Computational Methods and Experimental Measurements XV

EDITORS
G.M. Carlomagno & C.A. Brebbia

WIT PRESS

Computational Methods and Experimental Measurements XV

WIT PRESS

WIT Press publishes leading books in Science and Technology.
Visit our website for new and current list of titles.
www.witpress.com

WIT eLibrary

Home of the Transactions of the Wessex Institute.
Papers presented at Computational Methods and Experimental Measurements XV are archived in the WIT eLibrary in volume 51 of WIT Transactions on Modelling and Simulation (ISSN 1743-355X). The WIT eLibrary provides the international scientific community with immediate and permanent access to individual papers presented at WIT conferences.
Visit the WIT eLibrary at www.witpress.com.

FIFTEENTH INTERNATIONAL CONFERENCE ON
COMPUTATIONAL METHODS AND EXPERIMENTAL
MEASUREMENTS

CMEM XV

CONFERENCE CHAIRMEN

G. M. Carlomagno
University of Naples "Federico II", Italy

C.A. Brebbia
Wessex Institute of Technology, UK

INTERNATIONAL SCIENTIFIC ADVISORY COMMITTEE

Organised by

Wessex Institute of Technology, UK
University of Naples "Federico II", Italy

Sponsored by

WIT Transactions on Modelling and Simulation

WIT Transactions

Transactions Editor

Carlos Brebbia
Wessex Institute of Technology
Ashurst Lodge, Ashurst
Southampton SO40 7AA, UK
Email: carlos@wessex.ac.uk

Editorial Board

D N Riahi University of Illinois at Urbana-Champaign, USA

B Ribas Spanish National Centre for Environmental Health, Spain

K Richter Graz University of Technology, Austria

S Rinaldi Politecnico di Milano, Italy

F Robuste Universitat Politecnica de Catalunya, Spain

J Roddick Flinders University, Australia

A C Rodrigues Universidade Nova de Lisboa, Portugal

F Rodrigues Poly Institute of Porto, Portugal

C W Roeder University of Washington, USA

J M Roesset Texas A & M University, USA

W Roetzel Universitaet der Bundeswehr Hamburg, Germany

V Roje University of Split, Croatia

R Rosset Laboratoire d'Aerologie, France

J L Rubio Centro de Investigaciones sobre Desertificacion, Spain

T J Rudolphi Iowa State University, USA

S Russenchuck Magnet Group, Switzerland

H Ryssel Fraunhofer Institut Integrierte Schaltungen, Germany

S G Saad American University in Cairo, Egypt

M Saiidi University of Nevada-Reno, USA

R San Jose Technical University of Madrid, Spain

F J Sanchez-Sesma Instituto Mexicano del Petroleo, Mexico

B Sarler Nova Gorica Polytechnic, Slovenia

S A Savidis Technische Universitat Berlin, Germany

A Savini Universita de Pavia, Italy

G Schmid Ruhr-Universitat Bochum, Germany

R Schmidt RWTH Aachen, Germany

B Scholtes Universitaet of Kassel, Germany

W Schreiber University of Alabama, USA

A P S Selvadurai McGill University, Canada

J J Sendra University of Seville, Spain

J J Sharp Memorial University of Newfoundland, Canada

Q Shen Massachusetts Institute of Technology, USA

X Shixiong Fudan University, China

G C Sih Lehigh University, USA

L C Simoes University of Coimbra, Portugal

A C Singhal Arizona State University, USA

P Skerget University of Maribor, Slovenia

J Sladek Slovak Academy of Sciences, Slovakia

V Sladek Slovak Academy of Sciences, Slovakia

A C M Sousa University of New Brunswick, Canada

H Sozer Illinois Institute of Technology, USA

D B Spalding CHAM, UK

P D Spanos Rice University, USA

T Speck Albert-Ludwigs-Universitaet Freiburg, Germany

C C Spyrakos National Technical University of Athens, Greece

I V Stangeeva St Petersburg University, Russia

J Stasiek Technical University of Gdansk, Poland

G E Swaters University of Alberta, Canada

S Syngellakis University of Southampton, UK

J Szmyd University of Mining and Metallurgy, Poland

S T Tadano Hokkaido University, Japan

H Takemiya Okayama University, Japan

I Takewaki Kyoto University, Japan

C-L Tan Carleton University, Canada

E Taniguchi Kyoto University, Japan

S Tanimura Aichi University of Technology, Japan

J L Tassoulas University of Texas at Austin, USA

M A P Taylor University of South Australia, Australia

A Terranova Politecnico di Milano, Italy

A G Tijhuis Technische Universiteit Eindhoven, Netherlands

T Tirabassi Institute FISBAT-CNR, Italy

S Tkachenko Otto-von-Guericke-University, Germany

N Tosaka Nihon University, Japan

T Tran-Cong University of Southern Queensland, Australia

R Tremblay Ecole Polytechnique, Canada

I Tsukrov University of New Hampshire, USA

R Turra CINECA Interuniversity Computing Centre, Italy

S G Tushinski Moscow State University, Russia

Computational Methods and Experimental Measurements XV

EDITORS

G.M. Carlomagno
University of Naples Federico II, Italy

C.A. Brebbia
Wessex Institute of Technology, UK

WIT PRESS Southampton, Boston

Editors:

G.M. Carlomagno
University of Naples Federico II, Italy

C.A. Brebbia
Wessex Institute of Technology, UK

Published by

WIT Press
Ashurst Lodge, Ashurst, Southampton, SO40 7AA, UK
Tel: 44 (0) 238 029 3223; Fax: 44 (0) 238 029 2853
E-Mail: witpress@witpress.com
http://www.witpress.com

For USA, Canada and Mexico

Computational Mechanics Inc
25 Bridge Street, Billerica, MA 01821, USA
Tel: 978 667 5841; Fax: 978 667 7582
E-Mail: info@compmech.com
http://www.witpress.com

British Library Cataloguing-in-Publication Data

A Catalogue record for this book is available
from the British Library

ISBN: 978-1-84564-540-3
ISSN: 1746-4064 (print)
ISSN: 1743-355X (on-line)

The texts of the papers in this volume were set individually by the authors or under their supervision. Only minor corrections to the text may have been carried out by the publisher.

Printed in Great Britain by Martins the Printers.

Preface

This volume contains most of the papers presented at the 15th International Conference on Computational Methods and Experimental Measurements (CMEM 11), held at the Wessex Institute of Technology in the New Forest, UK. The Conference, which is reconvened every couple of years, started in Washington DC in 1981. Since then, it has taken place in many different locations around the world. The key objective of the Conference is to offer the scientific and technical community an international forum to analyse the interaction between computational methods and experimental measurements and all associated topics with special consideration to their mutual benefits.

The constant development of numerical procedures and of computers efficiency, coupled with their decreasing costs, have generated an ever increasing growth of computational methods which are currently exploited both in a persistently expanding variety of science and technology subjects, as well as in our daily life. As these procedures continue to grow in size and complexity, it is essential to ensure their reliability. This can only be achieved by performing dedicated and accurate experiments. Current experimental techniques at the same time have become more complex and elaborate and require the use of computers for running tests and processing the resulting data.

This book presents a significant number of excellent scientific papers, dealing with contemporary research topics in the field. The contributions have been grouped in the following sections:

- Computational and Experimental Methods
- Fluid Flow
- Heat Transfer and Thermal Processes
- Stress Analysis
- Composite Materials
- Detection and Signal Processing

- Advances in Measurements and Data Acquisition
- Multiscale Modelling
- Ballistics
- Railway Transport

The Editors are very grateful to all the authors for their valuable contributions and to the Members of the International Scientific Advisory Committee, as well as other colleagues, for their help in reviewing the papers in this volume.

The Editors
Wessex Institute Campus
The New Forest, UK, 2011

Contents

Section 3: Heat transfer and thermal processes

Section 4: Stress analysis

Section 5: Composite materials

Section 6: Detection and signal processing (Special session organised by Prof. A. Kawalec)

Section 7: Advances in measurements and data acquisition

Section 8: Multiscale modelling

Section 9: Ballistics

Section 10: Railway transport

Section 1
Computational and experimental methods

This page intentionally left blank

The exponentially weighted moving average applied to the control and monitoring of varying sample sizes

J. E. Everett
Centre for Exploration Targeting,
The University of Western Australia, Australia

Abstract

The exponentially weighted moving average (EWMA) can be used to report the smoothed history of a production process, and has some considerable advantages over a simple moving average (MA). Discussion of these advantages includes comparison of the filter characteristics of the EWMA and MA in the frequency domain. It is shown that the EWMA provides a much smoother filter than does the MA, and the corresponding implications of this difference are examined in the time domain. In smoothing a production process, the successive entities being smoothed commonly have varying "weights", where the weights may be such quantities as tonnage, value or time interval. Standard textbook treatments of moving averages and exponential smoothing are generally confined to equal spaced data of equal weight. Adapting the average to cope with items of varying weight is shown to be trivial for the case of MA, but is not so obvious for the EWMA. This paper shows how the exponential smoothing constant has to be adapted to provide a consistent EWMA. Applications of the EWMA in process control are discussed, with particular reference to quality control in the mining industry.
Keywords: quality control, forecasting, exponential smoothing, sample size.

1 Introduction

It is common to consider a series of observations, x_n, where each observation is equivalently spaced in time or distance or some other relevant dimension.

For forecasting and for system control purposes, it is useful to have some summary of the performance up to the n^{th} observation.

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110011

The summary could be calculated as the mean (M) of all the observations since the first one:

$$M_n = {}_1\Sigma^n x_m/n \tag{1}$$

Usually we are mainly interested in recent history, so a straight average over the entire history of observations. Two approaches are to consider either a Moving Average (MA), applying equal weight to the past k observations, or an Exponentially Weighted Moving Average (EWMA), where successively declining weights are applied as we go further back in history.

1.1 Moving average (MA)

Usually we are mainly interested in recent history, perhaps over the past k observations, so a moving average (MA) over those k observations would be more appropriate:

$$MA_n = {}_{m=0}\Sigma^{k-1} x_{n-m}/k \tag{2}$$

Figure 1 shows the uniform weights of 1/k that are applied to the past k observations.

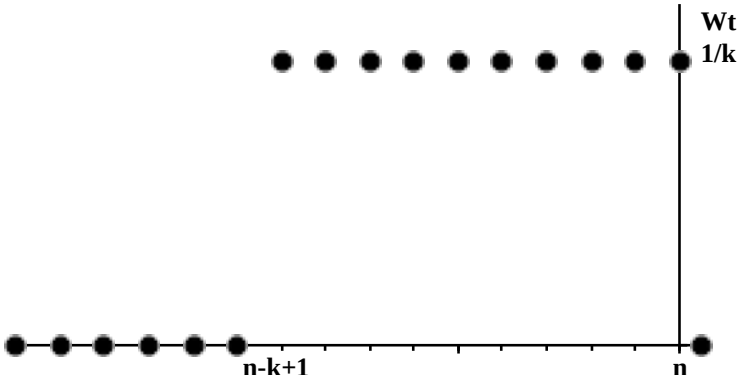


Figure 1: Moving average (MA) weights applied to recent data.

The moving average has the disadvantage that, for the first k intervals, each of the observations is treated as being of equal importance, but then is suddenly disregarded, as soon as it falls off the end of the data being averaged. This discontinuity has several disadvantages that will be discussed more fully in a later section.

1.2 Exponential smoothing (EWMA)

Exponential smoothing, across an "exponentially weighted moving average" (EWMA), provides a smoother means of averaging, where data becomes gradually less influential as it ages.

$$EWMA_n = S_n = (1-\alpha)S_{n-1} + \alpha x_n$$
$$= (1-\alpha)((1-\alpha)S_{n-2} + \alpha x_{n-1}) + \alpha x_n$$
$$= {}_{m=0}\Sigma^{Infinity}\alpha(1-\alpha)^m x_{n-m} \quad (3)$$

Figure 2 shows how the weights applied to earlier data die off exponentially as we go back through the data history.

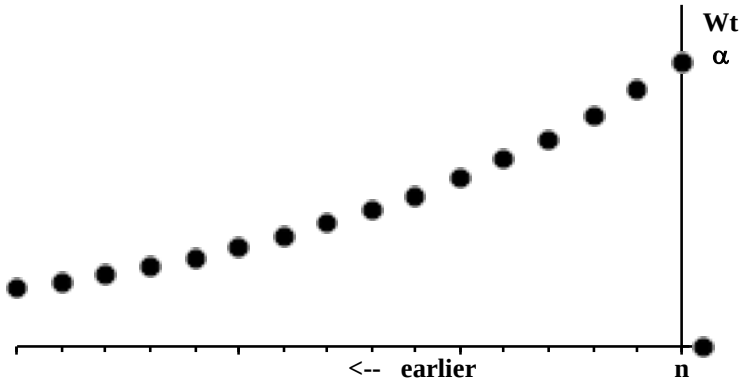


Figure 2: Exponential smoothing (EWMA) weights applied to recent data.

Exponential smoothing is commonly used in forecasting, and is generally described in forecasting textbooks. Diebold [1]) provides a good description.

It can be shown that the EWMA is a minimum mean squared error predictor when the true data generating process is ARIMA(0,1,1).ARIMA processes cover the very wide field of "Autoregressive Integrated Moving Average" processes, identified by Box and Jenkins [2]. As its parameters imply, an ARIMA(0,1,1) process is not autoregressive, but is first-order integrated moving average. Ramjee et al. [3] show that the EWMA method can also provide simple, yet useful, forecasts for other types of ARIMA processes.

Treatments in the literature are generally confined to equally spaced observations of equal weight, so that each new observation is of equal importance. However, it is commonly the case that the desired quality control or forecasting relates to observations that are of varying weight.

An example of this situation, with observations of varying weight, would be a mine's production, shift by shift, of ore of varying tonnage and grade. In this

particular example, the objective may be to forecast the grade of the next shift's production. Alternatively, the purpose may be to summarise the grade of the recent production so that the next shift's ore can be selected so as to restore the smoothed grade back to its target value. Everett [4] provides an example of this type of EWMA application in the iron ore mining industry.

Whether the averaging is being used to generate a forecast or to control a production system, it is being used to summarise recent behaviour. In doing so, it needs to respond to sustained changes in the data, but not be over sensitive to short-term variations. The averaging process is therefore being required to act as a low-pass filter.

Sections 2 and 3 will discuss more fully the advantages of an exponentially weighted EWMA over an MA. Comparing the Fourier transforms of the filters enables their performance as low-pass filters to be evaluated, and clearly demonstrates the advantages of the EWMA over the MA.

Adjustment for varying sample size is comparatively straightforward for the MA. For the EWMA, the adjustment for varying sample size is not so obvious, and appears to have been neglected in the literature.

Section 4 will consider the appropriate treatment of data where sample sizes vary.

Both for MA and EWMA, the choice of weighting constant, and the consequent length over which the data is averaged, depends upon the purpose for which the average is being used.

Section 5 considers the choice of the alpha constant for an EWMA, and its relation to the length of a comparable MA.

2 MA and EWMA compared

Figure 3 shows a signal x with a wavelet disturbance, first up and then down.

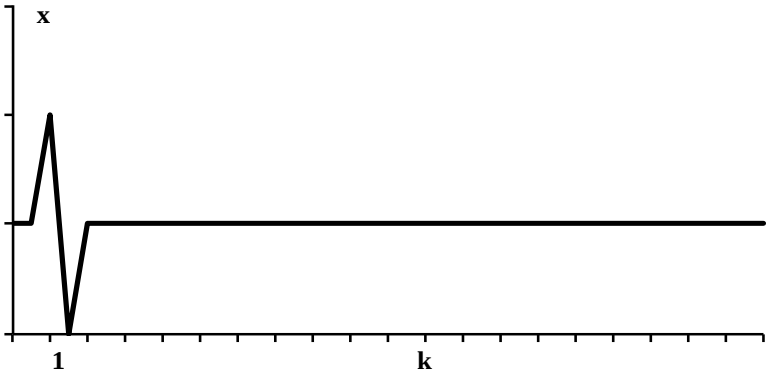


Figure 3: Signal "x" with a wavelet disturbance.

Figure 4 shows the effects of applying a Moving Average (MA) and Exponential Smoothing (EWMA) to this signal x.

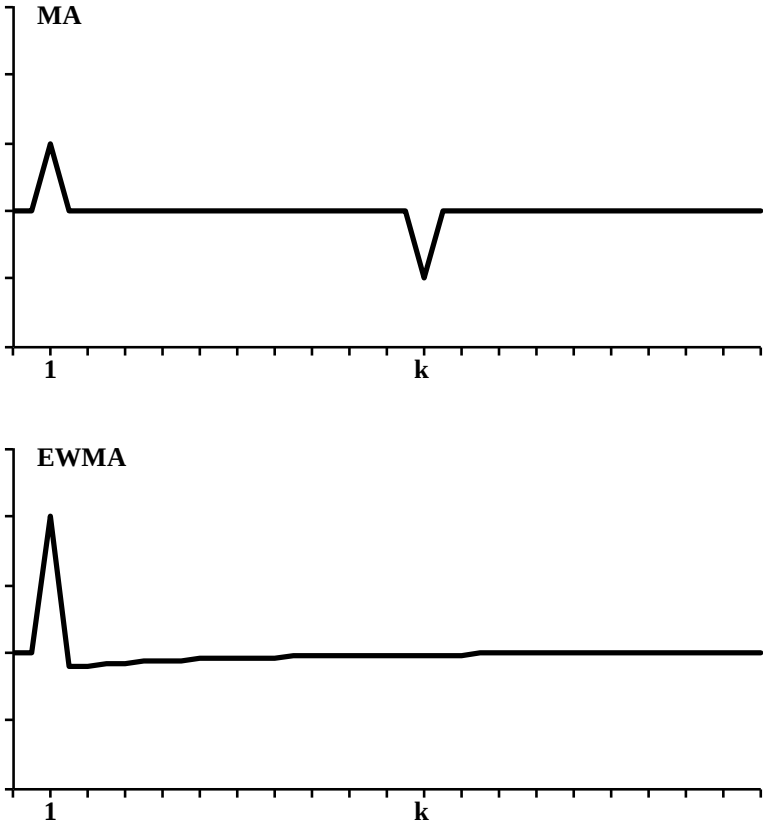


Figure 4: Signal "x" with a wavelet disturbance.

In both cases the disturbance appropriately appears in the smoothed trace at the time it occurs in the signal x.

With the MA an equal and opposite disturbance appears at a delay equal to the length of the Moving Average. This delayed rebound effect is spurious, since its occurrence is dependent solely on the length of the MA and has no relation to the wavelet disturbance.

The EWMA, by contrast, is well behaved, with a gradual return to normal after the disturbance.

3 MA and EWMA considered as low-pass filters

3.1 The Fourier transform

Fourier analysis provides a standard procedure for converting data from the time or distance domain to the equivalent frequency domain [5].

Consider a set of N data values, x_n, equally spaced in time or distance. Their Fourier transform generates N points in the frequency spectrum. These N points in the frequency spectrum carry exactly the same information as the N points in the time or distance domain. The lowest frequency has a period equal to the data length.

Fitting cosine and sine waves of this wavelength to the data generates the real and imaginary components of this fundamental frequency.

Further, fitting cosine and sine waves of each multiple of the fundamental frequency generates its real and imaginary components, up to the "Nyquist" frequency. The Nyquist frequency is N times the fundamental frequency and has a wavelength equal to twice the data interval. Any signal frequency higher than the Nyquist frequency cannot be detected, but will "fold back" to add to the amplitude of a corresponding lower frequency.

Each frequency value can be expressed either as real and imaginary components (the cosine and sine fits), or as an amplitude and phase

The Fourier transform converts the N values in the time (or distance) domain to the equivalent N values in the frequency domain.

Applying the Fourier transform in turn to the frequency domain data converts them back to the time (or distance) domain.

For real-world data, the time (or distance) values are strictly real, while the frequency values will have real (sine wave) and imaginary (cosine wave) components corresponding to their amplitude and phase.

If the data length N is a power of 2 (i.e. $N = 2r$, where r is an integer), the very efficient Fast Fourier transform algorithm can be used. Cooley and Tukey [6] first publicised this algorithm in 1965 (although it was discovered by Gauss in 1805).

Sequentially averaging a set of data is equivalent to applying a low-pass filter to the frequency data.

Applying averaging weights as in equations (2) or (3) to the time (or distance) data is a "convolution" operation. Multiplying the frequency spectrum of the filter weights by the frequency spectrum of the data set is exactly equivalent to convolving the time (or distance) domain data. The Fourier transform of the resulting product of the two frequency spectrums gives the same result as is obtained by convolving the corresponding MA or EWMA with the time (or distance) data.

MA and EWMA each act as low-pass filters, so it is instructive to compare the frequency spectrums.

3.2 Frequency spectrum for the moving average (MA)

The amplitude of the frequency spectrum for the Moving Average filter of Figure 1 is shown in Figure 5.

The amplitude is the square root of the summed squares of the cosine and sine Fourier components. (The phase would be the arctangent of the ratio of the sine and cosine Fourier components, but is not being considered here).

The amplitude spectrum of the MA filter is seen to have side lobes. Instead of the low-pass filter steadily reducing the amplitude of higher frequencies, it

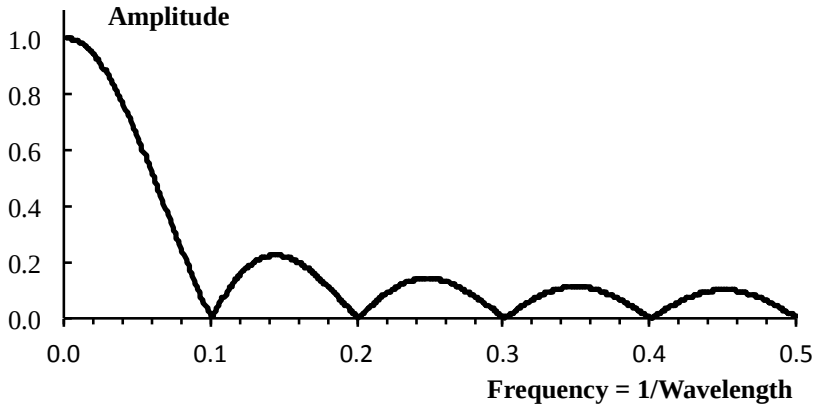


Figure 5: The amplitude spectrum for an MA filter of length = 10.

completely cuts out frequencies of 0.1, which corresponds to a wavelength of 10, the length of the Moving Average filter in Figure 1.

As we increase the frequency, the amplitude rises again, before again falling to zero at a frequency of 0.2 (5 units wavelength). This behaviour is repeated, allowing through ever diminishing side lobes, with complete cut-off at each harmonic of the filter length.

So, as we consider frequencies increasing from the fundamental lowest frequency, they will alternately be filtered out, allowed through, filtered out, and so on repeatedly, with the proportion of signal amplitude allowed through steadily diminishing for each side lobe.

