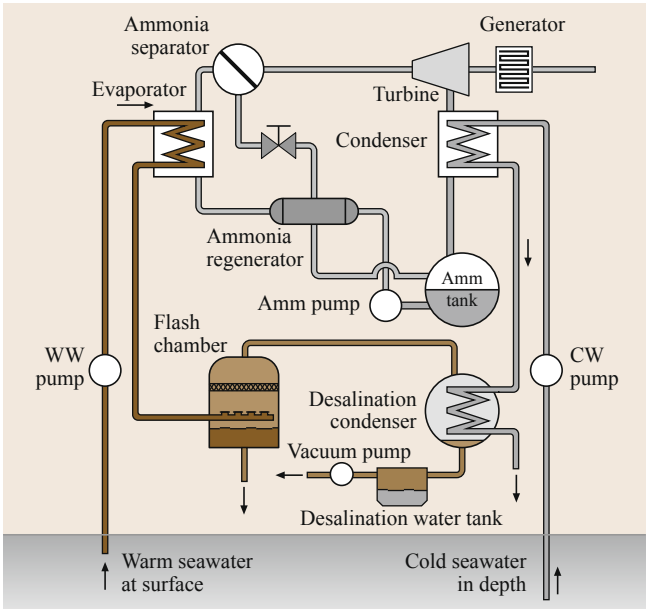


Fig. 48.5 Diagram of the hybrid OTEC process

Advanced Power Cycles

OTEC power cycle utilizes a major portion of net power for pumping the sea water and working fluid. As it work with a small temperature difference, the overall efficiency of the power cycle is very less. There are many attempts to improve the overall efficiency of the OTEC power cycle as discussed below.

Kalina Cycle. Among several OTEC systems developed, the leading ones are the Rankine cycle, Kalina cycle, and Uehara cycle. The working fluid used in the Rankine Cycle is pure ammonia where as other cycles use mixtures such as ammonia–water. American Scientist Dr. A. Kalina developed a system using ammonia–water mixture for the working fluid, providing a significant efficiency boost over the earlier Rankine cycle (Fig. 48.6). In the Kalina cycle, after the working fluid (vapor) expands through the turbine and in the phase before it goes into the condenser, it is liquefied using an absorber. Also, in the phase before the working fluid (liquid) goes in to the evaporator, the

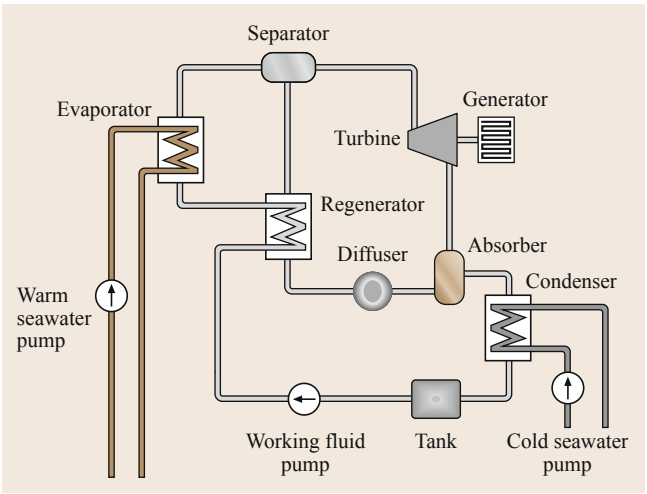


Fig. 48.6 Schematic diagram of the Kalina cycle

vapor that emerges from the evaporator and the liquid that was unable to evaporate is separated by a separator, and the liquid is mixed with the leftover liquid at the regenerator to improve efficiency. Kalina cycle uses a working fluid mixture which has the drawback of lower performance in the evaporator and condenser because some parts easily evaporates or condenses and others do not.

The heat efficiency of the Rankine cycle is about 3% under the conditions of 28 °C surface water and 4 °C deep-sea water, but the Kalina cycle showed astonishingly 5% under the same conditions. But there are

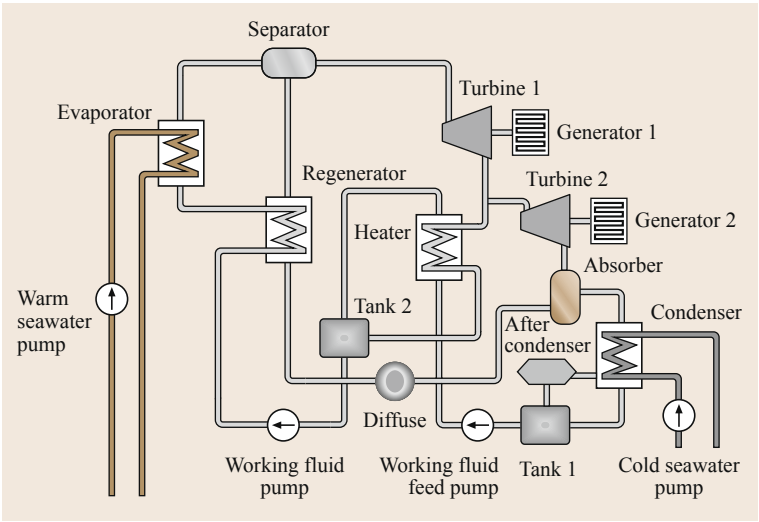


Fig. 48.7 Schematic diagram of Uehara cycle

problems too such as, as compared to the Rankine cycle, the fairly large loss of efficiency of both evaporator and condenser, which was due to the working fluid of ammonia–water mix.

Uehara Cycle. Uehara cycle provides an additional turbine and lessens the load on the condenser by extracting the vapor from the first turbine [48.9]. Extraction means drawing the vapor out of the turbine after it has done its work but has not completely expanded out of the turbine (drawing some of the vapor out of the turbine). In addition, the extracted fluid and the liquefied one by the condenser are heated by a heater in the phase before it enters the regenerator, thereby increasing efficiency. The cycle also has an *after condenser*, which liquefies the vapor that could not be liquefied in the condenser (Fig. 48.7).

Comparison of Cycles. The Kalina cycle has added a separator, a regenerator, and an absorber to improve its efficiency. Uehara cycle has added a second turbine, a heater, and an after-condenser to improve its efficiency much more. In the Kalina cycle, because of the heavy load on the evaporator and condenser, the intake volume of seawater could not be reduced. In the Uehara cycle, the intake volume could be reduced, making it possible to reduce the diameter and weight of the intake pipe. While the Kalina cycle has only one turbine Uehara cycle has two turbines. Vapor that comes out of the primary turbine is extracted to reduce the load of the condenser. The Uehara cycle shows an improvement in efficiency over the Kalina cycle, but the capital cost of the plant is higher due to the addition of another turbine and heater. Studies of Prof. Uehara show that the thermal efficiency of the Kalina cycle 4.9% and that of Uehara cycle 5.4%. This is 10% higher than that of the Kalina cycle and 30% higher than that of the Rankine cycle. The estimation is done under the operating conditions of 28 °C warm water inlet temperature and 6 °C cold-water inlet temperature.

48.2 History of OTEC Installations Worldwide

The overall research and developments of OTEC in the past are discussed by *Avery* [48.10]. The concept of OTEC was first recognized by the French engineer Arsene d' Arsonval in 1881. However, the first experimental system was designed and built by George Claude in 1929 and was field-tested in Matanzas Bay, Cuba. This OTEC plant generated 22 kW (gross) of power. In Claude's cycle, the surface water is flash-evaporated in a vacuum chamber. The resulting low-pressure steam is used to drive a turbine-generator and the relatively colder deep seawater is used to condense the steam downstream from the turbine. The plant failed to achieve net power production because of a poor site selection and a mismatch of the power and seawater systems. However, the plant did operate for several weeks. This was followed by the design of a 2.2 MW floating plant for the production of up to 2000 tons of ice for the city of Rio de Janeiro. Claude was able to deploy his plant, housed in a ship, about 100 km from the harbor, but unfortunately he failed in his numerous attempts to install the long vertical cold-water pipe required to transport the deep ocean water to the ship, and

Table 48.1 The major events in OTEC History

1881	French scientist J. D'Arsonval presented concept of OTEC
1926	Initial study for commercialization by French scientist Claude
1933	Claude built a floating 1200 kW OTEC plant, but it failed
1970	Research association of New Power Generation System started the experiment as one of the theme
1972	OTEC development program initiated in the USA
1973	Saga University started experiments
1974	OTEC was a priority subject for *Sunshine Project* USA launched their OTEC study under Energy Research and Development Association (ERDA) project First International OTEC Conference in USA
1977	Saga University succeeded in 1 kW OTEC experiment
1979	Mini-OTEC by USA succeeded in generating 50 kW OTEC
1980	The proposal of a 25 MW floating OTEC plant off Tamil Nadu
1981	OTEC-1 experiments to test 1 MWe heat exchangers
1981–1985	Long term bio fouling and corrosion experiments at STF, Hawaii
1981	Tokyo Electric Co. and others succeeded in 120 kW OTEC test in Nauru Republic
1982	Kyushu Electric Co. and others succeeded in generating 75 kW
1982	France proposed to build a 5 MWe shore based plant in Tahiti
1985	Saga University built 75 kW OTEC pilot plant
1990	IOA (International OTEC Association) was organized by Taiwan, USA, Japan and others
1993–1998	USA built 210 kW pilot plant, that is open cycle on Kona
1997	National Institute of Ocean Technology, India and Saga University signed memorandum on technical collaboration
1999	International OTEC/DOWA (Deep Ocean Water Associaton) Conference was held in Imari city, near Saga University
2001	1 MW floating OTEC plant was constructed by NIOT, but failure in cold water pipe deployment.
2002	Forum on desalination using renewable energy in Palau (FDE in Palau) was held in October
2003	New Pilot Hybrid OTEC plant is to be constructed in Saga and 3rd WWF in Kyoto was held in March
2010	10 MW proposed closed cycle OTEC plant by Lokheed Martin at Hawaii

had to abandon his enterprise in 1935. His failure can be attributed to the absence of an offshore industry, and of ocean engineering expertise, as it is today. His biggest technological challenge consisted in installing the required seawater pipes at sea. The situation is markedly different now in that there is a proven record for such operations.

In the 1950s, a French corporation planned the construction of an OTEC plant at Ivory Coast of Africa, generating 3.5 MW of electricity, but later it was dropped due to political reasons. The Gulf oil crisis of 1973, ignited the search for substitutes for petroleum, and re-examined, along with other options like wind energy, solar energy, bio-mass, etc. D'Arsonval's concept was successfully demonstrated in 1979, when a small plant mounted on a barge off Hawaii (Mini-OTEC) produced 50 kW of gross power for several months, with a net output of 18 kW.

This closed-cycle plant was sponsored by private industry and the State of Hawaii. Heat exchangers for a 1 MW OTEC plant, that is, with a thermal rating of about 35 MW, were successfully tested and known as OTEC-1. In OTEC-1, ammonia was used as a working fluid and a polyethylene pipe was used to pump cold water from 670 m depth. Subsequently, a 120 kW gross power, land-based plant was operated for 2 years in the island nation of Nauru by a consortium of Japanese companies. In 1982, a 50 kW closed-cycle land-based OTEC plant was established with plate heat exchangers on the Island of Tokunoshima for 2 years. These plants were primarily used to prove the concept, and were too small to be scaled to commercial size systems. The OTEC plant in a small OC-OTEC land-based

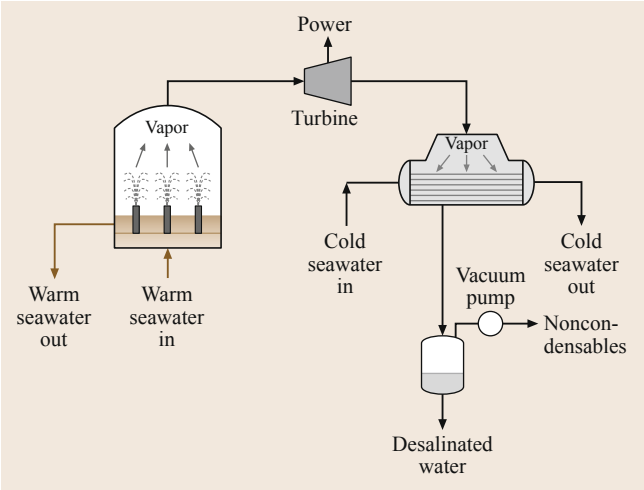


Fig. 48.8 The 210 kW open cycle OTEC plant (1993–98)

Table 48.2 Pilot-scale OTEC systems tests

Serial number	Major project	Salient features	Remarks
1	U.S. Mini-OTEC Program	A complete OTEC system, including a 670 m cold water pipe (CWP), was constructed, deployed, and tested in the Mini-OTEC program, in 1978–1979, The program was conducted with private funding by an industrial team headed by the Lockheed Missiles and Space Division. Mini-OTEC was constructed in 28 months, using off-the-shelf components.	Net power of 18 kW was produced, in close agreement with the predicted performance, The Mini-OTEC program demonstrated OTEC floating system feasibility in 4 months of at sea operation, including first time production of net OTEC power.
2	OTEC-1 Program This program was conducted by DOE to demonstrate OTEC heat exchanger operation at 35 MWt scale in an at sea environment.	A simulated OTEC power system was installed in which the turbine-generator was replaced by a throttle valve that simulated the pressure drop through the turbine. The equipment included a 35 MWt shell and tube heat exchanger, a 700 m-long, 2 m-diameter plastic CWP and pumps. Provision was made for testing bio fouling control methods such as Amertap and chlorine injection. The CWP was assembled on shore then towed to a site roughly 10 miles off-shore from Kailua-Kona, Hawaii, where it was upended and inserted into a gimbal on the ship. The gimbal allowed movement through an angle of 30°. Significant accomplishments demonstrated during the 4 month at sea operation.	(a) The system performed well with a maximum wave height of 4 m, maximum subsurface currents of 3 knots. Wind gusts up to 45 knots and ship roll angle of 18°. (b) The heat exchanger cycle operated in close accordance with predicted performance. Bio fouling control was accomplished with injection of 60 ppb of chlorine, 1 hr per day. (c) Deployment and operation of the CWP were successful. The pipe survived a sea state that caused the pipe to move to a 30° angle at the gimbal.
3.	Japanese Shore-Based 100 kW OTEC Pilot Plant in cooperation with the Republic of Nauru, Tokyo Electric Power Co. and the Toshiba Corpn in 1980 constructed a 100 kW (gross) OTEC plant at a site on the island of Nauru	Power-generating operation was performed from October 1981 to December 1981. Heat-loop operation was continued to July 1982. The Nauru test was designed to provide accurate experimental data on the performance of the complete power cycle as well as demonstrate total pilot-plant construction and operation. The goals were all successfully accomplished.	The plant provided the first demonstration of land based OTEC net power generation and established a record for total net power production, at 31.5 kW. It was also the first demonstration of OTEC power connection to a utility grid.
4	Japanese Mini-OTEC Program. A Mini-OTEC plant was constructed for at-sea tests off the coast of Shimane Prefecture (1979).	The tests included deployment of a steel CWP, 190 m long, made of 12 m-long tubular sections of 26.7 cm outer diameter (OD) and 0.45 cm wall thickness, joined by flanges. The pipe was supported by a float and held in place by a steel cable attached to the pipe and anchored at a depth of 300 m. The OTEC power plant and water pumps were mounted on the research vessel, which was attached to the float.	The program demonstrated successful deployment and operation of the system, including heat transfer measurements.

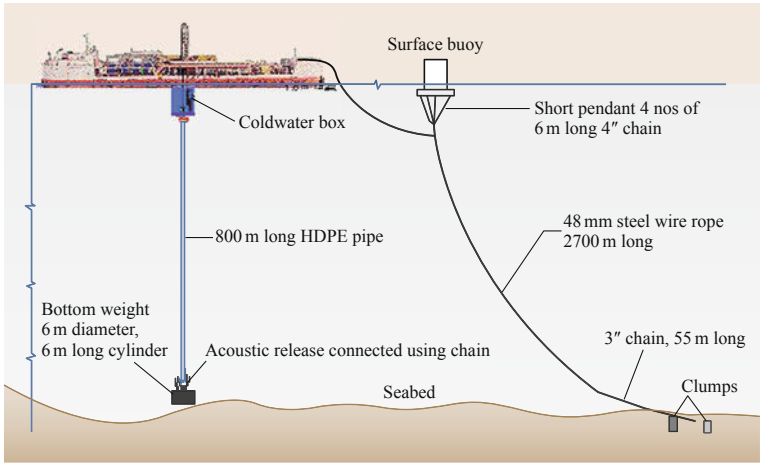


Fig. 48.9 1 MW floating OTEC plant

experimental facility, sponsored by the US Federal Government and the State of Hawaii, at NELHA used a turbine generator that was designed for an output of 210 kWe when the surface seawater temperature is 26 °C and the deep seawater temperature is 6 °C. As of September 1993, the highest surface temperature available was 27.5 °C and correspondingly, 255 kWe (gross) was produced, the largest OTEC output ever recorded (Fig. 48.8). The plant was operational until it was decommissioned in 1998. The major events in the OTEC history are summarized in Tables 48.1 and 48.2.

48.2.1 Indian Experiences

Research on OTEC was intimated in Indian Institute of Technology, Madras, during 1985 with site surveys, feasibility studies, etc. An active work on OTEC was started with the formation of National Institute of Ocean Technology (NIOT), Chennai, in 1993.

Ocean Thermal Energy Conversion

NIOT under the sponsorship of Ministry of Earth Sciences (MoES) designed and developed a 1 MW (gross) floating OTEC plant to be operated between 29 °C and 7 °C using ammonia as the working fluid. The power plant components were assembled on a barge which is to be moored at 1000 m depth using a high-density polyethylene (HDPE) pipe of 1.0 m diameter (Fig. 48.9). The HDPE pipe carries deep-sea cold water to the barge with specially designed moon pool systems. The power cycle of a closed Rankine cycle consists of components such as evaporator, turbine-generator, condenser, and pumps for seawater and working fluid. The warm sea water exchanges heat in the plate heat exchangers with the working fluid, which undergoes a phase change in to vapor and this vapor is allowed to expand in the four-stage axial turbine producing mechanical energy. This turbine drives an AC synchronous generator of 1 MW rating at 440 V and 50 Hz. The vapor from the turbine is allowed to condense in the plate condensers, which is again pumped back to the evaporator using the working fluid pump. Inside the condenser, the heat is exchanged with the cold seawater pumped from the deep sea at 1000 m water depth. The details of the research program are discussed and project could not be commissioned due to failures in the intake joint [48.11–13]. All the precommissioning tests were carried out satisfactorily during 2001–2004 [48.14].

Ocean Temperature Driven Desalination

As spin-off, desalination utilizing the ocean temperature difference was experimented and succeeded in 2004. In low-temperature thermal desalination, the sea surface water at about 28–29 °C is flash evaporated in a chamber under vacuum conditions. The deep sea

Fig. 48.10 100 m^3 per day desalination plant at Kavaratti Island (courtesy of National Institute of Ocean Technology)

water at about 12–14 °C is utilized to condense the evaporated vapor to generate the fresh water. The main attraction of this technology is the operational simplicity and the ease of maintenance. This was commissioning at Kavaratti Island, Lakshadweep group of islands with a capacity of 100 m^3/day as shown in Fig. 48.10. It is producing potable quality water since May 2005 and supplying to the islanders [48.15]. Similar two more plants were installed later in the islands, namely, Agatti and Minicoi during 2010–2011.

48.3 Current Status of OTEC Technologies

The recent trends and technological challenges on OTEC are discussed in the following sections.

48.3.1 Developing Technologies on OTEC

The oil crises in the 1970s gave a boost for OTEC research and there were many international attempts to demonstrate the technology. Many commercial plants were designed based on selected site conditions. After the 1980s, the primary aim was to develop OTEC technologies for commercialization such as new power cycles and new materials for the OTEC power plant. There are many institutions across the world pursuing OTEC research. Many Asian and African countries are interested in OTEC. The climatic conditions required for the operation of an OTEC plant are satisfied in about 100 countries. They are located in a range of 40° north and 40° south latitude. Pilot projects are underway for a 3–10 MW OTEC plant to be built in the Pacific Region. Countries like Korea, Philippines, Sri Lanka, New Guinea, are interested in OTEC.

48.3.2 Research on OTEC in International Scenario

Argonne National Laboratory (ANL) in USA proposed the hybrid OTEC system that has the advantages of both open and closed systems while eliminating the disadvantage of both. Tests of different types of heat exchangers such as shell and tube, plate-frame and plate-fin were conducted at ANL test facility. Vertical double-fluted-aluminum-tube evaporator, sprayed bundle, horizontal shell and tube evaporator, etc. were tested with different working fluids to evaluate the heat-transfer characteristics. ANL also tested an Alfa Laval, titanium-plate-type evaporator with chevron type passages. Special braced aluminum heat exchangers known as Trane heat exchangers were developed for OTEC application and tested up to a range of 1 MW.

Institute of Ocean Energy (IOES) has been continuing their work on OTEC research for more than 30 years. IOES is devoted to the areas of research such as experimental studies on superior power cycles, working fluids, heat-transfer studies by special coating, spray-flash desalination system, production of hydrogen and storage, extraction of lithium and storage, and experimental work on deep-sea water, etc.

USA and Japan are working on various large-scale research projects to develop cheaper and more compact OTEC power systems. Pacific International Center for High Technology Research (PICHTR) was established in 1986 in Hawaii for the US–Japan joint program on OTEC technology. Its main objective is to provide an operational testbed for OTEC system components. The products expected are electric power, fresh water, aquaculture products, and air-conditioning. They have installed a 1 m diameter pipeline at Natural Energy Laboratory of Hawaii (NELHA) for OTEC research.

48.3.3 Commercial Designs for OTEC Plants

Research and development in OTEC technology have been witnessed sporadically in several other countries such as France, Japan, Taiwan, and Indonesia, and also in India where several feasibility studies and preliminary design studies were carried out after eighties. There are site specific, detailed commercial designs available ranging from 25 to 200 MW OTEC plants. It includes land-based plant, shelf-mounted plant, and floating plants. The power utilization is also discussed in some of the reports. The consolidated list of the past designs for commercial OTEC plants is given in Table 48.3. The high cost of capital, coupled with the absence of operational data from a full-fledged pilot plant, has been the major deterrent hampering the commercialization of OTEC technology in the past.

48.3.4 Industrial Proposals for OTEC Commercialization

There are many international agencies available for the commercialization of OTEC technology. Most of them have contracts with international expertise in the field of power plant installation, power distribution, deep-sea operation, etc. Sea Solar Power (SSP), a government-supported consortium, was made up of many of America's most prominent industrial giants. SSP seeks long-term contracts to sell electricity, fresh water, and mari-culture products from OTEC plants

Table 48.3 Conceptual designs of OTEC systems

Serial Number	Major project	Salient features	Remarks
1.	Conceptual Designs of Japanese 100 MWe Floating OTEC Plants during the period 1974–1977, a program, named the *Sunshine Project*, was established in Japan (1980).	First-phase study of a 1 MWe, land-based plant and proceeded to a study of a 100 MWe floating plant. Design of platform structure, station keeping, riser cable design, and heat exchanger improvements. Ammonia as the working fluid. Plate heat exchangers are incorporated in 25 MWe power modules.	Estimated cost 3550 US$/kW (max.)
2	General Electric (GE) 40 MWe Shelf-Mounted Hawaii Plant (1982)	GE proposed a tower-mounted OTEC plant, which would be installed 2200 m offshore at a depth of 100 m, at Kahe Point, Oahu, Hawaii.	Power produced is to be carried by underwater cable to the Hawaiian Electric Company (HECO) plant for distribution to its customers. Total estimated cost – 320 M US$ (8000 $/kW)
3	Puerto Rico Electric Power Authority (PREPA) – a tower-mounted 50 MWe (net) OTEC plant (1985)	PREPA proposed a tower-mounted 40 MWe (net) OTEC plant to be sited at the edge of the continental shelf at 75 m depth approximately 200 m offshore from Punta Tuna, Puerto Rico. Electric power transmission by underwater cable.	Cold water is drawn from a depth of 1000 m at a distance of approximately 1500 m from the tower. Estimated cost – 378 M US$
4	Taiwan 50 MWe (net) shore-based OTEC plant by Taiwan Power Company (1985)	Three shore sites on the east cost of Taiwan were selected. Four 12.5 MWe (net) power modules using ammonia as working fluid. Shell and tube heat exchangers with titanium tubes. A 10 m dia 15 m long CWP made of concrete to be deployed.	Estimate of cost 4920–5690 $/kW
5	French 5 MWe OTEC Plants at Tahiti. Shore-based and floating OTEC plants to be sited in Tahiti. Sponsorship – French Centre National Pour l'Exploitation des Oceans (CNEXO) (1985)	Detailed site studies, evaluation of closed- and open-cycle options, methods of constructing and deploying the cold water pipe and estimates of costs of favored alternatives. Preliminary surveys led to selection of a site at the port of Papeete for the projected plant.	Plant layouts for closed-cycle and open-cycle systems were developed. Total cost Open cycle – 547 M FF Closed cycle – 523 M FF
6	OSEA 100 MWe OTEC Power Plant. A floating OTEC plant of the SSP design that would supply power to the Puerto Rico Water and Power Authority (PREPA). (1985)	(a) Design for submerged operation so that the major power plant components can be placed in the ocean outside the hull (b) A four-stage vapor cycle to improve the average temperature difference between the water flows and working fluid R.22 A *stockade* construction of the CWP with sections of steel pipe	Details of the construction, deployment, and operation are not available. Estimated cost 250 MUS$ (2500 $/kW)

Table 48.3 (continued)

Serial Number	Major project	Salient features	Remarks
7	SOLARAMCO Concept of OTEC 40 MWe (net) Grazing Ammonia Plant ship (1990) Working fluid ammonia	Location of Hawaii where a temperature difference of 22–23 °C would be available. The net OTEC electric power for water electrolysis to produce hydrogen for on-board ammonia synthesis (125 MT /day)	The proposal included provisions for testing four 10 MWe power modules equipped with different heat exchanger types. Estimated cost 250 MUS$ (6250 $/kW)
8	200 MWe (nom) OTEC floating plant Methanol. (1990)	Oxygen and hydrogen to be produced by water electrolysis in combination with low cost coal to produce a low-pollution liquid vehicle fuel. The power plant includes four power modules at each end of the barge. CWPs, 7–10 m diameter made of fiber reinforced plastic (FRP) or lightweight concrete, are supported by a universal joint or ball-and-socket mounting.	Energy transfer is done by production of 1750 tons/day of methanol. Estimated cost (Max.) 1313 MUS$ (6560 $/kW)

with 100 MW net electric power and 400 MLD capacity desalination plants. SSP claims that four independent engineering evaluations have been carried out and their OTEC design is technically sound and economically feasible. Xenesis is a Japanese-based company having technical collaboration with Saga University, Japan.

Makai Ocean Engg. was involved in the design, development of deployment plan, identification of materials, and fabrication in various sea water intake installations in the past. They were also involved in the engineering of sea–water system for Mini OTEC. Design and deployment of pipes ranging from 300 to 1400 mm were carried out. In 1987, Makai has deployed the 1000 mm HDPE pipe made of two sections of 915 m length and 610 m length to a depth of 701 m for Hawaii Ocean Science and Technology (HOST) Park. A pipe of 450 mm diameter was installed in the following year to replace a 300 mm pipe installed 6 years earlier. In 2001, Makai completed the installation of a 1.4 m diameter HDPE sea water supply system from a depth of 915 m with an approximate length of 3050 m to pump about 1705 kg/s seawater at 4 °C. This is the largest and deepest cold water intake system in the world and was installed for the State of Hawaii. These pipelines were for OTEC and aquaculture research. However, the Hawaii OTEC plant does not work for a number of sound reasons, but cold water is still being pumped up from the bottom of the ocean to support the businesses that flourish from the use of the cold water. The OTEC industrial park has grown from 300 acres to 547 acres because of this successful use of the cold nutrient-rich water (Fig. 48.11).

Fig. 48.11 Aqua-culture farm at the Natural Laboratory of Energy Hawaii (courtesy of NELHA)

48.4 Design Considerations for Future OTEC Plants

There are successful OTEC demonstration plants in the past and the commercialization of OTEC technology is expected in the nearby future. Research and development works are continuing and OTEC is mature enough to establish commercial plants. However, there are constraints and limitations which need further attention. The following discussion is on the selection of working fluids, equipment, materials on OTEC, and the utilization of power generated.

48.4.1 Choice of Working Fluids

The selection of the working fluid is important in an OTEC plant. The selection of the equipment, size of the pipe, working conditions, etc. depends on the selection of working fluid. The following are the important characteristics to be considered while choosing the working fluid:

1. Low-temperature boiling and condensation between 25 and 10 °C
2. High heat transfer coefficient for liquid fluids
3. Material compatibility: metals, oils, gaskets, etc.
4. Environmental friendly
5. Low-specific volume and high heat of vaporization
6. Safety: Nontoxic, nonflammable, and nonexplosive.

The following working fluids are commonly suggested:

1. Ammonia
2. Fluoro carbon (R-22)
3. Tetra fluoro ethane (HFC-134a)
4. Propane (C_3H_8)
5. Propylene (C_3H_6) (propane).

All fluoro carbons are ozone depleting and hence cannot be considered for any future OTEC projects. Ammonia is easily available with better thermodynamic properties; it is corrosive especially with copper and its alloys. Also, it is highly toxic for storage and handling. Ammonia cannot wash away any grease deposits on the pipeline and heat exchanger surface. Propane is combustible and each liters of propane contains 26.2 MJ energy. It is easy to detect for leak and its volume increases approximately 270 times when it is vaporized. There is no possibility of water or air pollution with propane.

Thermodynamic and physical properties of HFC-134a, coupled with its low toxicity and nonflammability makes it a very efficient and safe refrigerant. HFC-134a has long-time stability with copper, steel, and aluminum. It has the zero ozone depletion potential (ODP) and has the global warming potential (GWP) 1200. HFC-134a possesses no acute or chronic hazard when exposed below recommended exposure limits such as 1000 ppm, 8 or 12 hours' time. However, inhaling high concentration may cause temporary nervous system depression. HFC-134a in the liquid form can freeze skin or eyes on contact causing frostbite. However, it is not an eco-friendly refrigerant. Hence if safety is an important issue the choice is HFC-134a. But ammonia is less costly, easily available and has superior thermodynamic properties such as heat of vaporization, pressure ratio, etc. Experimental studies show that the contamination of ammonia by water adversely affects the performance of the plant. About 1% of water by weight can reduce the heat transfer coefficient by 8–10% and the resulting power loss. Hence, it is advisable to keep water concentration below 0.1% when ammonia is used as the working fluid [48.16]. A comparison of different working fluids is shown in Table 48.4.

Table 48.4 Properties for working fluids for OTEC

	Ammonia		Propane		HFC – 134a	
Chemical formula	NH_3		C_3H_8		CF_3CH_2F	
Molecular weight	17.03		44.09		102.0	
Temperature	25 °C	10 °C	25 °C	10 °C	25 °C	10 °C
Density **[kg/m^3]**						
Liquid	603.68	625.0	496.08	516.45	1205.9	1259.8
Vapor	7.903	4.907	19.42	13.57	32.359	20.243
Vapor pressure, saturated, kPa	1003.5	615.20	918.62	633.10	666.06	414.92
Heat of vaporization [kJ/kg]	1165.21	1225.03	337.50	364.02	178.0	404.5
Specific heat **[kJ/(kg K)]**						
Liquid	4.82	4.695	2.760	–	1.44	–
Vapor	3.088	2.694	1.703	2.52	0.204	–
Conductivity **[W/(m K)]**						
Liquid	0.477	0.512	0.091	0.103	0.0824	–
Vapor	0.031	–	0.015	–	0.0145	–

Safety Consideration in the OTEC Plant

In an OTEC plant, large amount of working fluid is to be stored and handled (Fig. 48.12). At low concentration in air, ammonia vapor irritates the eyes, nose, and throat. Inhalation of high concentrations produces a sensation of suffocation, quickly causes burning in the respiratory tract, and may result in death. Liquid anhydrous ammonia causes severe burns on contact with the skin, while its effect on being inhaled would cause severe corrosive action in the mouth, throat, and stomach. Safety guidelines to be followed as large quantity of working fluid is to be handled in the OTEC plant during the following operations such as transfer to storage tank, venting and draining, testing the components, and plant operation. A detailed checklist should be prepared before each operations and the entire team is to be briefed about the operation in detail. Continuous monitoring of the pressure in the pipeline should be done. In the case of emergency, the operation personnel are required to evacuate through escape route marked on both sides of the platform to ensure the safety for them.

48.4.2 Choice of Materials

The material selection for OTEC heat exchangers is based on two criteria: (i) corrosion resistance to marine environment and (ii) compatibility with working fluid. Various materials are considered such as aluminum, cupro-nickel, stainless steel, and titanium. The following material combination can be considered for OTEC plant heat exchanger with the working fluids:

- Titanium – Ammonia
- Cu–Ni – Propane
- Cu–Ni – HFC-134a
- Aluminum – Propane
- Aluminum – HFC-134a

Fig. 48.12 Ammonia transfer to OTEC plant (courtesy of National Institute of Ocean Technology)

Aluminum and Aluminum Alloys

Aluminum grades 5052, 5083, 5086, and 6061 are suitable for seawater applications. This metal can be safely used with any working fluid considered for OTEC. It has a high thermal conductivity, low density, and good malleability but poor weldability. It is subjected to pitting corrosion at sea water velocities greater than 3 m/s, but is quite inert to seawater at velocities less than 2.2 m/s. This is because the passive protective layer on its surface, formed due to oxidation, is not stripped away by seawater at low velocities. Other metals, which may be used along with aluminum in a marine environment, are those that acquire a similarly oxidized surface condition in the course of time acquiring corrosion resistance. Special technologies are to be developed to control the galvanic corrosion. Titanium must be well insulated from aluminum if the latter is not to be destroyed by galvanic corrosion. Suitable cladding is to be done when aluminum is used. Heat-exchanger tubes can be made of aluminum with proper cladding. ANL carried out extensive studies for the use of aluminum as the heat-exchanger material. The special heat exchanger developed and tested at ANL is known as Trane heat exchanger [48.17]. Studies at IIT Madras showed that Al alloys requires 15% less tonnage compared to titanium and 2% less floor area. The cost of titanium is nearly 10 times that of Al alloys.

Stainless Steel

The oxygen content of seawater is sufficient to maintain a passive film of chromic oxide on the surface provided the seawater velocity is high enough to prevent debris settling on the surface. The recommended lower velocity limit is 1.5 m/s. If debris deposition or fouling occurs, differential aeration corrosion would set in causing pitting. It should be noted, however, that higher seawater side velocities would result in higher pressure drop in the exchanger, which should be avoided. Where stainless steel is employed, an alloy of greater chromium content, such as 355 S 80, should be used, since the behavior of lower percentage chromium alloys in marine environment is less reliable. Also, stainless steel containing molybdenum (326 S 16) has a greater resistance to corrosion at low seawater velocities. Steel belonging to the austenitic grade is used in marine application (Type 316 and 304). The mechanism of corrosion protection for stainless steels differs from that for carbon steels, alloys steels, and most other metals. In these cases, the formation of a barrier film of true oxide separates the metal from the surrounding atmosphere. Stainless steel grades AL-6XN and 29-4C are reported to be having excellent sea-water resistance compared to conventional duplex steels. The high nickel chromium

content provides a corrosion resistance against chloride stress-corrosion cracking.

Austenitic grades of steel are susceptible to pitting and crevice attack. These will be minimal if velocities are maintained below 1.5 m/s. The addition of molybdenum is usually used in high-velocity service systems such as pump impellers. 3% molybdenum has been used in the impellers of the OTEC plant. Molybdenum provides longer induction period before pitting or crevice attack. Deep-ocean experiences for these steels were varied with slight edge attack or tunneling. Stainless steel is compatible with any working fluid and has good machinability and weldability.

Cupro-Nickel

The most common alloy combination are the 90/10 and 70/30 cupro-nickels. All cupro-nickels are incompatible with ammonia. However, they have various other advantages. They are corrosion resistant below a seawater velocity of 4.5 m/s. The 90/10/1.5 copper–nickel–iron alloy is also very resistant to fouling by marine organisms. Chlorination of seawater, to prevent fouling, does not increase the corrosion rate of cupro-nickel. But seawater polluted with H_2S enhances the corrosion rate. Titanium and cupro-nickel can be used together without appreciable galvanic corrosion.

Titanium

It is virtually immune to seawater due to the self-healing ability of its protective film. Titanium cannot be used in conjunction with various copper alloys, mild steel, and aluminum as these undergo galvanic corrosion. Titanium is compatible with most working fluids. In most marine applications, a thin oxide film forms on titanium and provides complete protection. In spite of titanium's excellent corrosion behavior, some problems encountered in the marine environment are: (i) pitting at oxygen-starved crevices when seawater is heated at 250 °C (ii) stress-corrosion cracking in the presence of tensile stresses and surface flaw. These types of effects would not be encountered in the OTEC plant as the plant operates at ambient temperatures of 28 °C. Titanium has the ability to withstand high velocities with negligible attack. It is also the most expensive of the materials considered.

In conclusion the titanium and high-alloy stainless steel (AL-6XN and 29-4C) can have a life of 30 years. Cu-Ni 90/10 alloy is a good candidate for sea water but not compatible with ammonia. Aluminum alloys are suitable for OTEC applications but mechanical cleaning is to be limited to prevent erosion. Pitting corrosion was observed with deep-sea cold water in aluminum alloys. An average life expectancy of 12–15 years can be considered for aluminum.