The non-monotonic behaviour of the MA amplitude spectrum is a direct consequence of the MA filter's discontinuity in the time (or distance) domain that we saw in Figure 1.

The operational implication is that some high-frequency disturbances will pass through the filter, while lower-frequency disturbances will be completely blocked if they happen to be close to one of the harmonic frequencies.

For this reason, we must conclude that the Moving Average (MA) filter is unsatisfactory.

3.3 Exponential smoothing (EWMA)

The amplitude of the frequency spectrum for an Exponentially Smoothed filter of Figure 2 is shown in Figure 6.

The amplitude spectrum now has no side lobes, but declines steadily and exponentially. So the EWMA filter is much better behaved than the MA filter. The EWMA filter monotonically decreases the amplitude passed as the frequency increases.

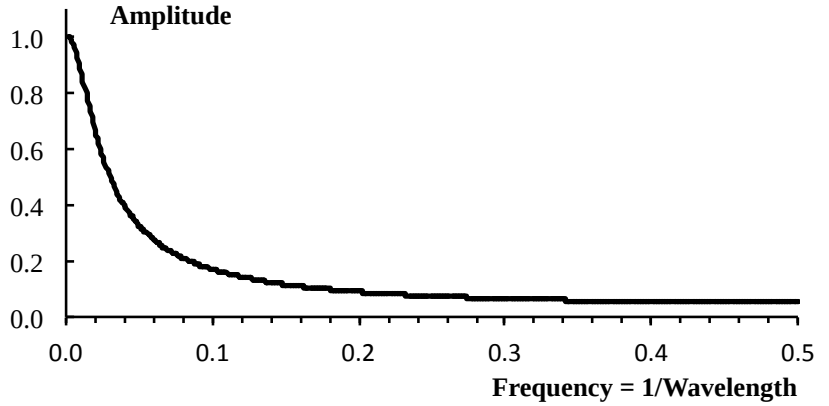


Figure 6: The amplitude spectrum for an EWMA filter with alpha = 0.1.

4 Adjustment for varying sample size

The treatment so far has assumed that the data are of equal importance. However, in many real situations, successive observations may need to exert varying influence. For example, if we are forecasting the grade of ore from previous shifts of varying tonnage, the higher tonnage shifts should have more influence than those of lower tonnage.

We will now consider such a situation of varying tonnage, so that observations x_n relate to tonnages w_n.

4.1 Moving average (MA)

If the MA is to be taken as the moving average over a total tonnage T, then equation (2) becomes:

$$MA_n = {}_{m=0}\Sigma^{k[n]} w_{n-m} x_{n-m}/T, \text{ where } {}_{m=0}\Sigma^{k[n]} w_{n-m} = T \quad (4)$$

For a Moving Average, the length k[n] over which the average is taken will therefore have to be varied so that it encompasses the same tonnage (or as nearly as possible, the same tonnage).

4.2 Exponential smoothing (EWMA)

The treatment for exponentially smoothing over observations with varying tonnages is not so immediately obvious.

It is clear that the appropriate alpha value is a function of the tonnage: if the tonnage w increases we should use a larger $\alpha[w]$, so that a larger tonnage has more influence on the smoothed grade.

Consider two scenarios. Under the first scenario, two successive shifts have identical grade x and equal tonnage w.

Under the second scenario a single shift delivers ore of twice the tonnage, 2w but again with the same grade x.

If we start with a smoothed grade S_O, it is clear that under either scenario we should end up with the same grade, which we shall call S_F.

Under the first scenario, where each of the two shifts has grade x_n and tonnage w_n:

$$S_F = (1- \alpha[w])((1- \alpha[w])S_O + \alpha[w]x) + \alpha[w]x$$
$$= (1- \alpha[w])^2 S_O + \alpha[w](2- \alpha[w])x \quad (5)$$

Under the second scenario, the single shift has grade x and tonnage 2w:

$$S_F = (1- \alpha[2w])S_O + \alpha[2w]x \quad (6)$$

Equating the coefficients of S_O and of x in equations (5) and (6) appears to give rise to two conditions that have to be satisfied.

For the coefficients of S_O in equations (5) and (6) to be the same:

$$(1- \alpha[2w]) = (1- \alpha[w])^2 \quad (7)$$

For the coefficients of x in equations (5) and (6) to be the same:

$$\alpha[2w] = \alpha[w](2- \alpha[w]) \quad (8)$$

We see that these two conditions are in fact identical, both being equivalent to:

$$\alpha[2w] = 1 - (1- \alpha[w])^2 \quad (9)$$

By induction, the condition can be extended to:

$$\alpha[nw] = 1 - (1- \alpha[w])^n \quad (10)$$

If w = 1, unit tonnage, then:

$$\alpha[W] = 1 - (1- \alpha[1])^W \quad (11)$$

Equation (11) has the satisfactory properties that $\alpha[0]$ is zero, and also that $\alpha[W]$ tends to 1 as W becomes very large.

5 How large should alpha be?

We have seen that alpha for an observation of tonnage W should be a monotonically increasing function of the tonnage W, and of $\alpha[1]$, the alpha for unit tonnage.

The question remains as to the appropriate choice for $\alpha[1]$. Clearly, this must depend upon the purpose for which the exponentially smoothed grade or other variable is being monitored.

In the control system discussed by Everett [4], ore was selected for each shift so that the expected grade of the selected ore, exponentially smoothed into the shift history, gave a grade on target. The ore was being blended onto stockpiles of 200 kilotonnes. So if a Moving Average (MA) were being used, it would be appropriate average over a tonnage T = 200 kt, as in equation (4), so the averaging weight applied to each kilotonne is 1/T.

For Exponential Smoothing, the choice of $\alpha[1]$ is not so clear cut. One criterion is to consider the average "age" of the sample. For a moving average, or for a completed stockpile of tonnage T, the average age is T/2. For an exponentially smoothed average to have the same average age of sample:

$$T/2 = {}_{m=0}\Sigma^{Infinity} m\alpha[1](1-\alpha[1])^m \tag{12}$$

$$= (1-\alpha[1])/\alpha[1] \tag{13}$$

$$\alpha[1] = 2/(2+T) \approx 2/T \tag{14}$$

So the starting weight for an EWMA should be about twice that of an equivalent MA, as shown in Figure 7:

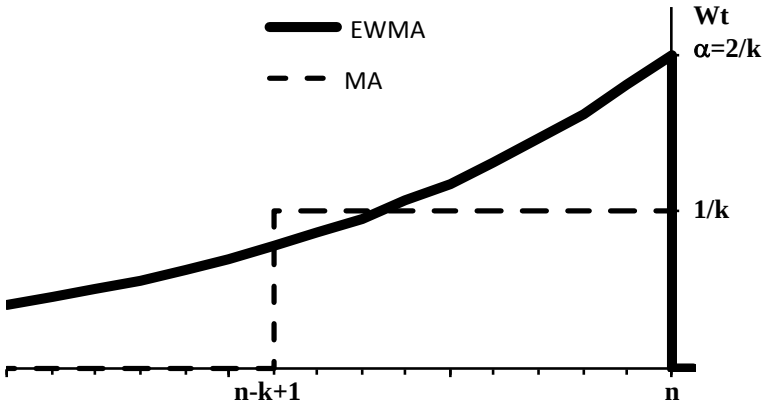


Figure 7: Equivalent EWMA and MA weights applied to recent data.

In a production process, such as a mining operation, T would be the subsequent blending tonnage, achieved either by blending directly onto stockpiles or inherent in the processing and transportation system.

6 Conclusions

By considering both the time (or distance) domain and the frequency domain, this paper has shown that Exponential Smoothing (EWMA) has considerable advantages over Moving Averages (MA).

The problem of varying sample sizes has been considered, and we have shown that the appropriate exponential smoothing factor for a sample of size w is given by equation [11], $\alpha[W] = 1 - (1- \alpha[1])^W$, where $\alpha[1]$ is the exponential smoothing factor to be applied to samples of unit weight.

We have further shown, in equation (14), that $\alpha[1]$ should be approximately 2/T, where T is the comparable MA tonnage, or the blending tonnage in a production process.

References

[1] Diebold, F.X. *Elements of Forecasting.* Fourth ed. Mason, OH: South-Western, 2008.
[2] Box, G. & Jenkins, G. *Times Series Analysis: Forecasting and Control.* San Francisco, CA: Holden-Day, 1970.
[3] Ramjee, R., Crato, N. & Ray, B.K. A note on moving average forecasts of long memory processes with an application to quality control. *International Journal of Forecasting*,**18,** pp. 291–297, 2002.
[4] Everett, J.E. Computer aids for production systems management in iron ore Mining. *International Journal of Production Economics*,**110/1**, pp. 213-223, 2007.
[5] Marks R.J. Handbook of Fourier Analysis and Its Applications. Oxford University Press, 2009.
[6] Cooley, J.W. & Tukey, J.W. An algorithm for the machine calculation of complex Fourier series, *Mathematics of Computation*,**19,** pp. 297–301, 1965.

This page intentionally left blank

Experimental and analytical study on high-speed fracture phenomena and mechanism of glass

H. Sakamoto[1], S. Kawabe[1], Y. Ohbuchi[1] & S. Itoh[2]
[1]*Department of Mechanical Engineering and Materials Science, Kumamoto University, Japan*
[2]*Okinawa National College of Technology, Japan*

Abstract

The new high-speed crushing technique of glass bottles was proposed for recycling. The proposed system uses the underwater shock-wave by explosive energy. This method is excellent compared with the conventional mechanical method in the crushing time, crushing efficiency, collection ratio of glass "cullet" and using simple crushing apparatus etc. In this study, using commercial beer bottle, the behaviors of underwater shock-wave by explosive energy and the high-speed fracture phenomena of glass bottle were clarified by the experiment and analytical method.
Keywords: glass-cullet, underwater shockwave, explosive enemy, high-speed fracture, recycle.

1 Introduction

Glass bottles are widely used as containers of drinks, food and medicine due to their characteristics for sealing up, transparency and storage stability. Many of these glass bottles after use are reused as returnable bottles or recycled as the raw material of glass container which is called "cullet' that are crushed to small fragments [1, 2]. The authors paid attention to this raw material recycling process of generating "cullet". In the conventional cullet generation method, a mechanical crushing one is used [3]. In order to recycle them by using this method, these bottles need to be washed inside before melting. As the bottles' shapes vary greatly, this washing of the bottles' inside takes a lot of time and it is


WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110021

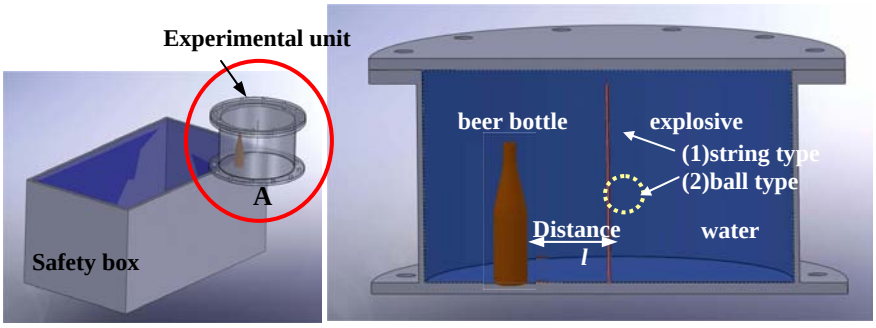
difficult to do this operation automatically. In addition, the noise of this process is large and some of glass fragments also disperse.

Then, the new cullet generation technique which used the underwater shock-wave by explosive energy was proposed for overcoming these problems. This proposed technique has a simultaneous crushing and pre-washing function. In this report, the behavior of underwater shock-wave by explosive energy and the high-speed fracture phenomena of glass bottles were clarified by the experiment and analytical method for the recycling of glass containers.

2 Experiment

2.1 Experimental method

The schemes of experimental apparatus are shown in Fig.1. The experiment was done in an explosion-proof steel container, which had been filled with water. The glass bottles were given underwater shockwave by two kinds of explosive, string type (1) and ball type (2) shown in Fig.1. The glass fragments called 'cullet' were all collected and the weight of each fragment size was measured after the explosive experiment.



(a) Explosion proof container (b) Detail A

Figure 1: Scheme of experimental apparatus.

2.2 Specimens and experimental conditions

We used a commercial beer bottle (500 cc, H=250 mm, D=72 mm, 430 g) as the experimental glass bottle (shown in fig. 2). The inside of the bottle was charged with air and the bottle and explosive were set (as shown in fig.1).

The distance l shown in fig. 1 of the explosive and the bottle in order to examine the effect on the fragment grain sizes have been changed. The six distances from explosive to bottle were determined as follows: l = 100 mm, 50 mm, 40 mm, 30 mm, 20 mm and 10 mm.

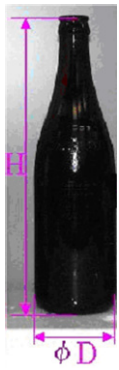


Figure 2: Specimen (beer bottle).

The PETN(explosive rate: 6308 m/s) as explosive and the electric detonator as ignition were used and explosive shape.

2.3 Observation of high-speed fracture process

In order to visualize the behavior of underwater shock-wave and the bottle fracture process, a high-speed photograph system was used, which consisted of a high-speed camera (IMECOM 468) and two Xenon flashlights (H1/20/50 type). The optical observation system is shown in fig.3

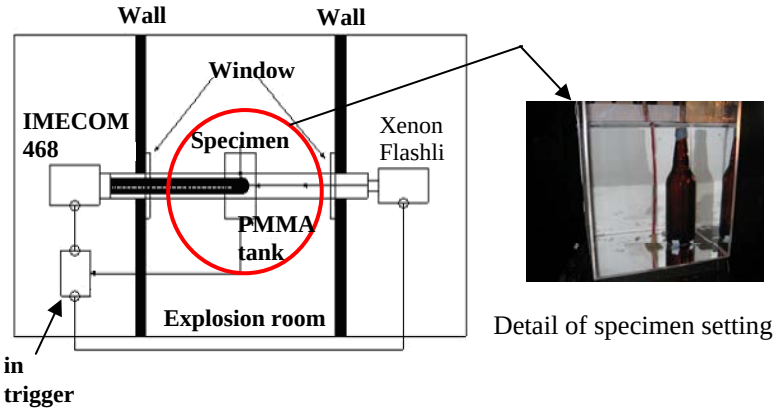


Figure 3: Optical observation system.

3 Experimental results

3.1 Cullet distribution

The fragments of glass bottle were collected and these 'cullets' were classified by three kinds of sieves (4.75 mm, 2 mm, 1 mm) and each weight was measured.

The weight ratio in the case of string type explosive is shown in fig.4.

From fig.4, it is found that the weight of 1 mm or less cullet size decreases as the distance increases and the weight ratio of 1 mm or less grain size increases as the distance decreases. Here, it is interesting that the weight ratio of the 1〜2 mm cullet size is almost constant regardless of the distance *l*.

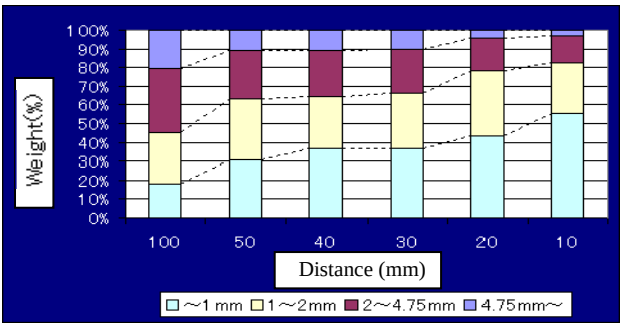


Figure 4: Effect of distance between the explosive and the bottle on cullet size (String type explosive case).

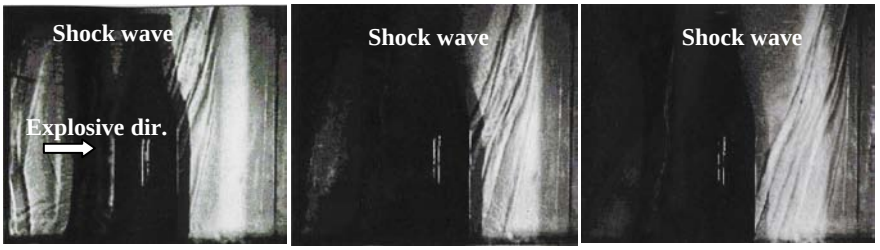
3.2 Fracture process observation

The glass bottle was charged the underwater shock wave by explosive shown in Fig. 1. Here, the behaviors of underwater shock-wave by explosive energy and the high-speed fracture phenomena of glass bottle were observed by a high-speed photography method. The photograph was taken using an Image Converter High-speed camera (IMECOM468) and Xenon flashlight (H1/20/50type) as lighting.

Figure 5 shows the behaviors of shock wave in the case of string type explosive. Figure 6 shows the one in the case of ball type explosive. The appearance of the shock wave propagation can be clearly observed from these figures.

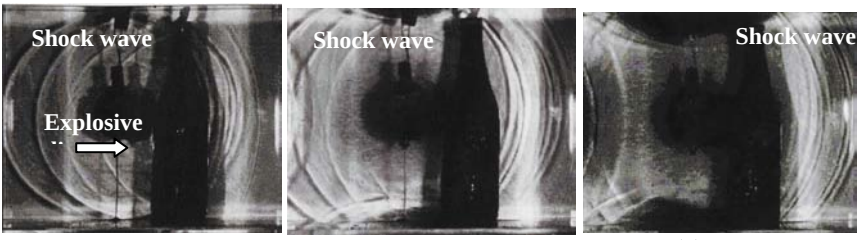
From figs. 5 and 6, the appearances of shock -wave propagation in the case of string type and ball type were illustrated in fig.7 (a), (b), respectively.

Next, we show some examples of fracture framing photographs. Figure 8 shows the fracture process in the case of string type explosive. In order to clarify the appearance of the deformation and fracture process, white paints are painted on the bottle surface. After the shock-wave propagated, the generation of the explosion-gas was seen from the photographs.



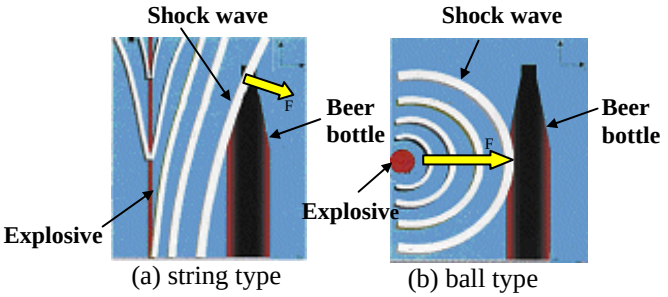
(a) 80μs (b) 100μs (c) 120μs

Figure 5: Shock-wave propagation (string type explosive).



(a) 80μs (b) 100μs (c) 120μs

Figure 6: Shock-wave propagation (ball type explosive).



(a) string type (b) ball type

Figure 7: Illustration of shock-wave propagation.

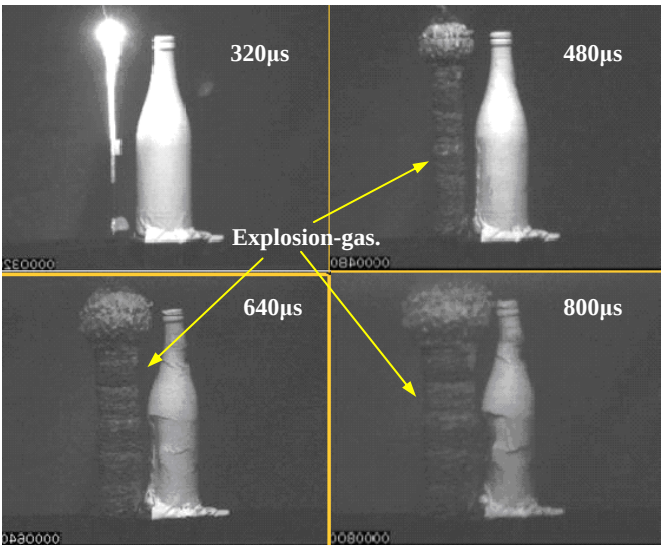
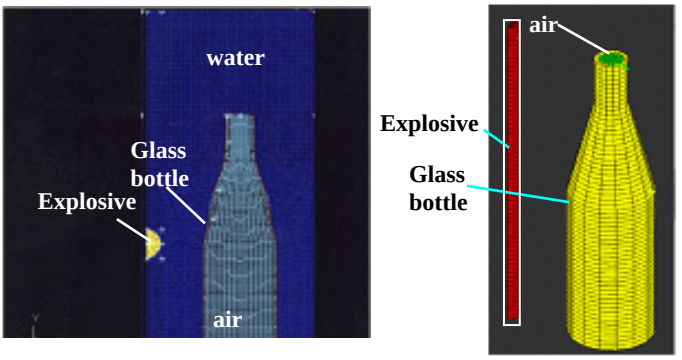


Figure 8: The behaviors by underwater shock wave.

4 FEM simulation

The fracture behaviors of bottles submitted to underwater shock-wave was simulated using the FEM analysis (FEM code: LS-DYNA) [4].

In this simulation, 2-D model (shock-wave propagation analysis) and 3D-model (shock-wave propagation analysis and fracture analysis) were employed. 2-D and 3-D analysis models are shown in fig.9 (a) and (b), respectively.



(a) 2-D analysis model (b) 3-D analysis model

Figure 9: The behaviors by underwater shock wave.

The 2-D simulation result is shown in fig.10. The behavior of shock-wave propagation agrees well with experimental results shown in fig.5.

Next, 3-D simulation results are shown. Figure 11 shows the fracture process of glass bottle by string type explosive. These simulation results are corresponding to the observation ones obtained by high-speed photography method.

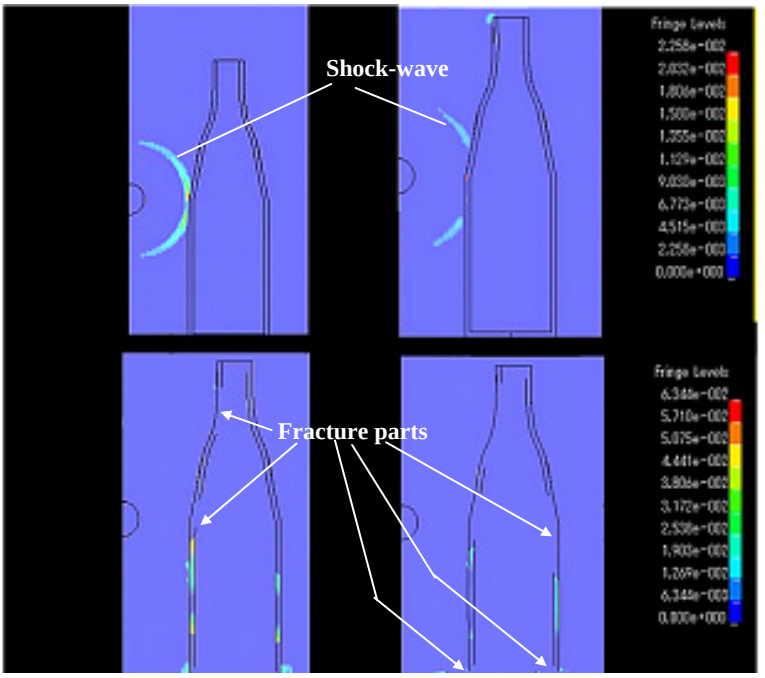


Figure 10: The behaviors of underwater shock wave by 2-D FEM simulation.