Materials for the Structures

Carbon steel is the most commonly used material for marine structural applications. The most significant parameter that influences the behavior of this metal is the availability of oxygen needed to depolarize cathode regions, which can affect the rate of attack. Normally a rust coating, or a calcareous scale, or a layer of bio fouling will develop within a few months on the surface and will act as an interfering barrier. Deep ocean corrosion rates for carbon steels tend to be lower compared to surface seawater. The rate of attack tends to decrease with time because of the low temperatures. Additions of 0.1–0.5% nickel are effective in promoting increased resistance to the marine environment. Addition of 1–2% of chromium results in the decrease of corrosion rate by half. Proper epoxy-based coating suitable to marine environment should be provided for all marine structures made of carbon steel. Composite technology is being developed and this can be a good alternative for marine structures of the OTEC plant due to the light weight and corrosion resistance.

48.4.3 Selection of Equipment

Major cost of the OTEC project is on the power module components such as heat exchangers, turbine, seawater pumps, storage systems. The following section is a description on the principles of selection of these major components.

Selection of Heat Exchangers

Heat exchangers are the most critical components of the OTEC plant. Keeping in view the stringent requirements of low approach and maximum heat recovery, counter current type plate heat exchangers can be considered both for evaporators and condensers. OTEC heat exchangers work on large heavy duty due to the low thermal efficiency of power cycle even as low as ≈ 2.5%. Seawater on the one side and a two-phase mixture of ammonia on the other side makes the process complex. The capital cost of the heat exchangers is nearly 30–50% of the overall cost of the plant. The heat transfer surface required is 7–8 m^2/kW (net) for the evaporator and condenser. The conventional shell and tube heat exchanger design, with a low ratio of heat transfer to exchanger volume, would require very large volumes (1.7–1.9 times) of the plate heat exchanger. The heat transfer area for the double-fluted-type exchanger is about 1.5 times as large as when using the plate type one. This would increase the material inventory and the fabrication costs. A qualitative comparison between the shell-tube exchanger and the plate-fin exchanger on various points of interest is set out in the Table 48.5.

Table 48.5 Techno feasibility comparison of different heat exchangers

No.	Aspect	Shell-tube	Plate-fin	Plate heat exchanger
1.	Present level of technology in marine conditions	Well-developed	More developments/ work required	Developments are in progress
2.	Volume of heat exchanger	Large	Comparatively smaller	Least
3.	Material inventory	Large	Small	Least
4.	Seawater side heat transfer coefficient	Lower	Higher	Lower
5.	Seawater side fouling	Lower	Higher	Lower
6.	Frequency of use of cleaning procedures	Low	High	Medium
7.	Heat transfer area required	More	Less	Least
8.	Heat transfer area per unit volume of exchanger	Low, about 200 m^2/m^3	High, around 500 m^2/m^3	About 300 m^2/m^3
9.	Fabrication processes	Already existing	Need to be developed	Already existing
10.	Pressure drop in exchanger	Lower	Higher	Medium
11	Maximum working pressure	60 bar	25 bar	15 bar

More experimental data are required on fouling rates, fouling control measures, and development of more economical fabrication processes. Fabrication costs can be greatly minimized if a leak-tolerant system rather than a leak-intolerant one is adopted. In the former, costly and precise welded joints are replaced by rubber seals. Each plate is fitted with a gasket, which seals the channel and directs the fluids into alternate channels. The overall heat transfer coefficient for high-angle plates is about 33% greater than that of low-angle plates under the similar operating conditions. However, the water side pressure drop for high-angle plates is nearly four times greater because of the higher turbulence. The number and size of plates are determined by the flow rate, physical properties of the fluids, pressure drop, and temperature program. The test results of plate heat exchangers in Indian OTEC plant are as shown in Fig. 48.13. Some of the reported values of overall heat transfer coefficient (K) are listed in Table 48.6.

Experimental studies at ANL and Saga University proved that a special coating on the ammonia side can improve the K-value by 30–40% or more. The tiny cavities act as nucleation points and enhance the vaporization process. As OTEC heat exchangers work on low temperature of approach, coating makes appreciable difference. The coating is achieved by spraying molten stainless steel on the sand-blasted plate surface using compressed air (Fig. 48.14).

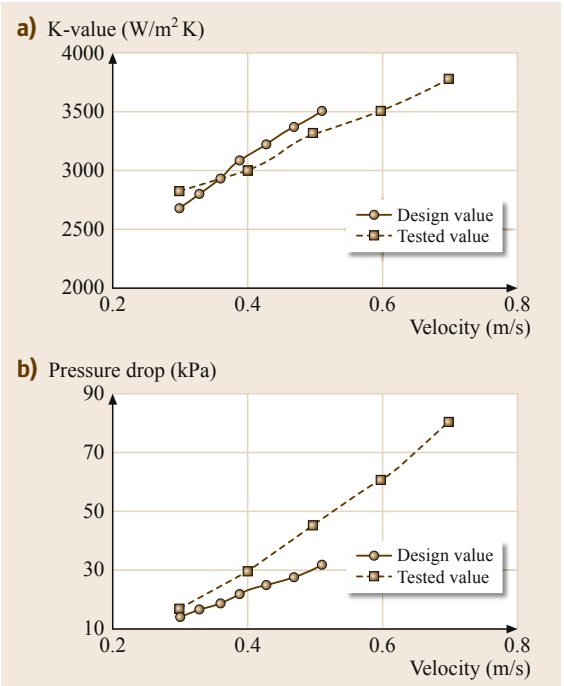


Fig. 48.13a,b The test results on plate heat exchanger. **(a)** Test results of overall heat transfer coefficient with varying flow channel velocity for the thermal spray coated plate heat exchanger; **(b)** pressure loss in the plate exchanger is measured for different channel velocities. These results are compared with design correlations

Selection of Turbine

The selection of the turbine depends on the type of the cycle chosen. Higher capacity plants are to be based on the closed cycle. Axial flow turbines will give a better adiabatic efficiency and part load efficiency compared with radial flow turbines. However the construction cost will be higher. The unit comprising of a turbine-gearbox-alternator is expected to give an overall efficiency of 80%. The main components namely the turbine disk and rotor blades can be made of a special material 17-4PH semi austenitic stainless steel. The 17-4PH SS is known for its resistance to corrosion. The possibility of crevice corrosion occurring in the assembly is ruled out because of the close fits. The material 17-4PH SS can be selected for the turbine disk can withstand high stresses, high resistance to crack propagation and proven resistance to stress-corrosion cracking in marine atmosphere. The shaft can also be made of 17-

Table 48.6 Reported values of overall heat transfer coefficients

Sea water Channel velocity [m/s]	K-value [W/m^{2}K]	Ammonia side of evaporator	Source of data [48.10, 11, 17]
0.61	4801	With coating	Saga University
0.61	4431	With coating	Reported by Alfa Laval
0.60	3145	Without coating	C.B. Panchal's analysis
0.64	3000	Without coating	Reported by Avery
0.60	3662	With coating	Tests on NIOT project

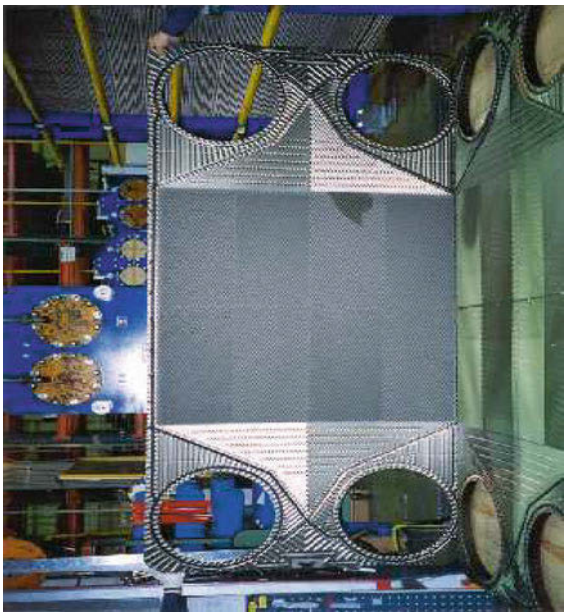


Fig. 48.14 Plate coated on ammonia side (courtesy of National Institute of Ocean Technology)

4 PH Stainless steel H1025 39-42 HRc. The material for all the turbine rotor blades is 17-4PH stainless steel at H1025 condition, which provides greater transverse ductility, lower transition temperatures, and more useful levels of fracture toughness. By pass mode of governing can be adopted for the OTEC turbine. A quick shut-off valve can be provided to the inlet to the turbine as a safety measure.

Selection of Sea Water Pumps

For the OTEC application, the pump required should have characteristics so as to meet the demand of large discharge and low head, which limits the choice to mixed flow centrifugal pumps. Centrifugal pumps have also been considered for pumping ammonia in the working fluid cycle. Seawater pumps would consume about 30–40% of gross power produced by the plant. The choice of the material, strength, and endurance should be taken into consideration at the design stage due to ship motion (for floating plant), fluctuations in submergence, and seawater corrosion. Materials compatible in the marine environment such as stainless steel for impellers and shafts, Ni cast iron for casing are selected based on the superior corrosion resistance exhibited by these materials and their stability in chlorinated seawater adopted for fouling control. A special roller thrust bearing is to be provided to account for about 5% pitching motion of the ship for floating plants. Impellers of pumps are made of ASTM A351 grade materials. The inner walls of the pump casing can be coated with copper oxide and covered by a layer of epoxy. Chlorine is to be introduced into the caisson below each feed pump to prevent the buildup of fouling on the casing and the impellers. Certain steel like the Duplex steel show high pitting probabilities at high potentials and a significantly lower pitting probability as the potential is reduced.

48.4.4 Sea Water Intake System and Deployments

One of the critical components of an OTEC system is the pipe which will carry the cold water from a depth of 1000 m to the OTEC platform. These seawater intake pipes have to handle huge quantities of water and hence are generally large in diameter. The pipes must withstand forces of the ocean waves, currents as well as corrosion, and bio fouling over a long period of time say, 25–30 years. The pipes in general will have length from 1 to 4 to 5 km depending on the type of the plant. The design of the connection of the pipe to the floating platform and the study of the pipe's behavior under various environmental conditions is are challenging task. The sea–water systems of the land-based plants have special requirements for the discharge pipe and the surf zone transition for all the pipes of the system. In floating plants the pipe can be a part of mooring as in the mini-OTEC plant. In another way, the mooring is separated from the pipe and the pipe is vertically suspended from the barge itself. OTEC-1 experiments used this configuration.

Materials for Intake Pipe

The selection of most economic and technically suitable pipes with a desirable life expectancy is a problem by itself. The choice of pipes must be made with regard to its capacity to withstand load, resist pressure, min-

imum increase in temperature during the flow of cold water from intake to sump and corrosion, which may be caused, by the seawater. In addition to the conventional steel pipes, there are several other pipe materials like cast iron, concrete, HDPE, FRP. Most of the experimental plants, installed till date, have used either HDPE or polyethylene pipes. For the proposed plants, HDPE has been selected as the material for bringing cold and warm water. Table 48.7 gives the properties of various types of pipes available [48.18].

Station Keeping Systems

There are two alternatives to position: a floating platform at a point such as dynamic positioning system and the mooring system. The dynamic positioning system with a thruster is more costly than a mooring system. However, mooring system for deep waters is still a technical challenge. The selection of a suitable mooring system for offshore platforms depends on several factors such as water depth, environmental conditions, type and size of platform, platform motions, period of operation, whether temporary or permanent system is required, maximum allowable drift, cost. There are three types of mooring systems (Fig. 48.15):

1. Single-point mooring in which the platform is held by a single wire or a chain system. Mini OTEC proved that even the HDPE pipe can be used for mooring.
2. Vertical tension mooring in which some tension is introduced in the mooring lines.
3. Multi-leg catenary moorings in which different wires are used to keep the platform in position.

Single-Point Moorings. Single-point moorings have weather-vaning capability and are economical for deep-water conditions as they always try to minimize the environments loads on the structure by heading into the weather. A wide variety of designs are available for single-point mooring system even though all of them perform the same function. Figure 48.9 is an example of single-point mooring system.

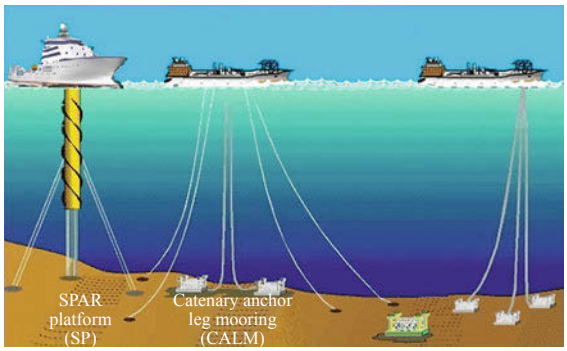


Fig. 48.15 Station keeping systems for OTEC barge

Turret Mooring. In this system, a number of catenary mooring legs are attached to a turret that is an essential part of the vessel/platform to be moored. The turret is provided with bearings to allow the vessel to rotate around the legs.

Catenary Anchor Leg Mooring (CALM). In this system, a large buoy is used to support a number of catenary chain legs anchored to the seabed. The flow lines coming from the sea (the cold water pipe lines in the case of OTEC plant) are connected to the bottom of the buoy and a hawser; typically, a synthetic rope connects the floating platform to the buoy. Since the response of the CALM buoy is totally different from that of the floating platform under the action of waves, this particular system has some limitations in withstanding severe environmental conditions. Hence provisions are made to disconnect the platform from the buoy during extreme environmental conditions. For cases where disconnection is not possible, rigid structural yokes with articulations are provided to tie the ship/platform to the top of buoy. The rigid yokes are designed in such a way that they totally eliminate the horizontal motion between the platform and the buoy.

Single-Anchor Leg Mooring (SALM). This system uses a vertical riser system that has a large amount of buoyancy near or on the surface and is held by pretensioned riser. A tubular articulated riser with a fixed yoke is employed. Chain risers with soft mooring connections can also be provided. The riser behaves like an inverted pendulum due to a large amount of vertical buoyancy available at the surface. Whenever the system is displaced from its original position, the pendulum action tends to bring back the riser to its old position. As mentioned in the earlier case the floating platform is connected to the top of the SALM buoy either by a rigid yoke or a flexible hose. The bottom portion of the riser is connected to a piled or dead weight concrete or steel structure on the seabed. In deep water, the SALM system usually has mid-span articulation (Fig. 48.16).

Dynamic Mooring Systems. Dynamic positioning (DP) is used either as a sole source of station keeping or used to assist a catenary mooring system. Dynamic positioning system consists of a position reference system with a thruster controlled using computer around the vessel. When dynamic positioning system is used, it is called as DP assisted mooring and thruster assisted mooring when thrusters are manually controlled. The use of dynamic position assisted mooring systems for

Table 48.7 Comparative statement on different types of pipe materials

Serial number	Properties	Cast iron	Concrete	Steel	HDPE	FRP
1	Strength	Fair	Fair	Very good	Fair	Good
2	Density	Heavy (7500 kg/m^3)	Light (2400 kg/m^3)	Heavy (7950 kg/m^3)	Lightest (955 kg/m^3)	Light (1500 kg/m^3)
3	Flexibility	Fair	Fair	Good	Very good	Very good
4	Tensile strength	196 N/cm^2	59 N/cm^2	24 500 N/cm^2	588 N/cm^2	9800 N/cm^2
5	Resistance to sea water	Affected	Less affected	Affected	Least affected	Least affected
6	Protection coating	Required	Not required	Required	Not required	Not required
7	Cathodic protection	Required	Not required	Required	Not required	Not required
8	Hydraulic properties					
	Friction	Fair	Good	Good	Best	Best
	Buoyancy	Least buoyant	Least buoyant	Least buoyant	Highly buoyant	Buoyant
9	Proven record	Good	Good	Good	Not available for long time usage	Not available for long time usage
10	Thermal conductivity	51.128 W/m^2/K	0.963 W/m^2/K	62.748 W/m^2/K	0.499 W/m^2/K	0.209 W/m^2/K
11	Indigenous availability	Yes	Yes	Yes	Yes	Not known

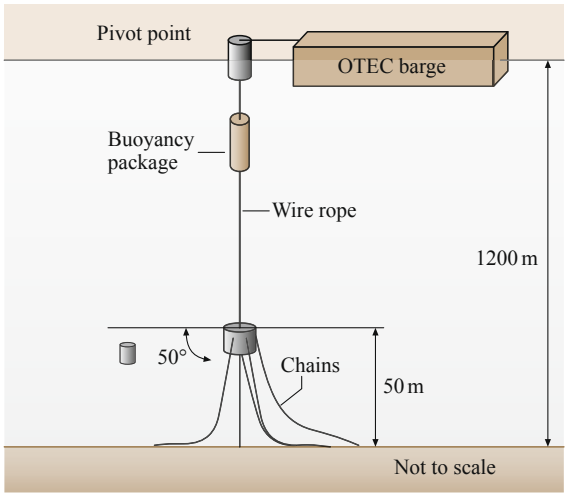


Fig. 48.16 Counter weight articulated mooring system

ship shaped vessels has drawn considerable attention in recent years. The ability to control vessel heading, damp vessel motions, and subsequently reduce mooring line loads, and improve operating efficiencies, are key factors for consideration.

For the economic design the realistic estimation of the load on the mooring system is to be carried out considering all wind, wave and current forces. The problem of fatigue and corrosion are major challenges in the mooring system for an OTEC plant.

Deployment of Pipe

The cold-water pipe is a critical component of an OTEC plant since it conveys the cold water to the power plant. Since the temperature of cold water required is around 6–7 °C, the pipe has to be about 1 km long. The long pipe has to be horizontally towed and vertically upended in a water depth of over 1000 m. This task, that is, the deployment of the cold water pipe is a complex operation. Fortunately, materials like HDPE floats in water hence an HDPE pipe will float by itself. Other materials may need flotation devices to make the pipe float. HDPE is flexible however it also has a limiting bend radius and during deployment it must be ensured that the bending stresses do not exceed the limiting values. After the pipe is towed to the site it has to be made vertical, for this a crane barge is required to lower the pipe at a slow rate and capable of handling all the weights attached to the pipe. The mooring has to be first deployed and subsequently after the pipe is made vertical, connection to the vessel need to be made. The upending of the pipe is a very delicate task and may involve air locks, which will lead to failure, and hence the entire procedure needs to be carefully planned and executed. Connections to vessel involve divers and several onboard equipment like winches and cranes are used for handling. Generally, a large barge fitted with handling equipment, a large vessel with positioning system for support and logistics, tugs, and several fishing craft are required for the deployment procedures.

48.4.5 Bio-Fouling and Control

Any solid surface in seawater begins to acquire a film of microbes in the form of detectable slime layer within a few hours. This film grows in thickness and impairs the performance of the heat exchanger by acting as a thermal insulator. The rate of micro fouling on a sur-

face is dependent on a number of factors such as the character of the surface, kind and amount of nutrients present near the surface, oxidizing capacity of the environment, pH of the seawater. With the fast flow of seawater past the heat exchanger walls, the loosely attached top parts of the film break away. However, with time the film becomes harder and dense decreasing the heat transfer coefficient even up to 60%. The resistance to heat flow is measured as the reciprocal of heat transfer coefficient and is designated as the fouling factor (R_f). The typical fouling factors generally considered for surface seawater is $0.00005\,m^2K/W$ and for deep water is $0.000015\,m^2K/W$. Experimental studies at NELHA over a period of 5 years showed that negligible bio-fouling of heat exchangers at deep-sea cold-water side [48.19]. Bio-fouling in the evaporator could be prevented by suitable chlorine dosing. The following are the general conclusions:

- Fouling characteristics of open seawater are comparable irrespective of site
- Fouling characteristics for near shore/coastal water and open seawater are different.
- A low level of intermittent chlorination may prevent fouling buildup of heat exchanger surfaces.

There are many bio-fouling control measures as discussed below.

Chlorination

In this system, chlorine is injected into the water intake in the form of NaOCl or with the help of a small electro-chlorine generator system. This injection can be continuous such that the concentration of chlorine never exceeds 0.5 ppm, or it can be intermittent at regular intervals with chlorine concentrations as high as 1.2 ppm.

Ultrasonic Cleaning

In ultrasonic cleaning, high-frequency sound waves remove the bio fouling from the heat exchanger surface. The power required to be transmitted is of the order $1\,W/cm^2$ of the heat exchanger surface area. Other methods such as Amertrap rubber sponges and mechanical cleaning using brushes have also been effective. Slurries of silica particles of various sizes and concentrations circulated past the heat exchanger surfaces were of limited effectiveness.

Tight control over heat exchanger cleanliness is required as the gap between successive plates in a plate exchanger is only 3–6 mm and hence fouling can easily impede flow. Based on preliminary evaluation liquid sodium hypochlorite was selected as a possible biocide for fouling control in OTEC operation based on its effectiveness, easy availability, low cost, and ease in mid sea replenishment of stock. Subsequently experiments were conducted in near-shore waters, at residuals of 0.2, 0.5, 0.8, 1.0, 1.2, and 1.5 mg/l in simulated flow chambers (Pederson device) and slime control was achieved at continuous residual levels of 1.2 mg/l. A log reduction 0.8 times in bacterial density between control and chlorinated plates was observed at this regime. Later on experiments were conducted at the OTEC site by changing the regime to an intermittent mode, with a view to reduce chlorine inventory on the OTEC barge. These studies showed an intermittent chlorination regime of 1.2 mg/l residuals (two hours interval) was effective for slime control. Since the control of fouling is critical for heat transfer efficiency, the dose and frequency evolved in coastal and oceanic waters using flow chambers were validated using a model plate heat exchanger at the oceanic site onboard a research vessel. A summarized result from the experimental studies is shown in Table 48.8.

An intermittent chlorination of 1.2 ppm with periodical mechanical cleaning would be the ideal method for bio-fouling control for the OTEC plant. However site specific bio fouling studies have to be carried out for establishing the chlorine dosing regime [48.20].

48.4.6 Power Transmission from Offshore OTEC Plants

The fact that OTEC plants, if built, would be stationed at sea, 4–100 km (or even more) off-shore, raises the question of how, and perhaps in what form, the energy produced by these plants will be made available to the user markets on land. The link between the OTEC plant and the land-based energy markets is an important part of the total system, as the overall credibility of the OTEC concept will depend upon the availability and economics of the transmission system. Power generated by OTEC plants moored less than 25 km off-shore may be transmitted through submarine cable as high-voltage AC ($\approx$ 150 kV). For longer distances, up to 200 km, high-voltage DC (direct current) ($\geq$ 250 kV) transmission has to be resorted to. Underwater bulk-power transmission at depths (1–2 km) and distances characteristic of OTEC conditions has not been attempted so far. This is bound to pose numerous engineering problems, particularly with regard to platform positioning

Table 48.8 Thickness of biofilm developed on the titanium

Chlorine dose regime	Biofilm thickness [μm]		
	1st day	3rd day	6th day
Continuous – 1.2 ppm	9	15	20
2 h interval – 1.2 ppm	12	18	32
4 h interval – 1.2 ppm	13	22	45

and cable-design and installation. The DC cable will have to be in one piece since no splices or joints are permissible in lengths exceeding 100 km.

For OTEC plants, which are more remotely sited than a few tens of kilometers from the coast, both engineering and economic considerations seem to favor strongly the conversion of the primary OTEC electrical energy to hydrogen by electrolysis of deionized seawater. The hydrogen thus produced could be transmitted as gas through submarine pipes if the distance is not large, or else converted in the OTEC plant itself to liquid hydrogen or liquid ammonia (NH_3), which could be shipped in special containers or tankers to the land terminals. The latter alternatives offer greater operational flexibility in the disposal of the energy product. Submarine cables are usually subject to much more onerous installation and service conditions than an equivalent land cable and it is necessary to design each cable to withstand the environmental conditions prevailing on the specific route. Subject to certain restrictions, paper insulated solid-type cables, oil-filled cables, gas-filled cables, and polymeric cables are all suitable for submarine power cable installations. There are under water cable installations up to a depth of 550 m to a length of 127 km (between Denmark and Norway).

48.5 Conclusion

The OTEC offers a great potential of energy for the generation of continuous power with renewable nature like conventional power plants. The ever-increasing energy demands combined with the alarming ozone layer depletion rate and the socio-environmental impact of fossil fuels necessitate focusing the attention to the ocean for the energy needs. There are many countries in the world, which are keenly interested in the development of OTEC. There are technological and economical hurdles to overcome. But building the confidence and reliability on the ocean energy technologies is the present challenge. For the main land applications rating of 25 MW and above the cost of power generation is expected to be comparable to the fossil fuel plants. However, OTEC power plants will be cost competitive to island communities, even today. In view of the enormous OTEC resource available in many countries and because of its renewable and pollution free nature, this technology development program of smaller power plants and generation of experimental data should be treated as investment for the future technologies and not expect to be cost competitive at present with conventional fuel plants.

References

48.1 M.M. El-Wakil: *Powerplant Technology* (Tata McGraw-Hill Education, New Dehli 1984)

48.2 N. Lenssen: Providing energy in developing countries. In: *State of the World*, ed. by L.R. Brown (Norton, New York 1993) pp. 228–235

48.3 C.B. Panchal, K.J. Bell: Simultaneous production of desalinated water and power using a hybrid-cycle OTEC plant, J. Sol. Energy Eng. **109**(2), 156–160 (1987)

48.4 M.A. Syed, G.C. Nihous, L.A. Vega: Use of cold seawater for air conditioning, Proc. OCEANS, Vol. 1 (1991) pp. 60–64

48.5 F. Matsuda, T. Tsurutani, J.P. Szyper, P. Takahashi: The ultimate ocean ranch, Proc. IEEE OCEANS'98, Vol. 2 (1998) pp. 971–976

48.6 H. Kobayashi, S. Jitsuhara, H. Uehara: The present status and features of OTEC and recent aspects of thermal energy conversion technologies, 24th Meet. UJNR Mar. Facil. Panel (2001) pp. 1–8

48.7 G.T. Heydt: An assessment of ocean thermal energy conversion as an advanced electric generation methodology, Proc. IEEE **81**(3), 409–418 (1993)

48.8 B.K. Parsons, D. Bharathan, J.A. Althof: *Thermodynamic Systems Analysis of Open-Cycle Ocean Thermal Energy Conversion (OTEC)*, NASA STI/Recon Technical Report N 86 23043 SERI/TR-252-2234 (Solar Energy Research Institute, Golden 1985)

48.9 H. Uehara, Y. Ikegami, T. Nishida: OTEC system using a new cycle with absorption and extraction processes, Phys. Chem. Aqueous Syst. (1995) pp. 862–869

48.10 W.H. Avery, W. Chih: *Renewable Energy from the Ocean: A Guide to OTEC* (Oxford University Press, New York 1994)

48.11 M. Ravindran: The Indian 1 MW floating OTEC plant – An overview, IOA Newsletter Vol.11 NO.2/Summer 2000, International OTEC/DOWA Association (2000)

48.12 V. Jayashankar, J. Purnima, M. Ravindran, M. Mitsumori, H. Uehara: The Indian OTEC program, Offshore Technol. Conf. (1998), OTC-8905-MS

48.13 M. Ravindran, R. Abraham, S. Zacharia: Environmental friendly energy options for India, Int. J. Environ. Stud. **64**(6), 709–718 (2007)

48.14 M. Ravindran, R. Abraham: The Indian 1 MW demonstration OTEC plant and the pre-commissioning tests, Mar. Technol. Soc. J. **36**(4), 36–41 (2002)

48.15 R. Abraham, T.R. Singh: Thermocline-driven desalination: The technology and its potential, Int. J. Nucl. Desalination **2**(2), 109–116 (2006)
48.16 C.B. Panchal, J.J. Lorenz, D.L. Hillis: *Effects of Ammonia Contamination by Water on OTEC Power System Performance, ANL/OTEC-PS-8* (Argonne National Laboratory Report, Lemont 1981)
48.17 C.B. Panchal, D.L. Hillis, J.J. Lorenz, D.T. Yung: *OTEC Performance Tests of the Trane Plate-Fin Heat Exchanger, ANL/OTEC-PS-7* (Argonne National Laboratory Report, Lemont 1981)
48.18 J.R. Vadus, B. Taylor: OTEC cold-water pipe research, IEEE J. Oceanic Eng. **10**(2), 114–122 (1985)
48.19 C.B. Panchal, L. Genens, D.L. Hillis, J. Larsen-Basso, S. Zaidi, T. Daniel: Bio fouling and corrosion studies at the seacoast test facility in Hawaii, Proc. OCEANS (1984) pp. 364–369
48.20 P. Sriyutha Murthy, R. Venkatesan, K.V.K. Nair, M. Ravindran: Biofilm control for plate heat exchangers using surface seawater from the open ocean for the OTEC power plant, Int. Biodeterior. Biodegrad. **53**(2), 133–140 (2004)

49. Offshore Wind Energy

Mareike Strach-Sonsalla, Matthias Stammler, Jan Wenske, Jason Jonkman, Fabian Vorpahl

In 1991, the Vindeby Offshore Wind Farm, the first offshore wind farm in the world, started feeding electricity to the grid off the coast of Lolland, Denmark. Since then, offshore wind energy has developed from this early experiment to a multibillion dollar market and an important pillar of worldwide renewable energy production. Unit sizes grew from 450 kW at Vindeby to the 7.5 MW-class offshore wind turbines (OWT) that are currently (by October 2014) in the prototyping phase.

This chapter gives an overview of the state of the art in offshore wind turbine (OWT) technology and introduces the principles of modeling and simulating an OWT. The OWT components – including the rotor, nacelle, support structure, control system, and power electronics – are introduced, and current technological challenges are presented. The OWT system dynamics and the environment (wind and ocean waves) are described from the perspective of OWT modelers and designers. Finally, an outlook on future technology is provided.

The descriptions in this chapter are focused on a single OWT – more precisely, a horizontal-axis wind turbine – as a dynamic system. Offshore wind farms and wind farm effects are not described in detail in this chapter, but an introduction and further references are given.

49.1 **Current Offshore Wind Turbine Technology** 1269
49.1.1 Rotor 1271
49.1.2 Nacelle 1274
49.1.3 Support Structure 1278
49.1.4 Control and Protection System 1282
49.1.5 Power Electronics 1284
49.1.6 Challenges of Current Technology 1285
49.2 **Fundamentals of Turbine Dynamics and the Offshore Environment** 1285
49.2.1 Wind 1287
49.2.2 Waves 1291
49.2.3 Aero-Hydro-Servo-Elastic Models for Horizontal-Axis Wind Turbines 1293
49.3 **Outlook on Future Technology** 1299
References 1301

The use of renewable energy sources increased rapidly since the middle of the last century. For example, the percentage of installed wind power in the European Union (EU) increased from 2.2% in 2000 to 11.4% in 2012 [49.1]. To foster this trend, the member states of the EU have agreed on the target of generating 20% of their energy consumption from renewable energy sources by the year 2020. There are mainly two reasons for this: independency from fossil energy sources, which have become more expensive in recent years, and environmental awareness.

Wind energy has a large share in energy transition. As of 2012, 282 430 MW of wind power capacity have been installed, of which 5410 MW are located offshore [49.2]. Although the installed offshore wind capacity corresponds to only 1.9% of the total installed wind capacity, it has to be stressed that the offshore wind industry is relatively young and has been growing rapidly only in the last decade; the installed offshore wind capacity rose from 4 MW in 2000 to 1188 MW in the member states of the EU in 2012 [49.3]. In several countries, both politics and industry are eager to push offshore wind energy. For example, [49.3] states that *31 companies have announced plans for 38 new offshore turbine models*, and the German government wants to install 25 000 MW of offshore wind capacity by 2030 [49.4].

In general, an offshore wind turbine (OWT) generates electricity from the power contained in the wind. The electricity produced is fed into the grid and transported to land where it is used by households and industry. In contrast to fossil energy sources, the electricity produced by an OWT cannot be controlled on demand; an OWT can only produce electricity based

on current wind conditions. The wind is random and cannot be collected and put aside for later use. For large penetrations of wind power, this is an issue when establishing grid stability and providing a continuous electricity supply. The problems arising from the intermittent electricity fed into today's grid is causing a shift in the way grids are thought of and how they need to be designed in the future. Currently, a lot of research is devoted to this field. However, the electricity grid is not the topic of this chapter, but it is referred to in the description of [49.5].

This chapter gives an overview of the state of the art in offshore wind turbine technology and introduces the principles of modeling and simulating an OWT. The OWT components are illustrated, along with how the wind is used to produce energy, how ocean waves and corresponding loads are described, and how an OWT can be modeled and simulated. The chapter is concluded with an outlook on future technology. The descriptions in this chapter are focused on a single OWT – more precisely, a horizontal-axis wind turbine (HAWT) – as a dynamic system. Offshore wind farms are not described in detail, as this would be beyond the scope of this chapter. An introduction and further references can be found in [49.5].

A last note regarding usage of *offshore wind turbine*s and/or OWTs: although not explicitly stated, most of the following explanations are valid for land-based wind turbines as well. However, to avoid confusion, OWT is used throughout the chapter.

49.1 Current Offshore Wind Turbine Technology

In this section, the technology of OWTs is described. The components of an OWT are introduced following the definition of the International Electrotechnical Commission (IEC) in [49.6] (shown in Fig. 49.1). Accordingly, the term *offshore wind turbine* describes a system consisting of a rotor-nacelle assembly (RNA) and support structure. The focus of this section is on bottom-fixed three-bladed horizontal-axis wind turbines like the ones shown in Fig. 49.2 as they are used in commercial offshore wind farms today. Examples of technologies that differ from this concept are introduced in Sect. 49.3. The RNA consists of the rotor, the hub and rotor blades, and the nacelle, containing the generator and the drivetrain with a corresponding support frame. The RNA is described in Sect. 49.1.1 (rotor) and 49.1.2 (nacelle). The term *support structure* includes the foundation, the substructure, and the tower (Fig. 49.1) and is described in Sect. 49.1.3. The control system and the power electronics are an important part of a wind turbine and their location varies among OWT concepts, which are described in Sect. 49.1.4 (control system) and Sect. 49.1.5 (power electronics).

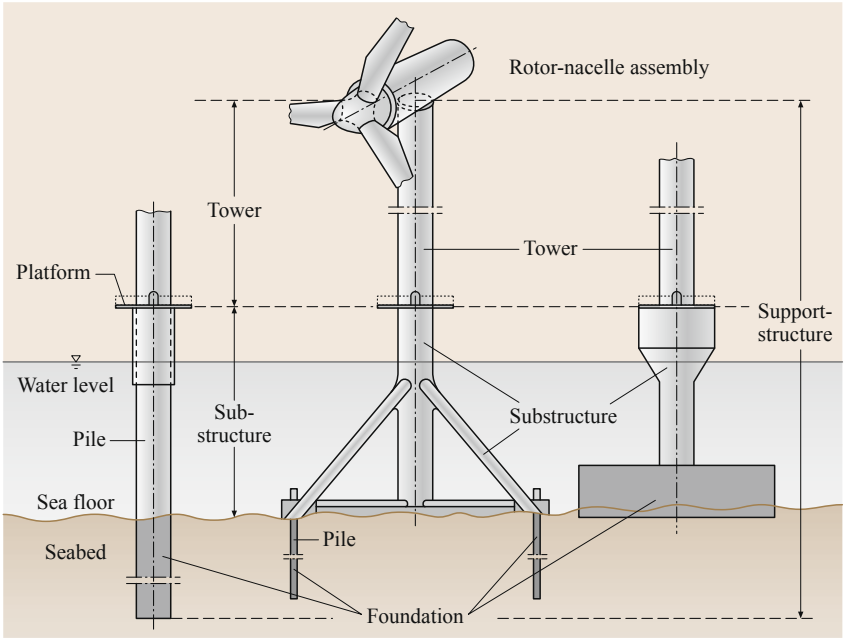


Fig. 49.1 Offshore wind turbine components according to IEC (after [49.6])

49.1.1 Rotor

The rotor converts the kinetic energy in the wind to rotational energy of the rotor and drivetrain by utilizing aerodynamic lift. Based on the arrangement relative to the wind direction, rotors can be divided into upwind and downwind rotors. Most of today's wind turbines have an upwind rotor with three blades; the rotor faces the wind in front of the tower.

The ratio between blade tip speed, the product of rotational speed Ω and rotor radius R, and the undisturbed wind speed v_{wind} is denoted as tip-speed ratio λ and is given in (49.1). The tip-speed ratio is a key parameter for the aerodynamic design of a rotor that is explained in standard wind energy textbooks like [49.5] or [49.7].

$$\lambda = \frac{\Omega R}{v_{\text{wind}}} \,. \tag{49.1}$$

Currently used OWTs are designed for tip-speed ratios of about 5–8 [49.8] and thus, deliver high generator speeds through a low torque. Also, most OWTs are variable speed, they can change the rotational speed of the rotor and can, therefore, operate at the optimal tip-speed ratio in a range of wind speeds. However, for variable-speed turbines, the electricity generated does not meet the grid frequency and has to be converted to be fed into the grid.

The power that is produced by the rotor has to be controlled to avoid violating limits like rated power. This approach protects the machinery and prevents overloading of the generator. A generic power output curve of a variable-pitch controlled wind turbine is shown in Fig. 49.3; the power produced is zero below the cut-in wind speed, increases cubically for wind speeds between cut-in and rated wind speed, is constant above rated wind speed, and becomes zero again after reaching the cut-out wind speed.