Figure 11: The fracture process by underwater shock wave (3-D FEM simulation in the case of string type explosive).

5 Conclusions

The relation between strength of underwater shockwave and fracture “cullet” grain size were discussed and the behaviors of shock-wave and fracture process by shock-wave were discussed. The results obtained are summarized as follows:

1) The weight ratio of the small ‘cullet’ grain sizes increases as the distance decrease and the weight ratio of the 1～2mm cullet size is almost constant regardless of the distance *l*.
2) The behaviors of shock-wave generated by explosive energy and the high-speed fracture process were clarified by using a high-speed photograph method and FEM simulation.

References

[1] Sakka, S., *The dictionary of glass, Asakura Press Ltd*., 1998.
[2] Kobayashi, A.S., *Experimental Techniques in Fracture Mechnics*, Society for Experimental Stress Analysis ,1973.
[3] Sakamoto, H., et al, WIT transactions on Modelling and Simulation, Vol.41, pp.497-503,2005.
[4] V.P.W. Shim, S. Tanimura, C.T. Lim, *Impact Response of Materials & Structure*, Oxford University Press, 1999.

Multiscale multifunctional progressive fracture of composite structures

C. C. Chamis[1] & L. Minnetyan[2]
[1]*NASA Glenn Research Center, Cleveland, USA*
[2]*Clarkson University, Potsdam, NY, USA*

Abstract

A new approach is described for evaluating fracture in composite structures. This approach is independent of classical fracture mechanics parameters like fracture toughness. It relies on computational simulation and is programmed in a stand-alone integrated computer code. It is multiscale, multifunctional because it includes composite mechanics for the composite behavior and finite element analysis for predicting the structural response. It contains seven modules; layered composite mechanics (micro, macro, laminate), finite element, updating scheme, local fracture, global fracture, stress based failure modes, and fracture progression. The computer code is called CODSTRAN (Composite Durability Structural ANalysis). It is used in the present paper to evaluate the global fracture of four composite shell problems and one composite built-up structure. Results show that the composite shells. Global fracture is enhanced when internal pressure is combined with shear loads.
Keywords: micro mechanics, laminate theory, thin shells, thick shells, built-up structures, non-linearities.

1 Introduction

The global fracture behavior of fiber composite structures has become of increasing interest in recent years, because of the multitude of benefits that composites offer in practical engineering applications such as lightweight airframes, engine structures, space structures, marine and other transportation structures, high-precision machinery, and structural members in robotic manipulators. Composite structures lend themselves to tailoring to achieve desirable characteristics such as a high strength to weight ratio, dimensional

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110031

stability under extreme thermal and hygral fluctuations, and the capability to allow controlled detectability such as in the Stealth technology. Because of the numerous possibilities with material combinations, composite geometry, ply orientations, and loading conditions, it is essential to have a reliable computational capability to predict the behavior of composites under any loading, geometry, composite material combination, and boundary conditions. A computational capability is also essential to design effective experiments for the further development of composite micromechanics theories, and to utilize existing experimental results in the most productive manner. In summary, the development of reliable computational simulation methods is necessary for the commercial maturation of composites technology.

The behavior of composites during progressive fracture has been investigated both experimentally and by computational simulation [1]. Recent additions to the computational simulation have enabled monitoring the variations in structural properties such as natural frequencies, vibration mode shapes, and buckling modes during progressive fracture [2]. Existing computational capabilities in the simulation of structural damage and fracture of composite structures have been implemented in the CODSTRAN (COmposite Durability STRuctural Analysis) computer program [3]. The ICAN (Integrated Composite ANalyzer) and MHOST computer codes [4–6] are coupled to form CODSTRAN. The description herein is mainly to show what can be done by progressive structural fracture. Details cannot be included because of space limitations by conference proceedings. However, references are cited for the interested readers.

2 Fundamental concept

It is instructive to briefly describe the fundamental concepts on the origin of CODSTRAN and the related concepts. The most obvious one is that classical

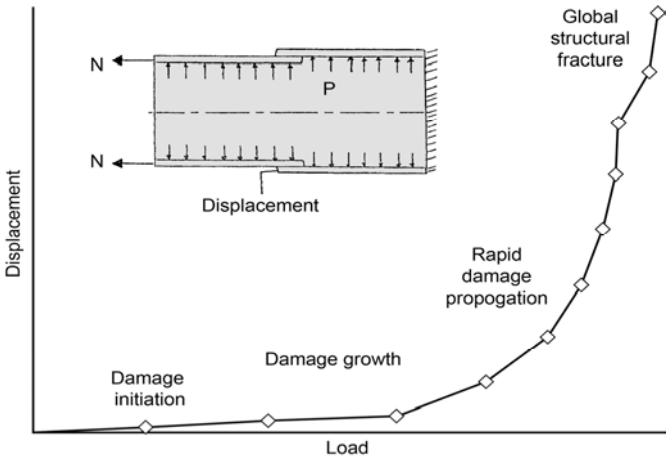


Figure 1: Overall CODSTRAN simulation.

fracture mechanics are not applicable to composite structural fracture. A physical consideration on how a structure will fracture is very important in describing any new approach.

It starts with the physical concept that a structure does not fail instantaneously, but requires some time for local damage to accumulate prior to structural fracture. Then the following must be happening during the process of damage accumulation. The process is depicted schematically in fig. 1. With fig. 1 as a background, this new approach to damage tolerance, structural fracture and risk is based on the following concepts:

1. Any structure or structural component can tolerate a certain amount of damage propagation (damage tolerance) prior to global structural fracture.
2. During damage accumulation, the structure exhibits progressive degradation of structural integrity (damage tolerance) as measured by global structural response variables such as loss in frequency, loss in buckling resistance or excessive displacements.
3. The critical damage can be characterized as the amount of damage beyond which the structural integrity and damage tolerance degradation is very rapid, induced by either (1) small additional damage or (2) small loading increase.
4. Structural damage tolerance degradation is characterized by the following sequential stages: (1) initiation, (2) growth, (3) accumulation, (4) stable or slow propagation (up to critical stage), and (5) unstable or very rapid propagation (beyond the critical stage) to collapse.

The global response variables are accurately predicted by finite element structural analysis [6]. The composite degradation is simulated by composite mechanics in ICAN [4]. The progressive degradation is predicted by an incremental approach with damage tracking and the corresponding stress redistribution. All these are included in CODSTRAN as is depicted schematically in fig. 2 for the CODSTRAN computational simulation of progressive fracture. It is interesting to note that at the bottom of fig. 2 the constituent composite properties are degraded based on environmental and load history effects.

2.1 Approach

The method of solution is depicted schematically in fig. 2. It starts at the micromechanics scale as is illustrated in fig. 3 where the local stresses and strains are described by appropriate equations programmed in [5]. Then, it continues to micromechanics of coupling a single fiber with a matrix with appropriate equation also programmed in [5]. It progresses to a mono fiber laminate; to a multiply laminate; to a finite element model of the structure; to structural scale that includes the boundary conditions, the environmental conditions, the loading conditions, and also the structural finite element mode [7].

The scales modeled span from constituent scale to structural scale which definitely is multiscale. The left side of fig. 2 is “called” the composite synthesis

to the single finite element node. The ICAN code is run for each ply in the laminate and saved so that each finite element node has its own ICAN code run to expedite decomposition. These ICAN runs are saved for the downward composite structural decomposition as noted in the right side of fig. 2.

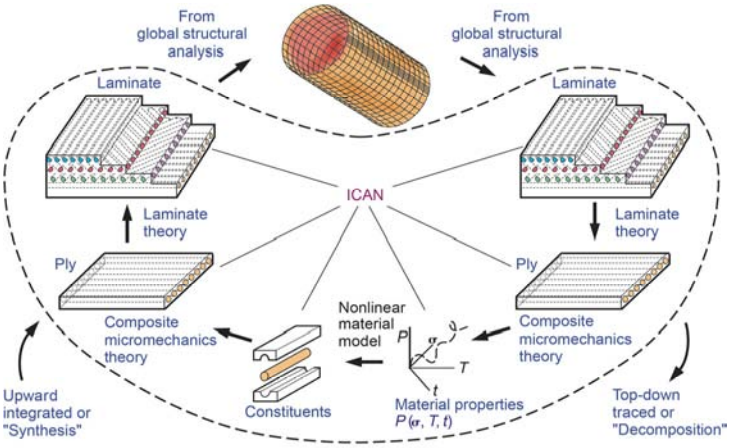


Figure 2: CODSTRAN simulation cycle.

Once the finite element solution of the first load increment with internal forces/displacements has been obtained at each node then the downward decomposition starts. It is noted that the finite element solution requires nodal information because it is computationally more expedient for the composite decomposition to be performed [8]. Then the decomposition is repeated by using the ply information stored in the synthesis process. The mono ply stresses/strains are then evaluated by using the schematic in fig. 3 where the local failures are identified. If any failures occurred at this level, the respective stiffness and fractured region are eliminated for the second simulation. The process continues until local incremental convergence has occurred. At this point the load is increased by the second increment. Loading increments are progressively larger at the beginning until local fracture is detected. Then the load increment is reverted back to the last increment and is progressively halved until convergence is achieved and the next load increment is applied by a value equal to the previous load increment. Fig. 4 illustrates this concept. Therefore, the solution is incremental from the micromechanics scale to the structural local/global convergent scale. The structural dynamics equations solved by the finite element in CODSTRAN, which have global variable convergence criteria, are summarized in the chart as is depicted in fig. 5.

These equations are solved at each load increment. There is another level of ply failure criteria. This is a stress failure criterion with a combined stress failure

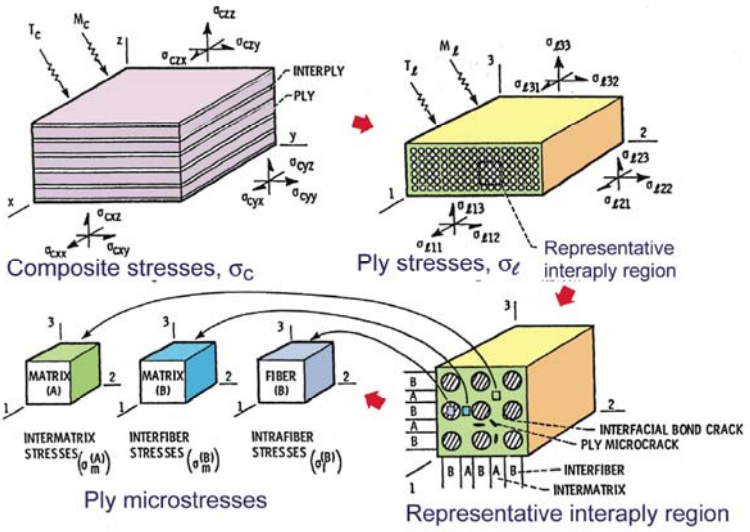


Figure 3: Ply micro-stresses through composite stress progressive decomposition.

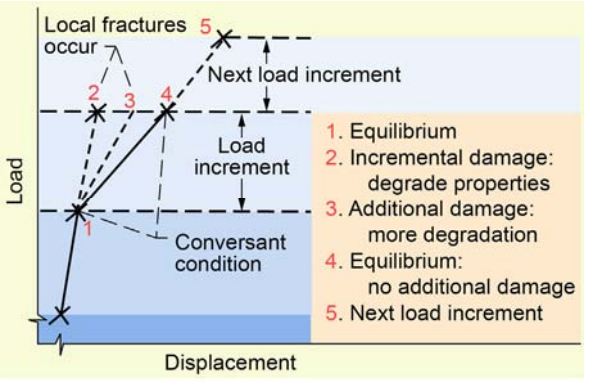


Figure 4: CODSTRAN load incrementation.

function as depicted in fig. 6. The combined stress failure criterion is applied first. Then ply dominant stress is identified and that stress is used in the constituents to identify which region has failed. Therefore, the solution is robust and quite reliable as will be illustrated subsequently by the solution results of the sample cases.

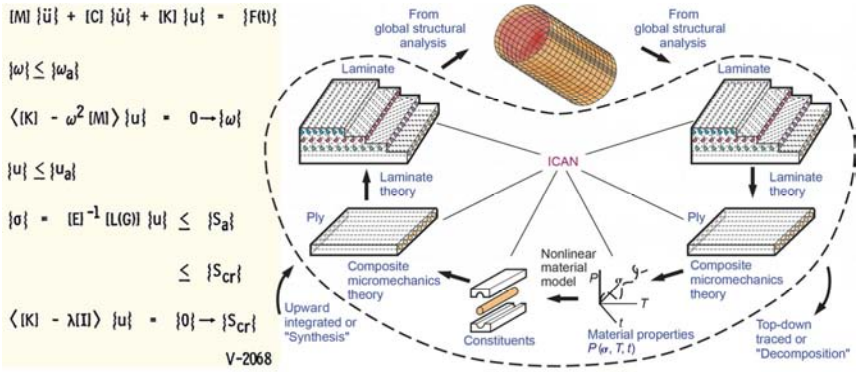


Figure 5: Structural behavior/response governing equations.

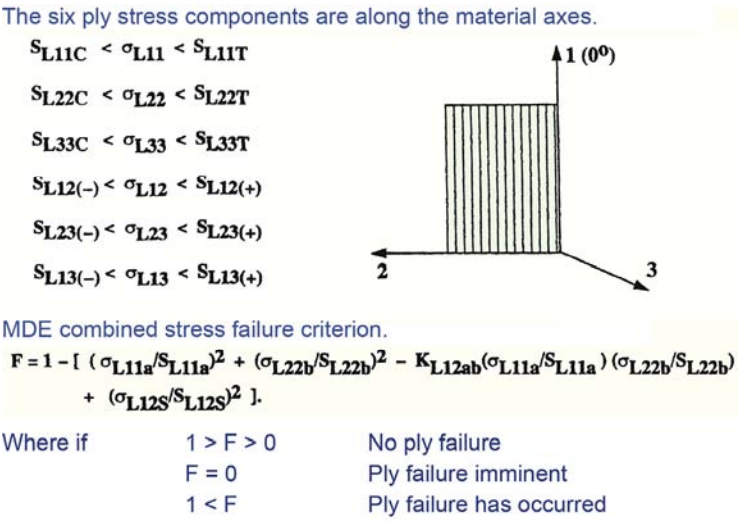


Figure 6: Play failure criteria, maximum stress failure criterion.

3 Sample cases results and discussions

3.1 Composite shell with defects

The first illustrative sample problem is a composite shell subjected to internal pressure and having structural conditions and a through-the-thickness longitudinal defect, as shown in fig. 7(a) [8]. The type of composite and the

laminate configuration of the shell are also shown in the title of fig. 7. Additional details are described in [8]. The environmental conditions are noted in the small table insert to the right of the shell. The results are plotted pressure versus damage percent in part (b) top right. Third vibration frequency versus pressure part (c), down left; and third vibration frequency versus damage percent (d), down right.

Each plot has six different curves, one each for environmental effects. The very top curve (○) is with no environmental effects. The second from the top curve (□) is room temperature and one-percent moisture content by volume. The third from the top curve (Δ) is for the temperature 99.9ºC (200°F). The fourth from the top curve (◊) represents the combined temperature moisture effects 200°F with one-percent moisture by volume. The fifth from the top curve (⌒) is for temperature 149ºC (300 °F) only. The last curve (∇) is for the combined environmental effects 148.9ºC (300 °F) with one-percent by volume moisture. Note that the 148.9ºC (300°F) temperature only curve shows the second greatest failure pressure. The reason is that that shell has the lowest residual stress that counteracts the temperature degradation effects. The important point to observe in these results is that the environmental effects have substantial structural integrity degradation effects. The curves plotted in fig. 7(c) show the significant degradation on the third vibration frequency. The structural degradation effects are also significant when the third vibration frequency is plotted versus damage percent.

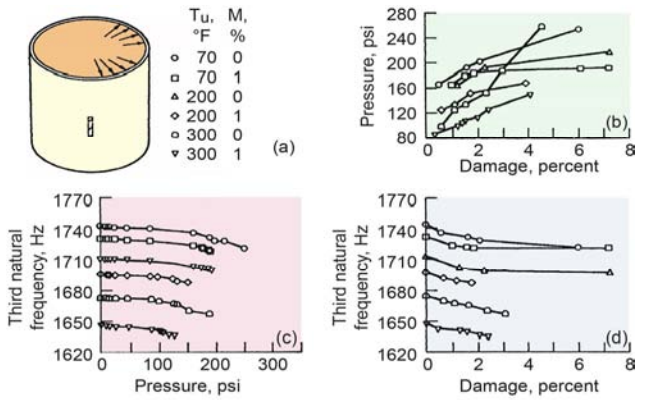


Figure 7: Hygrothermal effects – T300/epoxy $[90 \pm 15]_S$ shell with fiber volume ration 0.6 (a) Geometry and environment. (b) Pressure. (c) Vibration frequency. (d) Buckling load (1 psi = 6.9 Pa).

3.2 Composite shell with defects and external pressure

The second illustrative example is a composite cylindrical shell under external pressure with no defects, and with two types of defects—surface plies, and mid-thickness, fig. 8 [9]. The shell composite type and laminate configuration are shown at the bottom of the figure as well as defect sizes. The shell with the

surface defects showed limited damage tolerance of about 149 Pa (50 psi). The shell with the mid-thickness defects have rather insignificant effect compared to defect free shell which exhibited no damage tolerance at all. The important conclusion from the evaluation of this shell is that only surface defects show a limited damage tolerance when the shell is subjected to external pressure. The reason is that compression has an enhancement in small defects. Additional details for this composite shell are described in [9].

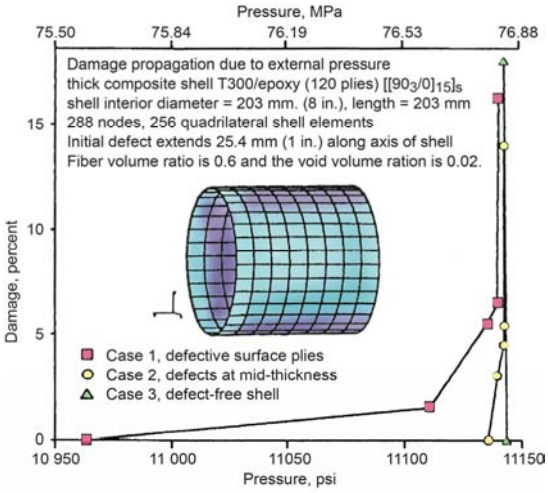


Figure 8: Thick shell fracture under external pressure.

3.3 Composite thick shell

The third illustrative example is a thick shell subjected to external pressure and with three types of defects as shown in fig. 9, where the composite type, laminate configuration and finite element model are also shown [10, 11]. The location and shape of the defects are shown in fig. 10. Results obtained are shown in fig. 11. This figure is very important because it shows the damage initiation and fracture of the defect free and defected shell. It is interesting to note the ply lay-up of the shell thickness in fig. 10. The damage initiation is shown as ratios of the defect free shell which is unity (1.0).

The damage initiation of the defect free shell is 0.84. This value indicates that the shell has16 percent damage tolerance from initiation to global fracture. The mid-thickness defects exhibited a 0.75 initiation and 0.77 global fracture. The damage tolerance of the shell with the mid-thickness defects had almost no damage tolerance. That is the shell exhibited a rather brittle behavior. The shell with the inner surface defects had a 0.45 damage initiation and 0.85 global fracture. This shell had the greatest damage tolerance of 40 percent from its initial damage to its global fracture. For additional details see [11].

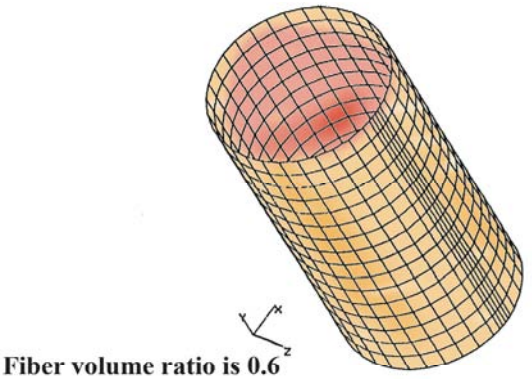


Figure 9: Shell structure evaluated (1 in.= 1.5 cm).

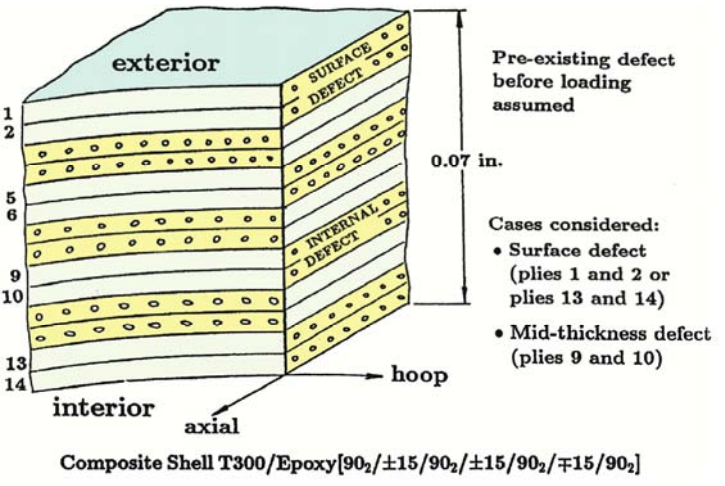


Figure 10: Shell laminate structure schematic indicating initial defects (1 in. = 1.5 cm).

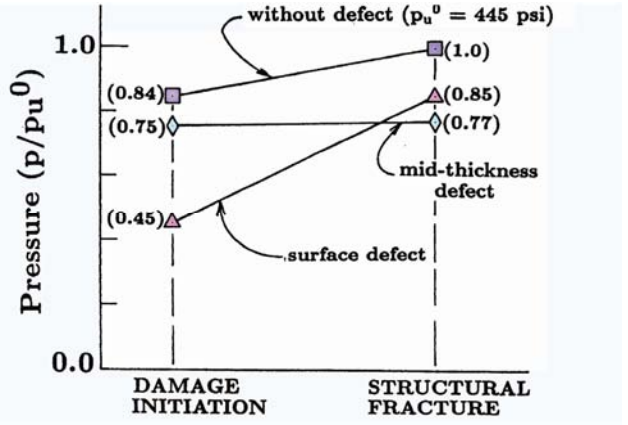


Figure 11: Summary of results. Composite shell T300/ Epoxy [90_2/ $\pm 15/90_2$ / $\pm 15/90_2$ / $\pm 15/90_2$.

3.4 Composite thick shell global degradation

The fourth illustrative example is another thick shell under external pressure as shown in fig. 12 where the shell composite system, laminate configuration, finite element model are also shown. This shell was analyzed for degradation in the frequency and the buckling load as the damage propagated along the longitudinal direction as shown in fig. 13. The buckling load did not degrade until the damage length was about 42cm (28 in). long. After that the buckling load degradation was relatively great with global collapse of about 80 psi down from 2.3KPa (340psi0 or a degradation of about 1.8KPa (260psi). This is a very interesting result because it indicates the buckling of a composite thick shell has a relative large damage tolerance with respect to buckling resistance. Though frequency degradations are not shown here, these degrade slower than the buckling load. Additional results are described in [11].

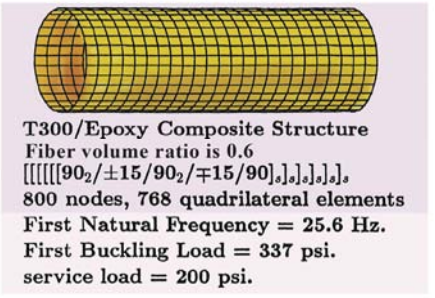


Figure 12: Composite shell: 576 plies (2.88 in. thick) diameter = 15 ft; length = 50 ft. (1 psi = 6.9 Pa; 1ft = 18 cm).

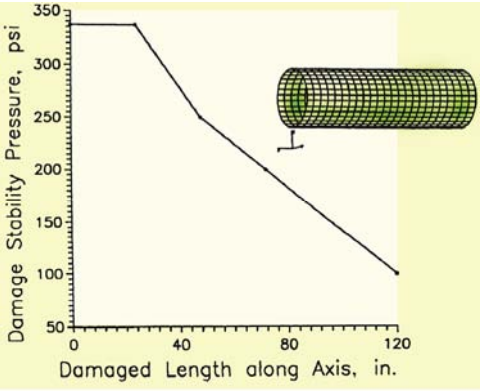


Figure 13: Damage tolerance for composite shell (1 psi = 6.9 Pa).

4 Concluding remarks

Up-dated computational simulation is one direct approach to evaluate fracture in composite structures. CODSTRAN is a stand-alone general purpose integrated multiscale/multifunctional computer code which consists of several modules including micromechanics through structural analysis. It is applicable to general classes of composite structures. Composite shells fracture investigated herein included defect free shells and shells with defects. The simulation results presented are from the microscale to global structural fracture. A built-up composite structure subjected to combined loads was evaluated from micromechanics fracture to global fracture. Results from all of the above problems indicate that shear load combined with tension or compression stabilize the solution as shown by the greater damage sustained at global structural fracture. Embedded defects have no influence in the global shell fracture when the shell is subjected to internal pressure.

Acknowledgement

The authors express their sincere appreciation to Dr. Subodh Mital whose review comments improved the readability of the article.

References

[1] Chamis, C. C. and Sinclair, J. H., "Dynamic Response of Damaged Angleplied Fiber Composites." NASA TM–79281, 1979.
[2] Minnetyan, L., Chamis, C. C. and Murthy, P. L. N., "Structural Behavior of Composites with Progressive Fracture," *Journal of Reinforced Plastics and Composites*, Vol. 11, No. 4, April 1992, pp. 413–442.
[3] Chamis, C. C. and Smith, G. T., "Composite Durability Structural Analysis." NASA TM–79070, 1978.

[4] Murthy, P. L. N. and Chamis, C. C., "Integrated Composite Analyzer (ICAN): Users and Programmers Manual." NASA Technical Paper 2515, March 1986.
[5] Murthy, P. L. N. and Chamis, C. C., "ICAN: Integrated Composites Analyzer," *Composite Technology Review*, Spring 1986.
[6] Nakazawa, S., Dias, J. B. and Spiegel, M. S., *MHOST Users' Manual*, prepared for NASA Glenn Research Center by MARC Analysis Research Corporation, April 1987.
[7] Irvine, T. B. and Ginty, C. A., "Progressive Fracture of Fiber Composites." NASA TM–83701, 1983.
[8] Minnetyan, L, Murthy, P.L.N. and Chamis, C.C., "Progressive Fracture in Composites Subjected to Hygrothermal Environment," *International Journal of Damage Mechanics.* Vol. 1, pp. 60–79, 1992.
[9] Minnetyan, L., Chamis, C.C. and Murthy, P.L.N., "Damage Tolerance of Thick Composite Shells Under External Pressure." Naval Air Warfare Center, Aircraft Division, Report No. NAWCADWAR-94096-60, Vol. II, pp. 85–95, April 1994.
[10] Minnetyan, L., Clarkson University. "Progressive Fracture of Composite Structures." NASA CR–210974, July 2001.
[11] Minnetyan, L. and Chamis, C.C., "Progressive Fracture of Composite Cylindrical Shells Subjected to External Pressure." *Journal of Composite Technology & Research.* Vol. 19, No. 2, pp. 65–71, 1997.

Enhancing simulation in complex systems

R. M. Alqirem
Department of Management Information Systems,
Al-Zaytoonah University of Jordan, Jordan

Abstract

Simulation is an important tool in understanding and designing physical systems, engineering systems and social systems. Because of its importance and broad range, it has been the subject of numerous research studies and books. Simulation is about techniques for using computers to imitate (simulate) the operations of various kinds of real world complex systems**.** It has been an accepted tool for the improvement of decision making through learning how to deal with the complexity of the real world. The complexity slows the learning loop and reduces the learning gained on each cycle.

This paper illustrates the importance of system thinking in enhancing the simulation process and providing the ability to see the world as a complex system, where "you cannot just do one thing" and that "everything is connected to everything else". It is a holistic worldview that enables people to act in consonance with the best interests of the system as a whole and thus enhance the learning loop through various system thinking tools. The case study in this paper illustrated the use of a system dynamics simulator to allow the financial manager in a firm to test different account receivables scenarios and the strategies to control these accounts. We found that this simulator has helped the manager to get a deeper insight into the effect of their decisions and the different interrelated variables that involve with setting a strategy to control account receivables.

Keywords: systems thinking, complex systems, systems dynamics, simulation.

1 Introduction

As the world becomes more complex, many people and organisations find themselves bombarded with lots of problems to solve, less time to solve them, and very few chances to learn from their mistakes. Managers will need to deal with complexity and with these changes. Also they need to develop their new

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110041

capabilities to be able to create an effective learning process in complex dynamic systems to overcome the different barriers to learning which are created by complex dynamics systems, and thus to get a significant insight into their simulated problems. This can be done by allowing the managers to develop new ways of thinking that allows deeply understanding of the simulation process, the relations between the variable, and the consequences of their experiments more effectively.

Lots of philosophers, scientist and managers have been calling for a fundamental new ways of thinking that improve the ability to see the world as a complex system. As Sterman [1] stated, "a system in which we understand that you can't just do one thing and that everything else is connected to everything else". He argued that it's crucial to develop new ways of system thinking. He states that "if people had a holistic worldview, they would then act in consonance with the long term best interest of the system as a whole, identify the high leverage points in systems, and avoid policy resistance".

This can be done by using new simulation methods that increase the capabilities of managers to experiment their decisions more effectively by using system dynamics simulation tools.

2 Systems thinking

The systems thinking concept was produced as the world of systems found that there is a need to shift from the more linear, analytic way of thinking that many of us are used to, to a non linear, dynamic and holistic thinking.

Moving to new the Paradigm in analysing complex problems enables the managers and the analyst to understand dynamic relationships and complexity that influence the behaviour of a system as shown below

2.1 Features of systems thinking

2.1.1 Dynamic and non-linear thinking

As discussed before, the static thinking assumes that causality runs only one way and any system's factors are independent which is quite primitive. Dynamic thinking offers effective alternatives to see and understand systems or problems. This creative thinking allows viewing the world with ongoing, interdependent relations, dynamic process. Each of the causes in the dynamic thinking is linked in a circular process to both the effect and to each of other causes. These circular processes are the feedback loops which enable us to better understand what is going on in the system; these circular loops represent a non-linear and dynamic thinking (Richmond [2]).

Taking into consideration this type of thinking, the analyst or the manager can understand the problem in a better way as the feedback process inside a firm clarifies the dynamic relations inside a firm, analysing the causes and effects and their interconnection and allows for observing the behaviour over time.

For example, if a firm decreases its product's price, this decision has an effect on the sales as it increases the sales, but on the other hand the firm's profits will

be less than usual, which affects the firm's pricing policy and push the firm to increase the prices.

2.1.2 Holistic thinking

Holistic thinking is one of the most significant features of systems thinking as it allows us to see the "Big Picture". So instead of examining each part of the system, the whole system is examined. Whatever the problem we are experiencing and searching for its source, we must always widen our focus to include that bigger system. Dealing with the wholes rather than parts is a very effective idea in system analysis. Each part or department in a firm is not isolated from other department, so trying to solve a problem in one process; we must first look the whole firm and the interconnections inside it to understand the nature and the reasons for such problem.

This research illustrated how systems thinking tools provides managers and analysts with a creative holism

2.1.3 Systemic thinking

In recent years, systems thinking has provided new effective methods for tackling issues in a systemic than a reductionist way. Systems thinking allow us to look for various patterns of behaviour, to seek underlying systemic interrelationships which are responsible for these types of behaviour and events.

A recent study by Bartlett [3] defines systemic thinking as a technique that provides a deeper insight into complex situation very quickly. It stated that Systemic thinking "combines analytical thinking (breaking things apart) with synthetical thinking (putting things together)" as the next figure shows.

This provides more effective holistic and systemic analysis of the system.

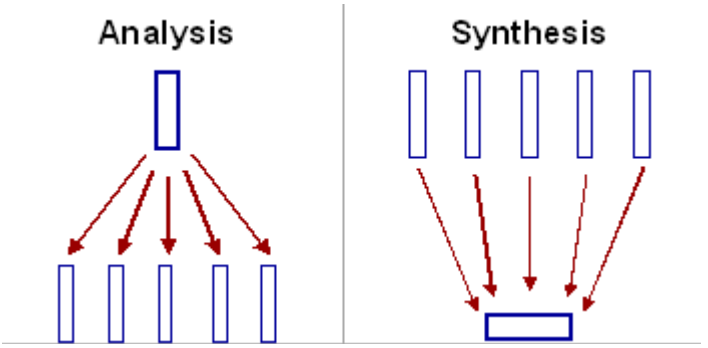


Figure 1: Analysis and synthesis (Bartlett [3]).

2.2 Systems thinking approaches

Systems thinking approaches can be divided into two categories, hard systems and soft system. Hard systems thinking is a generic title defined by Checkland [4] to various types of systems approaches for solving problems in the real world. The approaches related to hard systems are operational research, system

analysis and system dynamics. It is a useful way of thinking which gives the managers and the decision makers the ability to optimize and control the system's performance in pursuit of clearly identified goals.

Another approach is introduced to help to decide what purpose the organisation should pursue which is soft systems methodology and cybernetics.

In this paper, we will introduce the System Dynamics methodology.

2.2.1 System Dynamics

System Dynamics methodology is considered an important approach which has a sophisticated tools such as Virtual world (formal models, Microworld, management flight simulators, computer simulation) in which decision makers and managers can refresh decision making skills, test their scenarios and strategies, and conduct experiments through effective simulation by using system Dynamics simulation tools.

System Dynamics is a computer-aided method for analysing, framing, understanding, solving problems in complex real world systems. Jay W. Forrester's work at the Massachusetts Institute of Technology was the major move towards developing this field to extend the range of applied system thinking to more strategic problems in complex systems.

System Dynamics approach uses a perspective based on information feedback and delays to understand the dynamic behaviour of complex physical, biological, and social systems. It also helps the decision maker untangle the complexity of the connections between various policy variables by providing a new language and set of tools to describe. It even models the cause and effect relationships among these variables [1, 7, 8].

Furthermore, System Dynamics method enables the decision makers or the modeller via its tools in any system to identify the underlying structure of their system or issue and how this structure determines the system's behaviour as shown in the diagram in figure 2, The left arrow symbolizes the relationship while the right arrow indicates the deeper understanding that happens from analysing a system structure. System Dynamics can also be used to study the changes in one part of a system in order to observe its affect on the behaviour of the system as a whole (Martin [5]). Sterman [1] gives an insight that the real value of an SD model should be to eliminate problems by changing the underlying structure of the system rather than anticipating and reacting to the environment. This allows the model to interact with the environment and gives/alerts feedback for structure changes. This is what the term (Dynamics) refers to, i.e. the changes in the system's variables while interacting which stimulate changes over time.

Lyneis [6] stresses the importance of System Dynamic simulation and its power to forecast a market demand for instance and compares this with a statistical forecast. He mentions that an SD Model provides more reliable forecasts than the statistical (non-structural) models and tries to understand the underlying structure that created the data stream.

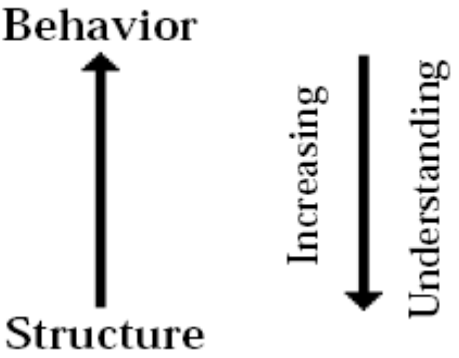


Figure 2: The link between structure and behaviour.

There is lots of user friendly System Dynamics software available now that allows conversion of causal loop diagram and stock and flows diagram into sophisticated computer simulation of the problems or issues being investigated. Examples of these different kinds of software are DYNAMO, STELLA, ITHINK, VENSIM, and POWERSIM (the latest is being used in my study).

Initial values are identified for the stocks, variables values are also identified for the relationships, and the structural relationships are determined between the variables using constants, graphical relationships, and mathematical functions where appropriate. The computer simulation software also facilitates the creation of Microworlds (or management flight simulators) which are kinds of System dynamics Simulators [7, 8] as shown in the case study.

3 Case study

In this section, a simplified generic system dynamics model of a small firm has been built. This model can be used to analyze the firm which sells product or services to its customers and control the money owed to the firm by its customers which is shown in its accounts as an asset called account receivables.

Figure 3 shows the account receivable structure which represents the total monies owed the firm by its customers on credit sales made to them. It depends mainly on the credit policy given to the customer to encourage them to increase their purchases from the firm.

Collecting the sales revenue from the customers, new loans, and the owners investment are the main income to the firm which increases the cash as shown in figure 4. The cash is the most liquid asset in the firm that is always ready to be used to pay bills, pay supplier, repayment of bank loans, and much more expected and unexpected outlays.

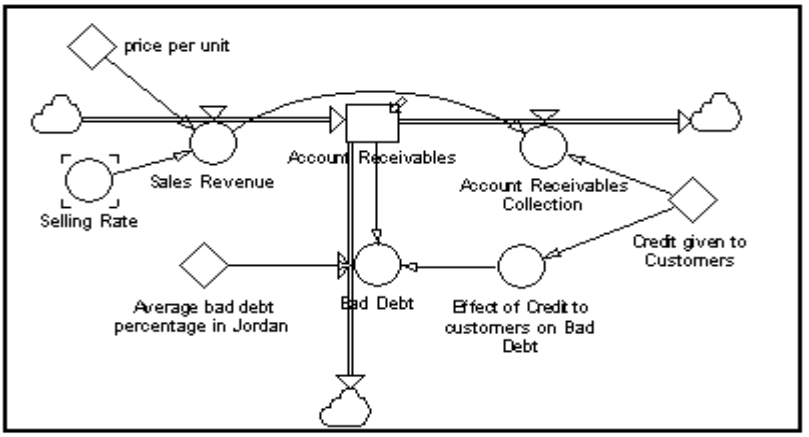


Figure 3: Account receivables structure.

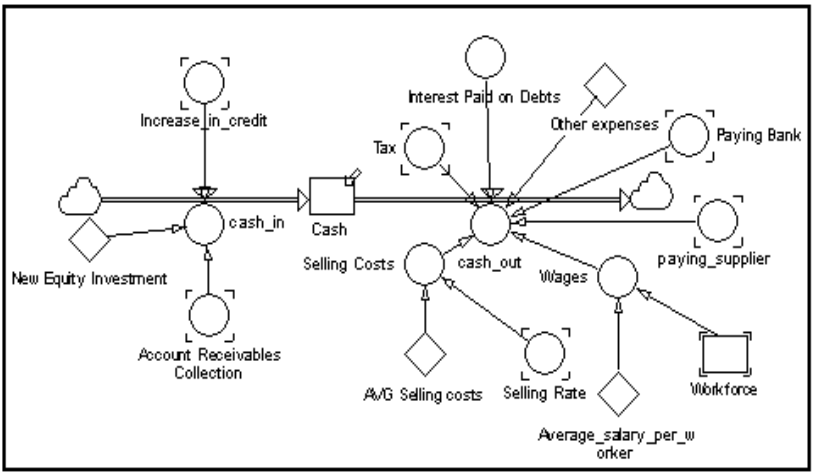


Figure 4: The cash structure.

3.1 The Simulator

The Simulator's friendly interface is designed to enable a financial manager to analyse a firm more effectively. The manager inputs the initial data to the simulator, interacts with various windows according to what variables or decisions he is analysing and selects the reports or graphs he needs. The manager is able to change his decisions about how long should the firm provide credit facilities to their customer, and observe the outcome feedback and result in another window. In this paper, we present only the account receivables window which has the system dynamics inside.

The manager has the opportunity to change the credit term allowed by the firm to its customers. This shows the analyst the behaviour of the firm if any changes in the credit terms happened and its effect on the monthly repayment of loan, interest rate, and the line of credit approved by the bank.

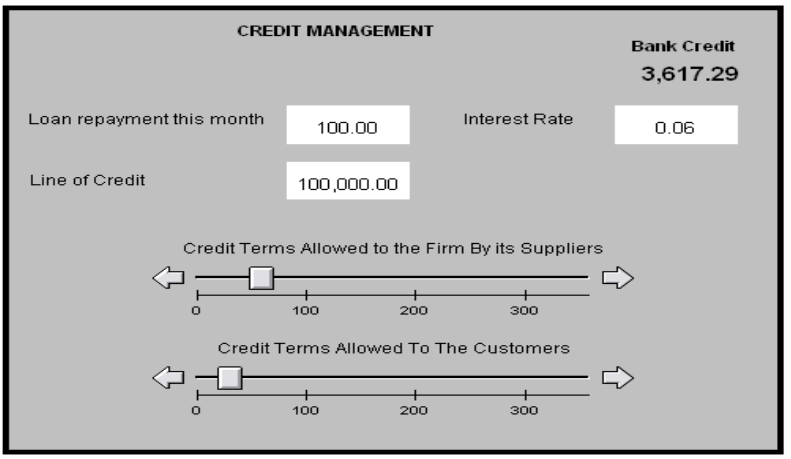


Figure 5: Credit Management Window.

3.1.1 Running the Simulator

If the firm is collecting its revenue from its customer immediately without allowing them to delay their payment, then it won't need to withdraw cash from its bank credit as shown in the manager first scenario.

This scenario increase the sales orders because of the open credit policies, as assumed in this scenario, the firm encourages the customers to buy its product with a credit offer to pay after 60 days. As a consequence of this decision, the cash will rapidly decrease and the firm will need to withdraw from its credit line as shown in figure 6 to cover the cash shortage.

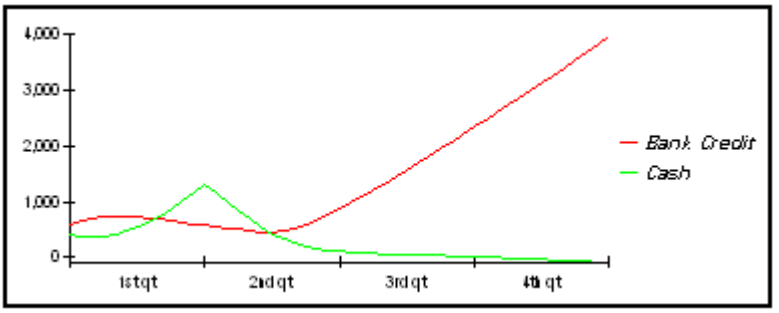


Figure 6: The behaviour of Cash and Bank Credit.

Analyzing the performance of the firm from the financial manager point of view does not concern only with these graphs, but most likely is to analyse the financial ratios and the changes that might happen to these ratio when simulating the scenario. Two ratios were used here to demonstrate these changes.

Time	Current Ratio	Avg Collection period
01/01/2007	1.43	7.75
01/02/2007	0.72	6.89
01/03/2007	0.80	4.68
01/04/2007	0.95	3.76
01/05/2007	1.21	3.73
01/06/2007	1.47	3.81
01/07/2007	1.71	3.71
01/08/2007	1.88	3.55
01/09/2007	2.08	3.43
01/10/2007	2.23	3.37
01/11/2007	2.40	3.29
01/12/2007	2.48	3.22
01/01/2008	2.59	3.15

Figure 7: Current and average collection period ratios before simulating the scenario.

Examples of some of two main ratios that have significantly changed when simulating the scenario are shown in figure 7 which shows the historical changes during one year when the days allowed to customer's payment =0, while figure 8 depicts the changes that happened to these ratios when the firm expand its credit policy. As assumed before, the firm offer the customers 60 days to pay for their purchases.

Time	Current Ratio	Avg Collection period
01/01/2007	1.43	7.75
01/02/2007	0.71	19.83
01/03/2007	0.80	22.56
01/04/2007	0.87	28.60
01/05/2007	0.95	40.63
01/06/2007	0.99	52.97
01/07/2007	1.00	61.33
01/08/2007	0.98	67.32
01/09/2007	0.96	73.31
01/10/2007	0.93	79.63
01/11/2007	0.90	84.88
01/12/2007	0.87	89.52
01/01/2008	0.83	93.51

Figure 8: Current and average collection period ratios with simulated scenario.

As a consequence of expanding the credit policy given to its customers, the risk of collecting cash increased in the firm which is a normal result of the new policy. The average collection period increased from 3 days to 93 days at the end of the year which means that the firm will collects its sales revenue within 93 days, which in turn decreases the cash balance in the firm, and reduces the current ratio which measures the firm's ability to meet its financial obligations.

Current ratios of 2.59 as on 1/1/2008 before simulating the scenario is occasionally cited as good ratio, while a current ratio of 0.83 as on 1/1/2008 after simulating the new policy would be normally considered unacceptable.

On the other hand, implementing this scenario might be useful as it encourages the customers to place more orders and increases the sales in the firm.

The analyst was able to test and observe whatever ratios and scenarios he found vital to analyse the firm's performance from different aspects with many times to try.

4 Conclusion

It is clear now that Systems Thinking tools especially System Dynamics are better than others because it can easily deal with non-linearities and time which are not considered by a static analysis.

By applying System Dynamics, one can enhance the usefulness of the model to address and analyse problems in a complex situations and provide more significant, rational and pertinent policy recommendations.