For the aerodynamic control of a rotor, variable-pitch-to-feather control is most common. For these rotors, the rotor blades are rotated around their longitudinal axis according to the wind conditions. In above-rated wind speed, the rotor speed is fixed and the pitch angle of the blades is increased to decrease the resulting inflow velocity's angle of attack and to keep the produced power constant. For variable-pitch-to-feather rotors, the connection between rotor blades and hub is rather complicated because of the necessary bearings and actuator systems.

Another possibility to aerodynamically control the power produced by the rotor is to utilize the stall effect, which can be done actively or passively. For passive-stall controlled rotors, the rotor is kept at a fixed rotational speed over a range of wind speeds. Furthermore, the rotor blades are fixed at a certain pitch setting angle and cannot rotate along their longitudinal axis. For an active-stall controlled rotor, the rotation speed of the rotor is varying and the rotor blades are pitched actively to stall. In both cases, the resulting inflow velocity's angle of attack is increased. This increase causes the flow on the blade to detach and thus decreases the lift on the blade and reduces the power produced. It is not easy to predict stall behavior. Furthermore, the damping and fatigue behavior is less favorable than that of variable-

Fig. 49.2 Wind turbines in the German offshore wind farm Alpha Ventus (courtesy of Urs Wihlfahrt)

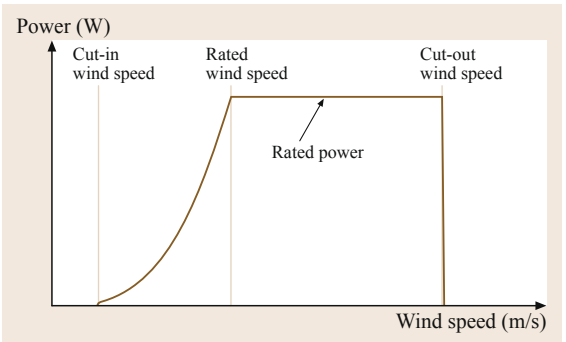


Fig. 49.3 Power output curve for a wind turbine

pitch rotors, resulting in higher blade costs than for variable-pitch-to-feather rotors [49.5], which is why the latter are more commonly used in current OWTs.

A typical OWT has three rotor blades. This configuration results in a constant moment of inertia with respect to the yaw axis (Sect. 49.1.2). More rotor blades would also result in this symmetry but would increase the weight of the rotor and, therefore, the cost of the whole turbine with very little gain in aerodynamic performance. Two or even one rotor blades result in more unfavorable dynamic properties of the rotor without sacrificing much aerodynamic performance [49.7]. Two-bladed rotors for OWTs are a topic of recent research projects.

Rotor Blades

The rotor blades are the most important components of a OWT as they actually extract the power from the wind. They are large flexible structures whose size is often underestimated when they are only seen up in the air as a part of the RNA. Rotor blades for a multimegawatt turbine compared to a truck are shown in Fig. 49.4. Both aerodynamic performance and structural strength are important for the design of a rotor blade.

An initial guess for the blade shape can be determined based purely on the aerodynamic performance following the optimal blade design according to Betz or Schmitz, which is explained in [49.5]. An example for an optimal blade design in terms of chord length, the width of the blade along its radius, is shown in Fig. 49.5. The cross-sections of a rotor blade are different airfoils that are *threaded* along the longitudinal axis. For rotors with high tip-speed ratios, i. e., for those of an OWT, the selection of these airfoils is of particular importance for the aerodynamic performance because the lift-to-drag ratio of the airfoils needs to be high. Also, the aerodynamic control of the rotor influences the choice of the airfoils.

The structural properties of a rotor blade are determined through a trade-off of dynamic properties against material usage. Thin, flexible blades might need less material but tower clearance has to be taken care of (the blades must not crash into the tower). Also, the shape of the blade root is often designed differently from the optimal aerodynamic design because of the loads that have to be transferred into the structure.

Fig. 49.4 Rotor blades for multimegawatt wind turbines compared to a truck (courtesy of Mareike Strach)

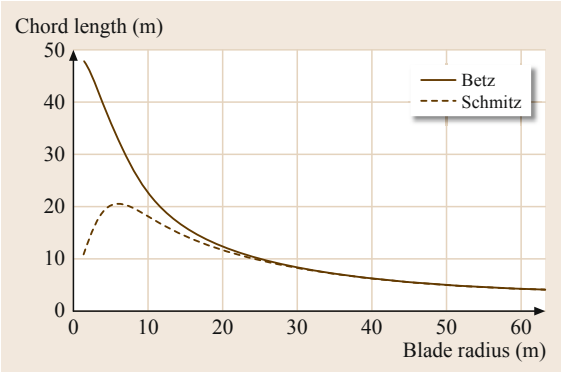


Fig. 49.5 Optimal blade chord length according to Betz and Schmitz [49.5] for a three-bladed rotor with a design tip-speed ratio of 8

Rotor blades are made from composite materials, mostly from glass-fiber-reinforced polyester. Large and flexible rotor blades, which are common for OWTs, also have carbon fibers. The most important processes in blade manufacturing are described in [49.7]. In addition, the automatization of the blade manufacturing process is a current research topic (Sect. 49.3).

Hub

The hub is the component that brings together the rotor blades and connects the rotor to the drivetrain. The most common form of hub for an OWT is rigid, which does not allow blade motions except for the pitch angle variation. That is, the hub and the rotor blades are connected through the pitch bearings and respective drives. There are other types of hubs that allow teetering of the rotor or hinging away of the rotor blades [49.7], but they are mainly used for rotors with two blades. Most hubs are made of steel and are either casted or welded structures.

49.1.2 Nacelle

The nacelle contains the drivetrain, generator, and other equipment such as hydraulic machinery, and protects all of it from the weather. It also provides the support frame for the drivetrain. To prevent salty air from corroding the components within the nacelle, most OWTs have a filtering system. For the commonly used upwind turbines and sometimes for downwind turbines, the nacelle needs to follow the current wind direc-

tion actively to allow power conversion and reduced loads on the support structure. This active control is achieved by a yaw system consisting of a yaw bearing that connects the nacelle with the tower and the yaw drives that turn the nacelle. Maintenance of OWTs is far more challenging than maintenance of land-based wind turbines (Sect. 49.1.6). Thus, some OWT nacelles are designed specifically to facilitate maintenance. For example, some designs ensure that there is enough spare space to store replacement components, or that the onboard cranes have sufficient capacity to unload transporting ships and lift heavy components, and workshops are placed within the nacelle. Additionally, maintenance personnel may have to stay in the nacelle for several days for bad weather or other reasons, therefore, survival packages, first aid boxes, and camp beds are stored there.

Drivetrain

The drivetrain is the rotating connection between the rotor and the generator. Currently, there are three major design philosophies for drivetrains: medium speed, high speed, and direct drive.

Early wind turbines had drivetrains with a three-stage gearbox that sped up the slow rotatory motion of the rotor. With such a gearbox, high-speed generators can be used. Standard high-speed generators work at rated speeds of 1500 rpm in a 50 Hz grid or 1800 rpm in a 60 Hz grid. Although this type of drivetrain has technical advantages, the main reason for combining the gearbox and high-speed generator in the early years of the wind energy industry was the availability of off-the-shelf industrial components.

As a result of several gearbox failures and the complexity of wind turbine gearbox design, German wind turbine manufacturer Enercon started to design directly driven, slowly turning multipole generators [49.9]. Although direct-drive nacelles are typically heavier than those with a classical drivetrain, they had superior reliability for several years.

With the emerging offshore wind energy industry, some companies sought to combine the advantages of the direct-drive and high-speed drivetrains by developing medium-speed drivetrains consisting of a two-stage gearbox and a medium-speed generator [49.10]. This design provides a relatively lightweight drivetrain without using the third (high-speed) stage of the gearbox. This is especially beneficial because the third stage is one of the main failure locations in common gearboxes.

In current OWTs, all three types of drivetrains are used. For several years, direct drives were said to be the dominating technology of the future; however advanced understanding of wind turbine gearboxes and the new medium-speed drivetrains made the race open again. Further insight on drivetrains is provided in [49.11] and [49.5]. A clear statement concerning the overall efficiency of the different wind turbine drivetrain concepts cannot easily be made. The annual yield for a given wind distribution at a particular site is a better benchmark value than the efficiency at the nominal operating point [49.12].

Bearings

Roller bearings are used in all drivetrain components, the pitch system, and the yaw system of an OWT. Bearing designers face several difficulties when developing bearings for wind turbine applications, and bearing failures were a main reason for reduced reliability of wind turbines in the early days of the industry [49.13].

In comparison to other industrial applications, wind turbine bearings face stochastic, varying loads. Within a high-speed drivetrain, rated speeds vary between 10 rpm on the rotor side and up to 1800 rpm on the generator side. Especially, in gearboxes that contain the entire range of speeds, a lubricant is always a compromise between the demands of the large low-speed and the small high-speed bearings. Only modern, synthetic lubricants with a powerful package of additives are able to fulfill these requirements.

Pitch and yaw bearings spend most of their life in standstill while experiencing huge bending moments. When they are turned, the rotating speed is rather low and the movement is short and sometimes oscillatory. This combination of loads prohibits the creation of a lubricating film that separates the different parts of a bearing. Wear instead of fatigue is the dominant damage mode. Bearing design for this application is especially challenging, and methods for testing and calculating the bearing lifetime are currently not available [49.13].

Generator

Today, three different kinds of main generators are used within OWTs: induction generators (IG), doubly-fed generators (DFIG), and electrically or permanent-magnet-excited-synchronous generators (EESG or PMSG). From a theoretical point of view, these types of generators can be combined with the three main drivetrain variants mentioned above. In practice, the whole bandwidth of application can only be found within the high-speed drivetrain designs. Induction generators have a higher weight than the other types and are, therefore, not used in medium-speed and direct-drive drivetrains.

The use of permanent-magnet material for field excitement seems to have the biggest advantages concerning efficiency, specific volume, and weight. However,

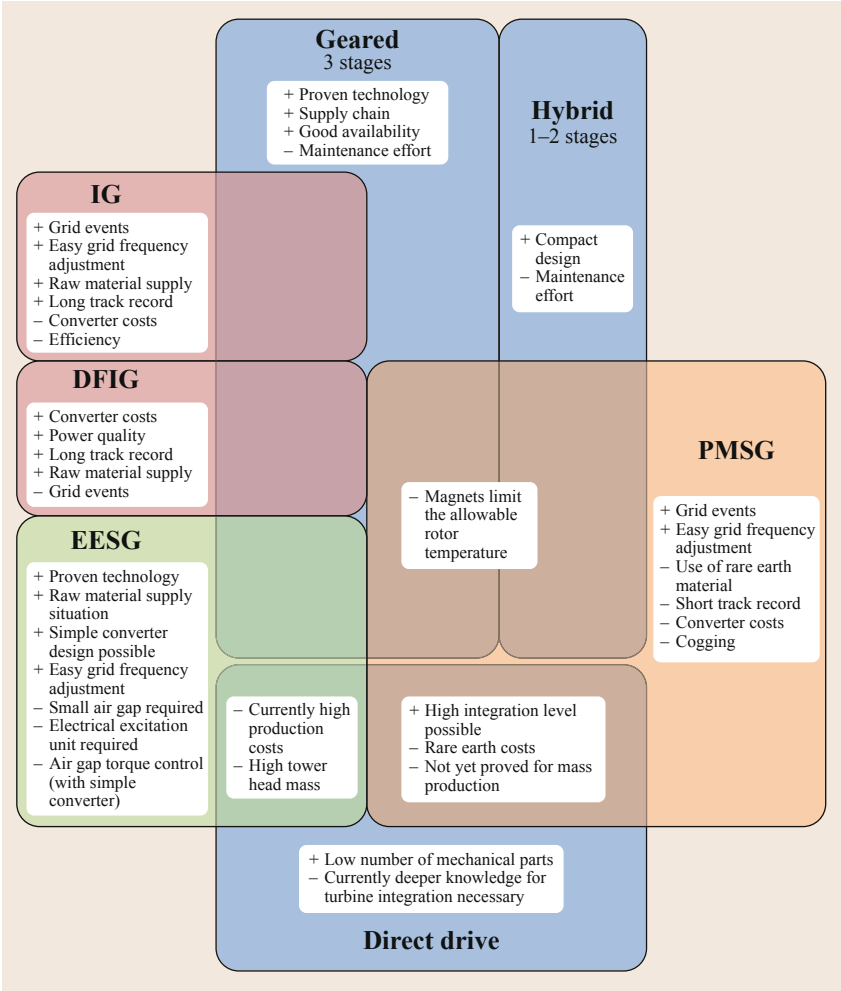


Fig. 49.6 Overview of drivetrain concepts and respective arguments

until now there has been no sufficient track record for direct-drive wind turbine applications, and volatile prices for rare earth materials can negatively influence costs. For example, in 2011, the price for 1 kg of raw neodymium started at 36 EUR, reached a top level of 195 EUR in July, and fell back to 110 EUR at the end of the year. In the same time span, the price for 1 kg of raw dysprosium varied between 243 EUR and 975 EUR, peaking at 1700 EUR. Current direct-drive generator designs need roughly 600–800 kg permanent-magnet material per megawatt of rated generator power, in which 30% of this weight is rare earth materials such as neodymium or dysprosium.

The EESG variants, which are based mainly on Enercon turbine technology, are generally very reliable, but have the disadvantage of a higher weight and a larger required installation space in comparison to the PMSG variants. The weight and size issue connected with various turbine concepts was much discussed in the past; however, it can be stated that modern direct-drive turbine developments by manufacturers such as Siemens are very close to achieving the latest benchmark-geared turbines (Vestas V90-3.0) in respect to their specific nacelle weight.

Among other things, the requirements for OWTs differ from the requirements for land-based wind turbines because of the different cost structure. Failures can lead to extensive downtime and cause high production losses. Consequently, ensuring high reliability and sophisticated service concepts are crucial (Sect. 49.1.6). As a result, direct-drive designs allow OWTs to get rid of gearbox reliability issues and gearbox maintenance. Most of the recently developed and announced offshore wind turbines by manufacturers such as Siemens, XEMC Darwind, and Alstom are direct-drive PMSG turbines. In contrast, other OWT manufacturers, like Sinovel and Repower, favor geared concepts with DFIG. Other companies, such as Ves-

tas and Gamesa, have developed OWTs with medium-speed PMSG. Nonetheless, a current trend towards direct-drive PMSG for OWTs is observable [49.12]. Figure 49.6 gives an overview of the above-mentioned concepts and the related technical arguments.

49.1.3 Support Structure

As mentioned earlier, the support structure represents the tower, substructure, and foundation of an OWT. In theory, all combinations of the different variations of each of these parts are possible [49.14]. Details on the substructure and foundation are provided below, with the focus on currently existing configurations. They are described together because the foundation and substructure cannot be separated uniquely for all configurations. An example of this would be a gravity-based substructure in which the substructure and foundation overlap to some extent.

Tower

Practically all of today's OWTs have tubular steel towers. Usually, they are built up from different tower segments (Fig. 49.7) that are bolted together at the site. A limiting factor for the tower segment diameter is a result of logistics (maximum height under bridges). For OWTs, this limitation can be avoided when the tower segments are manufactured close to the shore, making street transportation unnecessary. Towers usually have ladders and/or elevators as well as platforms on different levels inside to make the RNA accessible (for maintenance).

An important factor for the tower design is to prevent resonant vibrations by placing the first natural frequency of the tower far away from blade passing frequencies and/or rotational frequencies – typically once per revolution (1P) and three times per revolution (3P). The *tuning* of the first natural frequency is provided in greater detail in [49.5].

Fig. 49.7 Tower segments for multimegawatt offshore wind turbines (courtesy of Mareike Strach)

Substructures and Foundations

Offshore wind farms currently use monopiles, jackets, tripods, tripiles, and gravity-based substructures. They differ in appearance, how the loads are transferred to the soil, the way they are fixed to the seabed, and the water depths where they can be installed with respect to economic feasibility. So far, floating platforms for OWTs have not been used in commercial farms but in prototypes and are subject to research (Sect. 49.3). In general, when designing an OWT substructure, it is important to take into account the large overturning moment caused by the rotor thrust that is potentially much greater at rated wind speed than at cut-out or even the 50-year extreme wind speed. All substructures are equipped with so-called secondary steel, that is, with ladders, boat landings, j-tubes for cable guiding, and so on. Although secondary steel is essential for the operation and maintenance of an OWT and might induce vibrations and/or supplementary loads on the system, it is not primary, that is, load-carrying steel. Therefore, it is not discussed in detail here.

Monopiles. Monopiles are single steel tubes that are pile-driven or drilled into the seabed. A transition piece connects the monopile and the tower mainly to balance out possible inclinations in the pile and to provide possibilities for boat landing. The necessary diameter and mass of a monopile increases disproportionally with increasing water depth because the hydrodynamic loading dominates the overturning moment for great water depths [49.15]. Therefore, monopiles are usually uneconomic for water depths greater than about 30 m and are mainly used in shallow waters. Both vertical and horizontal loads are transferred to the soil through skin friction and shear forces, which requires a sufficient pile length and diameter.

Jackets. The jacket in offshore wind energy represents a three-dimensional (3-D) truss, usually with four inclined legs (the whole substructure is referred to as the jacket). Jacket designs for the German offshore wind farm Nordsee Ost are shown in Fig. 49.8. The spacing of the legs on the top of the jacket requires a transition piece to connect the tower and substructure. Just as for the monopile, this transition piece can also be used for boat landings. In theory, there is also the option of using a full-length jacket (without the transition piece and tower); however, this option has not been installed until now. There are two ways to fix a jacket

Fig. 49.8 Jacket substructures for the German offshore wind farm Nordsee Ost ready for transportation in the harbor of Bremerhaven, Germany (courtesy of Wolfhard Scheer)

Fig. 49.9 Tripod substructures for an offshore wind farm ready for transportation in the harbor of Bremerhaven, Germany (courtesy of Wolfhard Scheer)

to the sea bottom [49.5]: pile sleeves (piles are driven through tubes that are fixed at the bottom of the legs) and jacket stabs (plugs at the bottom of the legs are grouted into predriven piles). For a jacket, the loads are mainly transferred into the soil as axial loads in the legs. Also, a jacket has a higher resistance against overturning moments than a monopile because of the spacing of the legs at the sea bottom. Compared to monopiles, jackets can be considered as lightweight structures. Also, jackets can be assumed to be hydrodynamically transparent. Both load-bearing capacity and hydrodynamic properties make jackets suitable for greater water depths. However, as it is a large structure composed of multiple members that require a lot of welding, the manufacturing and logistics of jackets are more complicated and thus more expensive than for monopiles in shallower water.

Tripods. Currently installed tripod substructures are center-column tripods (Fig. 49.9). That is, they consist of a central tube that is supported by inclined tubes, which are connected through braces to provide sufficient stiffness. Transition pieces (Fig. 49.10) are used to connect the tripod and tower. Existing tripods are connected to the sea bottom using pile foundations. Because of the spacing of the legs, a tripod has a high resistance against overturning moments, similar to a jacket. However, the large central member can cause high hydrodynamic loads, and the geometry is even more complicated to weld than a jacket. In general, tripods can be used for the same water depths as jacket substructures.

Tripiles. Tripiles are patented by the BARD group and have only been used in the BARD Offshore 1 wind farm. Basically, a tripile consists of three piles that are

Fig. 49.10 Transition pieces for tripods used in combination with 5 MW OWTs in the German offshore wind farm Global Tech I (courtesy of Mareike Strach)

connected to the tower by a transition piece. For the installation, a concept similar to jackets is used: plugs at the bottom of the piles of the tripile are grouted in predriven hollow piles in the soil. The spacing of the piles provides the required stiffness and also makes the structure hydrodynamically transparent, as the piles have a relatively small diameter. Just like for the jacket, the loads are transferred to the soil through axial loads in the piles.

Gravity-Based Substructures. Gravity-based substructures (GBSs) are massive concrete structures in

which the distinction between the substructure and foundation is not as clear as it is for tripods or jackets. So far, two GBS concepts have been used for OWTs:

1. The base plate on the ground is tapered to meet the tower diameter at mean sea level.
2. A single pile is attached to a base caisson without a smooth transition.

The large weight and covered area of a GBS keep the OWT in place, and the loads are transferred to the soil through pressure on the underlying soil layers. Compared to branched structures like jackets and tripods, GBSs are very sensitive to overturning moments, which means that they are installed in shallower water.

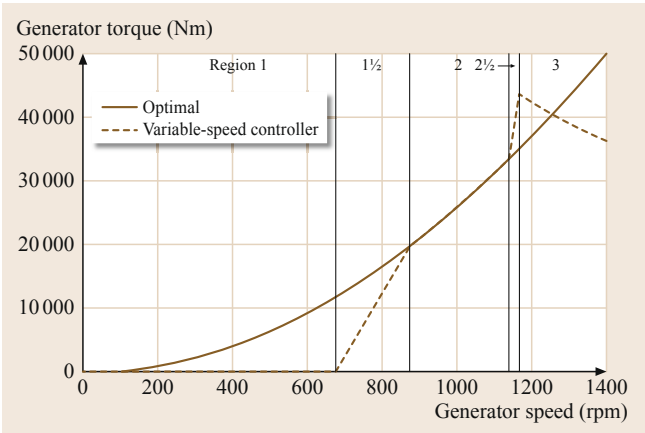


Fig. 49.11 Torque-speed-control curve for the National Renewable Energy Laboratory's 5 MW offshore reference turbine (after [49.16])

49.1.4 Control and Protection System

An OWT and its components include a large set of control systems and subsystems. Here, the term *control system* refers to the rotation speed and torque control as well as operating control. The latter is responsible for releasing and executing braking maneuvers, start-ups and shut-downs, and so on. In the following section, the focus is put on the former – the rotation speed and torque control for a pitch-controlled, variable-speed OWT.

The goals of the control system to maximize energy output while still working within operational limits like rated power, rated rotation speed, and cut-out wind speed, and to minimize static and dynamic loads on the OWT, as well to meet the grid requirements.

A pitch-controlled, variable-speed OWT has different operational modes. Above rated wind speed, the turbine is in fixed-speed mode; the pitch angle of the rotor blades is changed to keep the rotation speed of the rotor as constant as possible. Below rated wind speed, the OWT is in the variable-speed region with fixed pitch to ensure optimal energy production. Based on this principle, variations of control strategies are possible. The torque-speed-control curve of a generic 5 MW OWT is shown in Fig. 49.11. The solid line in the figure gives the aerodynamically optimal torque curve that would lead to an optimal energy output, which is a quadratic relationship between torque and generator speed. The dotted line in Fig. 49.11 represents the real torque-speed curve for the turbine. Region 1 is below the cut-in speed of the generator, i. e., the turbine is in an idling position, Region 1 1/2 is the start-up region, in Region 2 the controller follows the optimal torque-speed-curve, Region 2 1/2 is very steep and covers a wide range of wind speeds and is used to bring the turbine to generating the rated power (following the optimal curve would lead to too high rotation speeds that would cause too much noise), in Region 3 the turbine is in the variable-pitch operation mode.

Braking maneuvers, start-ups, and shut-downs are also controlled by varying the pitch angle and pitch rate of the rotor blades as well as the torque. Other control system functions include yaw control and safety and protection functions.

49.1.5 Power Electronics

Because current OWTs are variable-speed turbines, the electrical power (AC) produced by the generator is of variable frequency. In order for the frequency to be fed into the electrical grid, it needs to be within certain boundaries. For example, the target value would be 60 Hz, or 50 Hz for Europe. AC–DC–AC (AC: alternating current, DC: direct current) converters change the variable frequency power produced by the generator into DC and then back into AC power with the grid frequency.

For OWTs, usually insulated-gate bipolar transistor (IGBT) switches with reverse blocking voltage ability of 1200 V or 1700 V are used, with a DC-voltage level between 750 V and 1200 V and nominal AC-line voltages between 490 V and 690 V. Standard switching frequencies for those low-voltage converters are in the range of several kilohertz. Medium-voltage converters with a higher voltage blocking capability of 3.3 kV or 6 kV, respectively, and switching frequencies of several 100 Hz are increasingly used in OWTs. In general, this leads to lower current values at the same rated power with some inherent advantages such as lower losses, less copper, and higher reliability, but also higher investment costs [49.17].

In OWTs, converters are mostly placed within the nacelle for maintenance reasons and typically contain independent converters working in parallel. This quasi-redundancy allows for power production at a reduced level if one of the systems fail. Power converter reliability is crucial for the economically successful operation of OWTs [49.18]. As the circuit boards are very sensitive to even small interactions with the environment – for example, short circuits can be caused by flies landing on the boards – OWT converters are encased in a cabinet and have a water cooling circuit.

One aspect that is gaining importance with the increasing use of wind energy is the quality of the power fed into the grid and the wind turbine's reaction to grid events. Turbines using full converters have better capabilities in handling grid events than turbines with DFIG. Although the power delivered by the geared DFIG solution is of good quality, more sophisticated control effort and hardware protection systems are necessary to comply with most of today's grid codes. The biggest disadvantage of DFIG is the lack of a complete grid decoupling.

In general, high-speed and medium-speed PMSG/EESG have a higher efficiency than low-speed PMSG/EESG. This is mainly the result of design constraints like material usage, mechanical dimensions, air gap dimensions, and the cooling effort needed. However, the main advantage of the direct-drive concept is the mitigation of losses within the mechanical part of the drivetrain. On the other hand, gearbox turbines – especially those using DFIG technology – show slightly higher efficiency values in their optimal operating point. This is mainly caused by the significantly reduced installed power electronic (approximately 50–70% savings) compared to a full converter design (for PMSG applications).

49.1.6 Challenges of Current Technology

Although a lot of experience has been gained during the last years regarding installation and operation of OWTs, there are still challenges that need to be addressed in the future to reach the desired installed offshore capacity. One of the main challenges is the high cost of offshore wind energy compared to land-based wind energy.

Reasons for the increased cost can be found throughout the whole lifetime of an OWT. The planning and construction process takes more time and effort, mostly because of the substructure design and the grid connection. A big challenge is simply the resource management – that is, scheduling the arrival of parts to be assembled and installed at sea, given the delays in manufacturing or weather windows. The small number of suitable installation vessels can also be a reason for delays in offshore wind projects.

During the operation of an OWT, maintenance is more expensive than on land as the location can only be accessed by boat or helicopter. This limitation requires greater reliability of the condition monitoring system of an OWT [49.19]. There are also problems arising from *marinizing* existing land-based wind turbines (e.g., increased corrosion issues in the nacelle).

Despite these challenges, both research and industry have identified many bottlenecks and are working on ways to resolve them. Because there is already a solid sum of experiences from realized offshore wind projects, design standards are improving and design innovation is taking place (Sect. 49.3).

49.2 Fundamentals of Turbine Dynamics and the Offshore Environment

An OWT is a highly complex and dynamic system that is influenced by a number of linear and nonlinear effects. That is, an OWT is subject to static, cyclic, transient, and stochastic loads that originate from different meteorological, oceanographic, electrical, and structural sources [49.20]. Accordingly, both design and certification of OWTs heavily depend on accurate loads analysis with corresponding pre- and post-processing. Only with appropriate loads analysis can a reliable and cost-effective wind turbine design be achievable.

In addition, an OWT is subject to environmental conditions like turbulent wind, gusts, wind direction changes, wake effects, and ocean waves and currents. Apart from the environmental loading, the rotor of the OWT applies up to 10^9 load cycles of significant amplitude to the system during the lifetime – usually 20 years or more – of the OWT. The effects of these load cycles can be amplified by aerodynamic or structural imbalances of the rotor. Furthermore, the yaw and pitch systems, the mechanical brake, and the generator torque control system cause loads on the components of the turbine that must not be neglected when performing loads analysis.

The complexity of an OWT differs from conventional offshore structures used in the offshore oil and gas industry. For example, oil and gas structures are rather passive and high risk, because they are manned and carry a high pollution hazard. In contrast, OWTs can be considered as highly active structures because of their control system and the spinning ro-

tor, as well as low-risk structures concerning hazards for human life and the environment. Also, the design limit states for offshore oil and gas structures and OWTs are different; although conventional offshore structures are mainly designed against the ultimate and accidental limit state, the fatigue limit state is of higher importance for OWTs. Another goal is to achieve cost-effectiveness during design, manufacturing, and operation. For the latter, that means that an OWT has to withstand the before-mentioned combination of environmental and cyclic loading over its lifetime without causing high operation and maintenance costs. Furthermore, the natural frequencies of the support structures of OWTs have stricter requirements as they must not meet the rotational frequencies of the rotor or the blade passing frequency of the tower (and higher harmonics) to avoid resonance effects.

The extensive experience that exists for land-based wind turbines and for offshore oil and gas structures needs to be combined when designing OWTs; simply mounting a land-based turbine on an offshore structure is not sufficient, as this would neglect the load effects mentioned above and would not lead to a reliable and cost-effective design.

In the following sections, the main sources of environmental loading on an OWT, wind (Sect. 49.2.1) and waves (Sect. 49.2.2) are described. The inclusion of other sources of external loading (currents, ice loads, and seismic loads) are beyond the scope of this chapter. Instead, the reader is referred to the listing of the respective references in [49.20] and to the descriptions of the marine environment in [49.14] or [49.21], respectively. Furthermore, the aero-hydro-servo-elastic approach for modeling and simulating OWTs (for performing loads analysis) is introduced in Sect. 49.2.3.

49.2.1 Wind

Wind is both a source of energy and mechanical loading for an OWT. Stronger wind contains higher energy but also applies higher loading. Furthermore, the wind resources available are crucial for assessing possible locations for an OWT. An important characteristic of wind is its high variability; both wind speed and direction may change rapidly with time and position. It is, therefore, crucial to describe the wind in an appropriate way when performing loads analysis.

Nature of Wind

Wind itself is of a random nature because it varies greatly in both time and space. Following the categorization pattern of [49.7], variations in wind speed and direction can be categorized into *interannual* variations over time scales greater than 1 year, *annual* variations caused by seasonal changes (the wind is usually stronger in autumn and winter than in summer), *diurnal* variations depending on the time of the day, and *short-term* variations that describe changes within 10 min or less (gusts and turbulence).

Figure 49.12 visualizes an example for a wind speed spectrum that shows peaks corresponding to these variations. The so-called *spectral gap* between the diurnal and turbulent peak where the energy density is quite low indicates that short-term variations may be treated independently from variations on a larger time scale. However, in [49.5] it is argued that the wind spectrum shown in Fig. 49.12 is not *universally applicable*. Regardless of this, the *spectral gap* is the main reason for treating mean wind speed and turbulence separately in loads analysis (below).

In general, interannual variations are difficult to predict. For annual variations, however, the occurrence

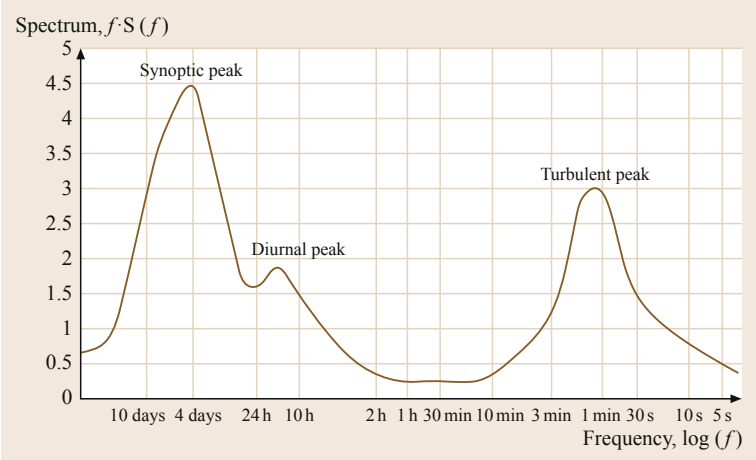


Fig. 49.12 Wind spectrum from Brookhaven, New York, USA (modified from [49.5], originally published in [49.22]. Modified with permission from [49.22]; courtesy of American Meteorology Society 1957)

probability of mean wind speeds can often be displayed by a Weibull distribution. The Weibull probability density function for a wind speed v depends on the two parameters A and k, where A is a scaling factor that determines the mean value of the probability density and k is a shape parameter that determines the width of the density.

$$p(v) = \frac{k v^{k-1}}{A^k} \exp\left(-\left(\frac{v}{A}\right)^k\right) \tag{49.2}$$

Examples of the dependency of (49.2) on different values for A and k are shown in Fig. 49.13.

With shorter time scales, the variation of wind speed is random and harder to predict. The probability distribution of turbulence can be assumed to be mainly Gaussian. This leads to the definition of a three-dimensional turbulence intensity I, which is the ratio between the standard deviation σ of the wind speed and the mean wind speed over 10 minutes, $\bar{v}$.

$$I = \frac{\sigma}{\bar{v}} \tag{49.3}$$

Turbulence is greatly influenced by terrain, with trees, bushes, and hills causing higher turbulence intensities; however, terrain-induced turbulence is of less importance for OWTs. A more important issue for wind turbines in an offshore wind farm are wake effects, in which the wake behind one turbine influences the flow of another turbine. An introduction to this topic can be found in [49.5].

Equation (49.3) is not valid for large changes in wind speed over a short time (for gusts). Therefore, for the loads analysis of OWTs, approximations of gusts according to different models are used (see below).

The wind speed also depends on the location. In general, wind speeds are higher and less turbulent offshore than onshore due to smaller surface roughnesses. The change in wind direction with height and atmospheric stability are also important factors when determining the wind speed, as well as the change in wind speed with height due to the atmospheric boundary layer, which is called wind shear. Examples of wind speed profiles based on logarithmic and power laws are shown in Fig. 49.14, and the corresponding equations can be found in standards and guidelines like [49.6, 23]. In addition, the shape of the wind speed profile over height depends on the surface roughness and the temperature distribution.

Energy in the Wind

Considering an idealized flow and constant air density ρ, the wind passing through a rotor disk with cross-sectional area A has a kinetic energy E, where m is the air mass moving with wind speed v_{wind}

$$E = \frac{1}{2} m v_{\text{wind}}^2 \,. \tag{49.4}$$

Differentiating the kinetic energy with respect to time yields the power P_{wind} that is contained in the wind

$$P_{\text{wind}} = \frac{1}{2} \rho_{\text{air}} A v_{\text{wind}}^3 \,. \tag{49.5}$$

The wind is slowed down when passing through the rotor disk, i. e., the rotor extracts energy from the wind; however, it cannot be stopped completely, as this would mean that the rotor disk is *blocked*, which would violate the continuity assumption in the flow. Thus, not all power contained in the wind can be used for energy production by a wind turbine. The maximum power that can be extracted from the wind theoretically, P_{Betz}, was derived by *Betz* in [49.24],

$$P_{\text{Betz}} = c_{\text{p,Betz}} \frac{1}{2} \rho A v^3 \,. \tag{49.6}$$

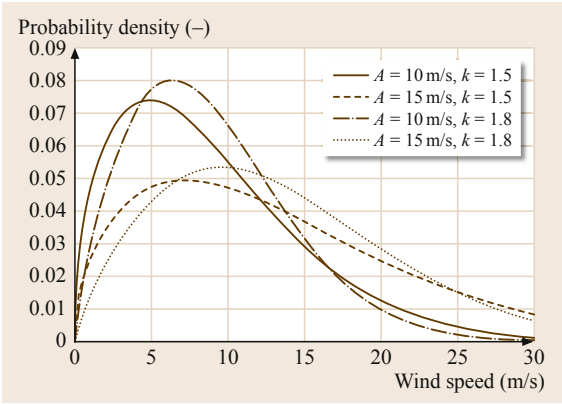


Fig. 49.13 Weibull probability density function for different values of k and A

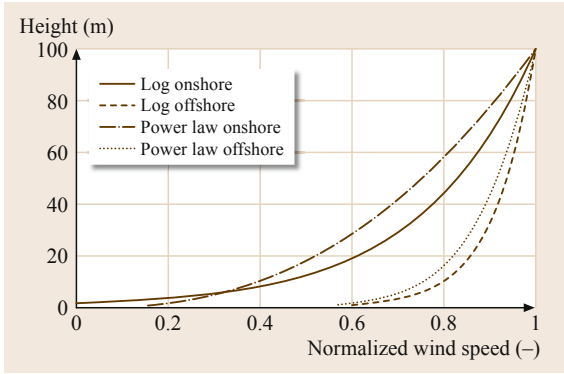


Fig. 49.14 Examples of land-based and offshore wind shear profiles according to logarithmic and power law

The Betz power coefficient $c_{p,Betz}$ equals $\frac{16}{27}$ or 0.59, which means that a maximum of 59% of the energy contained in the wind can be extracted. Practically, this value can not be reached, as friction and other losses reduce the power that is possible to extract.

Description of Wind in Loads Analysis

In loads analysis, standards and guidelines like [49.6] or [49.23] give recommendations on how to produce time series of wind velocities that can be used for the simulation of an OWT. Usually, 10 min time series are needed, for which a mean wind speed is assumed.

To choose an appropriate range for the mean wind speeds, measurements from a chosen site should be available. If this is the case, the measured mean wind speeds are sorted into bins of usually 2 m s^{-1} width and a Weibull distribution is fitted to the occurrence probabilities of the bins. In Fig. 49.15, an example of a wind speed histogram with corresponding Weibull distribution is shown. The Weibull probability density is needed for fatigue load calculation to obtain weighting factors that can be used to estimate the importance of simulated load time series.