In summary, the process is to "observe and identify problematic behaviour of a system over time and to create a valid diagrammatic representation of the system, capable of reproducing by computer simulation the existing system behaviour and facilitating the design of improved system behaviour. Existing business simulators are designed to allow the users to play a realistic role in management decision making. Users can make decisions and receive outcome feedback about their performance, by rehearsing strategies and observing results, the managers in the case study were able to discover how to make better decisions and hence improve their performance and reduce the risk of losing money and thus increase the firm's financial performance as illustrated.

References

[1] Sterman, J., *Business Dynamics: System Thinking and modelling for a complex world*, USA, 2000

[2] Richmond, B., Systems thinking: critical thinking skills for the 1990's and beyond. System dynamics review, 9(2), pp. 113-133, 1993

[3] Bartlett, G., Systemic thinking: a simple thinking technique for gaining systemic focus. *In the international conference on thinking, breakthroughs*, USA, 2001

[4] Checkland, P., *Systems Thinking: Systems Practice*, John Wiley & Sons, 1981

[5] Martin, L., *Mistakes and Misunderstandings*: *System Dynamics in Education Project*. System Dynamics Group, Sloan School of Management, Massachusetts Institute of Technology, 1997
[6] Lyneis, J., System dynamics for market forecasting and structural analysis. *System Dynamics Review*, 16(1), pp. 3-25, 2000
[7] Gonzalez, C., The use of Microworlds to study dynamic decision making. *Dynamic Decision Making Laboratory***,** Carnegie Mellon University, USA, 2004
[8] Langley, P. & Morecroft, J., Performance and learning in a simulation of oil industry dynamics. *European Journal of Operational Research*, (155), pp. 715-732, 2004

Computational aerodynamic analysis of flatback airfoils by coupling N-S equations and transition prediction codes

L. Deng[1], Y. W. Gao[1] & J. T. Xiong[2]
[1]*National Key Laboratory of Science and Technology on Aerodynamic Design and Research, Northwestern Polytechnical University, China*
[2]*Department of Mechanical and Aerospace Engineering, University of California, USA*

Abstract

Flatback (Blunt Trailing Edge) airfoils are adopted for the inboard region of large wind turbine blade due to their structural and aerodynamic performance advantages. Very limited experimental data at high Reynolds Number makes it difficult for wind turbine designers to design and use these section shapes because the wind tunnel experiments are limited by the Reynolds Number and the solid blockage. In this study, a 2-D Reynolds-Average Navier- Stokes Solver coupled with a transition prediction based on the e^N method is used to CFD computation of blunt trailing edge airfoils. A new coupling structure with a time-accurate transition prediction model taking the unsteady flow as a result of the bluff-body vortex shedding into account is developed. The computational grid is C-Grid generated by the tool of Gridgen, and the vertical angle at the blunt trailing edge is smoothed slightly to increase the grid quality. An airfoil of DU97-Flat modified by DU97-W-300 airfoil for wind turbine application is calculated and effects of grid points are investigated. The aerodynamic performance of DU97-W-300 is calculated and comparisons between the results from literature and wind tunnel experimental data are performed, and the results show that the method in present study can obtain the aerodynamics performance with much less grid numbers while agreeing better with the wind tunnel experimental data than the literature. One issue that requires attention is the prediction of maximum lift and the failure to accurately capture stall behaviour by the various computational techniques used in this study.

Keywords: wind turbine, airfoil, flatback airfoil, couple, transition prediction.

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110051

1 Introduction

In aerodynamic performance prediction and geometry design of horizontal axis wind turbine (HAWT), the airfoil data of lift and drag coefficient for the different airfoils applied along the span play a significant role. As a result, the designer will spend a lot of up-front time to prepare reliable airfoil aerodynamic data. It is believed that the errors in airfoil data tables are the single largest source of error in most rotor load and performance predictions [1–3].

Recently, the blunt trailing edge or called flatback airfoils have been proposed for the inboard region of large wind turbine blades [4–6]. Flatback airfoils provide several structural and aerodynamic performance advantages. Structurally, the flatback increases the sectional area and section moment of inertia for a given airfoil maximum thickness. Aerodynamically, the flatback increases section maximum lift coefficient and lift curve slope and reduces the well-documented sensitivity of the lift characteristics of thick airfoils to surface soiling [7]. But the flow separation and body-off vortex shedding in the flatback region increase the drag also. One of the problems with wind tunnel testing thickness airfoils is that these types of models tend to create a significant amount of solid blockage and wake blockage thereby affecting the measurements and the flow development in the wind tunnel test section. Solid blockage is typically kept at 5% or less, but this value limits the model chord length, which in turn limits attainable Reynolds Numbers. The Reynolds Numbers are also restricted by load limitations of wind tunnel pyramidal balance. As a result, the published experimental results on flatback airfoils are obtained at low Reynolds Numbers, or are for limited trailing edge bluntness. The lack of experimental data precipitates the analysis of flatback airfoils using computational fluid dynamics (CFD). Recent years, the analysis of flatback airfoils using CFD tend to using more and more grid numbers to capture the vortex structure, but the transition positions are generally obtained by a transition position prediction code and then used as a fixed transition position in the N-S solver and therefore that does not count for the unsteady nature of the flow field because of the blunt trailing edge [8–12]. In present study, several computational techniques are applied including using an N-S solver coupling a transition position prediction code and transition flow region model, and the transition positions are predicted counting the unsteady nature of flow.

2 Flow solver

The aerodynamic performance characteristics of flatback airfoils are calculated using CFD code based on Reynolds-average Navier-Stokes (RANS) equations. The code solves the compressible, two-dimensional, RANS equations. The governing equations are central differenced in standard second-order form and second- and fourth-order artificial dissipation terms are added for numerical stability. The code employs local time stepping and implicit residual averaging, multi-grid technique to accelerate convergence.

In present RANS solver, the one-equation Spalart-Allmaras (S-A) and the zero-equation Baldwin-Lomax (B-L) model are available. Although both models have been successfully applied to study flatback airfoils, the S-A model is selected because the results show that it gives a better agreement with the experimental data in separated flows.

3 The effects of grid

To reduce the calculation time, it is reasonable to keep the number of grid points at a minimum. This minimum is defined by the number of grid points, which is necessary to generate accurate results. In order to define the minimum number of grid points, different grid spaces are tested. A flatback airfoil of DU97-flat (see Fig. 1) is selected and that is created by adding thickness to the aft half of the airfoil of DU97-W-300, giving a blunt trailing edge with a width of 10% chord [10]. The lift coefficient at different grids numbers are calculated at Re of 3×10^6 and Mach number of 0.165, angle of attack of 0°, and at that condition the transition position are fixed at 0.343 (upper surface) and 0.320 (lower surface).

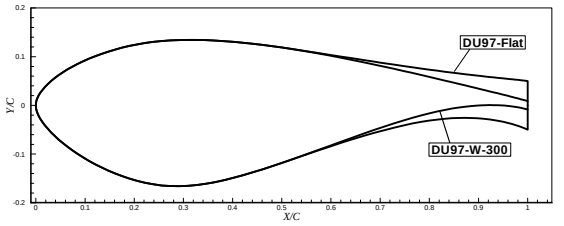


Figure 1: Airfoil shapes of DU97 and DU97-flat.

In previous study, in order to keep cell skewness minimal and the angles in the cells close to 90°, a C-grid works well with sharp trailing edges. When the trailing edge grows thicker, the cell skewness increases and the grid-shocks appear at the blunt base; in that case, the O- grid in contrast does not generate grid shocks and works well with blunt trailing edges [11]. But in the present study, in order to compare with sharp trailing edge, C-grid is generated using the grid generation tool called Gridgen. In grid generating, the vertical angle at the blunt trailing edge is smoothed slightly to increase the grid quality, and that is common for this kind of airfoils [13]. In this study, the RANS equations are solved coupling a transition position prediction code to obtain the pressure distribution of boundary layer. The accuracy of the parameters of the boundary layer influence the accuracy of the transition position prediction, therefore an initial spacing is 1.0×10^{-6} and the first 60 points are with equivalent spacing. Table 1 shows a series of grid numbers. Fig. 2 shows the C-grid of 352×96 that was used and a close-up of the trailing edge.

Table 1: Different computational grid.

Grid	wall	wake	vertical	far field/chord
256×80	96×2	32	80	12
352×96	128×2	48	96	20
400×104	152×2	48	104	20
448×112	160×2	64	112	20
496×112	184×2	64	112	30
544×128	192×2	80	128	30

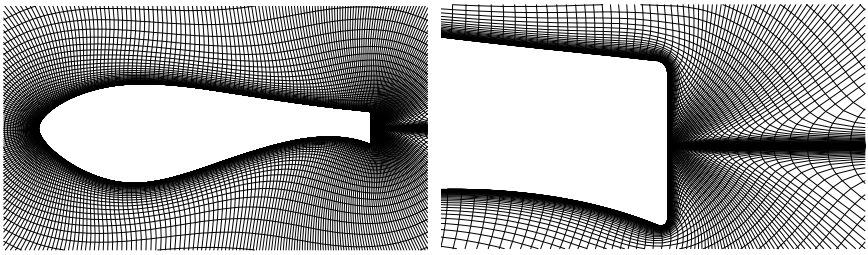


Figure 2: Computational grid of DU97-flat and close up.

Fig. 3 shows the calculated lift coefficient versus the number of grid points. We can see from the figure that with grid number increasing, the lift coefficient tends to converge toward a constant value, and in order to obtain enough accurate aerodynamic performance, we should increase the number of grid points but that brings a increase of computational cost; as a result, we should try to find a trade-off between them. In the following calculation, the number of grid is selected as 448×112.

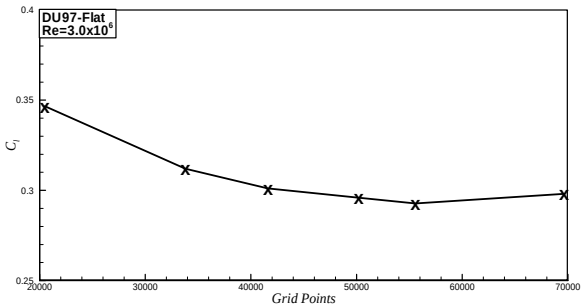


Figure 3: Comparison of Lift coefficient for different number of grid points.

4 Transition prediction and transition flow model

4.1 e^N method

The e^N method based on linear stability analysis is used as a transition criterion. For a set of specified dimensional frequencies, the amplification N factor is computed as:

$$N = -\int_{x_0}^{x} \alpha_i dx \tag{1}$$

here, x_0 corresponds to the location where amplification rate α_i is zero. A group of amplification curves (in present study, 15 curves) for different dimensional frequencies are computed and then to determine the first dimensional frequency that is referred to as the critical frequency which achieves the predetermined amplification factor corresponding to transition.

4.2 Transition flow model

In a previous study, usually the point transition is applied in the N-S solver, that is defined by merely switching on the turbulence model at transition onset. On one hand, this procedure results in an eddy viscosity production that yields rapid but not abrupt changes of the viscous layer properties; thus a small transition zone is created computationally. On the other hand, point transition represents a local perturbation due to the relatively sharp reduction in the displacement thickness in the vicinity of transition, which produces a strong viscous/inviscid interaction region with a remarkable upstream influence. Considerable perturbations in wall pressure and in viscous layer properties are present, which prevent the iteratively coupling N-S and e^N method computations to converge. To the contrary, the application of a finite length transition zone in N-S computations will be shown to reduce the strength of the local perturbation; correspondingly, the changes in wall pressure and viscous flow data will be more moderate. In present study, we use the Walker's model [14], which is initially defined as:

$$\mathrm{Re}_{\Delta X} = 5.2\,\mathrm{Re}_{X^T}^{3/4} \tag{2}$$

here, ΔX is the transition zone length, and X^T is the location of transition onset. However, Walker's transition length model is based on restrictive assumptions; such that the estimate is considered represent the minimum length. Walker proposed to use the double value of the minimum length model, and then the model is:

$$\mathrm{Re}_{\Delta X} = 10.4\,\mathrm{Re}_{X^T}^{3/4} \tag{3}$$

This model is applicable for flow situations where transition is predicted well upstream of laminar separation, but numerical experiments show that in cases small separation bubbles, this model is available as well.

5 Coupling of RANS and *eN* method

In published literatures on the aerodynamic performance computation of flatback airfoils, the transition locations are computed by a transition prediction code and then use the locations as fixed transition in RANS solver. But because the vortex shedding in the trailing edge, the calculated coefficients show cycle changes, thus the transition location prediction should count the unsteady of the flow field. In this study, the algorithm for the coupling of RANS and transition prediction works as follows:

a) The RANS solver is started with prescribed transition locations on upper and lower sides of the airfoil. In this moment, the locations are usually set very far downstream, such as 0.8 chord or even at the trailing edge (in the latter case, the RANS solver computes a fully laminar flow).

b) The RANS equations are iterated $cycle_{\max}$ times, in that moment the coefficients have come to the stage of cycle change.

c) The computation is continued next $\Delta kcyc$ times, and averaged pressure distribution of wall and boundary layer are computed, and the computed properties are used as inputs in the boundary layer equations solving. And then, the velocity profile and boundary layer parameters such as displacement thickness, momentum thickness and shape factor are computed.

d) The transition prediction module is called to transition onset coordinate of x^T. From the instability point to downstream, the e^N method is used to calculate the amplification factor N to check whether it is the onset point of transition. At every point, before the calculation of N factor, a check of whether the laminar separation has occurred; in the case of laminar separation, the separation point is used as an approximation of the transition location.

e) The current coordinate x^T is under-relaxed to use as the transition location. A new transition location $x^{*,T}$ is defined as:

$$x^{*,T} = C^T x^T \tag{4}$$

here, $C^T > 1$. The under-relax of the determined transition location prevents the case that at an unconverged stage, the transition location is determined to far upstream and might not be shifted to downstream again. According to Krumbein, the new location can be defined as[15]:

$$x_l^{*,T} = x_{l-1}^{*,T} - f\left[x_{l-1}^{*,T} - x_l^T\right] \tag{5}$$

here, f is 0.7, and l is current iteration step of transition prediction.

f) As convergence criterion, $\Delta x_l^{*,T} < \varepsilon = .01$, with $\Delta x_l^{*,T} = \left|x_l^{*,T} - x_{l-1}^{*,T}\right|$. In the case that criterion is satisfied, the iteration is finished; else the algorithm loops back to the station of c) using the transition location of $x_l^{*,T}$.

We use the DU97-flat airfoil as a validation at Re of 3×10^6 and Mach number of 0.165, angle of attack of 0°. The initial transition locations are set at 0.8 at both sides with $cycle_{\max}$ of 1500 and $\Delta kcyc$ of 200. Fig. 4 shows the curves of the transition locations versus iteration steps of transition location, and it shows that within 10 steps the transition locations converge on constant values. Fig. 5 shows that the lift and drag coefficient converge toward their converged values. From this case we can see that the coupling method in present study can be used in the computing of the aerodynamic performance of flatback airfoils.

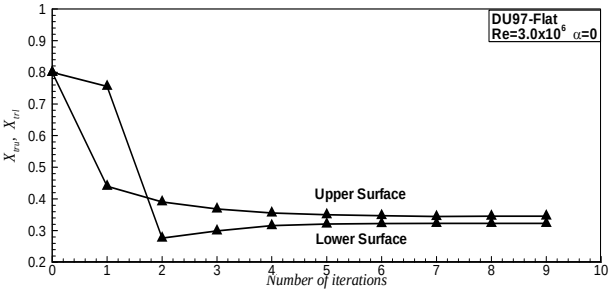


Figure 4: Convergence history of the transition locations.

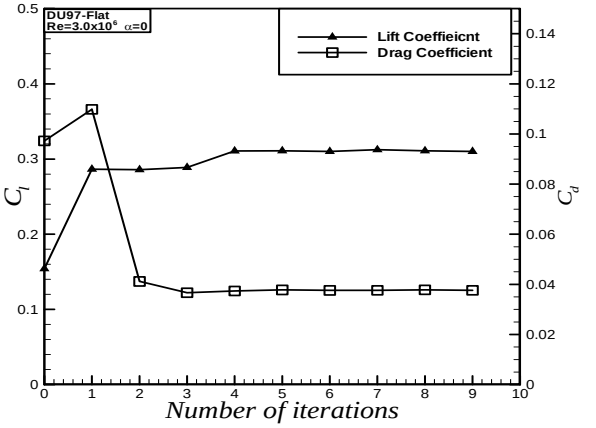


Figure 5: Convergence history of lift and drag coefficient.

6 Results and discussion

In 1999, Timmer published a wind tunnel measured data of airfoil of DU97-W-300 at Reynolds number of 3.0×10^6 and Mach number of 0.165 [16]. The DU97-W-300 is a wind turbine airfoil with a maximum thickness of 30% chord and a trailing edge thickness of 1.74% chord. Matthew computed DU97-W-300 with a RANS solver of SACCARA and used almost 4×10^5grid points, but the transition

locations were computed by the code of Xfoil and then used as fixed locations [10]. In the present study, we compute DU97-W-300 and compare the results with experimental data and Matthew's results. Fig. 6 is the comparison of transition locations computed by the coupling method in this paper and Xfoil code. We can see that at low angle of attack, the transition locations at lower surface agrees well with Xfoil results, but with some discrepancies appearing with angle of attack growing higher. The computed results in present study are smaller than Xfoil results; especially at angle of attack of 12 degree, the computed results show a separation at leading edge, but there is still 0.1 chord laminar flow in Xfoil results.

Fig. 7 compares the lift coefficient and moment coefficient. A better agreement is shown in lift coefficient between the computed and experimental lift results than the SACCARA results, especially in linear region; but just as Matthew's results show, the computed results did not capture the stall angle. And

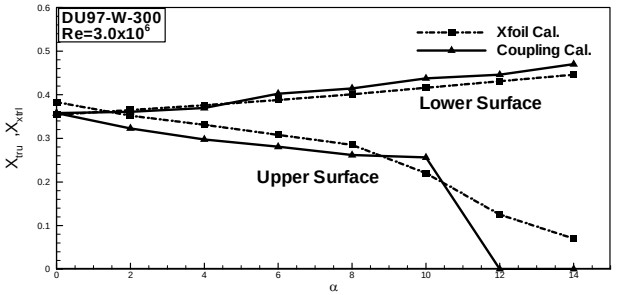


Figure 6: Comparison of transition locations.

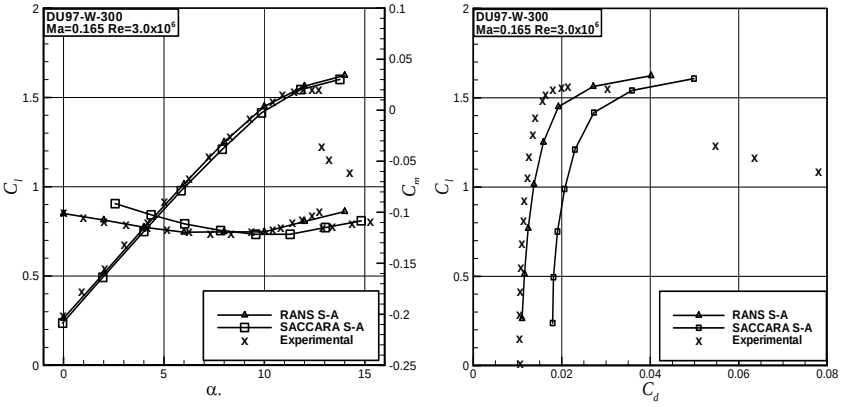


Figure 7: Comparison of lift and moment coefficient comparison of drag polar.

very good agreement is shown in moment coefficient before angle of attack 13 degree. Fig. 8 is the comparison of drag coefficients. The SACCARA' results are almost double of experimental data and the computed results show a good agreement with the experimental data but there are some discrepancies at high lift coefficient. And it is interesting that there is a constant difference between the SACCARA' results and the computed results in this study.

7 Conclusion

Flatback airfoils are considered for the inner regions of large wind turbine blades. The concept of blunt trailing edge is nothing new and has been previous investigated by CFD method. However, previous studies are typically based on a RANS solver using a fixed transition locations computed from a transition prediction code, which don't take the unsteady nature of flow field because of the vortex shedding at the trailing edge into account. From the investigation of the effects of grid points, a trade-off should be made between the growth of the number of grid points and the computational time. A RANS solver is coupled with a transition prediction code with a time-accurate transition model. The results show that the transition locations and the computed aerodynamic coefficient converge toward constant values within 10 iteration steps. The wind turbine airfoil of DU97-W-300 is computed by the coupling method and the results are compared with experimental results and the results from the literature. The comparisons show that the computed results agree better with the experimental data than the results in the literature while with less grid points. One issue that requires attention is the prediction of maximum lift and the failure to accurately capture stall behavior by the various computational techniques used in this study.

References

[1] Patrick J. AeroDyn Theory Manual, NREL/EL-500-36881, 2005.
[2] Simms D., Schreck S., Hand M., Fingersh L.J. NREL Unsteady Aerodynamics Experiment in the NASA-Ames Wind Tunnel: A Comparison of Predictions to Measurements. NREL/TP -500-29494, 2001.
[3] Tangler, J.L. The Nebulous Art of Using Wind-Tunnel Airfoil Data for Predicting Rotor Performance. NREL/CP-500-31243, 2002.
[4] TPI Composites, Innovative Design Approaches for Large Wind Turbine Blades – Final Report, SAND2004-0074, 2004
[5] Standish, K.J., and van Dam, C.P., Aerodynamic Analysis of Blunt Trailing Edge Airfoils, Journal of Solar Energy Engineering, **125(4)**, pp. 479-487, Nov. 2003
[6] Jackson K, Zuteck, M., van Dam, C.P., Standish, K.J., and Berry, D., Innovative Design Approaches for Large Wind Turbine Blades, Wind Energy, **8(2)**, pp. 141-171, 2005

[7] Van Rooij R.P.J.O.M., and Timmer W.A., Roughness Sensitivity Considerations for Thick Rotor Blade Airfoils, Journal of Solar Energy Engineering, **125(4)**, pp. 468-478, Nov. 2003
[8] Baker J., Mayda E., and van Dam C.P. Computational and experimental analysis of thick fatback wind turbine airfoils, AIAA Paper 2006-193, 2006.
[9] D.D. Chao, and C.P. van Dam, RaNS Analysis of an Inboard Flatback Modification on the NREL Phase VI Rotor, AIAA Paper 2006-195,2006
[10] Matthew F. Barone, Dale Berg. Aerodynamic and Aeroacoustic Properties of a Flatback Airfoil: An Update, AIAA Paper 2009-271, 2009
[11] TPI Composites. Computational Design and Analysis of Flatback Airfoil Wind Tunnel Experiment, SAND 2008-1782, 2008
[12] Christopher Stone, Matthew Barone, C. Eric Lynch, et al. A Computational Study of the Aerodynamics and Aeroacoustics of a Flatback Airfoil Using Hybrid RANS-LES, AIAA 2009-273, 2009
[13] T. Winnemöller, Design and numerical optimization of thick airfoils, AIAA 2006-238, 2006
[14] G.J. Walker, Transitional Flow on Axial Turbomachine Blading, AIAA Journal, **27(3)**, pp. 595-602, 1989
[15] A. Krumbein. Navier-Stokes Airfoil Computation with Automatic Transition Prediction Using the DLR TAU Code- A Sensitivity study, New Results in Numerical and experimental Fluid mechanics V., Springer Berlin/ Heidelberg, Vol. 92, 2006
[16] Timmer, W.A., and van Rooij, R.P.J.O.M. Design and Wind Tunnel Test of Airfoil DU 97-W-300, IW-98003R, Institute for Wind Energy, Delft University of Technology, 1999

Numerical investigation of dynamic stall phenomenon on a plunging airfoil

F. Ajalli[1] & M. Mani[2]
[1]*Department of Aerospace Research Institute, Amirkabir University of Technology, Iran*
[2]*Department of Aerospace Engineering and Center of Excellence in Computational Aerospace Engineering, Amirkabir University of Technology, Iran*

Abstract

The unsteady separated turbulent flow around an oscillating airfoil plunging in a sinusoidal pattern in the regime of low Reynolds number is investigated numerically, employing the *URANS* approach with advanced turbulence model, k-ω*SST* transitional. A comparison with experimental data shows that the Transition SST model is capable of predicting the flow characteristics for the increasing cycle while the main difficulty lies in the accurate modeling of the complicated separated flows during the decreasing stroke. The flow development of the dynamic stall is also discussed.
Keywords: dynamic stall, plunging airfoil, k-ω-SST transitional model.

1 Introduction

Dynamic stall has been widely known to significantly affect the performance of a large variety of fluid machinery, such as helicopters, highly maneuverable fighters, gas turbines, and wind turbines. It is well recognized that the dynamic stall process can be categorized into four key stages, i.e. attached flow at low angles of attack, development of the leading edge vortex (LEV), the shedding of the LEV from the suction surface of the blade and the reattachment of the flow [1]. Numerous experimental and computational investigations [2–5] have shown that the unsteady flow can be separating or reattaching over a large portion of the upper surface of the oscillating airfoil and that the predominant feature of the dynamic stall is the formation and rapid convection over the upper surface of the

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110061

airfoil of an energetic leading-edge vortex (LEV), also referred to as the dynamic-stall vortex, which induces a nonlinearly fluctuating pressure field and produces large transient variations in forces and moments that are fundamentally different from their steady-state counterparts.

Although the basic image of the phenomenon has been generally clarified, the physics of this strongly non linear unsteady flow phenomenon has not yet been completely understood and more efforts are needed to advance the knowledge to the level on which we could accurately predict and precisely control the dynamic stall [6]. Most of the previous researches have investigated flows at high Reynolds number (Re $\times 10^6$) or high Mach number (Ma$\geq$0.3) which fall into the compressible flow regime. However, dynamic stall at low Reynolds number has distinct features compared with those at high Reynolds number, such as flow transition process, laminar separation and reattachment, etc.

In the recent decades, due to the increased awareness of the environmental issues associated with the fossil fuel based power generation industry, wind industry is drawing more and more attention. Dynamic stall has been a critical phenomenon which has an important effect on the operation of both Horizontal Axis Wind Turbines (*HAWT*) and Vertical Axis Wind Turbines (*VAWT*).

The objective of the present study was to investigate the ability of the Unsteady Reynolds-Averaged Navier-Stokes (*URANS*) method with advanced turbulence model (k-ω SST transition model) in capturing the dynamic stall at low Reynolds number flows (Reynolds number based on the chord length of the airfoil Re_c is of the order of 10^5), and to provide a detailed two-dimensional analysis to gain a better understanding of the flow phenomenon. In order to validate the numerical results the experimental measurements of surface pressure distributions was compared with numerical ones at the same condition.

2 Numerical simulations

2.1 Case studied

The aerofoil employed in the numerical calculations is an E361 airfoil with a chord length of c=0.15m and maximum thickness of 12%c which in this case executes the sinusoidal plunging motion h=8(cm) +sin(ωt) with reduced frequency $k=\omega c/2U_\infty=0.14$. The free stream velocity is U_∞=10m/s with a turbulence intensity of 0.2% which corresponds to a chord Reynolds of $Re_c=1\times10^5$. The mean angle of oscillations was set up 12deg with in static stall angle. Numerical set up are based on the experimental tests in order to compare numerical results with experimental data. A more comprehensive description of the experimental setup is detailed in [7].

2.2 Numerical techniques

Firstly for the static flow field investigations the RANS approach with advanced turbulent model, namely the k-ω-SST transitional model were used, furthermore the k-ε RNG and the low Reynolds k-ω-SST model were employed as the

baseline turbulence models to predict the performance of low Re and fully turbulent approach without transition secondly for the dynamic simulations k-ω-SST transitional model were applied to the all cases. The transition SST Model is a four- equation turbulence model which is based on the coupling of the SST k-ω transport equations with two other transport equations, one for the intermittency and the other for the transition onset criteria, in terms of the momentum-thickness Reynolds number. This is a reasonably new model and it is expected to predict flows with massive separations more accurately. The FLUENT code used to solve the Reynolds averaged Navier–Stokes equations using finite volume discretization. Second order upwind discretization in space is used, and the resulting system of equations is then solved using the SIMPLE coupled solution procedure until convergence criteria are satisfied. The convergence rate is monitored during the iteration process by means of the residuals of the dependent variables of the governing differential equations. A convergence criteria of O(5) reduction in all dependent variable residuals is accepted as adequate for the present aerofoil study. Convergence is also checked using the relative differences between two successive iterations for each of the integrated force and moment coefficients. Free stream boundary conditions are used in the upstream, downstream and outer boundaries. No-slip boundary conditions are used at solid surfaces.

2.3 Grid design

A hybrid C-grid in figure 1 is used for the calculations and about 256 grid nodes are placed along the aerofoil and they are clustered close to the leading and trailing edges. The height of the first row of cells of the cells bounding the aerofoil is set to be 10^{-5}c which ensures y+≤1.0 for the cells immediately adjacent to the aerofoil so that the boundary layer flow can be properly resolved without using a wall function. The height of the cells expands with a growth factor of 1.2 towards the external boundary which is set to be at 25c from the aerofoil to eliminate the boundary reflections. The whole mesh sums up to be

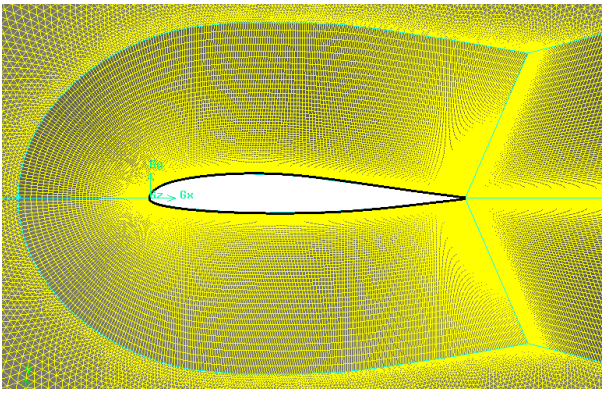


Figure 1: Hybrid grid quality close to the airfoil.

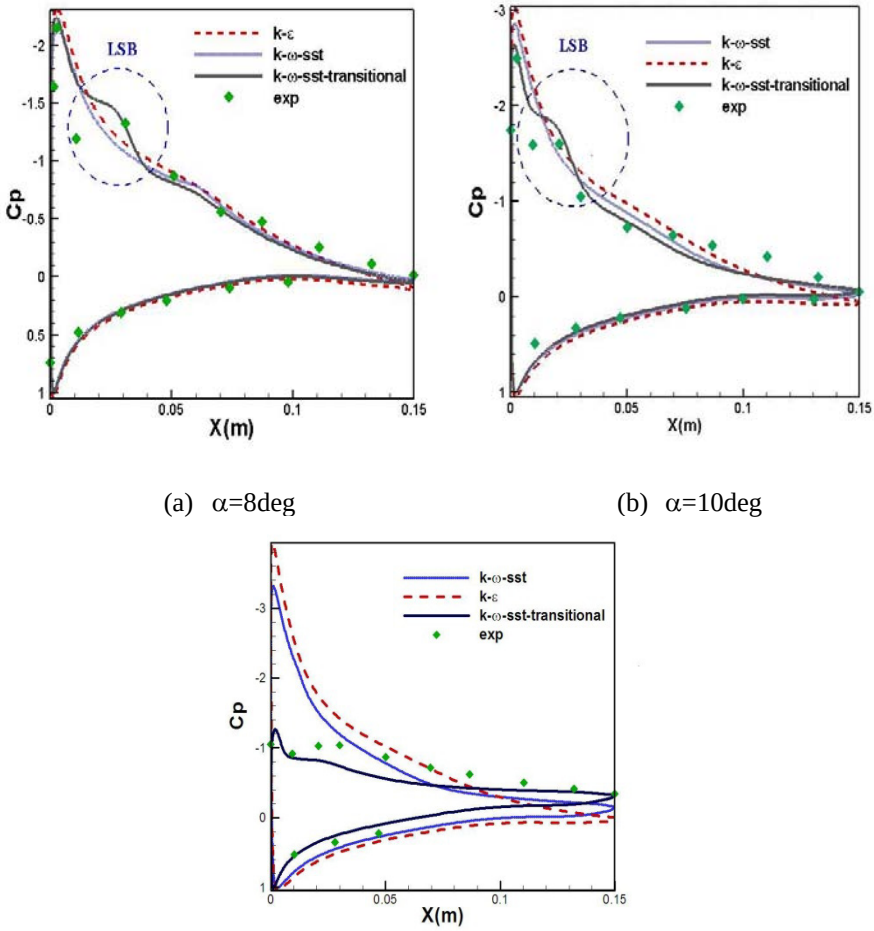
200,000. In order to simulate the sinusoidal plunging motion of the airfoil, the whole grid oscillates like a rigid body with the same sinusoidal mode as the airfoil. This is achieved by using the dynamic mesh technique [8] with a User Defined Function (*UDF*) subroutine developed and attached to the Fluent solver to control the movement of the grid points. The numerical time step size is set to be 0.001. Normally after three oscillation cycles, a perfect periodical solution can be obtained in this study.

3 Results and discussion

3.1 Static airfoil

In this part of the paper static flow field with increasing incidence at $Re=10^5$ is presented. The identification of static flow field, laminar separation bubble and trailing edge separation serves as a frame of reference for dynamic stall results. Figure 2 shows the computed static pressure coefficient obtained by using different turbulence modeling approaches, compared with the experimentally measured results [9]. There was good agreement between the experiment and k-ω SST transitional model at different angles of attack. According to the experimental data laminar separation bubble (LSB) was formed near the leading edge of the upper surface of the aerofoil, the existence of the separation bubble produced an approximate constant pressure plateau starting from the separation point to the transition region, followed by a sharp pressure rise in the turbulent bubble region blending in to the turbulent boundary layer. As shown in figure 2, k-ω SST transitional model predicted LSB properly but k-ε RNG and k-ω SST turbulence models missed the separation bubble; however there is not significant distinction between lower surface pressure coefficients of the different models. According to the experimental data the stall angle was recorded about 11 degrees whereas k-ε RNG and k-ω SST turbulence models over-predicted the stall angle while k-ω SST transition model under-predicted it. Lastly the stalled pressure coefficient of the airfoil are illustrated in figure 4c, accordingly, the pressure variations of the experimental data and k-ω SST transitional model included plateau as a result of stall phenomenon, however the computation data of k-ε RNG and k-ω SST demonstrated high suction pressure peak near the leading edge that indicated no stall characteristic of the aerofoil.

Information relating to the k-ω SST solution at $\alpha=4$ deg are provided in figure 3a, according to the results of this model, at 4deg angle of attack, laminar separation occurred on the upper surface at about 60% chord position from the leading edge. Because the adverse pressure gradient was modest, the separated flow was able to reattach to the surface. As the angle of attack increased, figure 3b, the adverse pressure gradient aft of the point of suction peak became stronger and the separation point moved toward the leading edge. The stronger pressure

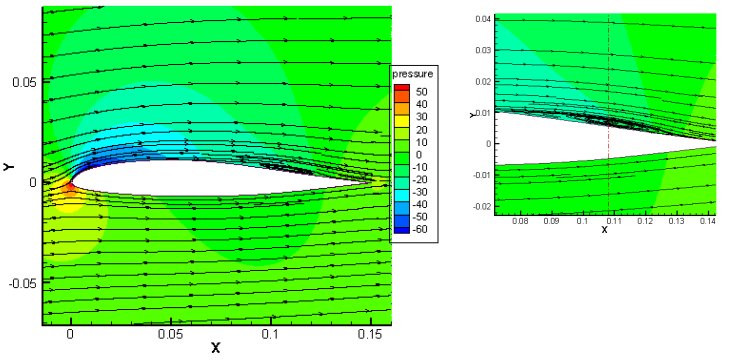


(a) α=8deg (b) α=10deg

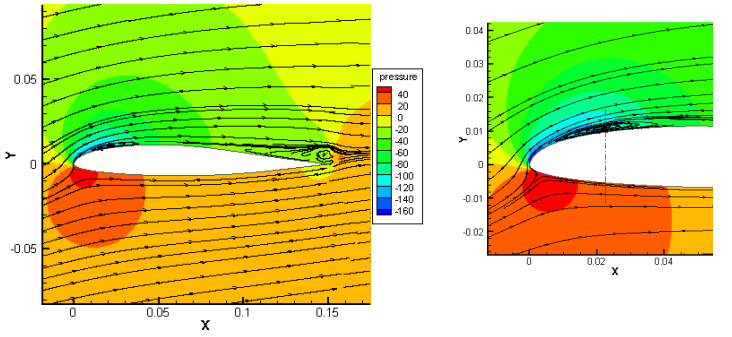
(c) α=12deg

Figure 2: Surface pressure distributions.

gradient amplified the disturbance in the separation zone and prompted transition and the bubble was shrinking in size. According to this model, at angle of attack of 10 deg, the separation position is at around 6% chord position and transition occurred at 14.2% of chord position. It is noticeable to say that, at this angle of attack the airfoil was close to stall therefore a large portion of the flow on the aerofoil was separated and the separated vortex formed near the trailing edge. Because of better prediction of k-ω SST transitional method, this model applied for the dynamic cases.



(a)



(b)

Figure 3: Static pressure and streamlines of k-ω SST transitional, α=4 and 10 deg.

3.2 Dynamic airfoil

Numerical unsteady investigations of dynamic stall phenomenon were conducted on a plunging airfoil at mean angle of attack 12deg with amplitude of 8cm. The reduced frequency is set k=0.141 and chord Reynolds number according to the static surveys is 10^5. Figure 4 illustrates the sinusoidal plunging airfoil displacements of the experimental tests which were transformed in to equivalent angle of attack using the potential flow transformation formula $\bar{\alpha}_{eq} = ik\bar{h}$ also the mean angle of attack would be added to the equivalent angle. As it is clear in figure 4, induced angle of attack has 90 degrees phase difference with vertical motion of the airfoil. It is noticeable to say that, the down stroke plunging displacement is chosen as the positive direction; besides there are four steps for the change of "h" and corresponding "α_{eq}" such as (i) downward displacement of "h" from 0-8cm that $\alpha_{max}<\alpha_{eq}<0$, (ii) upward displacement of "h" from 8cm-0,

correspond to $0<\alpha_{eq}<-\alpha_{max}$, (iii) upward displacement of "h" from 0-(-8cm) correspond to $-\alpha_{max}<\alpha_{eq}<0$, (iv) downward displacement of "h" from (-8cm)-0, conformable to $0<\alpha_{eq}<\alpha_{max}$ thus, according to these sequences first half of the plunging oscillation cycle is related to the decreasing equivalent angle and the second one attributed to increasing equivalent angle. As pictured (figure 4) α_{eq} is a maximum or minimum whenever h=0 during down-stroke and upstroke motions respectively. In order to better understanding of the flow field all the results are presented according to the equivalent angle.

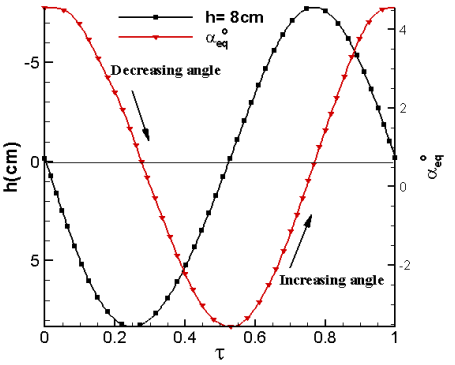


Figure 4: Time history of plunging motion with corresponding equivalent angle of attack.

Figure 5 (a-i) presents a chronology of the static fields superimposed on the instantaneous streamlines in order to depict the complicated vortex structures during the stall process. In the early stage of the increasing angle, the flow remained attached except in the trailing edge region, at $\alpha_{eq}\approx$ 11.76◦, Laminar Separation Bubble (LSB) close to the leading edge at x/c=1.37 was detected, as α_{eq} increases beyond the α_{ss} a thin flow reversal was observed in front of the trailing edge to leading edge, as one can observe from figures 5b and c upstream of the flow reversal the boundary layer remained attached on the upper surface of the airfoil and the position of the LSB was moved toward the leading edge (for example, figure 5b the center of the LBS changed to x/c=0.11) accordingly figure 5d (α_{eq}= 17.55deg) the turbulent boundary layer broke down and moved upstream and downstream, spillage the LSB, and initiated the formation of Leading edge vortex (LEV) afterward the LEV grew and covered rapidly downstream and covered the whole suction surface. During the further convection of the LEV due to the recirculation of this vortex, figure 5g-i, two secondary vortices near the leading edge were observed; aside from the vortices, a strong counter clockwise vortex appeared at the trailing edge toward the upper surface due to the low pressure region on the upper surface and high pressure difference on the trailing edge of lower surface which they pushed the flow from the lower surface and caused the CCW vortex; besides as α_{eq} increased the trailing edge vortex was driven upward.

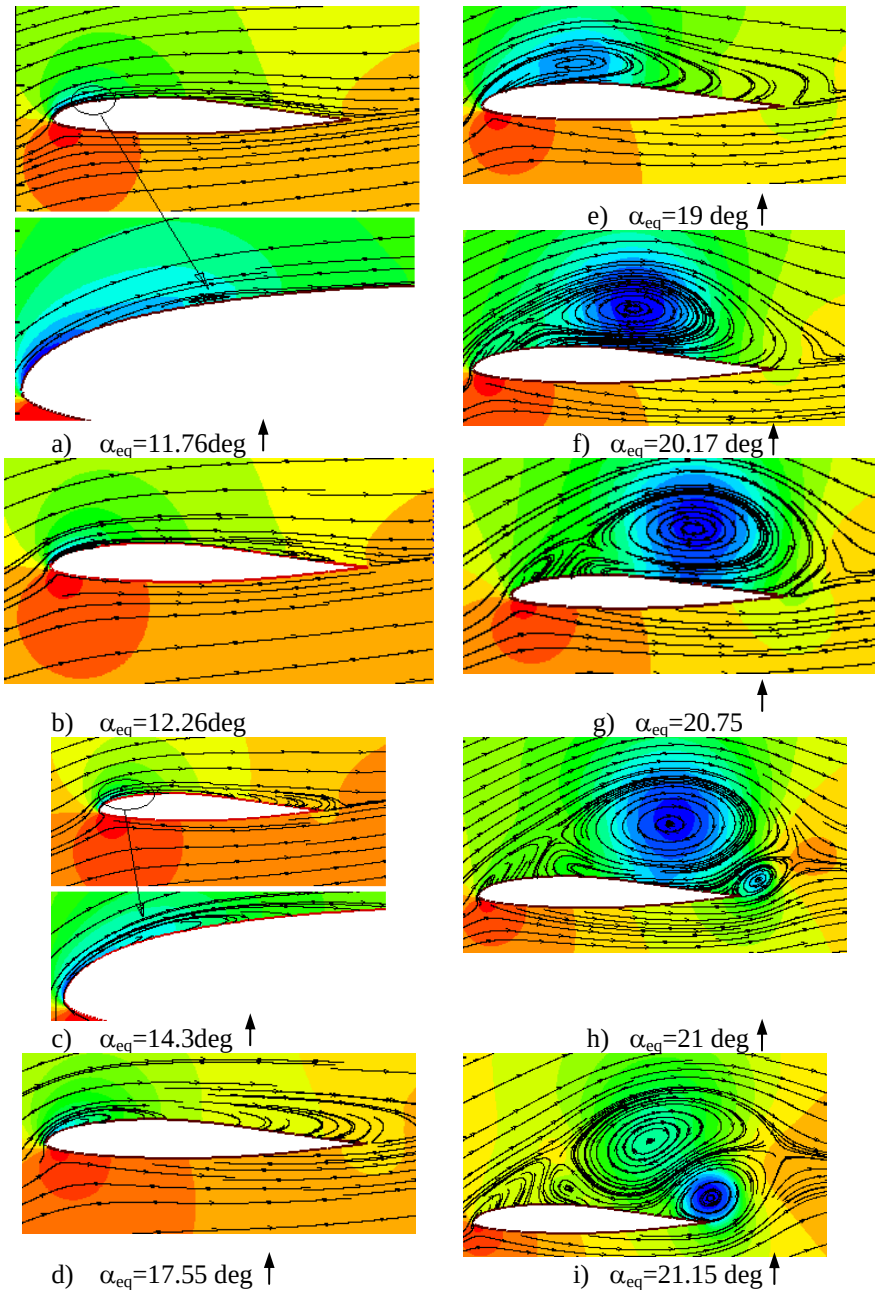


Figure 5: Pressure field superimposed on the instantaneous streamlines of increasing angle.