To obtain values for the turbulence intensity in each spatial direction, either measurements or recommendations from standards and guidelines are used. The variations caused by turbulence are then added to the mean wind speed so that the wind time series fluctuates around the mean wind speed. These turbulent wind time series are produced numerically using stochastic methods. The turbulence is derived from a wind spectrum for all three directions of wind speed at many points across the rotor plane, ensuring proper discretization and spatial coherence. Therefore, a turbulent wind field is random and different in each realization. The boundary conditions to obtain these random wind fields are given in the standards. Typically used models are the *Kaimal* spectrum [49.25] or the *Mann* model [49.26].

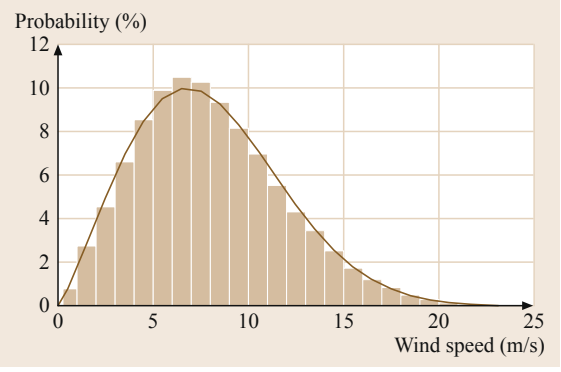


Fig. 49.15 Example of a wind speed histogram with fitted Weibull distribution

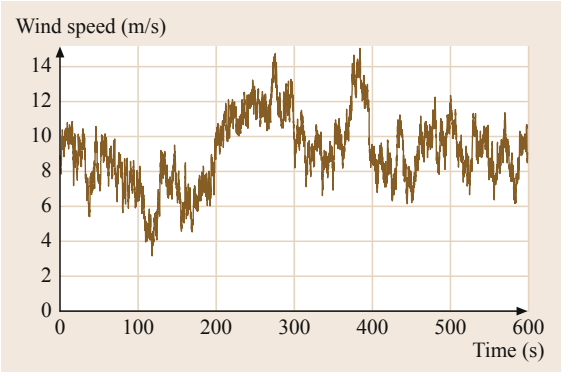


Fig. 49.16 Example of a turbulent 10 min wind speed time series with a mean wind speed of 9 m/s and 20% turbulence intensity

Based on these models, the wind field is generated. An example for a turbulent 10 min wind speed time series for the principle wind direction and a mean wind speed of 9 m s^{-1} and 20% turbulence intensity is shown in Fig. 49.16.

To display extreme wind speeds and gusts, a turbulent wind field is not sufficient. For extreme wind speeds, usually the 1 and 50-year wind speeds, that is, the maximum mean wind speed over a given time interval that statistically is exceeded only once in 1 and/or 50 years, is taken into consideration. Measurements over a long time are usually not available, such that the 1 and 50-year wind speeds can be determined using extreme value extrapolation or based on recommendations in guidelines like [49.23].

As mentioned above, gusts are commonly displayed using a predefined shape. Examples are the *one-minus-cosine* or the *Mexican hat* gust shapes [49.20]. Other shapes can be found in the respective standards and guidelines, like [49.6] and [49.23].

49.2.2 Waves

The description of ocean waves given in this section is purely focused on how waves are represented in offshore wind energy, particularly during loads analysis of OWTs. For more detailed descriptions of the different theories and methodologies that offer a wider focus, see other chapters of this book and/or standard offshore engineering textbooks like [49.27–29]. The basis for modeling ocean waves is sufficient long-term meteorological and oceanographical data. Usually, a joint probability distribution of significant wave height H_s, spectral peak period T_p, and wind speed v_{wind} is provided to select appropriate sea states for a mean wind speed. Such a distribution can be derived based on three-dimensional scatter diagrams for a given location.

Wind speed bin v_{wind} (m/s)		T_p (s)												Probability
		<0.5	1	2	3	4	5	6	7	8	9	10	11	
H_s (m)	4.5													
	4													
	3.5													
	3													
	2.5													
	2													
	1.5													
	1													
	0.5													
	<0.25													
Probability														

Fig. 49.17 Format of a part of a three-dimensional scatter diagram for significant wave height H_s and spectral peak period T_p for a given mean wind speed v_{wind}

An example of the format of a scatter diagram is shown in Fig. 49.17.

The theories used for modeling ocean waves in the context of loads analysis are given in standards and guidelines. According to [49.23], for almost every design load case (for almost every simulation) waves should be modeled as long-crested irregular waves. That is, a number of wave components with different frequencies and phase angles based on linear wave theory [49.27] are superposed to obtain an irregularly shaped wave elevation. This approach is no different to how turbulence at a location is derived from a wind spectrum; the wave kinematics are derived based on this irregular wave elevation. Wave spectra, which represent the energy density of a sea state depending on the wave frequency, are the basis for this description. For loads analysis of OWTs, two spectra are commonly used: the Pierson–Moskowitz spectrum [49.30] for fully developed sea states as well as the JONSWAP spectrum [49.31] for storms, which is derived from the Pierson–Moskowitz spectrum. Discretizing the wave spectrum gives the wave frequencies for the wavelets that are superposed. The phase angles are assumed to be uniformly distributed and are, therefore, of random nature. This approach makes each realization of a sea state different from another.

To display short-term events, such as extreme waves, the design wave method is commonly used, as described in [49.23]. Similarly to the n-year wind speeds, the n-year wave height with corresponding wave period is used to model the design wave. Using the 50-year wave height is recommended when modeling OWTs [49.6]. It is also common to use higher order wave theories to describe the wave kinematics for the design wave (wave theories that are based on Dean's stream function). In this way, the wave crest can be modeled larger than the wave trough, which is more realistic for high and extreme waves.

49.2.3 Aero-Hydro-Servo-Elastic Models for Horizontal-Axis Wind Turbines

The assessment of OWTs is based on aero-hydro-servo-elastic simulation models. Simulations of OWTs are performed in the time domain to capture the nonlinear dynamic effects and time-dependent events that are described in the introduction of this section can be accounted for. Aero-hydro-servo-elastic models display the complex interaction between aerodynamics (*aero*), hydrodynamics (*hydro*), the control system (*servo*), and structural dynamics (*elastic*) under the influence of external loads from the wind (including wake effects), waves, soil, electrical grid, and – where applicable – ice. This coupled approach is required by standards and guidelines to predict the dynamic response as well as extreme and fatigue loads of the OWT. The importance of the coupled models can be seen, for example, in the aerodynamic damping of the rotor, which can reduce the loads on the substructure significantly [49.14]. This effect would be lost when the wave loads are calculated separately from the wind loads in a so-called sequential approach.

To base an OWT design on simulations, the aero-hydro-servo-elastic models need to be detailed enough to predict load effects and dynamic response; however, because a large number of simulations need to be run, the time needed to run a simulation has to be considered when choosing the level of detail for the model.

Rotor Aerodynamics

Methodologies with different levels of detail that are used to calculate the aerodynamic loads on the rotor of an OWT are introduced below. The commonly used

blade-element-momentum (BEM) theory and the generalized dynamic wake (GDW) method are described. For a broader description of OWT rotor aerodynamics of OWT see [49.5] and [49.32].

Blade Element Momentum Theory. The BEM theory is the state-of-the-art method for calculating aerodynamic loads in OWT loads analysis. This method is fast and gives good results – assuming that appropriate aerodynamic airfoil data are available. Therefore, it is very common to use BEM theory with corrections in the wind industry.

In BEM theory, the rotor blade is divided into elements for which induction factors are determined. The induction factors describe how much the wind speed is reduced and redirected by each blade element. Based on the induction factors, the resultant velocities and angles of attack for the aerodynamic loads can be derived for each blade element. Each blade element is associated with aerodynamic airfoil data that are used to determine the aerodynamic loads. The aerodynamic loads influence the induction factors, which makes BEM theory an iterative method that is repeated until the induction factors converge. The calculations are based on the following principles:

- *Newton's law*: The calculation of the wind speed through the rotor disk is based on the equilibrium between aerodynamic loads on the blade elements and the change of momentum of the moving air. The change of momentum equals the change in velocity multiplied by the mass flow.
- *Bernoulli's theorem*: The loads at each blade element equal the average pressure difference on the rotor disk multiplied by the area swept by the rotor.
- *Airfoil theory*: The loads at each blade element can be determined using airfoil theory in which the loads are calculated using the angle of incidence of the wind speed.

BEM theory is based on assumptions that are often violated in reality. The method is, therefore, corrected to make up for these violations – for example, the flow is assumed to be stationary. This assumption requires a correction for instationary and skewed flow. Additionally, airfoil theory has to be modified using unsteady aerodynamic models, including dynamic stall (one possibility is the *Beddoes–Leishman* method [49.33]). It is further assumed that the flow is two-dimensional, which means that adjacent blade elements do not influence each other. This assumption requires the aerodynamic airfoil data to be corrected for 3-D effects. Moreover, the rotor is assumed to be a disk with an infinite number of blades, which makes a correction necessary that accounts for hub and tip losses (tip-loss correction according to Prandtl). Another correction that is often employed is a time filter to *pseudo dynamic wake.* Additionally, the flow reversion around the rotor is neglected. Therefore, for heavily loaded rotors, i. e., mostly for large axial induction factors, a correction according to Glauert is applied. Both BEM theory and the mentioned corrections are described in standard wind energy textbooks like [49.5].

Generalized Dynamic Wake Method. Similarly to the BEM theory, the GDW method provides fast results. The GDW method is a dynamic inflow method and is based on obtaining the pressure by solving the Laplace equation. Originally, it was developed for helicopters in [49.34].

The pressure and velocity field are described as superpositions of functions whose coefficients are obtained by solving a system of differential equations. The calculation of aerodynamic loads is done by applying airfoil theory. In contrast to BEM theory, which is based on a rotor disk, the GDW method allows for the localization of the position of the rotor blade within the velocity field. In this way, most of the above-mentioned corrections for BEM theory are obsolete for the GDW method. However, the dynamic stall method must still be applied.

Hydrodynamics

Similarly to the description of the waves in Sect. 49.2.2, this section focuses on the calculation of hydrodynamic loads for OWTs. For a general description of (moving) bodies in waves and the forces exerted from waves on offshore structures, the reader is referred to other chapters of this book or to standard offshore engineering textbooks like [49.27–29].

For aero-hydro-servo-elastic simulations of OWTs, the *Morison* equation [49.35] is commonly used to determine the loads applied from waves on the substructure. That is, the hydrodynamic forces are assumed to be a sum of inertia and drag loads. The inertia part consists of Froude–Krylov and the scattering term characterized by an added mass multiplied by the water particle acceleration based on the long wave-length approximation of the diffraction problem. Added mass based on member acceleration is typically included separately. The drag part accounts for viscous drag and is proportional to the squared water particle velocity, including structural motions. Other loads on an OWT result from the surrounding water include buoyancy, dynamic pressure, and slap and slam loads.

An obvious drawback of the Morison equation is the assumption of hydrodynamic transparency; it can only be used if the ratio between characteristic dimension and wave length is smaller than 0.2. That means

that radiation, reflection, and diffraction effects (the influence of the body on the wave field) are not accounted for. With the currently used substructures and water depths (Sect. 49.1.3), this is a reasonable assumption. However, with new technology arising (Sect. 49.3), the assumption of hydrodynamic transparency may no longer be applicable. Therefore, some OWT simulation tools already have the ability to account for the influence of the OWT substructure on the wave field. For example, this is done by using the *MacCamy–Fuchs* approximation [49.36] or by using the results from the frequency-domain analysis of the substructure in waves in the time domain, as described in [49.37].

Control System

In general, the control system is of high importance for the modeling of large and flexible wind turbines as it has a large influence on the magnitude of the loads. Without proper representation of the control system in the OWT model, it is not possible to achieve correct estimates of performance and loads. However, the control system is often also the most *secretive* part of the OWT, for which realistic parameters are hard to come by.

Using the example of a variable-pitch controlled OWT, it is necessary to display the time of activation, pitch rate, and maximum angle of pitch. Furthermore, the frequency of the control system must be included to display the time needed for a full control cycle (looking up the generator speed and returning the correct pitch angle). Safety functions of the control system also need to be implemented.

Structural Models

When conducting loads analysis of OWTs, structural models are used to determine load effects depending on the time for various components and positions of the turbine. Nonlinear effects like large deflections, nonlinear aeroelastic effects, nonlinear dynamic properties, and time-varying loads have to be accounted for [49.20]. In short, the structural model has to be coupled to the aerodynamics, hydrodynamics, and the control system model (introduction of this section). Furthermore, components of a wind turbine are coupled to each other, which has to be displayed in the model. For example, the wind-induced movement of the rotor blades can influence the drivetrain vibrations.

Usually, the rotor blades, tower, and drivetrain are modeled as flexible beam components. The nacelle, hub, gear box, and generator are often assumed to be rigid. The main bearing, yaw bearing, and clutches are usually modeled as connections between the respective components.

Choosing an approach for the structural models means doing a trade-off between accuracy of results and simulation time. Therefore, there are three approaches for modeling the structure of an OWT: multibody simulation (MBS), modal reduction approaches, and the finite-element method (FEM). However, combinations of these approaches are even more commonly used (e.g., a combination of MBS and modal reduction or a combination of MBS and FEM). Each approach and respective combinations are introduced below.

Multibody Simulations. In the time domain, the components of an OWT are displayed as rigid bodies that have a mass and are connected by massless joints. The joints can restrain the degrees of freedom (DOFs) of the bodies and elasticity of the joints can be modeled through spring–damper elements. The motion of the coupled bodies can be obtained using Lagrange formalism with generalized coordinates and masses.

Components for which a single rigid body is not sufficient – such as for the rotor blades and the tower – can be modeled through multiple bodies that are connected through springs and dampers. This approach increases the number of DOFs and, therefore, the simulation time. A modal reduction is often performed for these components.

Modal Reduction. It can be said that the higher the number of DOFs in a system, the more time is needed to solve the equations of motion. In a modal reduction approach, the number of DOFs is reduced and, therefore, the simulation time decreases. When simulating OWTs, the equations of motion are solved using an MBS approach; however, components like the rotor blades, tower, and drivetrain are modeled using a modal reduction approach.

Before simulating the whole system, the respective components are modeled using appropriate beam theories based on the geometry and material properties. The mode shapes of the OWT components are determined to be used as shape functions in the (often nonlinear) beam element formulation. The vibrations of the components in the OWT system are then modeled through superposition of a finite number of mode shapes. This method can reduce the number of DOFs drastically, although there is also a loss in accuracy. In loads simulation tools for OWTs, the following configurations are typically used for a modal-reduced model:

- The first three or four mode shapes are used for the rotor blades (two in flapwise, one or two in edgewise direction).
- Four mode shapes are used for the tower (two for bending out of the rotor-plane, two for bending in the rotor-plane).

- Three mode shapes are used for the drivetrain (one for torsion, two for bending).

Although there is a loss of accuracy with respect to the solution of the full system of equations, modal-reduced models agree with measurements. Therefore, they are validated and are used in a number of OWT loads analysis tools.

It should be clarified that modal methods in OWT simulation tools include many nonlinear terms that are important for modeling the dynamic response, such as axial shortening, coriolis and gyroscopic loads, and so on. A possibility is to model the rotor blades and the tower with a nonlinear beam formulation based on a limited number of shape functions. Because there is a limited number of shape functions, it is good to use mode shapes for the shape functions. This is often called a modal method, but it is not a linear approach.

Finite-Element Methods. Modal-reduction methods are efficient when it comes to simulation time. However, large deflections, as they become more important for sizeable and very flexible turbines, often cannot be displayed accurately enough with this approach. Instead, a FEM-based multibody approach can be used, which allows for the simulation of more complex deformation states of the OWT.

Similar to the modal-reduction approaches, the equations of motions of the OWT are solved using an MBS approach, but components like the rotor blades and tower are modeled through beam finite elements. Usually, beam elements are used for the components so that torsional, bending, extensional, and shear stiffness can be considered. This system of FEM-based dynamic MBS-equations can be solved iteratively for the combination of *elastic* DOF (coming from the finite elements) and *rigid* DOF (coming from the multibody connections).

FEM-based methods are slower than modal-reduction approaches in terms of the simulation time; however, they are already being used in some of the loads simulation tools for OWTs.

49.3 Outlook on Future Technology

The state of the art of OWT technology has been described in the previous sections. This section describes the technology that is in the early design stage, only exists in prototypes, and/or is currently subject to research. Technology types include floating offshore wind turbines and rotor configurations that differ from the previously described three-bladed horizontal-axis upwind rotors, new blade technology, and further developments related to analyzing OWTs through loads simulations.

A significant amount of research is currently being done to develop cost-effective and reliable floating OWTs. In these systems, the wind turbine is mounted on a floating platform with an appropriate station-keeping system instead of on bottom-fixed substructures (described in Sect. 49.1.3). It was shown by *Musial* et al. in [49.38] that floating OWTs could become economically feasible. As of today, there are four floating prototypes with multimegawatt turbines in operation. Two are installed in European waters (Norway and Portugal, shown in Figs. 49.18 and 49.19), while the other two are Japanese designs that are installed off Fukushima.

Both the design and modeling of floating OWTs are challenging. The prototypes that have been installed thus far use a traditional three-bladed horizontal-axis wind turbine, as described in Sect. 49.1, mounted on floating platforms that were designed based on experience from the offshore oil and gas industry. This approach has been adopted in many research projects. In contrast, the heavy weight of the RNA high above the floating platform may imply complications in terms of stability, therefore additional research projects are being conducted on vertical-axis floating OWTs. Another research possibility is to study the use of downwind rotors.

For the simulation of floating wind turbines, there are already a number of sophisticated tools available; these are described in [49.39]. However, there are also challenges to developing appropriate simulation tools; these are summarized in [49.40], such as the use of

Fig. 49.18 Hywind prototype: 2.3 MW wind turbine on a spar buoy platform (courtesy of Lars Christopher)

Fig. 49.19 WindFloat prototype: 2.0 MW wind turbine on a semisubmersible platform (courtesy of Untrakdrover, Wikimedia Commons)

more advanced hydrodynamics including second-order effects. Another current research interest is the validation of simulation models through model test data. Yet, this is a difficult area to study because of scaling issues and the generation of both wind and waves at the same time. Field data is rarely available because of the small number of available prototypes and their proprietary nature.

As mentioned in Sect. 49.1.6, the experience gained with OWTs has led to improvements in the certification guidelines. This is also the case for floating OWTs; both the IEC and certification bodies are developing guidelines dedicated to floating OWTs, see [49.23] or [49.41].

Restrictions in the development of land-based turbines do not necessarily apply to OWTs. For example, noise emissions are less of an issue offshore than on land, which makes higher tip speeds possible for OWTs. Furthermore, the ease of transportation offshore and the cost of the support structure, which has a large share of the overall OWT cost, leads to developments with higher rated power – possibly up to 10 MW. For this, larger blades are necessary, which leads to new developments in terms of materials and load reduction methods. A possibility for larger blades would be the use of so-called *smart blades* – blades with sensors and actuators and/or structural properties that allow for load reduction. Another possibility is the individual pitch control of each blade [49.42].

The challenge of making offshore wind energy more cost effective is being addressed by research dedicated to the automated manufacturing of components, such as the rotor blades and simplified logistics (self-installing substructures).

In summary, there is a lot of research dedicated to making offshore wind energy more reliable and cost effective in the future. This is done through the development of new concepts for the overall turbine layout as well as modifying components based on offshore requirements. Furthermore, logistics, guidelines, and aero-hydro-servo-elastic models are constantly being adapted and improved to meet the growth of the offshore wind industry.

References

49.1 EWEA: *Wind in Power – 2012 European Statistics* (European Wind Energy Association, Brussels 2013)

49.2 GWEC: *Global Wind Statistics 2012* (Global Wind Energy Council, Brussels 2013)

49.3 EWEA: *The European Offshore Wind Industry – Key Trends and Statistics 2012* (European Wind Energy Association, Brussels 2013)

49.4 Federal Government of Germany: *Das Energiekonzept für eine umweltschonende, zuverlässige und bezahlbare Energieversorgung* (Federal Government of Germany, Berlin 2010)

49.5 T. Burton, N. Jenkins, D. Sharpe, E. Bossanyi: *Wind Energy Handbook* (Wiley, Chichester 2011)

49.6 IEC: *Wind Turbines – Part 3: Design Requirements for Offshore Wind Turbines*, IEC 61400-3 (International Electrotechnical Commission, Geneva 2009)

49.7 J.F. Manwell, J.G. McGowan, A.L. Rogers: *Wind Energy Explained: Theory, Design and Application*, 2nd edn. (Wiley, Chichester 2010)

49.8 R. Gasch, J. Twele (Eds.): *Windkraftanlagen, 5th edn.* (Vieweg+Teubner Verlag, Wiesbaden 2007)

49.9 E. Hau: *Wind Turbines: Fundamentals, Technologies, Applications, Economics, 3rd edn* (Springer, Heidelberg 2012)

49.10 S. Siegfriedsen, G. Böhmeke: Multibrid technology – A significant step to multi-megawatt wind turbines, Wind Energy **1**, 89–100 (1998)

49.11 P. Jamieson: *Innovation in Wind Turbine Design* (Wiley, Chichester 2011)

49.12 J. Wenske: Special report direct drives and drive-train development trends. In: *Wind Energy Report Germany 2011*, ed. by S. Pfaffel, V. Berkhout, S. Faulstich, P. Kühn, K. Linke, P. Lyding, R. Rothkegel (Fraunhofer Institute for Wind Energy and Energy System Technology, Kassel 2011) pp. 59–63

49.13 M.N. Kotzalas, G.L. Doll: Tribological advancements for reliable wind turbine performance, Philos. Trans. R. Soc. A **368**(1929), 4829–4850 (2010)

49.14 M. Kühn: Dynamics and Design Optimisation of Wind Energy Conversion Systems, Ph.D. Thesis (Delft University of Technology, Delft 2001)

49.15 W.E. de Vries, V.D. Krolis: Effects of deep water on monopile support structures for offshore wind turbines, European Wind Energy Conference (2007)

49.16 J. Jonkman, S. Butterfield, W. Musial, G. Scott: *Definition of a 5-MW Reference Wind Turbine for Offshore System Development* (National Renewable Energy Laboratory, Golden 2009)

49.17 K. Fischer, T. Stalin, H. Ramberg, T. Thiringer, J. Wenske, R. Karlsson: *Investigation of Converter Failure in Wind Turbines – A Pre-Study (Elforsk report 12:58)* (Vindforsk, Stockholm 2012)

49.18 M. Wilkinson, B. Hendriks: *Deliverable D1.3: Report on Wind Turbine Reliability Profiles* (Reliawind, Brussels 2011)

49.19 C. Kupferschmidt, M. Strach, H. Huhn, F. Vorpahl: Offshore wind support structures. In: *Handbook of Technical Diagnostics*, ed. by H. Czichos (Springer, Heidelberg 2012) pp. 505–518

49.20 F. Vorpahl, H. Schwarze, T. Fischer, M. Seidel, J. Jonkman: Offshore wind turbine environmental loads simulation and design, Wiley Interdiscip. Rev. Energy Environ. **2**, 548–570 (2012)

49.21 J. van der Tempel: Design of Support Structures for Offshore Wind Turbines, Ph.D. Thesis (Delft University of Technology, Delft 2006)

49.22 I. van der Hoven: Power spectrum of horizontal wind speed in the frequency range from 0.007 to 900 cycles per hour, J. Meteorol. **14**, 160–164 (1956)

49.23 G. Lloyd: *Guideline for the Certification of Offshore Wind Turbines* (Germanischer Lloyd, Hamburg 2012)

49.24 A. Betz: *Wind-Energie und ihre Ausnutzung durch Windmühlen* (Vandenhoeck and Ruprecht, Göttingen 1926)

49.25 J.C. Kaimal, J.C. Wyngaard, Y. Izumi, O.R. Coté: Spectral characteristics of surface-layer turbulence, Q. J. R. Meteorol. Soc. **98**, 563–589 (1972)

49.26 J. Mann: The spatial structure of neutral atmospheric surface-layer turbulence, J. Fluid Mech. **273**, 141–168 (1994)

49.27 O.M. Faltinsen: *Sea Loads on Ships and Offshore Structures* (Cambridge Univ. Press, Cambridge 1990)

49.28 G. Clauss, E. Lehmann, C. Östergaard: *Offshore Structures – Conceptual Design and Hydromechanics* (Springer, Heidelberg 1992)

49.29 S. Chakrabarti (Ed.): *Handbook of Offshore Engineering* (Elsevier, Oxford 2005)

49.30 W.J. Pierson Jr., L. Moskowitz: A proposed spectral form for fully developed wind seas based on the similarity theory of S. A. Kitaigorodskii, J. Geophys. Res. **69**, 5181–5190 (1964)

49.31 K. Hasselmann, P. Barnett, T.E. Bouws, H. Carlson, E. Cartwright, D.K. Enke, A. Ewing, J.H. Gienapp, E. Hasselmann, D.P. Kruseman, A. Meerburg, P. Müller, J. Olbers, D.K. Richter, W. Sell, H. Walden: Measurements of wind-wave growth and swell decay during the Joint North Sea Wave Project (JONSWAP), Ergänzungsheft Dtsch. Hydrogr. Z. **12**, 7–95 (1973)

49.32 P.J. Moriarty, A.C. Hansen: *AeroDyn Theory Manual* (National Renewable Energy Laboratory, Golden 2005)

49.33 J.G. Leishman, T.S. Beddoes: A semi-empirical model for dynamic stall, J. Am. Helicopter Soc. **34**, 3–17 (1989)

49.34 D.M. Pitt, D.A. Peters: Theoretical prediction of dynamic-inflow derivatives, Vertica **5**, 21–34 (1981)

49.35 J.R. Morison, M.P. O'Brien, J.W. Johnson, S.A. Schaaf: The force exerted by surface wave on piles, Petroleum Trans. **189**, 149–154 (1950)

49.36 R.C. MacCamy, R.A. Fuchs: *Wave Forces on Piles: A Diffraction Theory* (Beach Erosion Board Corps of Engineers, Washington DC 1954)

49.37 J. Jonkman: Dynamics Modeling and Loads Analysis of an Offshore Floating Wind Turbine, Ph.D. Thesis (University of Colorado, Golden 2007)

49.38 W. Musial, S. Butterfield, A. Boone: Feasibility of floating platform systems for wind turbines, 23rd ASME Wind Energy Symposium (2004)

49.39 A. Cordle, J. Jonkman: State of the art in floating wind turbine design tools, Proc. 21st Int. Offshore Polar Eng. Conf. (2011) pp. 367–374

49.40 D. Matha, M. Schlipf, A. Cordle, R. Pereira, J. Jonkman: Challenges in simulation of aerodynamics, hydrodynamics, and mooring-line dynamics of floating offshore wind turbines, Proc. 21st Int. Offshore Polar Eng. Conf. (2011) pp. 421–428

49.41 Det Norske Veritas: *Design of Floating Wind Turbine Structures*, DNV-OS-J103 (Det Norske Veritas, Høvik 2013)

49.42 T.K. Barlas, G.A.M. van Kuik: Review of state of the art in smart rotor control research for wind turbines, Prog. Aerosp. Sci. **46**, 1–27 (2010)

Acknowledgements

A.6 Principles of Marine Corrosion

by Robert Melchers

The author acknowledges the continued provided by Australian Research Council Professorial Fellowship scheme (2004–2008, 2009–2013). He also acknowledges with gratitude the support he receives from a great team of co-researchers and from many in industry.

B.11 Highly Maneuverable Biorobotic Underwater Vehicles

by Promode Bandyopadhyay

The author's various works described in this review have been sponsored by the Bio-inspired Autonomous Systems Programs of the Office of Naval Research (ONR), Dr. Thomas McKenna (ONR 341), and also by ONR341 (Dr. Teresa McMullen, Dr. Harold Bright, and Dr. Linda Chrisey), and ONR33 (Mr. James Fein, and Dr. Kam Ng). The author's exposure to olivocerebellar dynamics owes a great deal to the fruitful interactions with neuroscientists Dr. Thomas McKenna of ONR, Professor Rodolfo Llinas of the Department of Neuroscience of the New York University Medical School, and Professor James Simmons of the Bat Laboratory of the Brown University Neuroscience Department. The author thanks his many co-workers who made the demonstrations described in the article possible over the last two decades.

B.20 Nested Autonomy for Distributed Ocean Sensing

by Henrik Schmidt, Michael Benjamin, Stephanie Petillo, Toby Schneider, Raymond Lum

The development of the fundamentals of the nested autonomy concept, and the MOOS autonomy system were performed at MIT with support from the Office of Naval Research Code 32 under the GOATS program (Thomas B. Curtin, Jeffrey Simmen, Randy Jacobson and Tom Swean, Program Managers) and the IvP extension was performed at MIT and NUWC under support from Dr. Don Wagner and Dr. Behzad Kamgar-Parsi from ONR Code 31. The operational paradigm and the associated MOOS-IvP configuration have subsequently been further developed and applied in several major field experiments under the ONR GOATS (Thomas B. Curtin, Ellen Livingston, and Bob Headrick, Program Managers), SWAMSI (Bob Headrick, Program Manager), and PLUSNet Undersea Persistent Surveillance (Thomas B. Curtin, Program Manager) programs, and most recently the DARPA DSOP Program (Andy Coon, Program Manager). The specific application of the concept to the adaptive tracking of the thermocline was performed under the ONR Adaptive Sampling and Prediction (ASAP) MURI (Thomas B. Curtin, Program Manager) and supported by the U.S. Department of Defense (DoD), U.S. Air Force Office of Scientific Research, and National Defense Science and Engineering Graduate (NDSEG) Fellowship 32 CFR 168a.

The first field demonstration of MOOS and the early version of the nested autonomy control paradigm were performed in the BP02 experiment in La Spezia, Italy, carried out jointly with the NATO Undersea Research Centre (NURC), who supplied engineering and operational support. The GLINT'08-'10 experiments were also carried out in collaboration with NURC, and the authors appreciate the assistance from NURC for their exceptional engineering and operational support, and the use of their OEX AUV for the thermocline mapping and multistatic acoustic experiments described here. The collaboration with Naval Undersea Warfare Center (Michael Incze, Scott Sideleau, and Dr. Don Eickstedt) Code 25 & the ONR Lightweight NSW UUV program in the Lake Champlain'09 experiment is highly appreciated.

B.21 Science of Autonomy: Time-Optimal Path Planning and Adaptive Sampling for Swarms of Ocean Vehicles

by Pierre Lermusiaux, Tapovan Lolla, Patrick Haley Jr., Konuralp Yigit, Mattheus Ueckermann, Thomas Sondergaard, Wayne Leslie

PFJL dedicates the results presented to Dr. Tom Curtin. Tom Curtin's vision and questions in the early 2000s on how to utilize the ocean currents as dynamic highways for autonomous vehicles were early seeds for our theoretical and applied research in marine autonomy. We are also very thankful to all members of the MSEAS group, past and present, for the many useful discussions. PFJL also thanks Drs. Tim Duda and Kevin Heaney, as well as Drs. Jim Bellingham, Russ Davis, Constanti-

nos Evangelinos, Dave Fratantoni, Glen Gawarkiewicz, Naomi Leonard, John Leonard, Sharan Majumdar, Nick Patrikalakis, Steve Ramp, and Henrik Schmidt for their collaborations on autonomy. We are grateful to the Office of Naval Research for support under grants N00014-09-1-0676 (Science of Autonomy, A-MISSION) and N00014-12-1-0944 (ONR6.2) to the Massachusetts Institute of Technology (MIT). We especially thank Dr. Marc Steinberg, as well as Drs. T. Paluszkiewicz and S. Harper, for their support and interactions. KY is very thankful to the Turkish Navy for support. MPU and PFJL thank the Natural Sciences and Engineering Research Council (NSERC) of Canada for the postgraduate scholarship partially supporting the graduate studies and research of MPU at MIT.

C.25 Physical Characteristics of Coastal Hazards

by Jennifer Irish, Robert Weiss, Donald Resio

This effort was supported by Virginia Tech and the University of North Florida. The authors wish to thank Dr. Celso Ferreira and Mr. Antonne Taylor for providing literature review materials on environmental impacts and storm waves, respectively.

C.26 Statistical Characterization of Hazards and Risk in Coastal Areas

by Donald Resio, Mark Tumeo, Jennifer Irish

This effort was supported by the University of North Florida and Virginia Tech.

D.36 Vortex-Induced Vibrations

by Michael Triantafyllou, Rémi Bourguet, Jason Dahl, Yahya Modarres-Sadeghi

The authors wish to acknowledge support by ONR, monitored by Dr. Thomas Swean, Jr., by the BP Major Projects Programme monitored by Dr. Michael Tognarelli and Mr. Pierre Beynet, and by Chevron's University Partnership Program, monitored by Mr. Yiannis Constantinides and Dr. Owen Oakley.

E.46 Ocean Current Energy Conversion

by Howard Hanson, James VanZwieten, Gabriel Alsenas

The Southeast National Marine Renewable Energy Center at Florida Atlantic University is supported by the State of Florida and by the US Department of Energy. The views expressed in this chapter are those of the authors and are not meant to represent the views of these public entities.

E.47 Harvesting Energy by Flow Included Motions

by Michael Bernitsas

The effort for the basic research and development of the VIVACE converter started in 2005 and several Federal, State, and private sponsors supported it. The following support is gratefully acknowledged:

1. DOE contract DE-EE0003644 to Vortex Hydro Energy with subcontract to the University of Michigan
2. DOE contract DE-EE0006780 to Vortex Hydro Energy with subcontract to the University of Michigan.

Besides DOE, present supporters include the University of Michigan, the State of Michigan, Vortex Hydro Energy, SPARK of Ann Arbor, and private investors to VHE. Previous supporters (2005–2014) include DOE (DE-SC-0010091), the National Science Foundation (IIP-0810426, IIP-1026367), the Office of Naval Research (N00014-03-1-0983, N00014-08-1-0601), the US Navy (N62583-09-C-0091), MUCI – Michigan University Commercialization Initiative, TAUW BV Deventer, The Netherlands, DTE Foundation, Wayne County Detroit Port Authority. The support of Mr. David Haynes of the Economic Development Alliance of St. Clair County has been most helpful in collaborating with the local authorities for field deployments.

The author would like to express his sincere thanks to the Editors of this Handbook, Drs. Manhar Dhanak and Nikolas Xiros for their thorough review of the manuscript. Since 2005, more than 100 students have been educated and trained in the MRELab including UROP, BSE, MSE, Ph.D. students, postdoctoral fellows, and international visitors. Their engineering innovative thinking has been a continuous source of challenge in pioneering a multifaceted research effort and developing a unique technology. In chronological order, the following collaborators, whose papers are cited in this chapter, have been major contributors in the research and development of the VIVACE converter: Kamal Raghavan Ph.D., Yaron Ben Simon P.E., Prof. Elisha M.H. Garcia Ph.D., Prof. R. Ajith Kumar Ph.D., Gustavo Simiao, Rebecca Macklem-Alter, Prof. Nikolas Xiros Ph.D., Jonghun Lee Ph.D., Che Chun Chang Ph.D., Wei Wu Ph.D., Hongrae Park Ph.D., Eun Soo Kim Ph.D., Austin Kana, Prof. Hai Sun Ph.D., Lin Ding Ph.D., Jianhui Liu Ph.D., Prof. O. Kemal Kinaci Ph.D. Undergraduate Research Opportunity (UROP) students George M. Hadjicharalambous and Ryan Ouderkirk contributed to the clarity of the document and referencing, respectively.

About the Authors

Raju Abraham

Naval Research Board
Pallickaranai Campus
Chennai, India
abraham@niot.res.in

Chapter E.48

Raju Abraham has worked at the National Institute of Ocean Technology, Chennai, on design, development, and operation of OTEC systems, desalination systems, and deep sea mining systems for 17 years. He worked on the installation of one MW OTEC plant in India, which was tested on a floating barge (1998–2003) and on a low temperature thermal desalination plant utilizing ocean temperature difference (2004–2007).

Gabriel M. Alsenas

Florida Atlantic University
Southeast National Marine Renewable Energy Center
Boca Raton, USA
galsenas@fau.edu

Chapter E.46

A founding member of FAU's SNMREC, Gabriel Alsenas' previous work focused on US Navy future projects and sensor systems. He earned BSc and MSc degrees in Ocean Engineering at FAU. He is Convenor of IEC/ISO's TC-114 AHG04 (tidal performance assessment), Chief US delegate for IECRE's certification and conformity ME-OMC, Treasurer for the USNC/IECRE, and a Judge at FIRST Robotics FRC competitions.

Pak-Cheung Edgar An

Florida Atlantic University
Dep. Ocean and Mechanical Engineering
Boca Raton, USA
pan@fau.edu

Chapter A.10

Dr Edgar An received his BS degree from the University of Mississippi in 1985, and his MS and PhD degrees from the University of New Hampshire in 1988 and 1991. He joined the Department of Ocean Engineering at Florida Atlantic University in 1995. His areas of interest are autonomous underwater vehicles, navigation, control, robotics, modeling and simulation, and neuro-fuzzy systems.

Palaniswamy Ananthakrishnan

Florida Atlantic University
Dep. Ocean and Mechanical Engineering
Boca Raton, USA
panantha@fau.edu

Chapter A.4

Dr P. Ananthakrishnan received his PhD in Naval Architecture and Offshore Engineering from the University of California at Berkeley. He joined Florida Atlantic University where he taught fluid mechanics, water wave mechanics, ship hydrodynamics, physical oceanography and computational methods. He has served as a visiting faculty member at the Indian Institute of Technology Madras, India and at Ecole Centrale de Nantes, France.

Alexia Aubault

Principle Power Inc.
Berkeley, USA
aaubault@principlepowerinc.com

Chapter D.34

Alexia Aubault attended Ecole Nationale Supérieure de Techniques Avancées, a Technology Institute of ParisTech, France. She earned a Master's degree in Ocean Engineering at the University of California, Berkeley. She is a Lead Engineer at Principle Power Inc., a renewable energy firm dedicated to offshore wind technology. A naval architect, Alexia has focused on global design of floating structures, complex hydrodynamic modeling, and hull loading.

Alexander Bahr

Ecole Polytechnique Fédérale de Lausanne
ENAC IIE DISAL
Lausanne, Switzerland
alexander.bahr@epfl.ch

Chapter B.14

Alexander Bahr is a Postdoctoral Associate at Ecole Polytechnique Federale de Lausanne. He holds the degrees of Diploma in Electrical Engineering from the Technical University Aachen and a PhD from the MIT/Woods Hole Joint Program in Applied Ocean Science and Engineering. His research interests include navigation, communications, and cooperation for autonomous underwater vehicles, and wireless sensor networks.

Promode R. Bandyopadhyay

Naval Undersea Warfare Center
Undersea Warfare Weapons, Vehicles, and Defensive Systems Dep.
Newport, USA
promode.bandyopadhya@navy.mil

Chapter B.11

Promode Bandyopadhyay's current research focuses on developing autonomous biorobotic technologies. Areas of recent research have been neuroscience-based control of unmanned underwater vehicles, microbial fuel cells, flapping fin propulsion technology, and analytical theories of wall-turbulence and control. His activities also include the mentoring of new and prospective hires and teaming with national and international performers.

David R. Basco

Old Dominion University
Dep. Civil and Environmental Engineering
Norfolk, USA
dbasco@odu.edu

Chapter C.30

Dr Basco is Professor of Civil Engineering and Director of the Coastal Engineering Centre at Old Dominion University, Norfolk, VA, USA.

Pierre-Philippe J. Beaujean

Florida Atlantic University
Dep. Ocean and Mechanical Engineering
Boca Raton, USA
pbeaujea@fau.edu

Chapter B.15

Pierre-Philippe J. Beaujean is an Associate Professor in the Department of Ocean and Mechanical Engineering, Florida Atlantic University. He specializes in underwater acoustics, with a particular interest in acoustic telemetry and networking, acoustic imagery, sediment acoustics, detection and classification of subsea objects, and machine reliability. Dr Beaujean is an active member of the Acoustical Society of America and a Marine Technology Society member.

James G. Bellingham

Woods Hole Oceanographic Institution
Woods Hole, USA
jbellingham@whoi.edu

Chapter B.16

Dr Bellingham received his PhD in Physics from MIT in 1988. He is a leader in the field of marine autonomous robotics, having created a large number of robotic systems and pioneered their use at sea. After 7 years as the Chief Technologist at MBARI, he joined WHOI as the founding Director of the Center for Marine Robotics.

Michael R. Benjamin

Massachusetts Inst. of Technology
Laboratory for Autonomous Marine Sensing Systems
Cambridge, USA
mikerb@mit.edu

Chapter B.20

Michael Benjamin is a Research Scientist at MIT in the Department of Mechanical Engineering and the Computer Science and Artificial Intelligence Lab. His research focus is on autonomy algorithms for unmanned marine vehicles. Dr Benjamin received BS and MS degrees in Computer Science and Cognitive Science from Rensselaer and MS and PhD degrees in Computer Science from Brown University.

Michael M. Bernitsas

University of Michigan
Dep. Naval Architecture & Marine Engineering
Ann Arbor, USA
michaelb@umich.edu

Chapter E.47

Michael M. Bernitsas, received his PhD in Ocean Engineering from MIT. He is the Mortimer E. Cooley Collegiate Professor of Engineering at the University of Michigan and the Director, of the Marine Renewable Energy Laboratory. He is a fellow of ASME and SNAME 1998. He is the founder and CTO of Vortex Hydro Energy, a company commercializing the VIVACE Converter.

Rémi Bourguet

University of Toulouse, CNRS
Inst. Fluid Mechanics
Toulouse, France
remi.bourguet@imft.fr

Chapter D.36

Rémi Bourguet received his PhD degree in Fluid Dynamics from the Université de Toulouse in 2008. He worked at MIT as a postdoctoral researcher. He has been a CNRS Research Associate at the Institut de Mécanique des Fluides de Toulouse since 2011. His research interests include fluid–structure interactions, turbulent flows, and their modeling and numerical simulation.

Mario P. Brito

University of Southampton
Centre for Risk Research
Southampton, UK
m.p.brito@soton.ac.uk

Chapter B.24

Dr Mario Paulo Brito is a Lecturer in Risk Analysis and Risk Management at the University of Southampton. He has worked as a risk and reliability analyst for several marine autonomous systems deployment projects. His main research interest is in the development of novel risk and reliability methods, particularly in exploring extreme environments.

Robert A. Brizzolara

Office of Naval Research
Sea Platforms and Weapons Division
Arlington, USA
robert.brizzolara@navy.mil

Chapter B.13

Dr Robert A. Brizzolara manages science and technology at the Office of Naval Research supporting small combatant craft. The program includes basic and applied research and technology development. He was previously Group Leader at the NSWC-Carderock Division, received his PhD degree in Physics from the University of Delaware, and performed his postdoctoral work at the Naval Research Laboratory.

Stefano Brizzolara

Massachusetts Institute of Technology
MIT Innovative Ship Design Lab
Cambridge, USA
stebriz@mit.edu

Chapter B.13

Stefano Brizzolara holds an MSc in Naval Architecture and Marine Engineering and a PhD in Numerical Hydrodynamics for Ship Design. He started his academic career in the Department of Naval Architecture of the University of Genova. In 2013 he joined MIT Sea Grant as Assistant Director for Research and MIT iShip, which researches the functional design of innovative high performance marine vehicles, fast ships, and offshore technologies.

Wendell S. Brown

University of Massachusetts – Dartmouth
School for Marine Science and Technology
New Bedford, USA
wbrown@umassd.edu

Chapters A.2, A.5

Wendell Brown received his ScM in Engineering from Brown University in 1967 and his PhD in Oceanography from MIT in 1971. After a post-doc at Scripps Institution of Oceanography, he became a faculty member at the University of New Hampshire. In 2000, he joined the newly-established School for Marine Science and Technology at the University of Massachusetts-Dartmouth. His area of expertise is coastal ocean physical oceanography and observing systems.

Andrew Cairns

AECOM
New York, USA
Andrew.cairns@aecom.com

Chapter C.31

Andrew Cairns is a professional engineer and leads AECOM's port and marine engineering practice in the Northeast United States. He began his port and harbor engineering career as an engineer diver, performing underwater investigations around the world. His experience includes planning, inspection, design, and construction supervision of port and marine projects worldwide, including industrial and military marine terminals, bulkheads, marinas, and shore side civil works.

Thomas J. Campbell

The Shaw Group
Coastal Planning and Engineering Inc.
Boca Raton, USA
sanddiver@icloud.com

Chapter C.29

Thomas J. Campbell is the President of Coastal Planning & Engineering. He has directed coastal engineering analysis, design, geotechnical and physical surveys, and numerical modeling for 75 beach restoration projects throughout the United States. He is a registered professional engineer in seven states, Vice President of the American Shore and Beach Preservation Association (ASBPA), and serves as the Co-Chair of the Science and Technology Committee.

John M. Carel

Estero, USA
johnmcarel@gmail.com

Chapter C.31

John Carel, a professional engineer and Vice President with AECOM ports and marine practice, received a Master's degree from Michigan State and has 40 years of marine and coastal experience, specializing instructural and geotechnical design of maritime structures. He has extensive experience throughout the metropolitan New York area, as well as the US East and Gulf Coasts, the Caribbean, and the Middle East.

Bruce Cathers

University of New South Wales
Water Research Laboratory
Manly Vale, Australia
b.cathers@unsw.edu.au

Chapter C.32

Bruce Cathers has more than 35 years of experience in the fields of coastal, estuarine and river engineering, and physical and numerical modeling. In particular, he has worked on ocean outfalls, water quality and hydrodynamic modeling, sediment transport simulations, breakwaters, and hydraulic structures. His teaching areas include fluid mechanics, hydraulics, air and water quality, and computational hydraulics

Ioannis K. Chatjigeorgiou

National Technical University of Athens
School of Naval Architecture and Marine Engineering
Athens, Greece
chatzi@naval.ntua.gr

Chapter D.38

Professor Chatjigeorgiou received his diploma and PhD from the National Technical University of Athens in 1990 and 1997, respectively. His research interests include linear and nonlinear hydrodynamics, wave resistance, violent slamming, hydroelasticity, and dynamics of slender structures with internal flow for marine applications. He has also been involved in the scientific computing of special functions and their applications to boundary value problems associated with Laplace, Helmholtz, and Stokes domains.

Cortis K. Cooper Chapter A.3 For biographical profile, please see the section "About the Part Editors".

Joseph A. Curcio

Gray, USA
jacurcio@mit.edu

Chapter B.23

Joseph Curcio received his MEng and MS in Ocean Engineering and Ocean Systems Management from MIT. While at the MIT Autonomous Underwater Vehicles Laboratory, he participated in the design, development, and deployment of early Odyssey class autonomous underwater vehicles (AUVs) and led a team in the design and development of CETUS, the world's first production level AUV capable of hovering and station keeping.

Jason Dahl

University of Rhode Island
217 Sheets Laboratory
Narragansett, USA
dahl@egr.uri.edu

Chapter D.36

Jason Dahl received his PhD degree in Ocean Engineering from MIT in 2008. He worked as a postdoctoral researcher at both the SMART Centre in Singapore and at MIT. He has been an Assistant Professor at the University of Rhode Island since 2011. His research interests include flow-induced vibrations, offshore renewable energy systems, and experimental methods for studying fluid dynamics.

Michael S. Dean

Naval Sea Systems Command
Washington D.C., USA
michael.s.dean@navy.mil

Chapter D.42

Michael Dean is an ocean engineer from Florida Institute of Technology. He has participated in and led numerous salvage, deep ocean search and recovery and underwater repair operations. As the Deputy Director of Ocean Engineering, he is the Navy's technical authority for salvage and underwater ship repair where he leads an outstanding team of civilian and military personnel conducting salvage and diving operations worldwide.

Robert G. Dean (deceased) Chapter C.29

Manhar R. Dhanak Chapters 1, A.7, E.44, E.45 For biographical profile, please see the section "About the Editors".

Alana E.S. Duerr

New West Technologies LLC
Englewood, USA
a.e.smentek@gmail.com

Chapter E.44

Alana Duerr received her BS in Naval Architecture and Marine Engineering from the Webb Institute and a PhD in Ocean Engineering from Florida Atlantic University. At Florida Atlantic University she focused on ocean current resource assessment for marine hydrokinetic devices. She currently supports the US Department of Energy's Offshore Wind research, development, and deployment efforts.

Gerald D'Spain

University of California San Diego
Scripps Inst. of Oceanography
San Diego, USA
gdspain@ucsd.edu

Chapter B.12

Gerald D'Spain is the lead Principal Investigator in the ONR-sponsored Flying Wing Autonomous Underwater Glider program, a partnership with the Applied Physics Lab, the University of Washington (APL/UW), with Peter Brodsky as Principal Investigator. The lead engineer is Richard Zimmerman. Other key engineering personnel include Scott Jenkins, Dennis Rimington and Dave Price at Scripps, and Jim Luby and Sean McPeak at APL/UW.

Frank Ehlers

Bundeswehr Technical Center for Ships
and Naval Weapons (WTD 71)
Kiel, Germany
drfrank.ehlers@gmail.com

Chapter B.22

Dr Frank Ehlers obtained his Dr rer nat degree in Theoretical Physics (1998) from the University of Kiel, Germany. Working at the Research Department FWG at the Bundeswehr Technical Center WTD 71, Germany, and at the NURC/CMRE, Italy (2006–2011), he conducts both application-oriented and fundamental research in data fusion, collaborative signal processing, autonomic behaviors, and coordinated distributed sensor systems for maritime surveillance.

Naomi Ehrich Leonard

Princeton University
Mechanical and Aerospace Engineering
Princeton, USA
naomi@princeton.edu

Chapter B.19

Naomi Ehrich Leonard is Edwin S. Wilsey Professor of Mechanical and Aerospace Engineering at Princeton University. She received her BSE in Mechanical Engineering from Princeton and her PhD in Electrical Engineering from the University of Maryland. Leonard's research focuses on control and dynamics in multi-agent systems from cooperative vehicle sensing networks to social groups of animals to human decision-making teams.

Karl Dietrich von Ellenrieder

Florida Atlantic University
SeaTech Inst. for Ocean Systems
Engineering
Dania Beach, USA
ellenrie@fau.edu

Chapter A.7

Karl von Ellenrieder received MS and PhD degrees in Aeronautics and Astronautics from Stanford University, in 1992 and 1998, respectively. He is currently a Professor in the Department of Ocean and Mechanical Engineering at Florida Atlantic University. His research interests include the development of unmanned surface vehicles, marine hydrodynamics, and design and control systems engineering.

R. Cengiz Ertekin Chapters D.34, D.35 For biographical profile, please see the section "About the Part Editors".

Merv Fingas

Spill Science
Edmonton, Canada
fingasmerv@shaw.ca

Chapter D.43

Merv Fingas is a scientist working on oil and chemical spills. Dr Fingas has a PhD in Environmental Physics from McGill University. He has published more than 875 papers and publications in the field. Merv has prepared eight books on spill topics. His specialities are oil spill behavior and fate, remote sensing, and in-situ countermeasures.

George Z. Forristall

Forristall Ocean Engineering, Inc.
Camden, USA
george@forocean.com

Chapter A.3

Dr George Z. Forristall formed Forristall Ocean Engineering, Inc. in 2004 to provide metocean services to offshore industries. He previously spent many years in various Shell companies specifying oceanographic design conditions for offshore structures worldwide. His work has involved fundamental research into kinematics and statistics of storm waves and the structure of oceanic circulation, as well as site specific investigations. He has a PhD in Mechanical Engineering from Rice University.

David Fries

University of South Florida
Inst. for Research in Arts
Tampa, USA
dfries@usf.edu

Chapter B.18

David Fries is Head of the Integrative Creative Technology Group at USF College of Arts and CTO of Spyglass Technologies. He has 35 patents issued, licensed 12 technologies, and over 60 publications in oceanography, analytical chemistry, medical technology, biotechnology, chemistry, microtechnology, electronics, and robotics. He has held a past appointment at Sandia National Laboratories/Lockheed Martin and holds BS/MS degrees in Chemistry.

Gwyn Griffiths

Autonomous Analytics
Southampton, UK
gwyn@autonomousanalytics.com

Chapter B.24

Gwyn Griffiths is an ocean technologist whose career began in the mid 1970s developing instrumentation for measuring ocean currents. In the late 1980s he first became involved with autonomous underwater vehicles through the UK Autosub project, working with scientists to develop applications and research programs. His interests now include underwater gliders and autonomous surface vehicles, and, above all, their reliability.

Patrick J. Haley Jr.

Massachusetts Institute of Technology
Dep. Mechanical Engineering
Cambridge, USA
phaley@mit.edu

Chapter B.21

Patrick Haley received a BS in Physics from Siena College in 1984 and a PhD in Applied Mathematics from Northwestern University in 1991. From 1991 to 2006, he worked in ocean modeling at Harvard University. Since 2007, he has been a Research Scientist at the Massachusetts Institute of Technology. His research interests include multidisciplinary ocean modeling, uncertainty, and path planning.

Andrew Hamilton

Monterey Bay Aquarium Research Institute
Moss Landing, USA
hamilton@mbari.org

Chapter D.40

Andrew Hamilton is an engineer at the Monterey Bay Aquarium Research Institute; his research interests include marine hydrodynamics and ocean wave-energy for oceanographic applications. Dr Hamilton received his PhD in Mechanical Engineering from the University of California, Berkeley, and has been involved in the design of a wide variety of oceanographic buoy and mooring systems.

Howard P. Hanson

Florida Atlantic University
Southeast National Marine Renewable Energy Center
Boca Raton, USA
hphanson@fau.edu

Chapter E.46

Hanson's previous role as Chief Scientist of Florida Atlantic University's Southeast National Marine Renewable Energy Center provided opportunities for his variety of experience at the University of Colorado at Boulder and at the Los Alamos National Laboratory in air–sea interaction and its role in the climate system to converge on a topic of importance to the future of clean energy.

John J. Holmes

Naval Surface Warfare Center
Underwater Electromagnetic Signatures and Technology Division
West Bethesda, USA
john.j.holmes@navy.mil

Chapter A.8

John Holmes received the PhD degree in Electrical Engineering in 1977 from West Virginia University. Since then, he has been with the Naval Surface Warfare Center conducting research to reduce the underwater electromagnetic field signatures of ships and submarines.

Jennifer L. Irish Chapters C.25, C.26 For biographical profile, please see the section "About the Part Editors".

Dong-Sheng Jeng

Griffith University Gold Coast Campus
Griffith School of Engineering
Griffith, Australia
d.jeng@griffith.edu.au

Chapter D.39

Dong Jeng received his PhD degree from the University of Western Australia in 1997, and since then has been an academic at Griffith University, the University of Sydney, and the University of Dundee. His primary research area is offshore geotechnics, with a focus on fluid–soil–structure interactions. Further research interests cover coastal engineering, groundwater hydraulics, porous flow, and offshore wind energy.

Scott A. Jenkins

University of California San Diego
Scripps Inst. of Oceanography
San Diego, USA
sjenkins@ucsd.edu

Chapter B.12

Scott A. Jenkins received a BS from Yale University in 1972 and a PhD in Oceanography from UCSD in 1980. He is author of 80 scientific papers, including the first generalized speed-to-fly glide theory in 3-D wind fields. He holds the three *Soaring Diamonds* from *Fédération Aéronautique Internationale* and was Chair of the ONR Underwater Glider System Study in 2003.

Jason Jonkman

National Renewable Energy Laboratory
National Wind Technology Center
Golden, USA
jason.jonkman@nrel.gov

Chapter E.49

Jason Jonkman is a Senior Engineer at National Renewable Energy Laboratory (NREL), which he joined in 2000, and leads the wind turbine multi-physics engineering tools development activities. He works on a variety of projects focused on developing, verifying and validating, and applying tools for offshore wind turbines. Jason holds a PhD in Aerospace Engineering Sciences from the University of Colorado.

James M. Kaihatu Chapters A.4, C.27 For biographical profile, please see the section "About the Part Editors".

Tae Won Kim

FMC Technologies Schilling Robotics
Davis, USA
twon.kim@gmail.com

Chapter B.17

Dr Kim works for FMC Technologies Schilling Robotics where he develops underwater devices and ROV simulators. He has served as a Co-PI of the Semi-Autonomous Underwater Vehicle for Intervention Missions project at the University of Hawaii, and a senior research engineer at Samsung Aerospace Ind., Korea.

William Kirkwood

Monterey Bay Aquarium Research Institute
Moss Landing, USA
kiwi@mbari.org

Chapter B.18

William Kirkwood is Senior Research and Development Engineer for the Monterey Bay Aquarium Research Institute. As lead engineer, he has managed several remotely operated vehicles, autonomous underwater vehicles and in situ instruments. He is an Adjunct Professor at Santa Clara University and Treasurer for the IEEE Ocean Engineering Society. He has over 65 publications on vehicles/mechatronics, optics, and instrumentation.

John J. Leonard

Massachusetts Institute of Technology
Dep. Mechanical Engineering
Cambridge, USA
jleonard@mit.edu

Chapter B.14

John J. Leonard is the Collins Professor of Mechanical and Ocean Engineering in the MIT Department of Mechanical Engineering. His research addresses the problems of navigation and mapping for autonomous mobile robots. He holds the degrees of BSEE from the University of Pennsylvania and DPhil in Engineering Science from the University of Oxford. He is an IEEE Fellow.

Pierre F.J. Lermusiaux

Massachusetts Institute of Technology
Dep. Mechanical Engineering
Cambridge, USA
pierrel@mit.edu

Chapter B.21

Pierre Lermusiaux has made outstanding contributions in data assimilation, ocean modeling, and uncertainty predictions. His research thrusts include understanding and modeling complex ocean dynamics and processes. With his group, he creates, develops, and utilizes new mathematical models and computational methods for ocean predictions and dynamical diagnostics, for optimization and control of autonomous ocean systems, for uncertainty quantification, and for data assimilation.

Wayne G. Leslie

American Meteorological Society
Boston, USA
wgleslie@mit.edu

Chapter B.21

Wayne Leslie is a Technical Editor at the American Meteorological Society. He has more than 30 years of world-class experience in ocean data analysis, oceanography, ocean modeling, and real-time forecasting. He is an expert at observational program management; data analysis, management and software development; physical and interdisciplinary ocean forecasting and modeling; and web-based dissemination of analysis and forecast products.

Xiao Li

Livingston, USA
Steven.li@aecom.com

Chapter C.31

Steven Li is a professional engineer with over 25 years of coastal engineering experience. The solid theoretical foundation and the wide range of experience in Steven Li's engineering career have enabled him to attain a high level of technical expertise related to analysis and modeling of coastal processes, design of sea defense structures, port design, ship mooring, and maneuvering study.

Tapovan Lolla

Massachusetts Institute of Technology
Dep. Mechanical Engineering
Cambridge, USA
ltapovan@mit.edu

Chapter B.21

Tapovan Lolla is a PhD candidate in the Department of Mechanical Engineering at MIT. He earned his Bachelor's degree from IIT Bombay in 2010 and a Master's degree from MIT in 2012. His research interests include robotic path planning, control systems, uncertainty quantification, and high-dimensional inference.

Raymond Lum

DSO National Laboratories
Singapore, Singapore
raylum@alum.mit.edu

Chapter B.20

Raymond Lum received the BEng (First Class Honors) and MSc degrees from Nanyang Technological University, and the PhD degree from Massachusetts Institute of Technology (MIT). During his time at MIT, he worked in the Laboratory for Autonomous Marine Sensing Systems. He is presently a Program Manager with DSO National Laboratories, where he leads teams of engineers to develop sonar surveillance systems for the Republic of Singapore Navy.

Patrick J. Lynett

University of Southern California
Sonny Astani Dep. of Civil and Environmental Engineering
Los Angeles, USA
plynett@usc.edu

Chapter C.27

Patrick J. Lynett is a Professor of Civil Engineering at the University of Southern California. His research interests are directed towards a better understanding of coastal processes, such as nearshore circulations, wave evolution from generation to the shoreline, multi-scale hydrodynamic interactions, and sediment transport. Amongst other awards, Dr Lynett received the Department of the Army Commander's Award for Public Service for his post-Katrina work.

Giacomo Marani

West Virginia University
WV Robotic Technology Center
Fairmont, USA
gmarani@gmail.com

Chapter B.17

Dr Marani serves as lead research scientist at the WV Robotic Technology Center, affiliated with the NASA Space Servicing Capabilities Office. In the past decade he served as researcher at the University of Hawaii at Manoa. He is Chair of the IEEE Marine Robotics Technical Committee and a member of the Editorial Board of the journal of Intelligent Service Robotics.

Spyros A. Mavrakos

National Technical University of Athens
School of Naval Architecture and Marine Engineering
Athens, Greece
mavrakos@naval.ntua.gr

Chapter D.38

Spyros Mavrakos is Vice Rector of NTUA, Director of the Laboratory for Floating Structures and Mooring Systems, and Professor of Offshore Structures. He has degrees in Naval Architecture and Marine Engineering, and in Offshore Hydrodynamics. His research activities include offshore hydrodynamics, mooring systems' analysis and design, hydrodynamics of wave energy converters, design of floating wind turbines, and multi-purpose floating structures for offshore wind and wave energy sources exploitation.

Robert E. Melchers

The University of Newcastle
Centre for Infrastructure Performance and Reliability
Callaghan, Australia
Rob.Melchers@newcastle.edu.au

Chapter A.6

Rob Melchers is Professor of Civil Engineering and ARC DORA Research Fellow at the University of Newcastle, Australia. He holds a BE and MEngSc from Monash University, and a PhD from the University of Cambridge, UK. He is a Fellow of the Australian Academy of Technological Sciences and Engineering. He specializes in corrosion modeling and structural reliability analysis.

Yahya Modarres-Sadeghi

University of Massachusetts
Mechanical and Industrial Engineering
Amherst, USA
modarres@engin.umass.edu

Chapter D.36

Yahya Modarres-Sadeghi received his PhD degree in Mechanical Engineering from McGill University in 2006. He worked as a postdoctoral researcher at MIT for 3 years. He has been an Assistant Professor at the University of Massachusetts, Amherst since 2009. His research interests include fluid–structure interactions, nonlinear dynamics, offshore wind energy, and biomimetics.

Stephanie M. Petillo

Woods Hole Oceanographic Institution
Woods Hole, USA
spetillo@whoi.edu

Chapter B.20

Stephanie Petillo is an Oceanographic Engineering PhD candidate in the MIT/WHOI Joint Graduate Program with a BS in Aerospace Engineering from the University of Maryland – College Park (2008). Her research focuses on the use of autonomous underwater vehicles for autonomous and environmentally adaptive sampling of the ocean environment, to improve efficiency and synopticity of data collection.

Roshanka Ranasinghe

UNESCO-IHE Inst. for Water Education
Delft, The Netherlands
r.ranasinghe@unesco-ihe.org

Chapter C.28

Dr Roshanka Ranasinghe is Associate Professor of Coastal Engineering at UNESCO-IHE and at Delft University of Technology, The Netherlands. He is also Visiting Professor at the Australian National University, Canberra, and strategy advisor on morphodynamic modeling at Deltares, The Netherlands. He is regularly invited to provide expert advice on coastal management and climate change impacts by national governments and international agencies.

Muthukamatchi Ravindran

Naval Research Board
Pallickaranai Campus
Chennai, India
ravinmv@gmail.com

Chapter E.48

Muthukamatchi Ravindran graduated in Mechanical Engineering from IIT Madras and served at IIT as Head of Ocean Engineering. He was the Founding Director of the National Institute of Ocean Technology, and initiated major national programs on ocean energy, the Deep Sea Mining and Data Buoy Programs. He has served on the International Oceanographic Commission, ISBA, and the Naval Research Board.

Dronnadula V. Reddy

Florida Atlantic University
Dep. Civil, Environmental and Geomatics Engineering
Boca Raton, USA
dreddy@fau.edu

Chapter D.33

Dronnadula V. Reddy is a Professor of Civil, Environmental, and Geomatics Engineering at Florida Atlantic University. His numerous publications evidence his wide research interests, which include computational, experimental, and structural geomechanics, offshore and coastal structures, concrete technology with the focus on supplementary cementitious materials; and corrosion and fire resistant structural concrete.

Donald T. Resio

University of North Florida
Dep. Civil Engineering
Jacksonville, USA
don.resio@unf.edu

Chapters C.25, C.26

Dr Donald T. Resio is Professor of Ocean Engineering at the University of North Florida and has 40 years of experience in wind, wave and surge prediction, and hazard/risk estimation. He has served as senior scientist for coastal work at the Army Engineer Research and Development Center and has led their post-Katrina effort to improve methods for assessing hazards/risks along US coastlines.

H. Ronald Riggs

University of Hawaii at Manoa
Dep. Civil & Environmental Engineering
Honolulu, USA
riggs@hawaii.edu

Chapter D.37

H.R. Riggs received his PhD degree from the University of California, Berkeley, in Structural Engineering in 1981. After 6 years in industry, he joined the University of Hawaii in 1987. Over the years he has worked extensively on hydroelasticity and very large floating structures. He is currently focusing on structural dynamic response to fluid and other loading.

George Rodenbusch

Rodenbusch Consulting
Dripping Springs, USA
george@rodenbusch.net

Chapter D.35

Dr Rodenbusch has more than 35 years of experience with the engineering and management of deepwater offshore oil and gas developments around the world. He holds a BS and an MME in Mechanical Engineering from Rice University and a PhD in Oceanographic Engineering from Massachusetts Institute of Technology & Woods Hole Oceanographic Institution. He is a Licensed Professional Engineer in Texas.

J.A. Dano Roelvink

UNESCO-IHE Inst. for Water Education
Delft, The Netherlands
d.roelvink@unesco-ihe.org

Chapter C.28

Dano Roelvink has more than 25 years of experience in coastal engineering and research. He has managed and contributed much to the development of the Delft3D system and has been actively involved in many EU-sponsored research projects on coastal morphodynamics. He has set up collaborative projects with the USGS, ONR, and many international institutes and leads the development of XBeach, an open-source model for storm impacts.

Salvatore Scaturro

Expert Water Modelling
Sydney, Australia
sal.scaturro@ewmworld.com

Chapter C.32

Salvatore Scaturro is a water engineer specializing in the analysis and design of pipelines, pumping stations, distribution networks, hydraulic structures, and marine intake and outfall systems. He is an experienced user of water hammer and computational fluid dynamics (CFD) modeling software and has also programmed his own models for a range of hydraulics applications.

Henrik Schmidt

Massachusetts Institute of Technology
Center for Ocean Engineering
Cambridge, USA
henrik@mit.edu

Chapter B.20

Henrik Schmidt is a Professor of Mechanical Ocean Engineering at MIT. His research focuses on underwater acoustic propagation and signal processing. In recent years Professor Schmidt has been pioneering the development of new underwater acoustic sensing concepts for networks of small autonomous underwater vehicles. He is a Fellow of the Acoustical Society of America, and the 2005 recipient of the ASA *Pioneer of Underwater Acoustics* medal.

Robert Seah

Chevron Energy Technology Company
Facilities Engineering Dep.
Houston, USA
RSeah@Chevron.com

Chapter D.41

Dr Seah has been involved in hydrodynamic research since 1998, initially at the National University of Singapore and the University of California at Berkeley, finally joining the Chevron Energy Technology Company in 2007. His past and present research include vortex breakdowns, drag reduction, ship roll damping, sloshing of LNG, and multi-floater response in waves. He is currently leading research into integrity management of FPSOs.

Thomas Sondergaard

TrackMan A/S
Vedbæk, Denmark
tsondergaard@gmail.com

Chapter B.21

Thomas Sondergaard is currently working as a Research and Development Engineer at TrackMan A/S. He has a Master of Science degree in Mechanical Engineering from the Massachusetts Institute of Technology, USA. Prior to that, he earned a Master of Engineering degree from Imperial College, London. His research interests lie in the fields of acoustics, computational data analysis, statistics, uncertainty quantification, nonlinear data assimilation, estimation theory, and signal processing.

Matthias Stammler

Fraunhofer Institute for Wind Energy and Energy System Technology IWES
Hannover, Germany
matthias.stammler@iwes.fraunhofer.de

Chapter E.49

Matthias Stammler studied Automotive Engineering in Germany and Mexico. He started his professional career as a drivetrain test engineer with Porsche. In 2011 he joined PowerWind as gearbox engineer for wind turbines. In 2013 he joined the Fraunhofer IWES as a Research Fellow. He further specializes in bearing design for wind turbine applications, focusing on pitch systems.

Milica Stojanovic

Northeastern University
ECE Department
Boston, USA
millitsa@ece.neu.edu

Chapter B.15

Milica Stojanovic graduated from the University of Belgrade, Serbia, in 1988, and received the MS (1991) and PhD (1993) degrees from Northeastern University in Boston, MA, where she is currently Professor of Electrical and Computer Engineering. She is also a Guest Investigator at the Woods Hole Oceanographic Institution. Milica is an Associate Editor and a Fellow of the IEEE.

Mareike Strach-Sonsalla

Fraunhofer Institute for Wind Energy and Energy System Technology IWES
Bremerhaven, Germany
mareike.strach-sonsalla@outlook.com

Chapter E.49

Mareike Strach-Sonsalla is Group Manager for Loads Analysis and System Dynamics at Fraunhofer IWES. She specializes in offshore wind energy turbine dynamics and modeling, particularly floating offshore wind turbines. She joined Fraunhofer IWES as a research fellow in 2011, after studying naval architecture and offshore engineering in Berlin and Trondheim.

Arisi S.J. Swamidas

Memorial University of Newfoundland
Faculty of Engineering and Applied Science
St. John's, Canada
aswamidas@mun.ca

Chapter D.33

Arisi S.J. Swamidas is a retired Professor of Civil Engineering at the Faculty of Engineering and Applied Sciences, Memorial University, St. John's, NL, Canada. Until 2014 he was active in supervising graduate students in computational/experimental analysis and testing of offshore and onshore structures/elements subjected to wind, wave, ice, and accidental impact forces. His publications include 2 books and more than 215 articles.

Peter M. Tate

Sydney Water Corporation
Parramatta, Australia
peter.tate@sydneywater.com.au

Chapter C.32

Peter Tate has more than 30 years' experience studying physical processes in the marine environment, specializing in the interaction of wastewater plumes and internal waves. He has worked in research, consulting, and government organizations, providing him with a broad perspective on marine outfalls. His interests lie in the design and execution of interdisciplinary monitoring programs to assess the impacts of wastewater discharges on the marine environment.

Krish P. Thiagarajan

University of Maine
Dep. Mechanical Engineering
Orono, USA
Krish.Thiagarajan@maine.edu

Chapter D.41

Krish Thiagarajan is the Correll Presidential Chair in Energy and Professor of Mechanical Engineering at the University of Maine. His research spans design and analysis of energy generating systems located offshore in the deep oceans. In particular, his area of research is dynamics and global performance of floating offshore systems for oil and gas, as well as renewable energy production.

Michael S. Triantafyllou

Massachusetts Institute of Technology
Dep. Mechanical Engineering
Cambridge, USA
mistetri@mit.edu

Chapter D.36

Faculty member in the Ocean Engineering Department since 1979 and in the Mechanical Engineering Department since 2005 at MIT, currently William I. Koch is Professor of Marine Technology and Director of the Center for Ocean Engineering. He received his degree from NTU Athens (1974) and SM (1997) and ScD (1979) degrees from MIT. His research interests include biomimetic robots, flow–structure interaction, and dynamics and control of marine vehicles.

Mark A. Tumeo

University of North Florida
College of Computing, Engineering & Construction
Jacksonville, USA
m.tumeo@unf.edu

Chapter C.26

Dr Mark A. Tumeo is Dean of the College of Computing, Engineering, and Construction, and Professor of Civil Engineering at the University of North Florida. He has over 25 years of experience in environmental consulting and research, environmental risk assessment, policy, and regulation. He is actively involved in several professional organizations and currently serves as a Corresponding Editor for the ASCE Journal of Professional Issues in Engineering Education and Practice.

Mattheus P. Ueckermann

Creare
Hanover, USA
mpu@creare.com

Chapter B.21

Dr Ueckermann received his Bachelor's degree from the University of Waterloo, and his MS and PhD from Massachusetts Institute of Technology in Mechanical and Computational Engineering. For his research, he developed computational schemes for regional ocean models. Presently he is working as an engineer at Creare, where he develops mathematical models and algorithms for machine vision, weather prediction, and thermodynamic applications.

James H. VanZwieten

Florida Atlantic University
Southeast National Marine Renewable Energy Center
Boca Raton, USA
jvanzwi@fau.edu

Chapters E.44, E.46

James H. VanZwieten is an Assistant Research Professor at Florida Atlantic University's Southeast National Marine Renewable Energy Center. His research areas include numerical simulation, resource assessment, and controller development for applications related to the extraction of energy from ocean currents.

Fabian Vorpahl

Leading Expert Offshore Engineering
Tower & Substructure Senvion GmbH
Osnabrück, Germany
fabian.vorpahl@senvion.com

Chapter E.49

Fabian Vorpahl has a degree in Mechanical Engineering from Karlsruhe Institute of Technology and has worked as a project engineer, project manager, group manager, and Head of Department at Fraunhofer IWES. His work is focused on research and development of global dynamics, structural dynamics, coupled simulation, modeling, and loads of (offshore) wind turbines.

Dirk-Jan R. Walstra

Deltares
Unit Hydraulic Engineering
Delft, The Netherlands
Dirkjan.walstra@deltares.nl

Chapter C.28

Dirk-Jan Walstra is a coastal morphologist and is involved in numerous Dutch and international projects in the coastal zone. Due to his specialist coastal modeling expertise he plays a key role in the development of morphological models such as Delft3D and Unibest. He is also a Research Associate at Delft University of Technology and an invited Lecturer at Utrecht University and Twente University.

Mick van der Wegen

UNESCO-IHE Inst. for Water Education
Delft, The Netherlands
m.vanderwegen@unesco-ihe.org

Chapter C.28

Since 1996 Mick van der Wegen's position at UNESCO-IHE has been related to capacity building, education and research in the field of coastal engineering, and coastal zone management in an international context. Specific fields of interest include estuarine morphodynamics on decadal to millennial timescales, salt intrusion and density currents, and process-based modeling (Delft3D).

Robert Weiss

Virginia Tech
Dep. Geosciences
Blacksburg, USA
weiszr@vt.edu

Chapter C.25

Dr Robert Weiss is an Assistant Professor of Geosciences at Virginia Tech with more than 12 years of experience and 30 peer-reviewed papers. Dr Weiss is a tsunami expert with expertise in numerical simulations of tsunami evolution and dynamics, sediment transport, and field reconnaissance surveys.

Jan Wenske

Fraunhofer Institute for Wind Energy and Energy System Technology IWES
Bremerhaven, Germany
jan.wenske@iwes.fraunhofer.de

Chapter E.49

Jan Wenske studied engineering at the Technical University of Clausthal. He worked as a project manager with STILL GmbH and as Head of Power Electronics Development at ESW Jenoptik AG. He joined Fraunhofer-IWES in 2011 as the Head of the Wind Turbine and System Technology Department, specializing in power mechatronics, accelerated testing, and advanced drive train control. He has been a Professor at the University of Bremen since 2013.