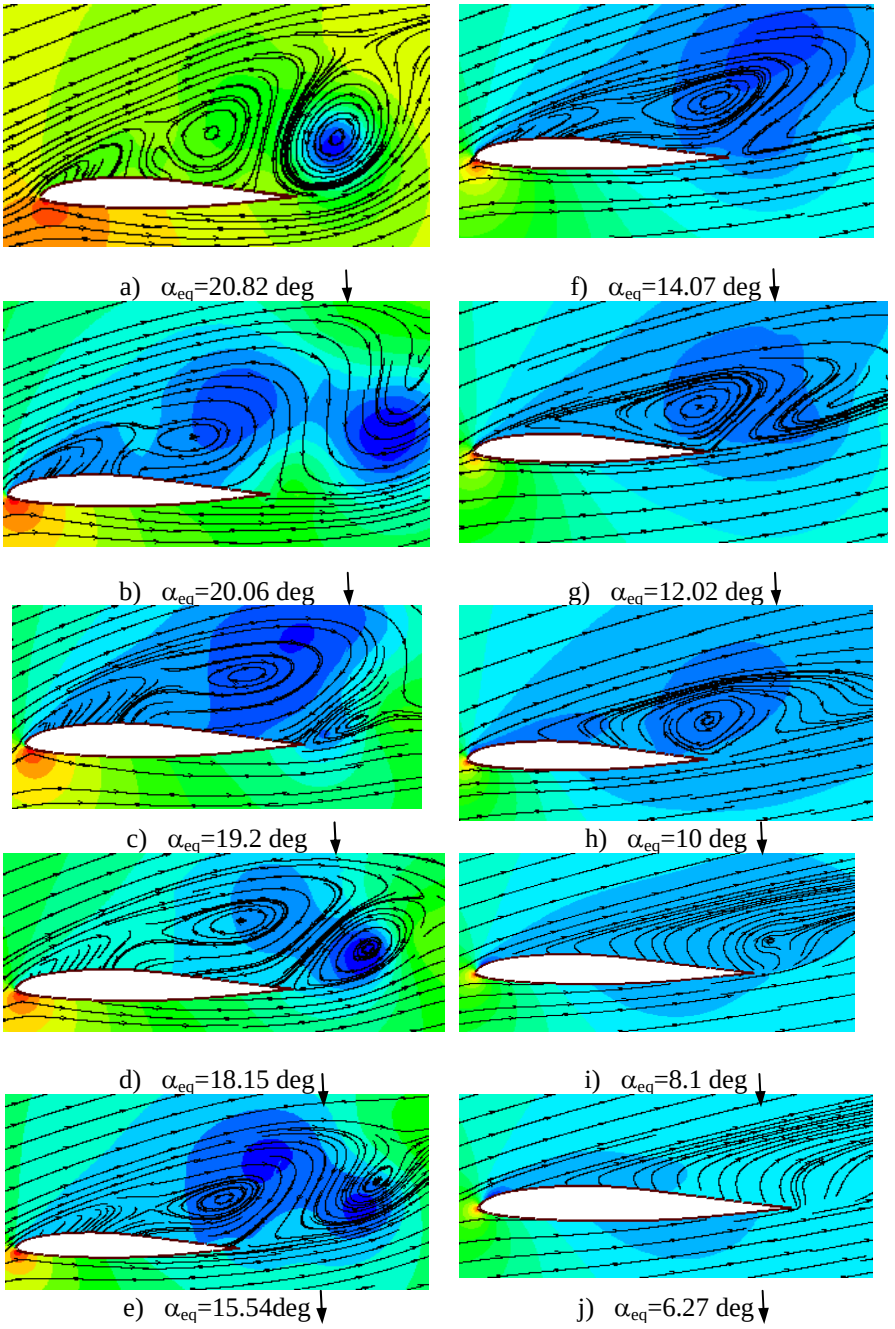
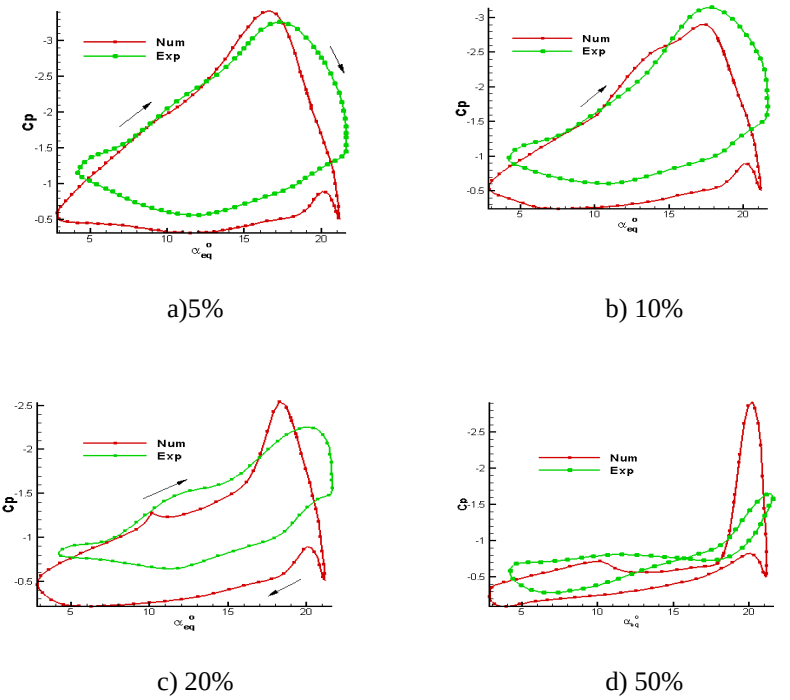


Figure 6: Pressure field superimposed on the instantaneous streamlines of decreasing angle.

Figure 6a-j shows the static fields and instantaneous streamlines during decreasing angle. According to figure 6a (α_{eq}= 20.82deg) in decreasing cycle, trailing edge vortex was grown and LEV became weaker, when α_{eq} decreased further the trailing edge vortex disappeared and shed in to the flow (figure 6b) then the clockwise main vortex (growth LEV) moved away from the surface and again induced trailing edge vortex (figure 6c, d and e) next the separated boundary layer flow reattached to the airfoil at lower incidences. It is significant to note that LEV shedding was forced by the motion in decreasing angle due to the inadequate time for fully develop of LEV before deceasing angle that had good agreement with experimental results [9].

Figure 7 compares numerical and experimental data at α_{mean}=12 deg and k=0.141, the hysteresis observed in the dynamic-pressure loops originated from the asymmetry in the locations or angles of attack, of the separation and there attachment. By inspecting the direction of the C_p hysteresis loops, it is seen that the direction of the C_p loops for both numerical and experimental data were clockwise, which means that the flow had lead phase. The pressure levels of the both results near the leading edge (figure 7a, b) were almost the same in increasing cycle and the Cp slopes were nearly identical but in decreasing stroke the results were deviated from the experimental data.This is mainly because of complicated separated flow and vortex shedding structure. Note that apart from the small divergence, these two graphs rendered same pattern in decreasing angle which indicted, this model could well predict boundary layer events, LEV, separation, detaching and reattachment, in addition the Transition *SST* model presents a too sharp drop-off of the C_p when the dynamic stall occurs giving an over-prediction of the strength of the stall. Figure 7c, d depicts the pressure distributions of the ports x/c=20% and 50%, according to this data narrower hysteresis (compared to the leading edge port) was recorded because of massive separation of the flow on the upper surface of the airfoil. In the aft portion of the airfoil figure 8d, the experimental pressure loops were similar to “figure 8 shape” moreover $|C_{Pmax}|$ of numerical method at this position was recorded 2.9 while it was 1.63 for the experimental case which means the numerical method predicted less dissipative LEV in process.

Figure 8 shows the numerical dynamic stall process of C_l versus α_{eq} at α_{mean}=12deg and k=0.141, the validity of these processes were published by McCroskey [3], McCroskey et al. [2], McAlister et al. [10], Lee and Gerontakos [11–12]. According to Lee and Gerontakos nomenclature point 1and 2 corresponded to the initiation and ending of the upward spread of the flow reversal respectively besides between the points 2 an 3 turbulence bread down, initiation and growth of the LEV were occurred and the lift coefficient underwent a sharp drop as soon as the LEV passed off the trailing edge, and the airfoil remained fully stalled point 3 and 4. At point 5 (decreasing α_{eq}=20.1) there was a slight increase in C_l owing to the presence and convection of the secondary vortex. According to explanations k-ω-SST transitional method could predict the boundary layer event with reasonable accuracy.



a)5% b) 10%

c) 20% d) 50%

Figure 7: Comparison of numerical and experimental pressure coefficient hysteresis loops.

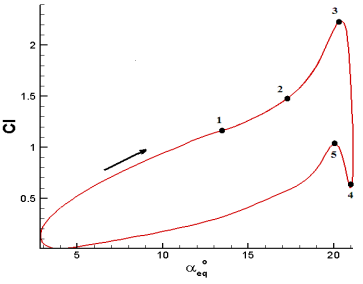


Figure 8: Variations of numerical dynamic lift coefficient.

4 Conclusions

In this paper static and dynamic flow field on an Eppler 361 airfoil at Re=10^5 were investigated. For a static airfoil different turbulence methods analyzed

consequently, k-ω SST transitional model was capable of predicting the flow characteristics such as LSB and trailing edge separation better than other models therefore, in general this model employed to simulate the fluid flow around an E361 airfoil executing a sinusoidal plunging motion. In the sequence of dynamic motion LSB, reverse flow, turbulence breakdown, spillage of LEV, secondary vortices and detachment of the fluid were captured accurately while the dissipation of the progressive LEV was less than experimental data. In the increasing cycle pressure levels of the both results near the leading edge were almost the same however in decreasing stroke the numerical data deviated from the experimental counterparts.

References

[1] P. Wernert, W. Geissler, M. Raffel, and J. Kompenhans, Experimental and numerical investigations of dynamic stall on a pitching airfoil, AIAA journal, vol. 34, pp. 982-989, 1996.

[2] McCroskey, W. J., McAlister, K. W., Carr, L. W., Pucci, S. L., Lamber, O., and Indergrand, R. F., *Dynamic Stall on Advanced Airfoil Sections*, Journal of American Helicopter Society, Vol. 26, July 1981, pp. 40–50.

[3] McCroskey, W. J., *Unsteady Airfoils*, Annual Review of Fluid Mechanics, Vol. 14, 1982, pp. 285–311.

[4] Ericsson, L. E., and Reding, J. P., *Fluid Mechanics of Dynamic Stall. Part I. Unsteady Flow Concepts*, Journal of Fluids and Structures, Vol. 2,Jan. 1988, pp. 1–33.

[5] Lee, T., and Basu, S., Measurement of Unsteady Boundary Layer Developed on an Oscillating Airfoil Using Multiple Hot-Film Sensors, Experiments in Fluids, Vol. 25, No. 2, 1998, pp. 108–117.

[6] S. Wang, L. Ma, D. Ingham, M. Pourkashanian, and Z. Tao, *Numerical Investigations on Dynamic Stall Associated with Low Reynolds Number Flows over Airfoils*, in The 2010 International Conference On Mechanical and Aerospace Engineering (CMAE 2010) Chengdu, China, 2010.

[7] Ajalli F., Mani M., Soltani M. *An Experimental Investigation of Pressure Distribution around a Heaving Airfoil*, The 5th International conference on Heat Transfer, Fluid Mechanics and Thermodynamics, South Africa, Spring 2007.

[8] Fluent 6.3,26 (Theory Guide),"

[9] M. Mani, F. Ajalli M. R. Soltani, An experimental investigation of the reduced frequency effects into pressure coefficients of a plunging airfoil, Advances in Fluid Mechanic 2008.

[10] McAlister, K. W., Carr, L. W. & McCroskey, W. J., *Dynamic stall experiments on the NACA 0012 airfoil*, NASA TP 1100. 1978.

[11] Lee, T., and Gerontakos, P., *Investigation of Flow over an Oscillating Airfoil*, Journal of Fluid Mechanics, Vol. 512, 2004, pp. 313–341.

[12] T. Lee , P. Gerontakos, *Investigation of flow over an oscillating airfoil*, J. Fluid Mech. (2004), vol. 512, pp. 313–341

Design optimization of a bioreactor for ethanol production using CFD simulation and genetic algorithms

E. R. C. Góis & P. Seleghim Jr.
Thermal and Fluids Engineering Laboratory,
University of São Paulo, Brazil

Abstract

The increase of the use of biofuels, like bioethanol, has presented new challenges to engineering problems. The optimization of bioreactors is crucial to improve the bioethanol extraction and to avoid stagnant zones in the flow that can compromise the chemical reactions involved in the process. This paper presents a solution using Computational Fluid Dynamics tools coupled with Genetics Algorithm to find an improved reactor structure. The preliminary results show that the influence of the height outlet tube alone is able to modify completely the flow pattern inside the reactor, improving the efficiency of the reactor.
Keywords: bioethanol, bioreactor, CFD, genetic algorithms, optimization.

1 Introduction

The search for new ways to provide fuel for the society has been constantly increasing. One of the great challenges for scientists and academic researchers is to provide fuels without jeopardizing the environment. An interesting alternative is the bioethanol production from sugar cane or other kinds of biomass, like corn, beet, *etc*.

The production of sugar cane ethanol is economically feasible only in large scale, such as 45 thousand liters per hour, for instance. At these scales, any efficiency improvement can result in a very significant overall optimization in terms of production rates and environmental impacts as well. Due to the difficulty to do experimental analyses, the expensive materials that are used and the time limitations, Computational Fluid Dynamics (CFD) tools are a useful alternative to study the flow inside reactors and bioreactors, and to better define

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110071

their design principles. CFD tools have been applied in different kinds of reactor and bioreactor simulations. For example, Patwardhan *et al*. [1] studied the relationship between geometry and operating conditions of an industrial gas-liquid stirred reactor and had a good prediction of the residence time distribution of the industrial scale reactor. CFD tools were applied to perform stirred solid-liquid reactors [3]. The study of suspension of solid particles and the influence of them in the flow were considered by Shao *et al*. [4].

Most of the studies, however, were developed for small-scale reactors. Large-scale bioreactors, usually applied at bioethanol production, must be studied more carefully because the rule of fluid velocity profile, shear stress, and other physical conditions may be considerably different compared with small-scale bioreactors. There are two possible ways of extracting ethanol from sugar cane: the first generation ethanol and the second generation ethanol.

First generation ethanol is produced directly from the sugar cane juice. The sugar cane juice is submitted to a fermentation process, which results in sugar or ethanol. In this case, a bioreactor is used because there is the necessity to use live organisms as a means of process acceleration. For the second generation ethanol, the principle is to reuse the bagasse resulting from the sugar cane juice. In this case, it is possible to use of a thermo-pressurized reactor with a mechanic process.

In both cases, the reactor should be able to provide optimum conditions to the chemical or mechanical process. The design of the reactor has a great influence in the flow conditions, with respect to mass transfer, shear stress, mixing, control of pH, temperature and substrate conditions [2].

Some flow conditions can be a bad influence on the chemical and mechanical reactions inside the bioreactor. Stagnant zones keep the sugar cane juice inside the reactor for longer or shorter than necessary. In both cases, stagnant zones are prejudicial to bioethanol production, consequently, the velocity flow inside the bioreactor must be controlled.

An alternative to control the flow inside the reactor and get yields improvements is to modify structural parameters. The geometry definition has a great influence in the flow and consequently in the chemical reaction. Several methods have been used in order to determine the best way to optimize and obtain the best structural parameters. Two classes of search methods have dominated the estimation for structural parameters: the deterministic and stochastic methods [5].

The deterministic methods have some limitations when applied in some recent engineering problems. This kind of methods, despite being simpler to implement than stochastic methods, requires the calculation of second order derivatives and are strongly dependents on the initial approximation for the solution. In some cases there is more than one objective function, in other words, more than one parameter to be optimized. Sometimes these parameters are conflicting. For example, Mokeddem and Khellaf [6] tried to find an optimal feeding profile in feed-batch reactors, intending to maximise the amount of a protein used while simultaneously minimizing the volume of inducer employed, minimizing the material costs. In these cases, the most convenient way to

evaluate the best structural parameters is applying a stochastic method, like genetic algorithms.

Genetic algorithms are a class of stochastic methods solving complex optimization problems of the real world [7]. These methods are based on mechanisms of natural selection and population genetics [8].

The aim of this paper is to provide an optimized structure design for a continuous reactor applied to ethanol production from sugar cane juice. For this, different heights outlet tube were tested, intending minimize the media of the velocities inside the bioreactor. To find the ideal structural parameters, the flow simulations are coupled with an optimization process based on genetic algorithms.

2 Methodology

The ANSYS CFX 13.0 commercial software was used to simulate the flow inside the bioreactor. To perform the flow inside the bioreactor, the selected fluid was initially water. The modelling equations were the Navier-Stokes equations, given by eqn. (1) and eqn. (2):

$$\frac{\partial \rho}{\partial t} + \nabla.\left(\rho U\right) = 0 \tag{1}$$

$$\frac{\partial (\rho U)}{\partial t} + \nabla.\left(\rho UXU\right) = -\nabla p + \nabla \tau + SN \tag{2}$$

where τ is given by:

$$\tau = \mu\left(\nabla U + \left(\nabla U\right)^T - \frac{2}{3}\delta \nabla.U\right) \tag{3}$$

where U is the normal velocity, p the pressure, t the time, ρ de density and SN the forcing terms.

Due the computational efforts required, some simplifications in the flow and the mesh were assumed. This simplification does not affect the objective of this study. A steady flow simulation was performed. The flow was isothermal and incompressible. Turbulence effects were taken in account with the K-ε model, and the Navier-Stokes equations were solved using the Finite Volume method. The advection scheme was solved using High Resolution Method.

To perform the first simulation, an initial structure was used. Figure 1 shows the mentioned structure, composed by one tube for fluid entrance, one tube for fluid exit, and a cylindrical reservoir. The tube diameter for flow entrance and exit are both 5 cm, and the reservoir volume is 887 liters. These dimensions were considered according to the requirements for ethanol sugar cane production. The reservoir height was 1.5 m and the initial outlet tube was 1m.

A tetrahedral mesh with about 2787 nodes and 13963 cells was used in this simulation, as shown in Fig. 2.The inlet velocity applied was 0.1m/s^2 and a steady analysis was performed, and the outlet pressure was defined as 0 Pa.

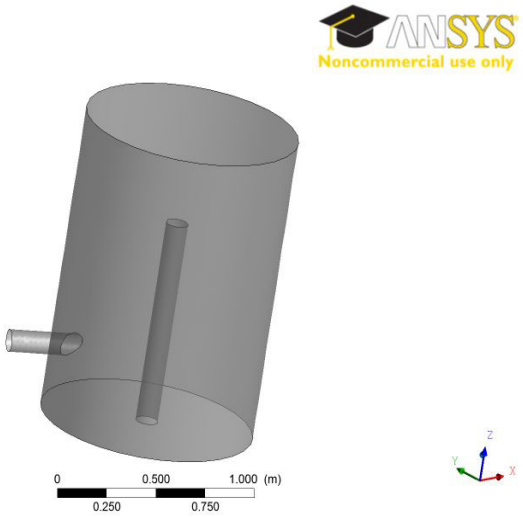


Figure 1: Structure used for initial tests.

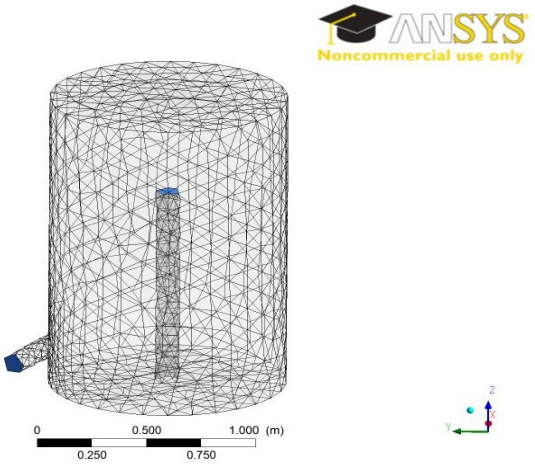


Figure 2: Domain discretization.

Besides the fluid analysis, the aim of this study was to optimize the bioreactor structure, improving the bioreactor efficiency by velocity controlling. As discussed previously, genetic algorithms are a useful tool in the optimizing process, allowing multi-objective optimization.

The results presented in this paper were performed applying the Multi-Objective Genetic Algorithm MOGA-II [9].MOGA II is an improved version of MOGA (Fonseca and Fleming [10]) and uses five different operators for

reproduction: selection, classical cross-over, directional cross-over, mutation and elitism. At each step of the reproduction process, one of the five operators is chosen and applied to the current individual. The directional cross-over assumes that a direction of improvement can be detected comparing the fitness values of two individual references, and this comparison is made in the same generation. This kind of cross-over helps the algorithm convergence. The MOGA II is implemented in the Mode FRONTIER 4.3 commercial software. This software allows the coupling between the Ansys CFX 13 solver and MOGA II, automating the process. Schematically, the coupling between CFX and Mode FRONTIER works in this order: structure generation, mesh generation, fluid solver, optimization and new parameters definition, new structure generation, and so successively, until the stop criterion is reached.

3 Results

MOGA II was employed to solve the optimization problems with the parameters shown in Table 1.

Forty-six height outlet possibilities were performed. All the height outlet tube possibilities and feasibilities are shown in Fig. 3.

Table 1: MOGA II parameters.

	Parameters
Number of generations	100
Probability of directional Cross-Over	0.5
Probability of selection	0.05
Probability of mutation	0.1

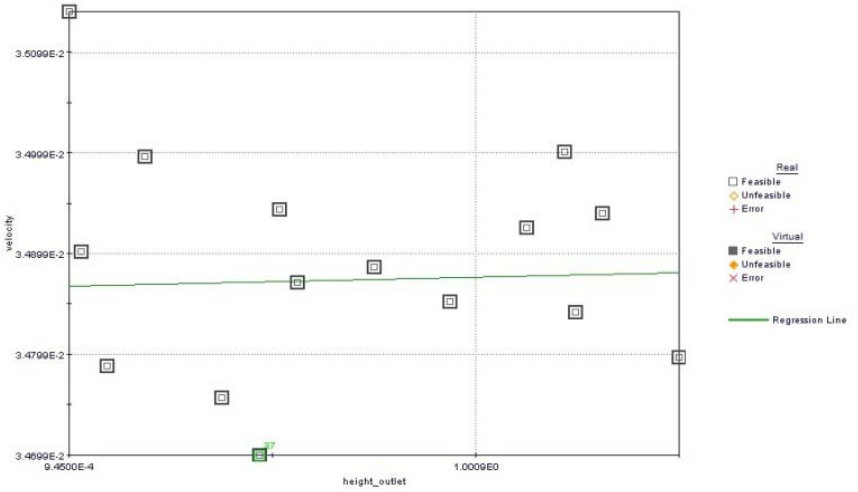


Figure 3: Height outlet tube and respective velocities.

After the evaluation of 46 different designs, the results applying MOGA II show that the ideal height outlet tube for minimize the media of velocities is 0.46 cm.

The streamlines for the optimized structure is shown in Fig. 4. The height outlet tube contributes for an organized flow inside the bioreactor.

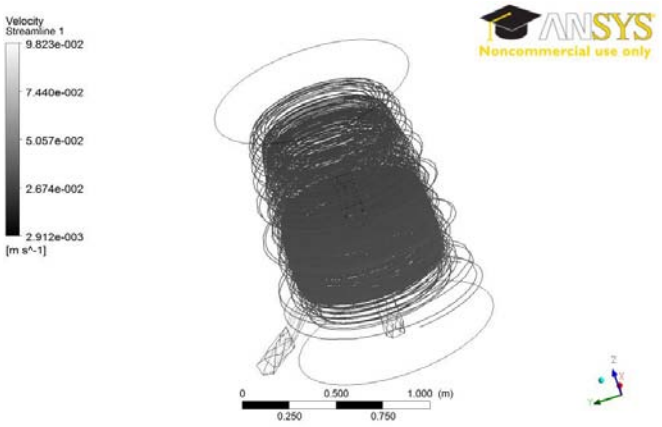


Figure 4: Streamlines for optimized case.