Nikolaos I. Xiros Chapters 1, A.9, A.10, E.45 For biographical profile, please see the section "About the Editors".

Konuralp Yigit

University of Illinois at Urbana Champaign
Dep. Aerospace Engineering
Urbana, USA
yigit2@illinois.edu

Chapter B.21

Konuralp Yigit obtained his Bachelor of Science degree from the Turkish Naval Academy in 2005 and his Master of Science degree from Massachusetts Institute of Technology (MIT) in 2011. He is currently a PhD student at the University of Illinois at Urbana-Champaign. His current research is centered on parallel computing and algorithms.

Solomon Yim

Oregon State University
School of Civil and Construction Engineering
Corvallis, USA
solomon.yim@oregonstate.edu

Chapter D.37

Solomon Yim received his PhD degree from the University of California, Berkeley. He has been teaching at Oregon State University since 1987 and is the Glenn Willis Holcomb Professor of Structural Engineering. Dr Yim's work focuses on analytical, numerical, and experimental modeling, and simulation of coupled fluid–structure interaction systems, high-performance computing, with applications in civil, naval, offshore, and renewable-energy engineering.

Junku Yuh

Korea Institute of Science and Technology
National Agenda Research Division
Seoul, Korea
yuh.junku@gmail.com

Chapter B.17

Dr Yuh, IEEE Fellow, is currently Endowed Chair Researcher at Korea Institute of Science and Technology (KIST). He served as the 5th and 6th President of Korea Aerospace University, US National Science Foundation Program Director, Head of NSF Tokyo Regional Office, Professor of Mechanical Engineering and Information and Computer Science at the University of Hawaii, and Editor-in-Chief for the International Journal of Intelligent Service Robotics.

Detailed Contents

List of Abbreviations XXIX

1 **Introduction**
Manhar R. Dhanak, Nikolaos I. Xiros 1
1.1 Enabling Maritime Design and Development 1
1.2 History 2
1.3 Basics 2
1.4 Applications 4
1.4.1 Automated Unmanned Systems 4
1.4.2 Coastal Design 5
1.4.3 Offshore Systems 6
1.4.4 Ocean Energy 6
1.5 Future Trends 7
References 8

Part A Fundamentals

2 **Elements of Physical Oceanography**
Wendell S. Brown 15
2.1 Atmospheric Processes 16
2.2 Ocean Structure 17
2.3 Oceanic Processes 18
2.4 Surface Gravity Waves 19
2.4.1 Surface Gravity Waves 19
2.4.2 Definitions 20
2.4.3 Wave Generation and Evolution 25
2.5 Wind-Forced Ocean Processes 28
2.5.1 Frictional Effects 28
2.5.2 Earth Rotation Effects 29
2.5.3 Hurricane Wind-Forced Ocean Response 29
2.5.4 Wind-Forced Upper Ocean Response with Earth Rotation Effects 33
2.5.5 Wind-Driven Currents: Ocean Basin Scale 34
2.5.6 Gulf Stream Ring Flows 37
2.6 Deep Ocean Currents 38
2.7 Coastal Ocean Currents 39
2.8 Ocean Surface Tides 40
2.9 Oceanic Internal Waves and Tides 43
References 45

3 **Metocean Extreme and Operating Conditions**
George Z. Forristall, Cortis K. Cooper 47
3.1 Quantifying the Metocean Environment 48
3.2 Overview of WWC Processes 49
3.2.1 Winds 49

3.2.2 Waves 51
3.2.3 Currents 53
3.3 Measurements 55
3.3.1 Historical Storm Databases 55
3.3.2 Satellite Databases 56
3.3.3 In Situ Measurements 57
3.3.4 Mobile Measurements 58
3.4 Modeling 58
3.4.1 Winds 59
3.4.2 Waves 59
3.4.3 Currents, Surge, and Tides 60
3.5 Joint Events 61
3.5.1 Response-Based Analysis 61
3.5.2 Load Cases 62
3.5.3 Environmental Contours 63
3.5.4 Inverse FORM 63
3.6 Operational Criteria 64
3.6.1 Probability Distributions 64
3.6.2 Persistence 65
3.6.3 Currents 65
3.7 Extreme Criteria 66
3.7.1 Risk and Reliability 66
3.7.2 The Historical Method 67
3.7.3 Synthetic Storm Modeling 68
3.7.4 Modeling Versus Measurements 68
3.7.5 Accounting for Physical Limits 69
3.7.6 Seasonality 69
3.7.7 Directionality 69
3.7.8 Combining Long and Short-Term Distributions 69
3.7.9 Rogue Waves 70
3.7.10 Extremely Rare Events 71
3.7.11 Quantifying Uncertainty 71
3.7.12 Stationarity 71
3.8 Conclusions 72
References 72

4 **Mechanics of Ocean Waves**
James M. Kaihatu, Palaniswamy Ananthakrishnan 77
4.1 Ocean Surface Waves 77
4.2 Wave Theories 78
4.2.1 Potential Flow Formulation 78
4.3 Properties of Small Amplitude Gravity Waves 80
4.3.1 Linear Dispersion Relation 80
4.3.2 Phase Speed 80
4.3.3 Group Speed 80
4.3.4 Amplitude Modulation of Water Waves 81
4.3.5 Average Wave Energy Density 81
4.3.6 Propagation of Wave Energy 81
4.3.7 Water Particle Trajectory 82

4.3.8 Spatio-Temporal Evolution of Waves 82
4.3.9 Shoaling and Refraction of Waves 83
4.3.10 Closing Remarks to the Section 83
4.4 Weakly Nonlinear Deep Water Wave Theories 83
4.4.1 Properties of Weakly Nonlinear Deep Water Waves 84
4.4.2 Evolution of Weakly Nonlinear Deep Water Waves 85
4.5 Shallow Water Wave Theories 87
4.5.1 Properties of Weakly Nonlinear Shallow Water Waves 88
4.5.2 Evolution of Weakly Nonlinear Shallow Water Waves 89
4.6 Transformation of Waves Approaching Land 90
4.7 Computational Method for Fully Nonlinear Waves 93
4.8 Wave Forces on Fixed and Floating Structures 94
4.8.1 Incident Wave Force: Froude–Krylov Force 94
4.8.2 Morison Force on a Stationary Body 95
4.8.3 Wave Diffraction over a Body 96
4.8.4 Wave Radiation Force on an Oscillating Body 96
4.9 Concluding Remarks 97
References 98

5 **Physical Properties of Seawater**
Wendell S. Brown 101
5.1 Hydrostatic Pressure 101
5.2 Temperature 101
5.3 Salinity 103
5.4 Density 104
5.5 Temperature–Salinity Relationships 105
5.6 Specific Heat 106
5.7 Freezing of Sea Water and Sea Ice 106
5.8 Coefficient of Thermal Expansion 106
5.9 Sound Velocity 107
5.10 Acoustic Ambient Noise 107
5.11 Light Transmission 108
References 109

6 **Principles of Marine Corrosion**
Robert E. Melchers 111
6.1 Chemical and Physical Composition of Seawater 111
6.2 Materials Used in Marine Environments 113
6.3 Marine Corrosion of Steel 113
6.4 Modeling Longer Term Corrosion of Steel 116
6.5 Other Influences on Steel Corrosion 118
6.6 Pitting Corrosion of Steel 119
6.7 Some Other Important Materials 121
6.7.1 Stainless Steel 121
6.7.2 Aluminum 121
6.7.3 Copper-Nickels 121
6.7.4 Reinforced Concrete 121
6.8 Conclusion 123
References 123

7 **Hydromechanics**
Karl Dietrich von Ellenrieder, Manhar R. Dhanak 127
7.1 Dimensional Analysis, Basic Estimation, and Model Testing 127
7.1.1 Drag Force on a Sphere 127
7.1.2 Physical Significance of the Dimensionless Π Groups 129
7.1.3 Similitude 134
7.1.4 Skin-Friction Drag 135
7.1.5 Estimating the Drag on a Ship from Model Testing 136
7.1.6 Hydrofoil Lift and Drag 138
7.1.7 Screw Propellers 143
7.1.8 Air and Wind Resistance 146
7.1.9 Hydrodynamic Characterization of Marine Surface Vessels 146
7.2 Fluid Statics 153
7.2.1 Pressure Forces on Surfaces 153
7.2.2 Static Stability 154
7.3 Hydrodynamics 155
7.3.1 Flow Kinematics 155
7.3.2 The Navier–Stokes Equations 156
7.3.3 Flow of an Ideal Fluid 159
7.3.4 Flow of a Viscous Fluid 169
References 174

8 **Ocean Electromagnetics**
John J. Holmes 177
8.1 Electromagnetism in an Ocean Environment 177
8.2 Electromagnetic Field Theory 178
8.3 Plane Wave Propagation 180
8.4 Reflection and Transmission of a Plane Wave at the Surface of Fresh Water 182
8.5 Plane Wave Incident on Seawater 184
8.6 Magnetic and Electric Dipoles in an Unbounded Ocean 186
8.7 Magnetic and Electric Dipoles in a Bounded Ocean 188
8.8 Electromagnetic Propagation in the Ocean at Optical Wavelengths 193
References 195

9 **Digital Signal Processing**
Nikolaos I. Xiros 197
9.1 Discrete-Time Systems 197
9.1.1 Discrete-Time Signals and Digital Systems 197
9.1.2 Signal Sampling 198
9.1.3 Analog Signal Reconstruction Using a Discrete-Time Signal 201
9.1.4 The Z-Transform 201
9.1.5 Discrete-Time LTI Systems 202
9.1.6 Continuous-Time System Mapping 203
9.2 Digital Filters 206
9.2.1 Important FIR Filter Structures 206
9.2.2 Important IIR Filter Structures 208
9.3 The Fast Fourier Transform (FFT) 211

9.3.1 Review of Integral Transforms 211
9.3.2 The Discrete Fourier Transform (DFT) 212
9.4 Waveform Analysis 216
9.4.1 Definitions for Waveforms and Random Signals 216
9.4.2 Signal Power and Power Spectral Density 218
9.4.3 Waveform Propagation Through a Linear, Time-Invariant System 219
9.5 Optimal Signal Estimation 220
9.5.1 System Identification 220
9.5.2 Discrete-Time Wiener–Hopf Equation over a Finite-Duration Window 221
9.5.3 Signal Estimation and the Wiener Filter 222
9.6 Concluding Remarks 225
References 225

10 **Control Theory and Applications**
Nikolaos I. Xiros, Pak-Cheung Edgar An 227
10.1 System Theory 227
10.1.1 Definitions and Fundamentals 227
10.1.2 The Laplace Transform 229
10.1.3 Linear Time-Invariant Systems 230
10.1.4 Multivariable Systems and State Space 233
10.1.5 Nonlinear Systems and Linearization 235
10.2 Analysis of LTI Systems 237
10.2.1 Block Diagrams 237
10.2.2 Stability 238
10.2.3 Controllability and Observability 240
10.2.4 Sinusoidal Steady-State Response and Bode Plots 241
10.2.5 Analysis of Second-Order Systems 244
10.3 SISO System Controls 247
10.3.1 Performance Criteria 247
10.3.2 ON/OFF Control 248
10.3.3 PID Control 249
10.3.4 Ziegler–Nichols' Methods for PID Controller Tuning 254
10.3.5 Digital Controller Implementation 256
10.3.6 The Root Locus Technique 259
10.4 Pole Placement of LTI Systems 261
10.4.1 Input–Output Decoupling 261
10.4.2 Full-State Feedback 264
10.4.3 Pole Placement Design 264
10.5 Course-Keeping Autopilots 267
10.5.1 The Vessel in the Control Loop 267
10.5.2 Surface Vessel State-Space Model 269
10.5.3 PID Autopilots 271
10.5.4 State Observers and Use in Autopilots 274
References 275

Part B Autonomous Ocean Vehicles, Subsystems and Control

11 **Highly Maneuverable Biorobotic Underwater Vehicles**
Promode R. Bandyopadhyay 281
11.1 Biorobotics 281
11.1.1 Flapping Fin Propulsion Technology 284
11.2 Theoretical Foundation of Animal-Inspired Hydrodynamics and Control 286
11.2.1 Hydrodynamics 286
11.2.2 Animal-Like Motion Control Laws and the Principles of Integrated Design 287
11.3 Description of Biology-Inspired Vehicles of Emergent Maturity 289
11.4 Reliability, Low Power Consumption, and Disturbance Rejection of Bio-Inspired Propulsion 296
11.5 Demonstrated Maneuverings of NUWC Bio-Inspired Vehicles 296
11.5.1 BAUV 296
11.5.2 SPLINE 297
11.5.3 RAZOR 297
11.6 Discussion 297
11.7 Concluding Remarks 298
11.8 Nomenclature 298
References 299

12 **Autonomous Underwater Gliders**
Scott A. Jenkins, Gerald D'Spain 301
12.1 Concept 301
12.2 Hydrodynamics of Wings Versus Propellers 305
12.3 Underwater Glider Attributes and Limitations 306
12.3.1 Depth Unlimited Roaming 307
12.3.2 Depth Limited Roaming 307
12.3.3 2-D Station Keeping 307
12.3.4 Payload/Cargo Delivery 307
12.3.5 Level Flight Hybrids 307
12.4 Optimal Size and Shape for Horizontal Transport Efficiency 308
12.4.1 Net Transport Economy 308
12.4.2 Size Factors 310
12.4.3 Shape Factors 311
12.4.4 Glide Polar 313
12.5 Thermal Glider 318
12.6 Discussion and Conclusions 319
References 320

13 **Autonomous Sea Surface Vehicles**
Stefano Brizzolara, Robert A. Brizzolara 323
13.1 Platforms 324
13.2 Autonomous Maneuvering and Navigation 326
13.3 Naval Architecture of AUSV Design 327
13.3.1 Froude Number and Hull Typologies 327
13.4 Optimized Class of Autonomous Unmanned Surface Vehicles 330
13.5 Conclusions 337
References 339

14 **Autonomous Underwater Vehicle Navigation**
John J. Leonard, Alexander Bahr 341
14.1 Sensors 343
14.1.1 Depth 343
14.1.2 Compass 343
14.1.3 Gyroscopes 343
14.1.4 Attitude Heading Reference Systems 344
14.1.5 Inertial Navigation Systems 344
14.1.6 GPS 344
14.1.7 Doppler Velocity Log (DVL) 344
14.1.8 Acoustic Ranging Methods 345
14.2 Algorithms 346
14.2.1 Dead-Reckoning and Inertial Navigation 346
14.2.2 Acoustic Navigation 347
14.2.3 Geophysical Map-Based Navigation 349
14.2.4 Simultaneous Localization and Mapping 350
14.2.5 Cooperative Navigation of Multiple Vehicles 351
14.3 Summary 352
14.3.1 Glider with Very Low Power Sensor Suite 352
14.3.2 Low-Cost AUV Sensor Suite 352
14.3.3 Standard AUV Sensor Suite 353
14.3.4 High-End AUV 353
14.3.5 Special-Task AUV Using Visual SLAM 353
14.4 Conclusion 353
References 354

15 **Acoustic Communication**
Milica Stojanovic, Pierre-Philippe J. Beaujean 359
15.1 A Brief History 360
15.2 Current and Emerging Modem Applications 360
15.3 Existing Technology 361
15.3.1 System Requirements 361
15.3.2 Commercially Available Modems 362
15.3.3 Field Tests 363
15.4 Propagation Channel 364
15.4.1 Attenuation and Noise 364
15.4.2 Multipath Propagation 365
15.4.3 Time Variability: Motion-Induced Doppler Distortion 369
15.4.4 Time Variability: Random Effects (Fading) 370
15.4.5 System Constraints 372
15.5 Point-to-Point Links: Signal Processing 374
15.5.1 Noncoherent Modulation/Detection 374
15.5.2 Coherent Modulation/Detection 375
15.5.3 Data Link Reliability 378
15.5.4 Turbo Equalization 378
15.5.5 Adapting to the Environment 379
15.5.6 Networks 379
15.5.7 Channel Sharing 381
15.5.8 Routing and Cross-Layer Integration 382

15.6 Future Trends 383
References 383

16 **Autonomous Underwater Vehicle Docking**
James G. Bellingham 387
16.1 Technical Elements of Docking 388
16.2 AUV Characteristics 390
16.2.1 Vehicle Configuration and Controllability 390
16.2.2 Vehicle Size 390
16.3 Sensors For Homing 391
16.3.1 Acoustic Homing 392
16.3.2 Optical Homing 393
16.3.3 Electromagnetic Homing Systems 394
16.3.4 Constraining Vertical Position 395
16.4 Capture and Connection Mechanisms 395
16.4.1 Directional Docks: Funnels 396
16.4.2 Omnidirectional Docks: Poles 397
16.4.3 Soft Docking Approaches 398
16.4.4 Other Docking Mechanisms 398
16.5 Coupling Power and Communications 399
16.5.1 Establishing a Communications Link 399
16.5.2 Power Transfer 400
16.6 AUV Control Considerations 400
16.6.1 Navigation, Currents, and Arriving at the Dock 400
16.6.2 Managing State: AUV Control 401
16.6.3 Managing State: Coordinating AUV and Dock 401
16.6.4 Detecting and Recovering from Failures 403
16.7 Conclusions and Future Prospects 403
References 404

17 **Underwater Vehicle Manipulators**
Tae Won Kim, Giacomo Marani, Junku Yuh 407
17.1 Underwater Vehicles for Intervention Missions 407
17.2 Dynamics of Underwater Vehicle Manipulators 409
17.3 Teleoperation of Underwater Vehicle Manipulators 410
17.4 Sensor-Based Manipulator Control 411
17.5 Coordinated Motion Control of Underwater Vehicle-Manipulator Systems 413
17.6 Underwater Autonomous Manipulation 414
17.6.1 Phase 1: Undock from the Pier and Navigate to a Search Area 417
17.6.2 Phase 2: Search for the Submerged Platform 417
17.6.3 Phase 3: Navigate and Dive Toward the Platform 417
17.6.4 Phase 4: Hover (Station Keeping) 417
17.6.5 Phase 5: Hook a Recovery Tool to the Target Object (Autonomous Manipulation) 418
17.6.6 Phase 6: Return to the Pier 418
17.7 Conclusions 419
References 419

18 **Non-Acoustic Sensors**
David Fries, William Kirkwood 423
18.1 Non-Acoustic Ocean Sensors: Sourcing and Classification 423
18.2 Classical Non-Acoustic Ocean Sensors 424
18.3 Chemical Sensor Systems 426
18.4 Biological Sensor Systems 429
18.5 Physical Sensor Systems 432
18.6 AUV-Based Physical Sensors – Horizons 435
18.7 AUV-Chemistry Sensors – Horizons 435
18.8 AUV-Based Biological Sensors – Horizons 436
18.9 Autonomous Sampling Systems – Extending Real-Time AUV Sensors 436
18.10 Non-Acoustic Sensor Packaging 436
18.11 The Essential Need for Sensors 437
References 437

19 **Cooperative Vehicle Environmental Monitoring**
Naomi Ehrich Leonard 441
19.1 Motivation 441
19.2 Background and History 443
19.3 Advances in Cooperative Vehicle Ocean Monitoring 445
19.3.1 Cooperative Gliders in AOSN II 445
19.3.2 Cooperative Gliders in ASAP 448
19.4 Recent Developments and Future Directions 452
References 454

20 **Nested Autonomy for Distributed Ocean Sensing**
Henrik Schmidt, Michael R. Benjamin, Stephanie M. Petillo, Raymond Lum 459
20.1 Nested Autonomy 460
20.2 Concept of Operations (CONOPS) 461
20.2.1 Field Level 461
20.2.2 Cluster Level 462
20.2.3 Node Level 462
20.3 Autonomy 463
20.3.1 MOOS-IvP Autonomy Architecture and System 463
20.3.2 The Payload Autonomy Paradigm 463
20.3.3 The MOOS-IvP Autonomy Architecture 464
20.4 Acoustic Communication Infrastructure 466
20.5 On-Board, Real-Time Signal Processing 467
20.6 Application Examples 467
20.6.1 Unified Command, Communication, and Control Infrastructure 467
20.6.2 Adaptive Thermocline and Acousticline Tracking 469
20.6.3 Bistatic Target Tracking (GLINT'10) 474
20.7 Conclusion 478
References 479

21 **Science of Autonomy: Time-Optimal Path Planning and Adaptive Sampling for Swarms of Ocean Vehicles** 481
Pierre F.J. Lermusiaux, Tapovan Lolla, Patrick J. Haley Jr., Konuralp Yigit, Mattheus P. Ueckermann, Thomas Sondergaard, Wayne G. Leslie
21.1 Time-Optimal Path Planning for Swarms of Ocean Vehicles 482
21.1.1 Canonical Steady Flows 483
21.1.2 Time-Dependent 2-D Flows 483
21.1.3 Maintain Swarms Formations 485
21.1.4 Forbidden Regions 485
21.1.5 Uncertain Flow Fields 486
21.1.6 Realistic Ocean Conditions: Three-Dimensional Multi-Scale Flows 487
21.1.7 Realistic Ocean Conditions: Swarms in Multi-Scale Flows with Complex Geometries 489
21.2 Adaptive Sampling for Swarms of Ocean Vehicles 489
21.2.1 Early Results: Approximate Schemes for Realistic Applications 490
21.2.2 Recent Progress: Towards Rigorous Schemes for Realistic Applications 492
21.3 Conclusions and Outlook 494
References 496

22 **Cooperative Vehicle Target Tracking** 499
Frank Ehlers
22.1 General Theoretical Framework 499
22.1.1 Control and Estimation Approach 501
22.1.2 Multiagent Approach 502
22.1.3 Cellular Automata Approach 502
22.1.4 Probabilistic Graphical Models Approach 502
22.2 Distributed Sensing, Control, and Decisions 503
22.3 Multistatic Sonar 505
22.4 Maritime Surveillance 510
22.5 Effective Coordination Schemes 511
22.5.1 The *Understanding Phase* 511
22.5.2 The *Trading Methodology* 511
22.6 Conclusions and Recommendations 512
References 513

23 **Rules of the Road for Unmanned Marine Vehicles** 517
Joseph A. Curcio
23.1 COLREGS 517
23.2 Sensing the World 518
23.3 Proper Behaviors Yield Compliance 519
23.4 Integrating UMVs into Public Water Space 523
23.5 Developing Standards 524
23.6 The Road Ahead 525
23.7 Conclusion 525
References 526

24 **Autonomy: Risk Assessment**
Mario P. Brito, Gwyn Griffiths 527
24.1 Risk Management Process for Autonomous Ocean Vehicles 528
24.2 Risk of Failure 529
24.2.1 Reliability Estimation 529
24.2.2 Reliability Modelling 530
24.3 Risk of Collision 532
24.3.1 Risks of Collision on or Near the Surface 532
24.3.2 Risks of Collision Underwater 533
24.4 Risk of Unavailability 534
24.5 Risk of Loss 535
24.5.1 Expert Judgment Elicitation Processes 536
24.5.2 Survival Prediction 540
24.6 Legal Risks 541
References 542

Part C Coastal Design

25 **Physical Characteristics of Coastal Hazards**
Jennifer L. Irish, Robert Weiss, Donald T. Resio 549
25.1 Types of Coastal Hazards 549
25.1.1 Coastal Storms 549
25.1.2 Tsunamis 551
25.1.3 Climate Variability, Climate Change, and Sea-Level Rise 553
25.1.4 Anthropogenic Activities 554
25.2 Coastal Impacts 555
25.2.1 Winds 555
25.2.2 Wind-Generated Waves 556
25.2.3 Flood Levels and Runup 557
25.2.4 Environmental Impacts 560
25.3 Summary 561
25.4 Nomenclature 561
References 562

26 **Statistical Characterization of Hazards and Risk in Coastal Areas**
Donald T. Resio, Mark A. Tumeo, Jennifer L. Irish 567
26.1 Overview of Risk and Uncertainty 567
26.1.1 Definitions of Basic Terms 568
26.2 Quantifying Coastal Hazards/Risks 570
26.3 Historical Perspective 574
26.3.1 The Development of Deterministic Methods 575
26.3.2 The Development of Probabilistic Methods 577
26.3.3 The Development of the Historical Storm Method for Estimating Coastal Extremes 581
26.3.4 The Development of Alternative Methods 583
26.3.5 Probabilistic Analyses of Extratropical Storms 586
26.3.6 Future Directions and Final Comments 586
26.4 Summary 587
26.5 Nomenclature 587

26.A Appendix: Glossary of Probability and Risk Terms 587
References 593

27 **Modeling of Coastal Waves and Hydrodynamics**
Patrick J. Lynett, James M. Kaihatu 597
27.1 Wind Wave Modeling 598
27.1.1 Linear, Analytical and Semi-Empirical Approaches 598
27.1.2 Spectral Modeling: Phase-Averaged and Phase-Resolving Approaches 600
27.1.3 Depth-Integrated and Boussinesq-Type Approaches 602
27.1.4 Navier–Stokes Equation-Based Approaches 603
27.2 Modeling Long Waves 604
27.3 Coupled and Nested Techniques 605
27.4 Summary of Model Properties 606
27.5 Conclusions 608
27.6 Nomenclature 608
References 608

28 **Modeling of Coastal Morphological Processes**
J.A. Dano Roelvink, Dirk-Jan R. Walstra, Mick van der Wegen, Roshanka Ranasinghe 611
28.1 Types of Coastal Models 611
28.1.1 Physical Scale Models 611
28.1.2 Analytical Models 612
28.1.3 (Semi-)Empirical Models 612
28.1.4 Process-Based Models 612
28.2 Principles of Process-Based Morphodynamic Modeling 613
28.2.1 What Is a Process-Based Morphodynamic Model? 613
28.2.2 Mass Balance Equation 613
28.2.3 Bed Celerity 613
28.2.4 Development Toward Equilibrium 614
28.2.5 Types of Process-Based Models 614
28.2.6 Upscaling Techniques 617
28.3 Modeling Approaches 619
28.3.1 Storm Event Modeling 619
28.3.2 Modeling Cross-Shore Breaker Bar Cycles 623
28.3.3 Long-Term Modeling of Estuaries 626
28.4 Future Directions 629
28.4.1 Ensemble Modeling 629
28.4.2 Biogeomorphology 629
28.4.3 Dune Modeling 630
28.4.4 Coral Reefs 630
28.4.5 Modeling Stratigraphy 630
28.4.6 Morphological Modeling of Tsunami Deposits 630
28.5 Nomenclature 631
References 631

29 **Beach Nourishment**
Robert G. Dean, Thomas J. Campbell 635
29.1 Advantages of Beach Nourishment Over Other Approaches 636
29.2 Methods of Delivery of Sand for Beach Nourishment 636

29.2.1 Characteristics and Considerations in Design of Beach Nourishment Projects 637
29.2.2 Significance of Good Quality Sand 638
29.2.3 Various Settings for Beach Nourishment 638
29.3 Role of Structures in Beach Nourishment 639
29.3.1 Terminal Structures 639
29.3.2 Groins 640
29.3.3 Detached Breakwaters 640
29.4 Design and Prediction Approaches and Methods 640
29.4.1 Pelnard Considère Methodology 641
29.4.2 One-Line Numerical Models 643
29.4.3 N-Line and more Detailed Models 643
29.5 Additional Design Considerations 643
29.5.1 Erosional Hot Spots 643
29.5.2 Nearshore Placement 644
29.5.3 Ad-Hoc Transformation for Modeling 645
29.6 Legacy Beach Nourishment Projects 645
29.6.1 Examples and Discussions of Legacy Projects 645
29.7 Other Beach Nourishment Projects 648
29.7.1 Seabright to Manasquan Inlet, NJ 648
29.7.2 Captiva Island, FL 648
29.7.3 Treasure Island, FL 649
29.8 Summary and Conclusions 649
29.9 Nomenclature 650
References 650

30 **Storm Hazard Mitigation Structures**
David R. Basco 653
30.1 Design Criteria, Philosophy, and Constraints 655
30.1.1 Water Levels 655
30.1.2 Wave Conditions 656
30.1.3 Surf Conditions on the Structure 656
30.1.4 Probabilistic Design 657
30.1.5 Coastal Risk 657
30.1.6 A Coastal Storm Severity Index 657
30.1.7 Design Constraints 657
30.2 Coastal Armoring Structures 658
30.2.1 Types and Purpose 658
30.2.2 Functional Design 658
30.2.3 Structural Design 663
30.3 Shoreline Stabilization Structures 674
30.3.1 Types and Purposes 674
30.3.2 Functional Design 675
30.3.3 Structural Design 681
30.4 Websites and Sea Level Rise Trends 681
30.4.1 Websites 681
30.4.2 Sea Level Rise Trends 681
References 681

31 **Port and Harbor Design**
Andrew Cairns, John M. Carel, Xiao Li 685
31.1 Port and Harbor Layout and Design 685
31.1.1 Types of Marine Terminals 687
31.1.2 Vessel Overview 689
31.1.3 Harbor Operational Limits 690
31.1.4 Harbor Navigation 691
31.1.5 Sediment Transport Considerations 692
31.2 Structure Types 692
31.2.1 Breakwaters and Wave Attenuation 692
31.2.2 Open Pile Platforms 693
31.2.3 Solid Structures 694
31.2.4 Drydock Facilities 698
31.2.5 Floating Structures 701
31.2.6 Swinging Moorings 701
31.2.7 Ice Breakers 703
31.3 Loads on Structures due to Vessel Mooring and Berthing 703
31.3.1 Vessel Berthing Loads 703
31.3.2 Mooring Loads 704
31.3.3 Loads on Structures due to Ocean Environmental Conditions 706
31.3.4 Tide and Storm Surge 707
31.3.5 Ice 707
31.3.6 Tsunamis 707
31.4 Suggested Reading 707
31.5 Notation 709
References 709

32 **Marine Outfalls**
Peter M. Tate, Salvatore Scaturro, Bruce Cathers 711
32.1 Terminology 712
32.2 Governance 713
32.2.1 Drivers for a Marine Outfall 713
32.2.2 Wastewater Treatment 714
32.2.3 Data Collection for Outfall Design 715
32.3 Predicting Near-Field Dilutions 716
32.3.1 Physical Models 717
32.3.2 Positively Buoyant Jets and Plumes 717
32.3.3 Negatively Buoyant Jets 718
32.3.4 Model Validation 718
32.3.5 Far-Field Numerical Modeling 721
32.3.6 Data for Running the Models 721
32.3.7 Conceptual Design for Positively Buoyant Discharges 722
32.4 Hydraulic Analysis and Design 723
32.4.1 Governing Hydraulics 723
32.4.2 Diffusers – Hydraulic Design 725
32.4.3 Flow Variability 729
32.4.4 Hydraulic Integration 730
32.4.5 Air Entrainment 732
32.4.6 Sedimentation 733

32.5 Outfall Construction 734
32.5.1 Construction Materials 734
32.5.2 Construction Methods 734
32.5.3 Some Considerations 735
32.6 Environmental Monitoring 736
32.6.1 Change Versus Impact 736
32.6.2 Pre- and Post-construction Monitoring 736
32.6.3 Long-Term Monitoring 738
32.6.4 Summary 739
References 739

Part D Offshore Technologies

33 **Offshore Platforms**
Arisi S.J. Swamidas, Dronnadula V. Reddy 745
33.1 Relevance 745
33.2 Types of Offshore Platforms 747
33.2.1 Fixed and Compliant Offshore Platforms 748
33.2.2 Floating Offshore Platforms 750
33.2.3 Subsea Systems and Pipelines 750
33.3 Future Trends and Developments in Offshore Platforms 751
References 752

34 **Stability of Offshore Systems**
Alexia Aubault, R. Cengiz Ertekin 755
34.1 Stability Criteria 756
34.2 Fundamentals 757
34.2.1 Static Movements and Hull Position Float 757
34.2.2 Dynamic Movements 758
34.2.3 Some Geometric Definitions 758
34.2.4 Center of Buoyancy and Gravity 758
34.2.5 Irregular Shapes and Numerical Integration 759
34.2.6 Unit Systems 761
34.3 Hydrostatic Forces and Moments 761
34.3.1 Buoyancy and Displacement 761
34.3.2 Equilibrium of Forces and Moments 762
34.3.3 Shifting of Weight and Volume 763
34.4 Stability 763
34.4.1 Righting Arm, Righting Moment, and Metacentric Height for Small Angles 763
34.4.2 Transverse Metacentric Radius 764
34.4.3 Trim and Longitudinal Initial Stability 764
34.4.4 Weight Addition, Removal, and Shift 765
34.4.5 Effects of Liquid Free-Surface in Internal Tanks 766
34.4.6 Scribanti's Formula 767
34.4.7 Stability at Large Angles of Inclination 767
34.4.8 Dynamical Stability: Energy to Incline 768
34.4.9 Hydrostatic Stiffness Coefficients 769
34.4.10 Stability of Submersibles 770

34.4.11 Purpose 770
34.4.12 Industry Practice 773
34.5 Loads 776
34.5.1 Environmental Loads 776
34.5.2 Dead Loads 778
34.5.3 Live Loads 778
34.5.4 Operational Loads 778
34.5.5 Accidental Loads 779
34.6 Lightship Parameters 779
34.6.1 Technical and Commercial Significance 779
34.6.2 Regulatory Approach 779
34.6.3 Prediction 780
34.6.4 Determination 780
34.6.5 Control 781
34.7 Subdivision 781
34.7.1 Purpose and Criteria 781
34.7.2 Regulatory Requirements 781
34.7.3 Damaged Stability and Residual Stability 782
34.8 Analysis 784
34.8.1 Static- and Quasi-Static Analysis 784
34.8.2 Dynamic-Response-Based Analysis 784
References 785

35 **Wave, Current and Wind Loads**
R. Cengiz Ertekin, George Rodenbusch 787
35.1 Wave Loads 787
35.1.1 Linear Waves 787
35.1.2 Nonlinear Waves 790
35.1.3 Shallow-Water Waves 791
35.1.4 Random Waves 791
35.1.5 Large Bodies 794
35.1.6 Slender-Member Bodies 798
35.2 Current Loads 802
35.2.1 Nonuniform Currents 802
35.2.2 Wave–Current Interaction 803
35.2.3 Wave Current Kinematics 804
35.2.4 Current-Induced Forces 804
35.2.5 Vortex-Induced Vibrations 805
35.3 Wind Loads 806
35.3.1 Wind-Speed Profile 807
35.3.2 Wind Spectra and Gusts 808
35.3.3 Steady-State Forces 809
35.3.4 Unsteady Forces 809
35.4 Model Tests 810
35.4.1 Principles and Similarity Laws 810
35.4.2 Scaling of Loads 812
35.4.3 Elastic Structures 812
35.5 CFD Tools 814
35.6 Extreme Response Estimation 814
References 816

36 **Vortex-Induced Vibrations**
Michael S. Triantafyllou, Rémi Bourguet, Jason Dahl, Yahya Modarres-Sadeghi 819
36.1 VIV Prediction of Wide-Span Rigid Cylinders in Uniform Cross-Flow 820
36.1.1 Flow Instability and the Formation of the Karman Street 820
36.1.2 VIV of Flexibly Mounted Cylinders in Cross-Flow 822
36.1.3 Forced Cylinder Vibrations 823
36.1.4 Lift Coefficient, Velocity and Amplitude 825
36.1.5 Correlation Length 828
36.2 VIV Prediction of Flexible Structures in Nonuniform Flow 831
36.2.1 Traveling Structural Waves and Multimodal Responses in Sheared Flows 831
36.2.2 Lock-In and Fluid-Structure Energy Transfer Within Shear Flow 835
36.2.3 Synchronization of In-Line and Cross-Flow VIV 838
36.3 Experimental Studies and Fatigue Analysis 838
36.3.1 Riser Orbital Motions and Excitation Region 839
36.3.2 Higher Harmonic Strain and Acceleration Components in Flexible Structures 839
36.3.3 Periodic (Type-I) and Chaotic (Type II) Signals 840
36.3.4 Response Reconstruction Using Experimental Data 841
36.3.5 Fatigue Calculations 841
36.3.6 VIV Prediction Tools 842
36.4 Effectiveness of Vortex Canceling Devices 843
36.4.1 Helical Strakes 843
36.4.2 Fairings 844
36.4.3 Other VIV Suppression Devices 844
36.5 Multiple Interfering Bluff Bodies 845
36.6 Effect of Reynolds Number 845
References 846

37 **Structural Dynamics**
H. Ronald Riggs, Solomon Yim 851
37.1 Single Degree-of-Freedom System 852
37.1.1 Equation of Motion 852
37.1.2 Response in the Time Domain – Free Vibration 854
37.1.3 Response in the Time Domain – Harmonic Loading 854
37.1.4 Response in the Frequency Domain 856
37.1.5 Structural Damping 856
37.1.6 Time Domain Response from Frequency Domain Response 858
37.1.7 Fluid Contribution 858
37.1.8 Nonlinear Systems 860
37.2 Multi-Degree of Freedom Systems 861
37.2.1 Equations of Motion 861
37.2.2 Modal Superposition 862
37.2.3 Numerical Time-Domain Solution 863

37.3 Linear Hydroelasticity for Inviscid Fluid Flow 864
37.3.1 Finite Element Structural Modeling 864
37.3.2 Hydrostatic Stiffness 864
37.3.3 Added Mass, Damping, and Exciting Forces 865
37.3.4 Reduced Basis Solution 865
37.3.5 Mapping from Structural Mesh to Fluid Mesh 866
37.3.6 Linear Response to Regular Waves – RAOs 866
37.3.7 Time-Domain Response 867
37.4 Linear Response to Random Seas 867
37.4.1 Specification of Long-Crested Random Seas 867
37.4.2 Specification of Short-Crested, Directional Random Seas 868
37.4.3 Response in Random Seas 869
37.4.4 Estimation of Extreme and Fatigue Response in Random Seas 869
37.5 Nonlinear Hydroelasticity (Nonlinear FSI) 870
37.5.1 Nonlinear Frequency Domain Analysis 871
37.5.2 Nonlinear Time-Domain Analysis 872
References 873

38 **Cable Dynamics for Marine Applications**
Ioannis K. Chatjigeorgiou, Spyros A. Mavrakos 875
38.1 Mathematical Formulation 876
38.1.1 Derivatives in Time and Space 877
38.1.2 Dynamic Equilibrium 878
38.1.3 Final System 878
38.1.4 Reduced Model Omitting Bending Stiffness 879
38.1.5 Equivalent Linearization 880
38.2 The Eigenvalue Problem of a Catenary Mooring Cable 882
38.3 High Tension Cables; Snap-Slack Conditions 885
38.4 Dynamics of Catenary Moorings 888
38.4.1 A Single Mooring Cable 888
38.4.2 Effect of Attached Intermediate Buoys 890
38.4.3 Mooring Line-Induced Damping 892
38.5 Second-Order Nonlinear Dynamics of Cables Formulated as Euler–Bernoulli Beams 894
38.5.1 Background 894
38.5.2 The Governing 2-D Nonlinear System 895
38.5.3 Perturbation Series Expansion 895
38.5.4 The Frequency Domain Approach 896
38.5.5 Numerical Solution 898
38.5.6 The Bottom–Structure Interaction Approach 899
38.5.7 Results of Second-Order Double-Frequency Effects 900
References 903

39 **Offshore Geotechnics**
Dong-Sheng Jeng 907
39.1 Basic Models 907
39.1.1 Simplified Models 908
39.1.2 Biot's Poroelastic Models 909
39.1.3 Poro-Elastoplastic Models 915

39.2 Mechanisms of Seabed Dynamics 916
39.2.1 Oscillatory Soil Response 917
39.2.2 Residual Soil Response 918
39.2.3 Integrated Model 919
39.2.4 Wave-Induced Residual Liquefaction 921
39.3 Wave(Current)-Induced Soil Response in Marine Sediments 922
39.3.1 Theoretical Formulations 922
39.3.2 Effects of Currents 923
39.3.3 Liquefaction of Seabed Under Combined Nonlinear Wave and Current Loading 924
39.4 Seabed Stability Around Caisson Breakwaters 925
39.4.1 Theoretical Models 925
39.4.2 Verifications 926
39.4.3 Dynamic Response of Seabed Around Caisson Breakwaters 928
39.4.4 Wave-Induced Liquefaction Around Breakwaters 929
39.5 Remarks 931
References 932

40 **Buoy Technology**
Andrew Hamilton 937
40.1 Buoy and Mooring Types and Design Considerations 937
40.1.1 Buoys 938
40.1.2 Moorings 940
40.1.3 Failures and Hazards 942
40.2 Buoy and Mooring System Components 942
40.2.1 Strength Members 942
40.2.2 Terminations and Strain Relief 946
40.2.3 Buoys and Floats 947
40.2.4 Hardware 947
40.2.5 Anchors 948
40.2.6 Acoustic Releases 949
40.3 Analysis Techniques 949
40.3.1 Environmental Conditions 950
40.3.2 Static Analysis 950
40.3.3 Frequency Domain Analysis 952
40.3.4 Nonlinear Analysis 954
40.4 Example Designs 956
40.4.1 Chain-Catenary Mooring 959
40.4.2 Auto-Detection Mooring 959
40.4.3 Deep-Water Mooring with Inductively Linked Instrumentation 961
References 961

41 **Liquefied Natural Gas Carriers**
Krish P. Thiagarajan, Robert Seah 963
41.1 Types of LNG Carriers 963
41.1.1 LNG Carrier Containment Systems 964
41.2 Thermodynamics of LNG 968
41.2.1 Properties of LNG 968

41.2.2 LNG Stages 969
41.2.3 LNG Onshore Storage Considerations 969
41.3 Environmental Challenges 970
41.3.1 Open Ocean Voyages 970
41.3.2 Boil-Off 971
41.4 Fluid Structure Interaction of LNG Systems 971
41.4.1 Wave-Induced Motions 971
41.4.2 Sloshing Impact and Consequences 972
41.4.3 Scaling 974
41.4.4 Multiple Body Interaction in Shallow Water 977
41.5 Design Methodologies of LNG Containment Systems 978
41.5.1 Comparative Analysis 978
41.5.2 Model Testing 980
References 983

42 **Salvage Operations**
Michael S. Dean 985
42.1 The Casualty and Response 986
42.1.1 Contract Types 987
42.1.2 Salvage Personnel 987
42.1.3 The Salvage Engineer's Role 988
42.2 Introduction to Salvage Engineering 988
42.2.1 Offshore and Coastal 989
42.2.2 Harbor and Inshore 989
42.2.3 Cargo and Equipment Recovery 989
42.2.4 Pollution and Hazmat 990
42.2.5 Wreck Removal 990
42.2.6 Clearance 990
42.3 Data, Surveys, and Planning 991
42.3.1 Essential Data and Ship Information 991
42.3.2 Salvage Survey 992
42.4 Types of Operations 996
42.4.1 Free Floating 997
42.4.2 Grounded Ships 999
42.4.3 Sunken Ships 1001
42.A Appendix: Dewatering 1003
42.B Appendix: Common Formulas, Calculations, and References 1008
42.C Appendix: De-Beaching 1028
42.D Appendix: Refloating 1037
42.E Appendix: Wrecking in Place 1055
References 1065

43 **Oil Spills and Response**
Merv Fingas 1067
43.1 Frequency of Oil Spills 1068
43.2 Response to Oil Spills 1070
43.3 Typical Oils and Their Properties 1070
43.3.1 Oil Composition 1070
43.3.2 Properties of Oil 1071

43.4 Behavior of Oil in the Environment 1071
43.4.1 Evaporation 1072
43.4.2 Water Uptake 1072
43.4.3 Natural Dispersion 1073
43.4.4 Dissolution 1073
43.4.5 Photooxidation 1073
43.4.6 Sedimentation and Oil–Mineral Particle Interaction 1074
43.4.7 Biodegradation 1074
43.4.8 Tar Ball Formation 1074
43.4.9 Spreading and Movement 1075
43.4.10 Submergence/Sinking 1075
43.5 Analysis, Detection, and Remote Sensing of Oil Spills 1075
43.5.1 Laboratory Analysis 1075
43.5.2 Detection and Surveillance 1076
43.5.3 Remote Sensing 1076
43.6 Containment on Water 1078
43.6.1 Containment Booms 1078
43.6.2 Boom Failures 1079
43.6.3 Sorbent Booms and Barriers 1080
43.7 Oil Recovery on Water 1080
43.7.1 Skimmers 1080
43.7.2 Skimmer Performance 1082
43.7.3 Sorbents 1083
43.8 Separation, Pumping, Decontamination, and Disposal 1084
43.8.1 Temporary Storage 1084
43.8.2 Pumps 1084
43.8.3 Separators 1085
43.8.4 Decontamination 1085
43.8.5 Disposal 1085
43.9 Spill-Treating Agents 1086
43.9.1 Dispersants 1086
43.9.2 Surface-Washing Agents 1086
43.9.3 Solidifiers 1087
43.9.4 Biodegradation Agents 1087
43.10 In-Situ Burning 1087
43.10.1 Advantages 1087
43.10.2 Disadvantages 1087
43.10.3 Ignition and What Will Burn 1087
43.10.4 Burn Efficiency and Rates 1088
43.10.5 Use of Containment 1088
43.10.6 Emissions from Burning Oil 1089
43.11 Shoreline Cleanup and Restoration 1089
43.11.1 Fate and Behavior on Shorelines 1089
43.11.2 Types of Shorelines and Their Sensitivity to Oil 1090
43.11.3 Cleanup Methods 1092
43.11.4 Recommended Cleanup Methods 1092
References 1092

Part E Ocean Renewable Energy

44 **Marine Hydrokinetic Energy Resource Assessment**
Manhar R. Dhanak, Alana E.S. Duerr, James H. VanZwieten 1099
44.1 Wave Energy Resource 1100
44.2 Tidal and Ocean Current Energy Resource 1101
44.2.1 Tidal Currents 1102
44.2.2 Ocean Currents 1102
44.2.3 Power and Power Density of Tidal and Ocean Currents 1104
44.3 Assessment of Global Ocean Current Resources 1106
44.3.1 Gulf Stream Case Study 1110
44.4 Other Considerations 1113
References 1114

45 **Ocean Wave Energy Conversion Concepts**
Nikolaos I. Xiros, Manhar R. Dhanak 1117
45.1 Basic Concepts in Primary Energy Capture 1118
45.1.1 Point Absorbers 1118
45.1.2 Oscillating Water Column 1121
45.1.3 Submerged Pressure Differential Devices 1122
45.1.4 Oscillating Wave Surge Convertors 1123
45.1.5 Attenuator and Terminator 1125
45.1.6 Overtopping Devices 1126
45.1.7 Other Advanced Concepts 1126
45.1.8 Control 1126
45.2 Power Takeoff Systems 1127
45.2.1 Mechanical Transmission PTOs 1127
45.2.2 Pneumatic and Hydro Turbines 1129
45.2.3 Hydraulic Power Takeoff Systems 1130
45.2.4 Linear Generator 1131
45.2.5 Other Considerations 1132
45.A Appendix: Practical Applications of Wave Energy Conversion Technologies 1133
References 1143

46 **Ocean Current Energy Conversion**
Howard P. Hanson, James H. VanZwieten, Gabriel M. Alsenas 1147
46.1 Fundamentals 1148
46.2 The Betz Limit 1149
46.3 Conversion Systems 1151
46.3.1 Axial-Flow Systems 1151
46.3.2 Cross-Flow Systems 1154
46.3.3 Tidal/Open-Ocean/In-Stream Differences 1155
46.4 Supporting Infrastructure 1156
46.4.1 Moorings 1156
46.4.2 Masts, Bases, and Platforms 1158
46.4.3 Other Components 1159
46.5 Beyond Engineering 1160
46.6 Summary 1160
References 1160

47 **Harvesting Energy by Flow Included Motions**
Michael M. Bernitsas 1163
47.1 Hydrokinetic Energy in Horizontal Flow 1166
47.1.1 Marine Hydrokinetic Energy 1167
47.1.2 Potential, Requirements, and Challenges 1168
47.1.3 Steady-Lift Technologies 1170
47.1.4 Alternating-Lift Technologies 1172
47.2 Alternating-Lift Technologies: The VIVACE Converter as a Case Study 1178
47.2.1 Concept, Scales, and Principles 1178
47.2.2 Scales, Models, Prototypes 1183
47.2.3 The Underlying Principles 1189
47.3 Methodology and Tools in Support of Development 1202
47.3.1 Experimental Facilities 1203
47.3.2 Damping Model 1209
47.3.3 Flow Visualization and Vortex Tracking 1210
47.3.4 Field Tests 1212
47.3.5 Computational Fluid Dynamics 1213
47.3.6 Mathematical Model of Harnessed and Dissipated Power 1225
47.3.7 Variable Added Mass Mathematical Model 1233
47.3.8 Physical Models and Model Equivalence 1234
47.3.9 Benchmarking 1236
47.4 Nomenclature 1237
References 1238

48 **Ocean Thermal Energy Conversion**
Muthukamatchi Ravindran, Raju Abraham 1245
48.1 OTEC Principles and Systems 1245
48.1.1 Spin-Off Technologies 1246
48.1.2 Types of OTEC Plants 1247
48.2 History of OTEC Installations Worldwide 1250
48.2.1 Indian Experiences 1253
48.3 Current Status of OTEC Technologies 1254
48.3.1 Developing Technologies on OTEC 1254
48.3.2 Research on OTEC in International Scenario 1254
48.3.3 Commercial Designs for OTEC Plants 1254
48.3.4 Industrial Proposals for OTEC Commercialization 1254
48.4 Design Considerations for Future OTEC Plants 1257
48.4.1 Choice of Working Fluids 1257
48.4.2 Choice of Materials 1258
48.4.3 Selection of Equipment 1259
48.4.4 Sea Water Intake System and Deployments 1261
48.4.5 Bio-Fouling and Control 1263
48.4.6 Power Transmission from Offshore OTEC Plants 1264
48.5 Conclusion 1265
References 1265

49 **Offshore Wind Energy**
Mareike Strach-Sonsalla, Matthias Stammler, Jan Wenske, Jason Jonkman, Fabian Vorpahl 1267
49.1 Current Offshore Wind Turbine Technology 1268
49.1.1 Rotor 1269
49.1.2 Nacelle 1270
49.1.3 Support Structure 1273
49.1.4 Control and Protection System 1275
49.1.5 Power Electronics 1275
49.1.6 Challenges of Current Technology 1276
49.2 Fundamentals of Turbine Dynamics and the Offshore Environment 1276
49.2.1 Wind 1277
49.2.2 Waves 1279
49.2.3 Aero-Hydro-Servo-Elastic Models for Horizontal-Axis Wind Turbines 1280
49.3 Outlook on Future Technology 1283
References 1284

Acknowledgements 1287
About the Authors 1289
Detailed Contents 1303
Index 1305

Index

2P mode 827
2S mode 827
2T pattern 828
3-D current 612
4-D ocean monitoring system 337
1980 equation of state of sea water 104

A

ABS strength assessment 978
absorption 364
– capture width 1120
accelerated low water corrosion 118
acceleration technique 617
accidental load 779
accounting for physical limits 69
acoustic
– communication 359, 466, 478
– homing 392
– imaging 393
– propagation 490
– release 949
acoustic Doppler current profiler (ADCP) 55, 425, 839, 950, 1151
acoustic Doppler velocimeter (ADV) 433
actuator disk model 145
adaptive
– and collaborative autonomy 460
– autonomy 473
– sampling 444, 467, 481, 489
– sampling and prediction (ASAP) 443
– thermocline tracking 469
– tracking 478
ADCP 57
added mass 97, 859
added-mass coefficient 796
added-weight method (AWM) 783
Aeolian transport 630
aero-hydro-servo-elastic
– model 1280
– simulation 1280
after condenser 1250
air
– cushion vehicle (ACV) 151
– lift 1021
– products propane precooled mixed refrigerant (C3MR) 969
airfoils 1270
airy wave theory 78
alternating current 218
alternating-lift 1174
– converter 1164
– technology (ALT) 1163, 1172, 1237
alternative performance standard 776
aluminum 1258
– corrosion 121
ammonia 1248, 1257
amplitude modulation of water waves 81
analog modulation (AM) 360
analog signal reconstruction 201
analog-to-digital conversion (ADC) 198
anchor 948
anchoring 1014
anemometer 57
angle
– of attack 315
– of wave incidence 624
annual
– exceedance probability (AEP) 571, 587
– mean wave power 1100
anomaly 505
anthropogenic activity 549
anti-submarine warfare unmanned surface vehicle (ACTUV) 324
apparent optical properties (AOP) 193
aquaculture 1246
Archimedes' principle 153, 761
Archimedes wave swing (AWS) 1122
areas of waterplane (Awp) 991
armored shoreline 653
array resolution 447
articulated tug barge (ATB) 689
aspect ratio 310
asphaltenes and resins (A/R) 1072
assimilated data 450
Atlantic meridional overturning circulation (AMOC) 1147
atmospheric processes 16
attenuation 364
attenuator 1125, 1142
attitude-heading reference system (AHRS) 344
auto-correlation function 792
automated
– guided vehicle (AGV) 5
– perception 474
– unmanned system 4
automatic
– control system 228
– feedback control 228
– identification system (AIS) 510, 524
– repeat request (ARQ) 372
autonomous
– benthic explorer (ABE) 443, 527
– light intervention vehicle (ALIVE) 414
– manipulation 414
– microbial genosensor 430
– ocean sampling network (AOSN) 387, 443
– ocean sampling network (AOSN) paradigm 387
– ocean vehicle (AOV) 527
– sampling system 436
– surface craft (ASC) 414
– systems laboratory (ASL) 415
– underwater glider 443
– underwater vehicle (AUV) 5, 55, 236, 301, 315, 341, 360, 387, 407, 423, 441, 460, 481, 493, 527, 1184
– wide aperture cluster for surveillance (AWACS) 487
autonomy 298, 463, 517
– mode hierarchy 465
autopilots with obstacle avoidance 525
Auto-Regressive Moving Average (ARMA) 203
Autosub 527

AUV
– based biological sensor 436
– based physical sensors-horizon 435
– characteristics 390
– chemistry sensor 435
– control 400, 401
– control consideration 400
– coordination 401
– docking capture mechanism 395
– hover capable 390
– operations 387
availability analysis 535
avalanching mechanism 620
average wave energy density 81
axial
– motion 884
– stretching 901
axial-flow 1151

B

back-EMF feedback 412
balance of moments 878
ballast 331
bandwidth 359, 364
bar
– amplitude 625
– cycle 623
– response 625
barrier island 620
base 1158
basic hull form
– displacement 149
– planing 149
– semidisplacement 149
basin demonstration 281
basis pursuit 377
BAUV 290
Bayesian aggregation 537
beach 611
– erosion 639
– erosion model 612
– nourishment 615, 635–650
beam-time record (BTR) 468
bearings 1271
bed
– composition 627
– evolution 613
– slope 616
behavior
– algorithm 519
– of oil 1071
behavioral aggregation 538
behavior-based decisions 463
bending stiffness 879
benzene, toluene, ethyl-benzene and xylenes (BTEX) 1071
Bernoulli principle 1150
Bernoulli's equation 159
berthing basin 686
Betz
– limit 1105, 1149, 1152, 1278
– model 1104
bilateral force feedback control 411
bi-normal motion 884
bio fouling 1263
bioaccumulation 738
biochemical oxygen demand (BOD) 714
biodegradation agent 1087
biofilm thickness 1264
biolocomotion 281, 282
bio-mimicry 282, 283
biorobotic autonomous underwater vehicle (BAUV) 282
Biot's poroelastic models 909
bioWAVE 1141
bistatic target tracking 474
blade element momentum theory 144, 1281
blended wing/body 312
block wall 696
blockage effect 1205
blow-out-preventer (BOP) 750
Bluefin 464
Bode plot 241
boil-off 963
– condition 971
– gas (BOG) 969
bollard pull 1031
Bond number 975
Bose-Chaudhuri-Hocquenghem (BCH) 378
bottom–structure interaction approach 899
boundary condition 613, 788, 900
– electromagnetic 180
boundary element method 328
boundary-element method 797, 872, 925
boundary-layer approximation 913
Boussinesq 602
– equation 88
– long wave theory 78
– model 172
box fish 283
braid vortices 821
breaker index 599
breaking
– index 92
– wave condition 90
– waves 53
breakwater 636, 654, 665–669, 672–678, 692, 929
Bretschneider spectrum 794
Brewster angle 184
brine 712, 723, 739
Broadband Environmental Networked Sensor (BENS) 475
broadband VIV 833
Brunt–Väisälä frequency 717
brute force simulation 616
Buckingham-Π theorem 128
buffeting 1164
bulkhead
– anchored 694
– cantilever 694
buoy 937, 1262
– surface 947
– system components 942
– technology 937
buoyancy 716, 761
– engine 304
– negative 712
– positive 712
buoyant
– negatively 714, 717, 718
– positively 714, 718, 722

C

cable dynamic 875
cabled observatory 452
caisson 663–665, 928
– floating gate 698
– gate miter 698
– wall 696
calcium carbonate 113
calm water wave drag 147
camber 317
camouflaging 283
cancelation frequency 891
canonical steady flow 483
canted strut 336, 337
capsizing 771
capture cone 389
capture mechanism 395
captured air bubble (CAB) 152
cargo load 778

carrier sensing multiple access (CSMA) 381
catamaran 150
Catenary Anchor Leg Mooring (CALM) 1262
catenary mooring cable 882
cathodic protection (CP) 696
Cauchy number 812
cavitation number 131
celerity 147
cellular automata (CA) 501
center
– of buoyancy (COB) 409, 997
– of gravity (COG) 705, 998
– of mass (COM) 415
CETO 1134
CFD tools 814
chain 942
– cutting 1059
chamber deaeration 723, 732, 733
channel 611, 1165
– estimation 379
– shoal pattern 626
chaotic
– (type II) signal 840
– response 840
characteristic current profile (CPC) 66
charge-coupled device (CCD) 412, 434
Chiksan loading arm 976
chlorination 1259
circular
– cylinder 801
– error probability (CEP) 413
circulation 317
closed cycle 1248
closed-loop
– stability 252
– system 238
clustering 512
cluster-level 462
clutter
– database 508
– in-situ clutter analysis 508
cnoidal wave 88, 791, 927
Coanda
– effect 316
– triplane 316
coastal
– design 5
– hazard 549, 551, 561, 567, 568, 570, 574, 577, 583, 587
– ocean 39
– storm 653, 657, 675
– storm impulse (COSI) 657
coastline 612
code-division multiple access (CDMA) 381
cofferdam 694
coherent modulation 375
cold water pipe (CWP) 1252
Colebrook–White equation 725
collaborative
– autonomy 473, 478
– robotic sensing 452
– sampling 467
– tracking 459
collective
– behavior 444
– intelligence 441
– motion 441
collision avoidance 521, 542
COLREGS compliance autonomy 522
COLREGS compliant behavior 525
communication 399, 467
– infrastructure 460
– link 399
– network 453
– peer-to-peer 503
– underwater 503
compact control language (CCL) 383
compatibility relation 878
competitive levelized cost of energy 1164
complex
– adaptive system 507
– conjugate 897
– frequency response function 856
– mission 454
– mode 831
– nonlinear system 511
– space 881
– velocity potential 159
composition of seawater 111
compressed natural gas (CNG) 968
computational
– grid 1216
– structural mechanics (CSM) 873
Computational Fluid Dynamics (CFD) 60, 313, 604, 787, 873, 971, 1165, 1213
concentration 711, 718–721, 734
concept of operations (CONOPS) 461
concept whisper 302
conduction current density 179
conductivity, temperature and depth (CTD) 58, 425, 468, 488, 722
connection mechanism 395
Conoco ConocooPhilips' optimized cascade process (CoPOC) 969
consolidation 907, 914, 915
constant-buoyancy method (CBM) 783
constitutive parameter electromagnetic 179
container terminal 687
containment
– boom 1078
– system 963
continuity equation 179
control 281, 474
– and estimation (CE) 501
– architecture for robotic agent command and sensing (CARACaS) 521
– infrastructure 467
– system 1275
controllability 239, 240
conversion system 1151
cooperation
– bio-inspired 504
– collaborative signal processing 501
– demand-based data sharing 510
– handoff mode 500
– self-organization 502
– swarming 502
– teamwork 499
– track and trail 500
cooperative
– control 441, 453
– exploration 452
– robotics 444
– vehicle environmental monitoring 452
coordinated motion 413, 450, 483
coordinated pattern 451
copper-nickel corrosion 121
coral reef 623
Coriolis
– force 53, 284
– parameter 50
– rotation 54
Cornell
– breaking wave and structure (COBRAS) 603, 922

– multi-grid coupled tsunami model (COMCOT) 604
– University long and intermediate wave model (COULWAVE) 603
correlation length 828
corrosion 3, 116
corrosion of steel 119
cost-effectiveness 1284
Coulomb friction 889
counterflooding 997
coupled oscillator dynamics 445
coupling power 399
Courant–Friedrichs–Lévy (CFL) 618
course-keeping autopilot 267
crest heights 52
critical
– damping 854
– nutrient 112
– regime 822
cross-country glider 301, 303
cross-flow (CF) 842, 1154
cross-shore process model 623
cruise ship terminal 688
cryogenic cargo transfer 976
CS1 membrane system 965
CTD measurement 470
cumulative distribution function (CDF) 217, 570, 587
current 49, 60, 65, 400, 447, 613, 800
– density 178
– effect 923
– load 777, 802
– loading 924
current-induced forces 804
curvilinear grid 621
cyclone 549, 555

D

damaged hull strength 1011
damaged stability 782
damping
– coefficient 97, 796, 892
– model 1209
– ratio 826, 861
Darboux vector 877
Darcy–Weisbach equation 724
Darrieus rotor 1155
data
– address generator 215
– assimilation 481
– fusion 503, 518
– rate 362
deadweight tonnage (DWT) 690
de-beaching 1028
decentralized partially observable Markov decision process (DEC-POMDP) 502
decision-feedback equalizer (DFE) 375
decontamination 1085
decoupled motion 890
deep
– ocean 38
– phreatic thermal explorer (DEPTHX) 533
– sea cold water 1246
deep water 83
– mooring 960
– wave 80
deformation response factor 855
degree of freedom (DOF) 407, 1282
delay-tolerant network (DTN) 382
demanded static load (DSL) 1015
deployable experimental multistatic undersea surveillance (DEMUS) 475
depth
– limited roaming 307, 314
– of closure 614
– unlimited roaming 307, 314
desalination 711, 724, 731–733
design
– criteria 653–655
– sloshing load 980
destabilization 887
detecting, classifying, localizing and tracking 462
dewatering 1003
– compressed air 1003
DEXAWAVE 1142
diaphragm wall 697
DIDSON 416
dielectric permittivity 178
differential global positioning system (DGPS) 417
diffraction 599–601, 795
diffuser 711–713, 718–726, 734
diffusion equation 909
digital
– acoustic telemetry system (DATS) 360
– audio broadcast (DAB) 377
– controller implementation 256
– filter 206
– video broadcast (DVB) 377
Digital Signal Processor (DSP) 211
digital-to-analog (D/A) 198, 256, 1200
dike 653, 662
dilution 716–723, 728–730
dimensional analysis 129, 810, 974
dimensional homogeneity 128
dipole 186
direct
– contact condenser (DCC) 1248
– current 188, 1159, 1264, 1275
– drive 1272
direct current 218
directional 69
– dock 396
– spreading function 868
direct-sequence spread-spectrum (DS-SS) 381
discharge coefficients 1018
discrete Fourier transform (DFT) 211, 858
discrete-time
– Fourier transform (DTFT) 211
– signal 197, 201
– system 197
– system stability 257
dispersion
– of surface waves 132
– relation 85, 789
displacement 761
– hull 327
– of caisson 929
– volume 282
dissipated MHK power 1225
dissipation rate 172
dissolved
– organic compound (DOC) 427
– oxygen 111
distance-aware collision avoidance protocol 382
distributed
– command and control 459
– lock-in 838
– roughness 1205
– surface roughness 1181, 1194
disturbance rejection 272, 296
– response 248
divergent waves 329
docking 388, 418
– AUV 387
– mechanism 398

Doppler 362, 369
– velocity log (DVL) 291, 341, 415, 471
doubly-fed generator (DFIG) 1271
downflooding angle 772
draft 758
– reading 994
drag 136
– air and wind resistance 146
– and inertia coefficients 801
– coefficient 50, 798, 881
– coefficientamplification 820
– crisis 845
– force 95, 409, 805
– frictional resistance coefficient 136
– lift hydrofoil 138
– residuary resistance coefficient 136
– skin friction 135
– surface vessel 135
– three-dimensional (3-D) coefficient 131
– two-dimensional (2-D) coefficient 131
– wave-making resistance 147
drift
– angle 757
– load 798
drivetrain 1271
dry bulk terminal 687
drydock 688
dual resonance 831
dual-frequency identification sonar (DIDSON) 412
dune 611, 620
dynamic
– amplification 855
– behavior 613
– calibration 981
– compact control language (DCCL) 462
– equilibrium 853, 878
– free surface boundary condition 79
– pore pressure 927
– positioning (DP) 1262
– pressure 80, 1150
– response 923
– response based analysis 784
– similitude 135
– stability 768
– stall 1281
– tension 889
dynamic model of an underwater manipulator 409
dynamically orthogonal stochastic PDE 486
dynamically tuned gyroscope (DTG) 343
dynamics of catenary mooring 888

E

Earth Resources Satellite (ERS1) 56
earthquake ground motion 853
ebb
– delta 617
– shoal 614
echo-repeater (ER) 475
economics metric 1109
ecosystem protection 990
eddy viscosity 172
edge wave 92
effect of liquid free surface 766
effective
– added mass 824
– metacentric heigth 767
effluent 736
eigenfrequency 882
eigenmode 882
eigenvalue 862
– assignment 264
– problem 882
Ekman flow 34
El Nino-Southern Oscillation (ENSO) 38
elastic
– eigenfrequency 891
– hose 945, 959
– structure 812
elasto-plastic model 921, 922
electric
– charge 178
– conductivity 178
electrically-excited-synchronous generator (EESG) 1271
electrolysis 1256
electromagnetic
– field theory 178
– homing system 394
– propagation 193
electromotive force (EMF) 411
elevated nutrient level 112
ELF band 181
empirical
– orthogonal function (EOF) 66
– simulation technique (EST) 573, 655
– track method (ETM) 583
energy
– density spectrum (EDS) 19
– dissipation 852, 892
– transfer 87
entrainment 173, 711, 729
entrapment 731
environmental
– condition 950
– contour 63
– load 776
– loading 1277
– monitoring 441
– sampler processor (ESP) 430
equation of motion 796, 852
– frequency domain 856, 861
– incremental 860
– linear 853
– SDOF system 852
– time domain 861
equatorial current 37
equivalent
– catamaran 336
– linearization 880
erosional hot spot (EHS) 643
error subspace statistical estimation (ESSE) 490
estuarine plan form 626
Euler
– angle 877
– number 812, 975
Euler-Bernoulli beam 876, 894
Euler's
– equation 159
– integral 79, 788
EU-MICORE 622
eutrophication 715
Exner equation 613
experimental harnessed power 1227
extended Kalman filter (EKF) 416
exteroception 519
extractable power 1106
extratropical storm 49, 570, 584, 586
extreme
– criteria 66
– response 793, 814

– value analysis 120
– value estimation 869
extremely low frequency (ELF) 178

F

fading 370
failure 657, 665, 681
– model end effect analysis (FMEA) 531
fairing 844
fast
– Fourier transform (FFT) 211, 375, 858, 1216
– tuning 1126
fatigue 841, 869
fault tree 531
feasibility study 1253
feedback 379, 446
– control 442
– interconnection 451
feet of seawater (FSW) 1017
ferry terminal 688
fiber
– optic gyroscope (FOG) 343
– reinforced plastic (FRP) 1256
field
– experiment 445, 450
– intensity 178
– of view (FOV) 412
finite difference 886
finite element 886
– analysis (FEA) 980
– method (FEM) 797, 852, 864, 923, 1282, 1283
finite impulse response 203
finite-duration
– impulse response (FIR) 203
– window 221
finite-volume method (FVM) 873
Finnegan 292
FIR filter structure 206
first-order
– hold (FOH) 201
– reliability method (FORM) 803
– system 896
fishing port 688
fixed tuning 1126
flame ionization detection (FID) 1067
flammable cargo 963
flapcase 1124
flapping-fin 281
flexibly mounted cylinder 822
float 947
floating
– liquefied natural gas (FLNG) 963, 969
– offshore wind turbine 1283
– plant 1247
– production, storage and offloading (FPSO) 6, 745, 757, 977
– regasification and storage unit (FSRU) 969
flood 550, 557–559, 574
flooding 577
flow
– calculation 726
– instability 821
– kinematics 155
– model 617
– separation 171
– visualization 1210, 1211
FlowCAM 431
FlowCytobot 432
flow-induced motion (FIM) 1096, 1163
– cylinder 1217
– in flow transition 1208
flow–structure interaction (FSI) 1096, 1163
fluid
– contribution 858
– mesh 866
– static 153
– structure interaction 872
– structure interaction (FSI) 860, 971
– viscosity 1024
fluke 306
flux density 178
flying wing 312
Folaga 468
force
– coefficient 131
– gravitational 811
– in phase with acceleration 823
– in phase with velocity 823
– inertial 811
– viscous 811
forced vibration 823
force–torque sensor 411
forcing mechanism 627
forecast 629
forward
– error coding (FEC) 378
– model 630
– perpendicular 765
fouling factor 1263
four-cylinder
– CFD 1222
– FIM 1222
Fourier
– series 881
– transform 51
– transformation 894
frequency
– Brunt–Väisälä 717
– coherence 368
– division multiple access 381
– hopped m-ary frequency-shift-key 374
– selectivity 369
– shift keying (FSK) 360, 468
frequency domain 888
– analysis 851, 949
– approach 896
– response 856
– to time domain 858
friction 1011
– velocity 172
Froude number (FR) 131, 327, 717, 811, 975
– densimetric 717
Froude scaling 136, 978
Froude–Krylov force 94
Froude's
– hypothesis 136
– scaling law 812
full dynamic model 913
fully nonlinear
– Boussinesq wave model (FUNWAVE) 603
– wave 93
fully turbulent unseparated boundary layer 310
functional design 653, 675
fundamental wave length 327

G

galloping 1175–1177
galvanic corrosion 1258
game theory 500
– Bayesian game 502
– partially observable stochastic games (POSG) 502
gap flow 1176, 1177

gas
– chromatograph (GC) 1075
– chromatography with mass spectrometric detection (GC-MS) 1067
gauge pressure 153
Gaussian Mixture Model (GMM) 492
geared drive 1272
gearing-sprocket generator powertrain 1128
general cargo terminal 688
generalized
– coordinates 862
– dynamic wake (GDW) 1281
– extreme value 577, 667
– Pareto distribution (GPD) 578, 982
generator 1271
genetic algorithm (GA) 492
geographic information system (GIS) 1070
geometric similitude 134
GhostSwimmer 292
girder design 1048
glass-reinforced plastic (GRP) 1003
glide
– path 314
– polar 314
glider 301, 445, 488
glider coordinated
– control system (GCCS) 446
– trajectory (GCT) 450
global
– drifter program (GDP) 58
– positioning system (GPS) 58, 228, 291, 341, 413, 446, 470, 720
– tropical and extratropical cyclone climatic atlas (GTECCA) 56
global ocean
– current resources 1106
– observing system (GOOS) 58
global warming 1246
– potential (GWP) 1257
GMM-DO Filter 493
Gorlov turbine 1169
GPS 344
gradient 446, 613
– climbing 442
– estimate 448
graving dock 698
gravity platform 749
gravity-based substructure (GBS) 1274
Green function method (GFM) 797
Greens theorem 93
Greenwich mean time (GMT) 449
grid resolution 617
groin 636, 643, 645, 668, 678–680
ground
– effect vehicle 152
– reaction 1008
– tackle 1031
grounding 779
groundwater flow 630
group
– speed 80
– velocity 790
groyne 614
Gulf Stream 54, 489, 1110
– ring 37
gulping 732
Gumbel
– distribution 120
– extreme distribution 120
– plot 120
gust 1279
gyroscope 343

H

Hagen–Poiseuille flow 169
Hankel transform 189
hard structure 653, 681
harmonic 894
harnessed MHK power 1225
Harvard Ocean Prediction System (HOPS) 450
hazardous material 990
hazardous material (HAZMAT) 990
head
– local 724
– piezometric 730–732
– total 723
headland 653, 674–678
– embayment 615
headloss 724–726
– local 724
heat
– exchanger 1259
– transfer coefficient 1257
heave 335, 758
– response 1119
heaving cylinder 1120
heel angle 757
height
– of the longitudinal metacenter 986, 1040
– of transverse metacenter above keel 991
helical vortices 305
hexapod 980
high
– density polyethylene (HDPE) 1253
– frequency (HF) 57, 254, 503
– frequency radar (HFR) 33, 57
– level platform 694
– lift 291
– pressure (HP) 1017
– sea state 297
– strength low alloy (HSLA) 114
– tow force (HTF) 325
– water mark (HWM) 585
higher
– harmonic force 830
– order wave load 871
high-level routing 482
highly tensioned cable 885
high-occupancy vehicle (HOV) 525
historical storm database 55
homing
– on patterns 394
– sensor 391
Hooke's law 910
horizontal
– axis wind turbine (HAWT) 1268, 1280
– electric dipole (HED) 188
– flow 1166
– magnetic dipole (HMD) 188
– transport efficiency 308, 320
hovering 285
hub 1270
hull
– deflection 995
– speed 148
– strength 1011
– stress 995
human–machine interface (HMI) 198, 1136
human–robot team 454
hump 614
– speed 334
HURDAT best-track data 56

hurricane 29, 549, 556, 558, 559, 569–571, 582, 584, 619
– Gloria 53
– Katrina 54
hybrid
– coordinate ocean model (HYCOM) 1106
– coordinate ocean model (HYCOM) current model 60
– cycle 1246
– drive 1273
– ROV 411
– strip theory 336
hydraulic
– accumulator 319
– jump 731
– mining 627
– power takeoff system 1130
– PTO 1130, 1131
hydraulics 711, 730–733
hydro turbine 1129
– overtopping WEC device 1130
hydrodynamic 281
– damping 859, 865
– force 409
– gap 977
– transparency 1282
hydrodynamics 155, 1281
hydroelasticity 851, 864
hydrofoil 138, 151
– small water area ship (HYSWAS) 150
hydrostatic pressure 80, 1028
hydrostatic stiffness 858, 864
– coefficient 769
hyperbolic navigation 348

I

ice load 777
ideal
– and viscous fluid flow 127
– sampler 198
idealized turbine 1149
IGC code 964
IHI-SPB self-supporting tank 964, 974
IIR filter structure 208
impact 711, 735–739
– pressure 966
impeller 1261
impressed current density 179
incident wave force 94
incipient breaking 90
incompressible fluid 787
independent self-supporting tank 966
indicative effective horsepower (EHP) 330
individual risk per annum (IRPA) 67
induced buoyancy 1006
induction generator (IG) 1271
inductive power module 389
inductively linked instrumentation 961
inertia
– coefficient 798
– force 95, 912
– of the waterplane (IT) 1037
inertial
– force 930
– measurement unit (IMU) 291
– navigation system (INS) 341, 344, 415
inferior-olive (IO) 287
infinite impulse response (IIR) 203
infinite-duration impulse response (IIR) 208
infragravity 620
– wave 92
infrared (IR) 326, 1077
initial consolidation 928, 930
initial stability 154
input reduction 619
input–output (I/O) 237
– decoupling 261
inspection, repair, and maintenance (IRM) 404
in-stream current 1155
insulated-gate bipolar transistor (IGBT) 1275
intake pipe 1261
integral transform 211
integrated
– model 919, 926
– perception 474
– tug barge (ITB) 689
integrator anti-windup 255
interaction index 974
inter-carrier interference (ICI) 377
intermediate buoy 890
internal
– diameter (ID) 976
– forcing 903
– tank 766
– wave 473
international
– practical temperature scale (IPTS) 102
– standards for risk 590
– system of units 103
interpolation function 864
inter-symbol interference (ISI) 368
interval programming (IvP) 463, 519
intervention-AUV (I-AUV) 414
intrinsic impedance 181
intrusion seawater 722, 735, 736
inundation 621
Invar membrane 965
inverse first-order reliability methods (IFORM) 63
irregular wave 91, 337
irrotational flow 787
Ishikawajima–Harima heavy industries (IHI) 967
Iver 471
Iver2 468
IvP-Helm 463

J

jacket-type platform 756
jet 171, 717–719, 732
joint probability
– density function 870
– method (JPM) 583, 655
JONSWAP spectral form 52
J–S curve 919

K

Kalina cycle 1249
Kalman (filter) 490
Kaplan–Meier estimator 540
Karman
– street 820
– vortex 283
keel 758
Kelvin wave pattern 147
Keulegan–Carpenter (KC) 799, 804
– number 95, 799
kinematic free surface boundary condition 79
Knudsen relation 103
Kolmogorov scale 172
Korteweg–de Vries equation 78, 88

Kutta–Jukowski lift theorem 166
Kvaerner–Moss self-supporting tank 964

L

Lagrangian 876
– coherent structure 453
laminar
– boundary layer 170
– flow 169
– unseparated boundary layer flow 310
– wake 171
land based plant 1247
land-installed marine power energy transmitter (LIMPET) 1137
Laplace
– equation 79, 159, 788, 908
– transform 229
– transformation 918
large eddy simulation (LES) 604
laser
– Doppler velocimetry 1203
– induced breakdown spectroscopy 428
– in-situ scattering and transmissometry (LISST) 434
– optical particle counter (LOPC) 434
law
– of moving centroids 763, 765
– of the wake 173
– of the wall 173
layered defense 509
leading-edge vortex (LEV) 283
length
– between perpendicular 690
– on waterline (LWL) 140
– overall (LOA) 140, 690, 963
level
– flight hybrid 314
– hybrid 307
– set 482
levelized cost of energy (LCOE) 7
Liberdade/XRay 304
lift coefficient 131
lift-to-drag ratio (L/D) 304
light detection and ranging 57, 326, 621
lightering 997
lightship 779
limit cycle oscillation (LCO) 249, 287
limited-time look-ahead 493
line vortices 305
linear
– and nonlinear waves 78
– dispersion relation 80
– equations of motion 769
– generator 1131
– hydroelasticity 864
– hysteretic damping 857
– time invariant (LTI) 203, 219, 230, 1192
– viscous damping 857
– wave 787
linearized
– AUV forward speed model 236
– drag coefficient 800
– mooring-induced damping coefficient 893
liquefaction 924
– zone 930
liquefied natural gas (LNG) 6, 54, 687, 743, 963
– carrier (LNGC) 963
– carrier containment system 964
– containment design methodology 978
– liquefaction process 969
– offloading 969
– production 968
– properties 968
– sloshing 978
– value chain 968, 969
– vessel 963
liquefied petroleum gas (LPG) 689
liquid
– bulk terminal 687
– chromatography mass spectrometry (LC-MS) 427
LISST-HOLO 434
list angle 757
live load 778
load analysis 1276
loaded mass 302
local control 485
localized corrosion 120
lock-in 827
– in shear flow 835
log law wind model 146
long baseline 341, 413
longitudinal
– metacenter 1039
– position of center of gravity 986, 1039
– position of the center of buoyancy 991
– stability 764
– strength 1001
longitudinal center
– of buoyancy (LCB) 331, 762
– of flotation (LCF) 758, 1039
– of gravity (LCG) 331, 762
longitudinal metacentric
– height (GML) 1040
– radius (BML) 765, 991, 1040
longshore
– current 133, 624
– forcing 624
long-term ecosystem observatory 444
Loop Current 54
Lorenz-95 model 493
lost-buoyancy (constant buoyancy) method 782
low
– aspect-ratio 298
– cycle high-stress fatigue 971
– density parity-check 378
– Earth orbiting 369
– frequency cooperative target tracking 507
– pressure 1017
– probability of detection (LPD) 381
– risk structures 1277
– speed 281
– turbulence free surface water channel 1203
– turbulence free-surface water 1165
lower explosive limit (LEL) 1027
LTI system
– discrete-time 202
– pole placement 261
LTI-MIMO 261
LTI–SISO 261

M

Maasvlakte-2 621
MacCamy–Fuchs approximation 1282
Mach number 975
machine learning 504
– clustering 512
– reinforcement learning 504
– winner-takes-all (WTA) 508

magnetic
– charge density 179
– dipole system 394
– permeability 178
Magnus effect 165
main vehicle computer (MVC) 464
manipulator 407
– control 411
Mariano Global Surface Velocity Analysis (MGSVA) 803
marine
– application 875
– autonomous systems engineering 415
– autonomy 522
– corrosion steel 113
– environment 113
– grade stainless steel 121
– hydrodynamics 127
– power 1167
– railway 700
– renewable energy (MRE) 1147, 1167
– spatial planning (MSP) 7
marine hydrokinetic (MHK) 1096, 1163
– energy 1099, 1147, 1163, 1167
– energy harnessing 1197
– power conversion 1226
– VIVACE power envelope 1230
maritime
– design 1
– hazard avoidance 520
Mark III
– membrane tank 964
Markov chain
– theory 534
– transition probability 534
mass
– flux 85
– ratio 826
– spectrometry (MS) 427, 1076
mast 1158
matching pursuit 377
material 113, 1258
– derivative 156, 877
mathematical
– aggregation 537
– approach 899
– formulation 876
– model 1214
matrix equation 897
maximum
– efficiency 1120
– possible intensity (MPI) 581
– power density 1108
Maxwell's equation 178
mean
– current 84
– high water (MHW) 680
– sea level (MSL) 50, 674
– set down 84, 85
– squared error (MSE) 222, 376
– time between failure (MTBF) 530
– water level 761
measure
– of effectiveness 501
– of performance 501
Mechanical Transmission PTOs 1127
medium
– access control (MAC) 380
– pressure 1017
membrane
– containment 979
– inlet mass spectrometry (MIMS) 427
– tank system 964
metacenter 763
metacentric
– height (GM) 332, 970, 997
– radius (BM) 1037
method of splitting tsunami (MOST) 604
metocean 47
metocean criteria 48
microbially induced corrosion (MIC) 118
micro-electro-mechanical system (MEMS) 343, 432
Microflow cytometer 432
MIKE 3 HD current model 60
Mike21SW 617
mild-slope equation 90
mineral particle interaction 1074
mini-hydroelectric scheme 731
minimum ground leg scope 1013
mission
– AUV 352
– oriented operating suite (MOOS) 463, 519
– planning 532
MIT VIV data depository 838
mixed Eulerian–Lagrangian (MEL) 93, 94
mixed integer linear programming (MILP) 490
mobile
– measurement 58
– offshore drilling unit (MODU) 774
– sensor 442
modal
– reduction 1282
– superposition 862
mode shape 862, 883
model
– physical 717
– scaling 974
– test 810
modeling 58, 474
– vs. measurements 68
modem 360
modified image theory 193
moment
– to change trim by one centimeter (MCTC) 765
– to change trim one inch 991
momentary liquefaction 928
momentum flux (MF) 671
monitoring
– environmental 736, 737
monochromatic motion 891
monopile 1273
Monte Carlo 715
Moody diagram 173
mooring 940, 1156, 1261
– auto-detection 959
– chain-catenary 959
– deep-water 961
– equipment guideline (MEG3) 708
– line damping 820
– line-induced damping 892
– system component 942
– type 937, 941
MOOS database 464
MOOS-IvP 459, 467, 522
– autonomy 463, 464
morfac approach 618
Morison equation 798, 860, 872, 1281
– method 95
Morison force 95
morphodynamic spin-up 627
morphological
– pattern 614
– tide 629
morphology 611
Moss–Rosenberg spherical LNG containment system 966

motion pattern 445
moving average 203
multi degrees of freedom (MDOF) 852
multi fluid cascade (MFC) 969
multiagent
– approach 502
– system (MAS) 501
multibody simulation (MBS) 1282
multi-buoy mooring (MBM) 708
multicarrier 368
– modulation 377
multidisciplinary simulation, estimation and assimilation system (MSEAS) 482, 493
multihopping 365
multi-input multi-output (MIMO) 234, 375, 379
– LTI system 234
multilayer architecture 505
multiobjective
– behavior coordination 464
– optimization 463
multipath 365
– spread 368
multiple
– access 362
– bar system 623
– body interaction 977
– cylinder interaction 1196
– objectives 453
multiple-access 379
– collision avoidance (MACA) 381
multiscale flows 487
multivariable LTI system 234
mutual information 493

N

nacelle 1270
narrowband VIV 833
NATO research vessel (NRV) 468
natural
– ocean current energy 1168
– tidal energy 1168
– wave energy 1168
natural frequency 824
natural frequency (NF) 854
– dry/wet 859, 866
natural gas (NG) 963
– liquid (NGL) 969
natural period
– heave 134
– roll 134
Navier–Stokes (NS) 603, 606
– equation 872, 1148
navigation 400
– aid (NA) 342, 352
– channel 686
Navy coastal ocean model (NCOM) 60, 450
nearly constant velocity (NCV) 476
near-resonant interaction 89
nearshore circulation 93
nested autonomy 459–461
net inlet pressure (NIP) 1021
net positive suction head (NPSH) 1006, 1020
net transport economy (NTE) 308
network 379, 460
– of autonomous platform 337
neutral loading point (NP) 1009
Newmark's method 863
Newtonian 894
nitrate 112
NO96 964
NOAA Wavewatch III model 1101
node-level 462
NOMAD 939
nominal reduced
– frequency 827
– velocity 827
nominal strength 1015
nonacoustic
– ocean sensor 423
– sensor packaging 436
noncoherent modulation 374
noncondensable gas 1248
nonhomogeneous seabed 912
nonlinear
– analysis 954
– hydroelasticity 870
– oscillator 1197, 1198
– property 84
– shallow water (NSW) 604
– system 860, 900
– time domain analysis 872
– wave 924
– wave theory 83
nonparametric model 530
nonuniform current 802
Noordwijk 623
normal motion 884
north equatorial current (NEC) 37
nozzle 712, 723–725, 732–736
number
– densimetric Froude 717, 728
– Reynolds 725
numerical
– integration 759–761, 863
– model 1215
– scheme 880
– solution 898
– solution time domain 863

O

objective analysis (OA) 444, 449
– mapping error 449
observability matrix 240
obstacle 484
– avoidance 533
occupational risk 589
ocean
– autonomy 482
– current 1102
– energy 6
– environment electromagnetism 177
– model 442, 450
– modeling 481
– observation system 459
– observing 481
– observing and prediction system 445
– power 1167
– prediction 450
– sampling 452
– structure 17
– surface tide 40
– temperature driven desalination 1253
– thermal energy conversion (OTEC) 746, 1096, 1132, 1245
– vehicle 481
ocean current
– energy potential metric 1110
– turbine (OCT) 1151
– wind-driven 53
Ocean Explorer (OEX) 468, 476
oceanic
– adaptive sampling 495
– conveyer belt 1147
– noise level 107
– path planning 495
oceanic internal
– tide 43
– wave 43

Oceanlinx 1139
oceanographic
– feature detection and tracking 469
– temperature 102
OEX AUV 470
offshore
– design 49
– environment 1276
– loading 976
– migration 625
– ocean exploration 746
– system 6
– terminal 685
– wind energy 1267
– wind turbine (OWT) 135, 1097, 1267
offshore platform
– fixed and compliant 748
– floating 750
– type 747
oil 714
– and hazardous substance (OHS) 990
– biodegradation 1074
– burning 1087
– cleanup methods 1092
– composition 1070
– containment 1078
– decontamination 1084
– detection 1076
– disposal 1084
– dissolution 1073
– evaporation 1072
– in-situ burning 1087
– natural dispersion 1073
– photooxidation 1073
– property 1071
– pump 1084
– recommended cleanup method 1092
– recovery 1080
– recovery skimmer 1081
– remote sensing 1076
– reserve 746
– restoration 1089
– rig quality (ORQ) 1059
– sedimentation 1074
– sensitivity 1090
– separation 1084
– shoreline cleanup 1089
– spill 1075
– spreading and movement 1075
– submergence 1075
– surveillance 1076
– tar ball formation 1074
– temporal storage 1084
– use of containment 1088
– water uptake 1072
oil and hazardous materials simulated environmental test tank (OHMSETT) 1173, 1208
oleophilic surface skimmer 1081
olivo-cerebellar
– control 287
– dynamic 283
omnidirectional dock 397
ON/OFF control 248
onboard routing 482
one-way travel-time (OWTT) 345
onshore
– LNG containment 970
– migration 625
open cycle plant 1248
open ocean
– current 1155
– density 105
open pile platform 693
operational load 778
optical
– homing 393
– wavelength 193
optimal
– blade design 1270
– control 483
– sampling (OS) 583
optimum propeller 335
optofluidics 432
organic
– nitrogen compound (DON) 427
– thin-film transistor (OTFT) 429
orthogonal frequency division multiplexing (OFDM) 377
oscillating water column (OWC) 1118, 1137
oscillating wave surge converter (OWSC) 1118, 1123, 1141
oscillatory
– pore pressure 918
– soil response 917
outfall 711–739
– shaft 725
– tunnel 725
overtide 626
overtopping device 1118, 1126, 1143
overturning moment (OTM) 61, 776, 1274
overwash 621
Oyster 1141
ozone depletion potential (ODP) 1257

P

Panchal–Bell cycle 1249
parallel
– online approach 618
– polarization 183
parallel-axis theorem 759
parametric model 334, 530
parbuckling 1045
partial differential equation (PDE) 483
partially observable Markov decision process (POMDP) 492, 502
particle
– acceleration 790
– image velocimetry (PIV) 1211
– velocity 789
passenger load 778
passive turbulence control (PTC) 1165, 1193
path planning 481, 508
payload 307, 463
peak
– enhancement factor 52
– wave frequency 52
peak-over-threshold 61, 67, 656
pectoral fin 283
Pelamis 1142
Pelnard Considère (P-C) 641
pEnvtGrad 470, 472
perception-driven control 475
performance
– criteria 247
– prediction 635
periodic (type-I) signal 840
permanent-magnet-excited-synchronous generator (PMSG) 1271
persistence 65, 499
perturbation 628
– expansion 788
– series expansion 895
phase
– shift keying (PSK) 360, 468
– speed 51, 80
phase-locked loop (PLL) 375
phase-locking 838
photomultiplier tube (PMT) 434

photonic inertial navigation system (PHINS) 415
photosynthetically active radiation 425
Pico plant 1138
pier 686
Pierson–Moskowitz (P–M) 794, 867
pipe outfall 713, 735
pipeline 712, 734, 750
pistoncase 1124
pitch 335, 758
– bearing 1271
pitting corrosion 114
plane wave
– incident seawater 184
– propagation 180
– reflection 182
– surface 182
– transmission 182
planetary boundary layer (PBL) 555, 585
planing hull 329
plastic volumetric deformation 921
plate heat exchanger 1253
platform 1158
plume 712, 717, 719–722, 737
PMSG/EESG 1276
pneumatic turbine 1129
point-to-point protocol (PPP) 362
pole placement design 264
pollution prevention 990
polyaromatic hydrocarbon 1071
polydimethylsiloxane (PDMS) 427
polymerase chain reaction (PCR) 431
polytropic index 975
poro-elastic theory 910
poro-elastoplastic 916
poro-elastoplastic (PZ3)
– model 908, 915
port 712, 724–729, 734–736, 758
position float 757
post-processing FIM data 1231–1233
potential flow
– Blasius integral 167
– circular cylinder 165
– D'Alembert's paradox 166
– doublet 164
– hand-drawn streamline 168
– irrotational line vortex 161
– Rankine body 162, 163
– source and sink of equal strength 163
– source/sink flow 161
– theory 78
– uniform flow 160
potential theory 794
power 362, 372
– augmented ram 152
– buoy 1133
– cycle analysis 1247
– electronics 1275
– in oscillator 1226
– math model 1225, 1226
– of single-cylinder VIVACE 1227
– series 84
– spectral density (PSD) 218, 364, 832
– takeoff 1118
– transfer 400
– transmission 1264
power density
– Florida Current 1111
– of ocean currents 1104
– of tidal currents 1104
power-envelope 1228
power-take-off 1178
practical salinity scale (PSS) 103
Prandtl number 975
pressure
– coefficient 130
– implicit with splitting of operators (PISO) 1215
– sensor 981
primary LNG containment 970
Princeton Ocean Model (POM) 61
prismatic tank 963
probabilistic design 653, 665
– tools for vertical breakwaters (PROVERBS) 665
probabilistic graphical model (PGM) 501
probability
– density function (PDF) 216, 371, 486, 511, 570, 587, 870
– distribution 64
– of failure 870
probable maximum hurricane (PMH) 576
profiling glider 301, 303, 309
programmable logic controller 1139
progressive
– flooding 773
– liquefaction 922
– wave 1117
project longevity 636
propagation
– of wave energy 81
– speed 181
proportional integral 252
proportional integral derivative 4, 13, 249
– autopilot 271
propulsive power 282
protection system 1275
PTC-to-FIM map 1181, 1193
publish-subscribe architecture 463
pump 1084
– discharge head 1019
– performance 1025
– suction head 1019
– theory 1018

Q

Qflex class 964
Qmax class 964
quadratic transfer function (QTF) 903, 977
quadrature amplitude modulation (QAM) 360
quadrature amplitude modulation (QPSK) 376
quasi-DC 188
quasi-static 190
– analysis 784
quay 686
QuikSCAT 57

R

radiation
– problem 795
– stress 92, 597
radio frequency (RF) 291, 399, 437, 461
rail mounted gantrie (RMG) 687
Raman spectrometer 424
random
– access 381
– signal 216
random sea 867
random wave 791
– model 91
rapid assessment of morphology 618

ray tracing 366
Rayleigh
– damping 861
– distribution 52, 91
RAZOR 291
reachable set 482
realistic ocean condition 487
real-time flow imaging and classification (REFLICS) 434
reanalysis wind 59
recoverable buoyancy 1004
recursive least square (RLS) 376
red muscle (RM) 282
reduced
– basis 865
– frequency 825
– model 879
– velocity 822
Reed–Solomon (RS) 378
reflection 367
refraction 83, 597
– angle 182
refraction/diffraction model (REF/DIF) 601
refrigerant 1257
Regional Oceanic Modeling System (ROMS) 450
regression analysis 62
reinforced concrete 122
reinforcement corrosion 122
relative sea level rise (RSLR) 654
reliability
– estimation 529
– index 64
– modeling 530
– trial 528
relieving platform 693
remote
– environmental monitoring units (REMUS) 492
– multi-mission vehicle (RMMV) 325
remotely
– operated vehicle (ROV) 297, 361, 388, 407, 428, 751, 944, 987
– operated vessel COLREGS compliance 524
REMUS 444
renewable
– energy 1267
– portfolio standard (RPS) 1166
Renyi entropy 506
representative condition 619
request-to-send (RTS) 382
reserve stability 771
residual
– buoyancy 1008
– liquefaction 921
– soil response 918
– stability 776, 782
resilience 653, 681
resistor and capacitor (RC) 201
resonance 950
resonant sloshing motion 972
response amplitude 952
– operator (RAO) 336, 797, 856, 939, 971
response in the time domain 854
response-based analysis 61
restricted in her ability to maneuver 524
retardation function 859
return
– interval 567, 579
– period 571, 579–581
revetment 653, 662
Reynolds
– effect on VIV 1204
– number 130, 283, 311, 717, 799, 811, 821, 975
Reynolds' scaling law 812, 813
Reynolds-averaged Navier–Stokes equation (RANS) 172, 329, 603, 814, 925
rhodamine 720
Ricean 371
rigging 1013
right of way 518
righting
– arm 763
– arm curve 775
– method 1045
– moment 763
– moment curve 771, 776
rigid
– hull inflatable boat (RHIB) 150, 324, 520
ring laser gyroscope (RLG) 343
riptide 133
riser 712, 723–726, 736, 839
risk 653–657, 673, 681
– assessment 567–570, 588
– factor 528
– mitigation 539
– of failure 529
– of loss 535
– of unavailability 534
risk of collision 532
robotic
– vehicle 441
– waypoint objective 521
Robotuna 292
rogue waves 70
roll 758
roll on roll off 149
roller bearing 1271
ROMS current model 60
rope
– synthetic 944
– wire 943
rosette 723–725
rotor 1269
– aerodynamics 1280
– blade 1270
rotor-nacelle assembly (RNA) 1268
routing 382, 490
rubber tire gantrie (RTG) 687
rubble mound 926
– breakwater 658, 671
– rock 663
rules of the road 517
runup 552, 559, 560, 621

S

sacrificial corrosion 114
safe
– navigation 517
– operation 518
safety factor 980
safety factor (SF) 574
sailing 140
– aerodynamic force 141
– course 141
– hydrodynamic force 142
– leeway angle 141
– true wind 140
salinity 101, 715, 729
– coastal 103
– ocean surface 103
salt marsh 629
salvage
– engineering 985
– foam 1007
– personnel 987
– planning 991
– pump 1022
– survey 992
salvage operation 985
– free floating 997

– grounded ship 999
– sunken ship 1001
sampling 450
– performance 449
– vehicle 493
sandbar 611
satellite
– communication (SAT) 503
– databases 56
Savonius rotor 1154
Saybolt
– furol second (SSF) 1025
– universal second (SSU) 1025
SBEACH 612
scalar (SISO) transfer function 237
scaling
– factor 980
– of loads 812
scarping 621
Schrödinger equation 86
science of autonomy 481
screw propeller 143
– apparent slip 145
– performance curve 144
– propulsive efficiency 144, 146
– Taylor wake fraction 146
– thrust coefficient 144
– torque coefficient 144
– true slip 145
Scribanti's formula 767
Sea Solar Power (SSP) 1254
sea water 106
– electrical conductivity 103
seabased WEC 1134
seabed
– dynamics mechanism 916
– stability 907
Seaglider 303, 444
seakeeping 335
sea-level rise (SLR) 549, 561, 597, 619, 626, 653, 674, 681
search algorithm 454
seasonality 69
sea-surface-temperature (SST) 553
seawall 653, 662
secant wall 697
second
– moment of waterplane area 765
secondary
– LNG containment 970
– steel 1273
second-order
– double-frequency effect 900
– hydrondynamic effect 1284
– nonlinear dynamic 894
– random sea 86
– system 896
sediment 713, 733–738
– balance 617
– compatibility 638
– concentration 624
– transport gradient 613
sedimentation 714, 723, 728
self-noise 306
self-organizing map (SOM) 66
self-propelled line array (SPLINE) 282
self-referential phase reset (SPR) 294
self-regulating nonlinear dynamic 281
self-supporting, prismatic IMO type B 967
semantic sensor network 506
semiautonomous underwater vehicle for intervention mission (SAUVIM) 415
semi-displacement hull 329
sensitivity analysis 626
sensor array 442
servicing 389
sewage 714–716, 728, 733, 739
SF6-water 981
shaft
– drop 731
– horsepower (SHP) 1032
– outfall 726–728
Shallow Water-06 487
shallow-water wave 80, 87
shear
– layer 171
– stress 923
– wave 93
SHEAR7 842
shelf-mounted plant 1247
ship
– bulk carrier 689
– Capesize 690
– container ship 689
– cruise 689
– deck barge 689
– general cargo 689
– tankers 689
ship-to-ship (STS) 964
shoaling 83, 597, 605
– coefficient 1101
shore protection 653, 674, 675
short baseline (SBL) 392
side-by-side ships 977
signal
– estimation 220
– processing 467
signal-to-noise ratio (SNR) 225, 364
significant wave 793
similarity
– law 810
– solution 170
similitude and model testing 127
Simpson's first rule 761
simulating waves nearshore (SWAN) 53, 600
simultaneous localization and mapping (SLAM) 341
single
– anchor leg mooring (SALM) 1262
– carrier 368
– degree-of-freedom 852
– input single-output 219, 230
– mooring cable 888
– point mooring (SPM) 702
sinkage rise 766
sinusoidal steady-state response 241
skimmer 1080
skin depth 394
– electromagnetic 181
slender structure 875
slender-member bodies 798
Slocum 303, 443
– glider 447
SLOSHEL joint industry project 972
SLOSHER 291
sloshing 971
small
– amplitude gravity wave 80
– buoy 938
– craft harbor 689
small waterplane area (SWA) 149
– twin hull (SWATH) 150, 323
smart blade 1284
smooth circular cylinder 800
snap load 977
snap-slack condition 885
Snell's law 83, 183, 599
soft
– docking 398
– galloping 1193
solar radiation 1245
solidifier 1087

solitary wave 88, 791
solution method 797
Sommerfeld condition 96, 795
Sommerfield integral 189
sonic detection and ranging (SODAR) 57
sorbent 1083
sound
– fixing and ranging (SOFAR) 3
– speed 366
source axisymmetric 718
south equatorial current (SEC) 37
Spalart–Allmaras model 1215
spatial scale 442
spatio-temporal evolution of wave 82
specific energy consumption 308
spectral
– model 598, 605
– wave model 617
spectrophotometric elemental analysis system (SEAS) 426
specular irradiance 193
SPERBOY 1140
spill-treating agent 1086
SPLINE 290
spray 303, 443
spreading 364
– function 794
spring 889
squall 50
stability 763
– at large angles of inclination 766
– energy criterion 775
– of submersibles 770
stabilized shoreline 653
stabilizer fin 332
stable energy flux 90
stagnation point 163
stainless steel 120
standard cubic feet (SCF) 1016
starboard 758
state-vector observer 274
static
– analysis 784
– bollard 1031
– stability 154
station keeping 294, 307, 417
stator 1131
steady state spectral wave model (STWAVE) 601
steady-lift
– converter 1164
– technology (SLT) 1164, 1237
– turbine 1171
steady-state
– flight 304
– response 248
steam power plant 1246
steel 114
Steward platform 980
still water level 133, 664, 761
stochastic
– game of life 492
– response 868
Stokes
– drift 85
– flow 170
Stokes wave 790, 927
– theory 78
storm 549–551, 557–559
stormsize effects 576
strain relief 946
strake 843
stranded ship 1028
stratification 712, 720–722
stream function 155
stress 1015
strong suppression (SS) 1193
Strouhal number (St) 131, 283, 805, 821
structural
– damping 854
– design 654, 663, 680
– dynamics 851
– mesh 866
– model 1282
– restoring force 852
– stability 653, 663
– velocity 887
strut canting angle 331
submerged pressure differential device 1118, 1136
subsea system 750
subsumption 519
subtropical gyre 1102
sunken ship stability 1037
supercritical regime 846
supervisory control and data acquisition system (SCADA) 1138
support structure 1268, 1273
surf
– beat 620
– zone 90
surface
– craft for oceanographic and unmanned testing (SCOUT) 522
– gravity wave 19
– roller 624
– seawater 1246
– wind 28
surface vessel
– motion 267
– state-space model 269
surface-washing agent 1086
surge 60, 553–559, 758
survival
– function 530
– prediction 540
suspended sediment concentration (SSC) 627
Swamee–Jain equation 725
SWAN 617
swarm 482
– formation 485
– time-optimal 482
sway 758
swim rule 284
swimmer delivery vehicle (SDV) 302
swirl 306
synergistic FIM 1196
synthetic
– aperture radar (SAR) 57
– storm modeling 68
system identification 1199
system identification (SI) 220

T

tactical scale UUV 298
tandem offloading 976
tangent stiffness matrix 863
tank sounding 995
task reconstruction 418
Taylor series 899
technology readiness level (TRL) 1186
teleoperation 410
temperature compensated crystal oscillator (TCXO) 352
temperature-salinity relationship 105
tendon 885
tension leg platform (TLP) 61, 747, 807, 853, 885

terminal
– homing 388
– structure 639
thermal
– coating 1260
– glider 316, 318
thermocline 469, 1245
– seasonal 102
– tracking 471, 472, 490
Thorp's formula 364
three-cylinder
– CFD 1220
– FIM 1220
thrust 285
tidal
– current 54, 1102, 1155
– forcing 626
– inlet 617
tide 60, 1102
tide-averaging approach 618
time
– reversal 379
– system mapping 203
time division
– duplexing (TDD) 381
– multiple access (TDMA) 381, 467
time domain 888
– analysis 851
– method 797
time-dependent flow 483
time-differencing bio-sonar 281
tip-flow effect 1235
tip-speed ratio (TSR) 131, 1152, 1269
Tollmein–Schlichting wave 135
tons per inch immersion (TPI) 1009
Torpedo-like AUV 390
torque-speed-control curve 1275
total
– alkalinity (TA) 426
– dynamic head 1020
– petroleum hydrocarbon (TPH) 1075
towing tank 1207
toxicity 737
TPC 766
TPI 766
tracer 719–721
transfer bridge 688
transfer function 793
– second order 871
transformation matrix 898
transient
– response 247, 855
– state 614
transition
– piece 1273
– to turbulence 135
translator 1131
transport peak 625
transportation efficiency factor (TEF) 338
transverse
– center of buoyancy (TCB) 762, 998
– center of gravity (TCG) 762
– metacentric height 764
– metacentric radius 764
trapezoidal rule of integration 760
trapped and leaky wave 92
traveling waves 831
triad interaction 89
triangular impulse 982
trim 764, 768
– angle 757
triplex laminate 965
tropical
– cyclone 569–571, 579, 580, 584–586
tropical storm 49
– Ernesto 489
tsunami 551–553, 559–561, 575, 587, 597, 604
tunnel
– outfall 712, 722, 734
– sea lion (TSL) 394
turbine 1260
– dynamics 1276
turbo equalization 378
turbulence 55, 603
– intensity 50, 1278
turbulent
– boundary layer 173
– flow 172
– mixing 173
– spot 135
– trip 138
turning
– ability 298
– basin 686
two-cylinder
– CFD 1219
– FIM 1218
– interference 1176
– near-wake 1218
two-dimensional (2-D)
– horizontal 604, 617
– unsteady, Reynolds-Averaged, Navier–Stokes 1213
– vertical 604
two-way communication 306
type A tank 966
type B tank 966
type C tank 966
types of shorelines 1090
typhoon 549, 569
typical oils 1070

U

Uehara cycle 1249, 1250
ullage
– gas 975
– pressure 964
– ullageocushioning 981
ultra high molecular weight-high density poly-ethylene 942
ultra large container ship (ULCS) 691
ultra large crude carrier (ULCC) 149, 691
ultrashort baseline (USBL) 343, 388, 414, 444
– acoustic system 388
ultrasonic cleaning 1264
unbounded ocean 186
uncertain flow 486
uncertainty 493, 567–570, 579–581, 592
– quantification 483
undertow 616
underwater 301
underwater (UW)
– bioluminescence assessment tool 433
– cable 1255
– communication 467
– excavation 1033
– surface vehicle 5
unified
– command 467
– model 475
uniform corrosion 116
uninterruptable power supply (UPS) 1133
unit system 761

unmanned
– aerial vehicle (UAV) 7, 443, 510, 517, 535
– boat 520
– compliance to COLREGS 523
– ground vehicles (UGV) 517
– marine vehicle 525
– marine vehicle (UMV) 517
– maritime vehicle system (UMVS) 524
– operation in public waters 518
– sea surface vehicle (USSV) 325
– sea surface–vehicle high speed 325
– surface vehicle 323, 520
– underwater vehicle (UUV) 287, 383, 396, 407, 541
– vessel classification 523
unsaturated soil 911
u-p 912
– approximation 907, 913–915, 926
up-over-and-down path 193
upper explosive limit (UEL) 1027
upwelling 449
Ursell number 53, 89, 812
USV master plan 324

V

vacuum insulation 976
validation 920, 1284
valve 722
– check 713, 730
– duckbill 713, 729
– nonreturn 716
variable
– added mass 1233
– frequency drive (VFD) 1138
– permeability 911
variable-pitch-to-feather 1270
variance of response 869
Varley-Seymour (VS) 911
vehicle
– configuration 390
– network 442
velocity
– potential 155
– self-cleansing 728, 734
– space 521
vertical
– center of buoyancy (VCB) 758, 991
– center of gravity (VCG) 758
– distance from the keel to the center of gravity 1001
– electric dipole (VED) 188
– magnetic dipole (VMD) 188
– riser 1262
very high frequency (VHF) 503
very large
– crude carrier (VLCC) 691
– floating structure (VLFS) 798, 812
Very Large Scale Integration (VLSI) 197
virtual
– body and artificial potential methodology (VBAP) 446
– control room (VCR) 450
– experiment 450
– instrument (vi) 1135
– oscillator 1199
viscosity solution 483
viscous drift loads 801
visual
– feedback control 412
– servoing 412
visualization by CFD 1223
VIV suppression 844
VIVACE
– component 1180, 1181, 1183
– converter 1178, 1180
– hydrodynamic principle 1189–1192
– model 1183
– principle 1189
– prototype 1187–1189
– scale 1183
VIVANA 842
void water 304
volume
– of fluid (VOF) 604, 670
– scattering function (VSF) 434
volume-averaged Reynolds-averaged Navier–Stokes equations (VARANS) 925
von Karman vortex street 132
von Karman's constant 50
vortex
– shedding 132, 336, 828, 1174
– tracking 1210
– wake 281
Vortex Hydro Energy (VHE) 1179
vortex-induced
– motion (VIM) 820
– vibration (VIV) 132, 742, 802, 819, 842, 876, 1096, 1163
– vibrations for aquatic clean energy (VIVACE) 1165
vorticity 305, 603
– transport equation 158

W

wake
– capture 827
– defect 174
– displacement 1177
– effect propeller slip 145
– frequency 1175
– interference 845
– stiffness 845
wall-shear stress 285
WAMIT 956
wastewater 711–723, 732–737
water
– level 655–658, 662–667, 673
– particle trajectory 82
– retension from wave splashing and rain 778
– tightness 771
– tracing (WT) 720
wave 549–554, 556–561, 571, 573, 613, 1279
– (current)-induced soil response 922
– body interaction 78
– boundary condition 79
– breaking 91
– condition 655–658, 670
– current interaction 803
– current kinematics 804
– diffraction 96
– dominated system 628
– effect 617
– evolution 25
– exciting force 796, 865
– force 78, 94
– form 216
– frequency 51
– frontage 1117
– generation 25
– height 571
– induced liquefaction 929
– induced pore pressure 910
– induced residual liquefaction 921
– information study wave model 600

– instability 78
– load 777, 787
– long-crested 788
– model 600
– model (WAM) 59
– nonlinear 790
– number conservation 82
– pattern 327
– potential perturbation 871
– power density 1100
– radiation force 96
– resistance 81
– runup 133
– seabed–structure interactions 925
– shallow-water 791
– skewness 616
– spectra 867
– spectrum 792, 1100, 1280
– theory 78
– wave interaction 89
Wave Dragon 1143
wave energy 147
– conversion (WEC) 81, 798, 892, 1117
– conversion actuator (WECA) 1139
– flux 1100
– resource 1100
Wave Information Studies (WIS) 656
Wave Modeling Project 59
WaveRoller 1141
waypoint 446
weakly nonlinear deep water wave theory 83
wear 1271
weathertightness 771
Weber number 131, 812, 975
Weibull distribution 982, 1278
weight
– addition 765
– removal 765, 1054
– shift 765
well 630
western boundary current 1104
wharf 686
WHOI Micromodem 471
whole effluent toxicity (WET) 737
wide band/narrow band response 833
wide sense stationary (WSS) 217
Wiener filter 222
wind 49, 59, 549–551, 553–556, 1277
– climate 629
– driven currents 34
– energy 1267
– gust 808
– load 776, 806
– pressure 809
– spectrum 51, 808, 1279
– speed 49
– speed profile 807
– stress rotation 54
– turbine 1269
– wave 98
winds, waves, and currents (WWC) 47
wing
– in ground-effect (WIG) 152
– induced drag 139
– wingtip vortix 140
winged body of revolution 303
working fluid 1257
World Ocean System Simulator 380
wreck
– grab 1061
– in place 1002
– removal operation 987
wrecking in place 1055

X

XBeach 619

Y

yaw 758
– bearing 1271
yawed swimming 284
yawing 285

Z

zero
– initial condition 230
– order hold (ZOH) 201
– perfect 103
– tension singularity 880
zone
– mixing 714