4 Conclusions

Propose new ways to obtain bioethanol from sugar cane and improve the existing ones are a great challenge for engineering applications. Considering a bioreactor applied in bioethanol production, this paper presents the results of an investigation about the influence of the structural parameters in the flow and, consequentially, in the chemical reactions in the bioethanol production. Were tested 46 different values of outlet tube height, and the best value was determined. To select the best parameters a Genetic Algorithm, MOGA II, was applied by Mode FRONTIER 4.3 commercial software coupled with Ansys CFX 13. Despite the initial tests have computational efforts limited, due the intention to execute the coupling between the CFD analysis and the optimization process, a good performance was obtained.

The knowledge about the optimized structure provides a bioreactor yield improvement, material economy and environmental impact reduction. Future works are concerned in tests validation and verification through an experimental prototype fabrication and improvements on CFD modeling.

References

[1] Patwardhan A. W., Joshi J. B., Fotedar S. & Mathew T., Optimization of gas-liquid reactor using computational fluid dynamics. *Chemical Engineering Science*.**60** , pp. 3081–3089 (2005)

[2] Xia J., Wang S., Zhang S. & Zhong J., Computational investigation of fluid dynamics in a recently developed centrifugal impeller bioreactor. *Biochemical Engineering Journal*.**38** , pp. 406–413 (2008)
[3] Micale G., Rizutti L. & Brucato A., CFD simulation of particle suspension height in stirred vessels. *Chemical Engineering Research and Design*.**82** , pp. 1204–1213 (2004)
[4] Shao X., Lynd L., Bakker A., LaRoche R. & Wyman C., Reactor scale-up for biological conversion of cellusosic biomass to ethanol. *Bioprocess Biosystem Engineering.* **33**, pp. 485-493. (2010)
[5] Rezende M. C. A. F., Costa C. B. B., Costa A. C., Maciel M. R. W. &Maciel R. F., Optimization of a large scale industrial reactor by genetic algorithms. *Chemical Engineering Science*.**63** , pp. 330–341 (2008)
[6] Mokeddem D. &Khellaf A., Optimal feeding profile in fed-batch bioreactors using a genetic algorithm. *International Journal of Production Research.* **48**, pp. 6125-6135. (2010)
[7] Kordabadi H. & Jahanmiri A., Optimization of methanol synthesis reactor using genetic algorithm. *Chemical Engineering Journal*.**108** , pp. 249–255 (2005)
[8] Dasgupta D., & Michalewicz, Z., (eds). *Evolutionary algorithms in engineering applications*, Springer: New York, 1997.
[9] Poloni A., Giurgevich A., Onesti L. & Pediroda V., Hybridization of a multi-objective genetic algorithm, a neural network and a classical optimizer for a complex design in fluid dynamics. *Computational methods in applied mechanics and engineering.* **186**, pp. 403-420. (2000)
[10] Fonseca, C.M. & Fleming P.J., Multi objective Genetic Algorithms. *Proc. of the IEE Coloquium on Genetic Algorithm for Control Systems Engineering*, IEEE: London UK, pp. 1–6, 1993.

This page intentionally left blank

Design of pipelines for high operating pressure by numerical simulations and experimental validation

Y. Theiner, H. Lehar & G. Hofstetter
Institute for Basic Sciences in Civil Engineering, University of Innsbruck, Austria

Abstract

For pipelines laid in difficult ground, as frequently encountered in Alpine regions, a pipe system, consisting of individual pipes, made of ductile cast iron, which are connected at joints at the construction site, are a favourable type of construction. The paper deals with the development of such pipelines for high operating pressures. Because the joints have to sustain both the high operating pressure and high axial forces assuring water tightness, they are a critical part of such pipelines. In this contribution the synthesis of numerical simulations and experimental validation will be presented as an efficient approach for developing such pipes. The dimensions of prototypes are determined on the basis of the results of fully three-dimensional FE-simulations. These prototypes are then used to check the design by ultimate load testes and to compare the numerical prediction with the measured response.

Keywords: penstock, pipelines, high operating pressure, ductile cast iron, sleeve joints, numerical simulation, material and geometric nonlinear behaviour, contact behaviour, load carrying behaviour, experimental validation.

1 Introduction

In Alpine regions, frequently difficult ground conditions are encountered for the laying of pipelines. A suitable construction type for difficult terrain is a pipe system, consisting of individual pipes made of ductile cast iron, which are connected at joints at the construction site.

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110081

The flexibility of the joints allows an easy adaptation of the course of the pipeline to irregular ground conditions, like bends in the trench (Figure 1). In the present case pipes of five meter length are connected at a joint by locking bars and water tightness is assured by a rubber sealing (Figure 2). There exists a broad range of applications for such pipelines including water supply lines for drinking water, for artificial snow machines [1] and for fire fighting services as well as penstocks for hydro power plants [2].

Figure 1: Laying of pipes in Alpine regions.

Figure 2: Connection of two adjacent pipes.

In collaboration with the Austrian company Duktus - Tiroler Rohrsysteme GmbH and the Unit of Strength of Materials and Structural Analysis of the University of Innsbruck, such pipelines are developed with the aims (i) to allow high operating pressures, (ii) to minimise material requirements, (iii) to develop geometric properties of the sleeve joints, allowing for an easy application of coatings and (iv) to permit simple connections of adjacent pipes at the construction site. The collaborative research project was funded by the Austrian Research Promotion Agency (FFG). This project follows a previous project on the development of pile-supported buried pipelines for sewage disposal at atmospheric internal pressure in regions with poor soil conditions [3, 4].

The aim of the present paper is to describe the employed approach for developing pipe systems for high operating pressures. The design approach is characterized by a synthesis of numerical and experimental investigations. The described pipe systems are developed for a large range of different diameters and wall thicknesses. For brevity, this paper exemplarily focuses on a particular pipeline, classified as DN 200 K13 according to the European standard ÖNORM EN 545 [5]. It is characterized by a pipe diameter of 200 mm and a wall thickness of 9.1 mm. The intended operating pressure is 100 bar.

2 Numerical model

The joints are the critical regions of such pipelines. Hence, it is important to properly reflect the behaviour of the connection of two adjacent pipes in a numerical model. In order to comply with this requirement, two adjacent pipes, denoted as spigot pipe and socket pipe in Figure 2**,** together with the locking bars and the rubber sealing are discretized by three-dimensional finite elements with quadratic shape functions. Because of two-fold symmetry only a quarter of the joint is discretized (Figure 3) and the respective symmetry conditions are applied as boundary conditions at the vertical faces of the FE-mesh.

The employed ductile cast iron is modelled as an elastic-plastic material of the von Mises-type with isotropic strain hardening. The Young's modulus is given as 154050 N/mm², the yield stress at 0.2% permanent strain and the tensile strength amount to 300 N/mm² and 620 N/mm^2, respectively, and the strain at rupture is about 7%. The given material properties refer to the respective test conducted on the joint DN 200 K13, which is described in the next section. In particular, the measured tensile strength exceeds the respective code requirement [5].

For describing the material behaviour of the rubber sealing the constitutive model by Mooney-Rivlin is employed. The latter is sufficiently accurate, because the strains in the rubber sealing do not exceed the limitations of this constitutive model. Contact between the pipes and the locking bars as well as between the pipes and the rubber sealing is taken into account by Coulomb-type friction laws. The complete 3D FE-model for the prototype DN200 K13 consists of about 1.2 Mio. degrees of freedom.

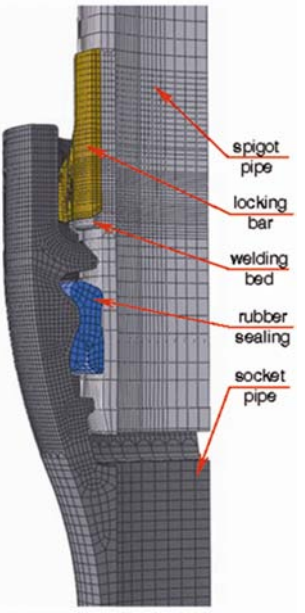


Figure 3: 3D FE-model of a quarter of the joint.

In the first step of the FE-analysis the installation of the rubber sealing is simulated. In this context, contact between the rubber sealing and the pipes has to be taken into account. The deformed configuration of the rubber sealing after installation, which was obtained from a simpler FE-simulation assuming axial symmetry, was integrated into the 3D FE-model.

In the second step the installation of the locking bars is simulated. In this context, contact between the locking bars and the pipes has to be taken into account.

Subsequently, the internal pressure is increased step by step. Figure 4 shows the predicted deformations of the joint (magnifying the displacements by a factor of 10) for pressure levels of 185 bar and 225 bar. According to Figure 4, the locking bar is clamped between the welding bed of the spigot pipe and the socket pipe. Failure of convergence of the numerical simulation occurs at an internal pressure exceeding 225 bar. Hence, the internal pressure of 225 bar indicates the ultimate load of the joint. At this pressure level, large domains of the joint are characterized by plastic material behaviour, in particular, the region in the vicinity of the contact area between the locking bar and the socket pipe and the region in the vicinity of the contact area between the welding bed of the spigot pipe and the locking bar.

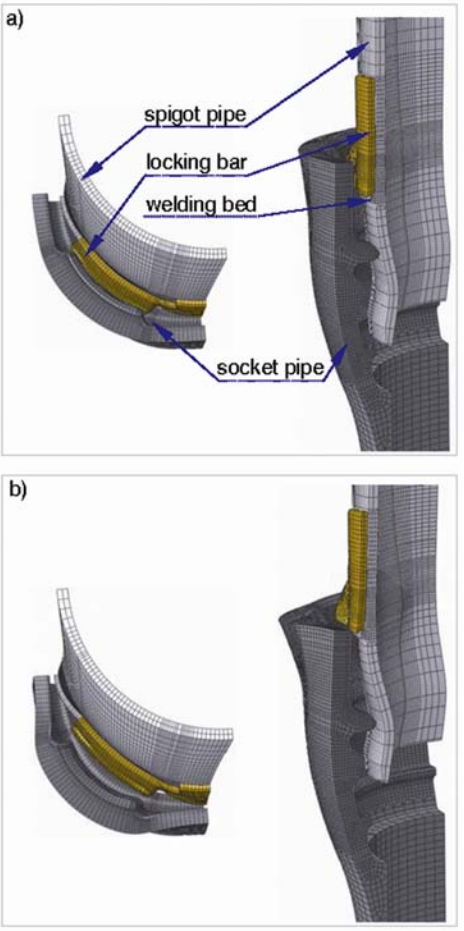


Figure 4: Predicted deformations of the joint at an internal pressure of a) 185 bar and b) 225 bar (displayed displacements are magnified by a factor of 10).

3 Experimental investigation

The test set-up for the experimental investigation, shown in Figure 5, was selected complying with the respective code requirements of the European standard ÖNORM EN 545 [5]. Two connected pipe segments of 1 m length each were supported by saddles. The latter were mounted on trolleys for reducing restraint effects due to relative displacements of the pipe segments at the joint, caused by the increasing internal pressure during the test.

Figure 5: Test set-up.

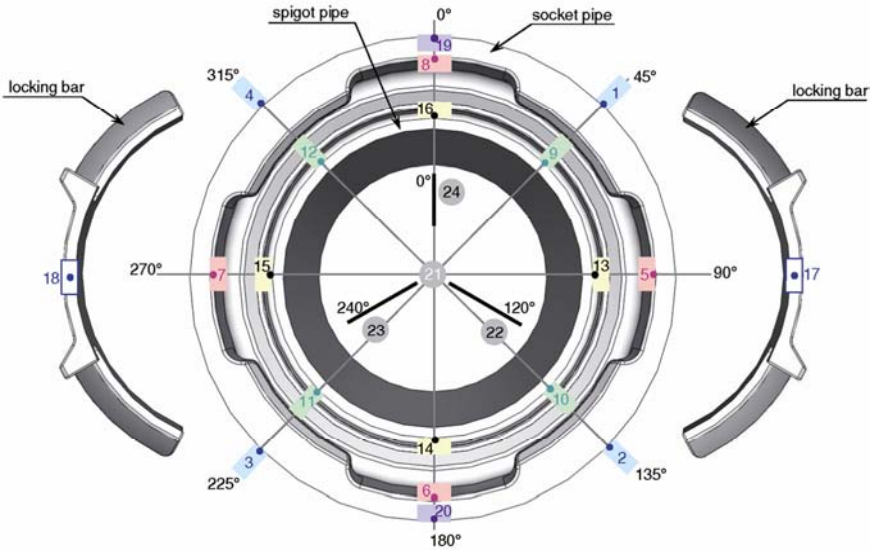


Figure 6: Arrangement of strain gauges and displacement transducers at the joint.

In order to monitor the structural behaviour of the joint with increasing internal pressure the pipe segments and the locking bars are equipped with strain gauges. Figure 6 shows the arrangement of the strain gauges applied at the socket pipe (measurement points no. 1 to 8), applied at the spigot pipe (measurement points no. 9 to 16) and applied at the two locking bars (measurement points no. 17 and 18), as well as the pressure transducer (measurement point no. 21) and three displacement transducers (measurement points no. 22 to 24). The latter are employed for measuring axial displacements of the socket pipe relative to the spigot pipe, thereby indicating a potential relative rotation of the pipe segments at the joint due to the application of the internal pressure during the test.

In the experimental investigation the internal pressure was increased up to 155 bar in three steps. This pressure was kept constant for one hour and, subsequently, it was increased further up to 185 bar. This pressure was maintained for two minutes, before the sample was unloaded. Failure of an air valve prevented the further increase of the pressure up to failure.

Figure 7 exemplarily shows the circumferential strains measured at the inner surface of the socket pipe at measurement points 5 to 8 (MP5 to MP8).

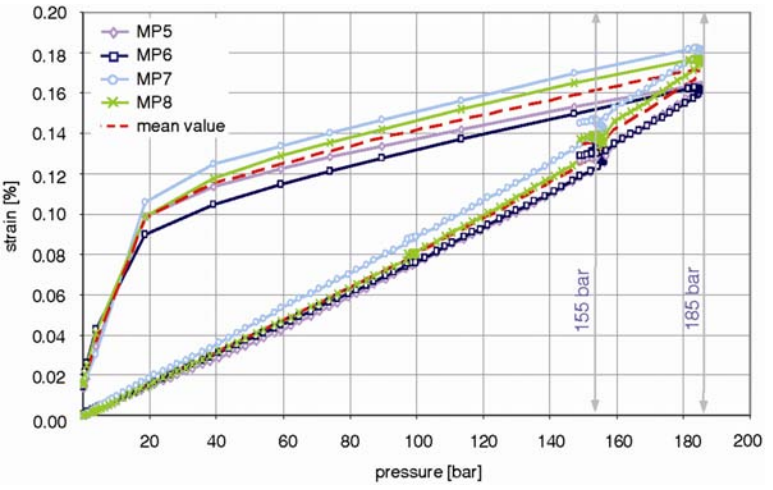


Figure 7: Measured circumferential strains at the inner surface of the socket pipe in terms of the applied pressure.

As can be seen from Figure 7, the differences of the measured circumferential strains at points MP5 to MP8 are relatively small. Hence, the measurement data confirm the desired uniform straining of the joint along its circumference. The diagram in Figure 7 shows permanent strains after complete unloading, which can be attributed to plastic material behaviour of the pipe segments at internal pressures, exceeding the design pressure of 100 bar considerably, and to friction forces and keying of the pipe segments.

In addition to the described experimental investigation with coaxial arrangement of the pipe segments, the same samples were used to carry out an experiment considering an intended kink of the axes of the two pipe segments of 4.8° at the joint.

4 Comparison of experimental and computed results

In order to assess the preliminary design of the joint the measured and the computed strains are compared. Here, for brevity, the comparison is restricted to the circumferential strains at the inner surface of the socket pipe at the pressure level of 185 bar. Figure 8 shows the respective computed strains for one quarter of the socket pipe segment. The two ellipses indicate the positions of the strain gauges MP 5 to MP 8 (see Figure 6). Hence, the mean value of the measured circumferential strains at the internal pressure of 185 bar, shown in Figure 7, can be compared with the computed strains at the locations, indicated by the ellipses in Figure 8. The distribution of the circumferential strains within the domain, enclosed by the ellipses, is characterized by a large gradient with a maximum value of 0.125%. Comparison of the mean value of the measured circumferential strains of about 0.17% (see the dashed line in Figure 7) with the computed ones shows satisfactory agreement, taking into account that the thickness of the tested pipe segments is somewhat larger than the minimum thickness considered in the numerical simulation.

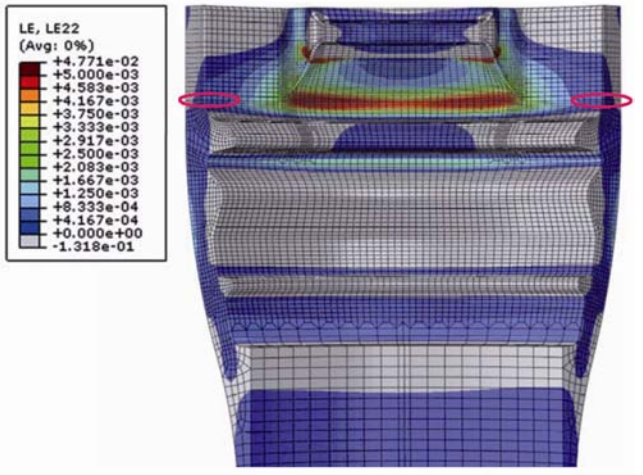


Figure 8: Computed circumferential strains at the inner surface of the socket pipe segment at an internal pressure of 185 bar.

5 Summary

This paper focused on the development of pipe systems for high operating pressures, consisting of individual pipes, made of ductile cast iron, which are connected at the construction site by joints, assuring water tightness by a rubber sealing. Since the joints are critical parts of such pipelines, a fully three-dimensional FE-model of a joint was developed for assessing the preliminary design of a joint by numerical simulations. It was employed for studying the load carrying behaviour of a joint up to failure. The numerical results, and, thus, the preliminary design, were evaluated by an experimental investigation on a prototype of a joint. In the test the structural behaviour was monitored by strain gauges and displacement transducers. Comparison of the computed and measured results and further numerical investigation of the impact of design modifications on the structural behaviour led to the final design of a joint.

Acknowledgements

Financial support of this applied research project by the Austrian Research Promotion Agency (FFG) is gratefully acknowledged. Furthermore, the work was supported by the Austrian Ministry of Science BMWF as part of the UniInfrastrukrurprogramm of the Forschungsplattform Scientific Computing at LFU Innsbruck.

References

[1] Titze, E., *Duktile Gussrohre für Beschneiungsanlagen,* Gussrohrtechnik, 37, pp. 13-17, 2003

[2] Titze, E., *Extreme Belastungen – Planung und Bau einer Turbinenrohrleitung aus duktilem Gusseisen unter Berücksichtigung bruchmechanischer Bemessungsverfahren. Gussrohrtechnik,* 32, pp. 58 ff, 1997

[3] Hofstetter, G., Lehar, H., Niederwanger, G., *Design of pile-supported buried pipelines by a synthesis of FE ultimate load analyses and experimental investigations,* Finite Element in Analysis and Design, 32, pp. 97-111, 1999.

[4] Lehar, H., Niederwanger, G., Hofstetter, G., *FE ultimate load analyses of pile-supported pipelines – tackling uncertainty in a real design problem*, in Analyzing Uncertainty in Civil Engineering, Eds.: Fellin, W., Lessmann, H., Oberguggenberger, M., Vieider, R., Springer-Verlag: Berlin, pp. 129-163, 2005.

[5] ÖNORM EN 545., *Ductile iron pipes, fittings, accessories and their joints for water pipelines. Requirement ant test method,* European Committee for Standardization (CEN), B-1050 Brüssel, 2007.

This page intentionally left blank

Mechanical behaviour of high metakaolin lightweight aggregate concrete

A. Al-Sibahy & R. Edwards
Department of Mechanical, Aerospace and Civil Engineering, University of Manchester, UK

Abstract

The work described in this paper forms part of a much larger investigation of the behaviour of a new developed type of lightweight aggregate concrete which would be suitable for use as load bearing concrete masonry units. The experimental work investigated the effect of high metakaolin (MK) content on the mechanical behaviour of newly modified lightweight aggregate concrete. 15% metakaolin and waste glass were used as a partial replacement for both ordinary Portland cement and natural sand. A medium grade expanded clay type Techni Clay was used as a coarse aggregate in the concrete mixes. Equal amounts of waste glass with particles sizes of 0.5-1 and 1-2 mm were used throughout this study. Unit weight, compressive and splitting tensile strengths were measured at various ages in accordance with the relevant British/EN standards. Fresh concrete properties were observed to justify the workability aspect. An assessment was carried out to indentify the pozzolanic activity of metakaolin material. The tests results were compared with the obtained results of controlled and lower metakaolin contents concretes which were previously studied. The tests results showed that metakaolin material had an explicit role in improving the strength and unit weight of modified lightweight concrete mixes. Compressive and splitting tensile strengths increase with an increase in the metakaolin content, while a counteractive behaviour was recorded for the unit weight aspect. The metakaolin material showed higher pozzolanic activity which was overcame there duction of compressive strength due to the negative effect of glass aggregate. However, the workability of concrete mixes degraded at higher metakaolin inclusion.
Keywords: lightweight aggregate concrete, metakaolin, waste glass, mechanical behaviour.

WIT Transactions on Modelling and Simulation, Vol 51,
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110091

1 Introduction

In recent years, much research has been undertaken with the aim of improving the mechanical properties of lightweight concrete. Improvements to the strength and unit weight are the major keys to evolution of the lightweight concrete behaviour. The strength characteristic represents the load bearing capacity of concrete to support the applied load. Unit weight is an indicator for its lightness and capability to afford thermal insulation when used within the internal and external building elements. This orientation also associated with environmental aspects in the construction processes, and the idea of environmental friendly solution showed up during the last few decades.

Due to huge amounts of waste glass produced in UK, crushed or ground glass aggregate is one of most effective environmental treatments to mitigate these wastes. The features of glass aggregates are granular particle shape, smooth surface texture and very low tendency to absorb the mixing water. These characteristics produce dry consistency and lower strength concrete mix. There is a slight increase in alkali–silica reaction ASR of glass aggregate concrete compared with normal aggregate concrete. However, this reaction can be reduced by using mineral by product materials.

In the concrete mixes, the crushed and ground glass aggregates are usually used as a partial replacement to the coarse or fine aggregate and Portland cement respectively [1].

Metakaolin is a versatile mineral by product material which can be used to improve the strength and durability of concrete mixes. It has a higher ratio of pozzolan and it is used as a construction material has sharply increased with consumption grows every year on a global basis. The production of metakaolin is by calcinations of kaolinitic clay at (650˚C–800˚C). It has surface area larger than Portland cement reach to about15 times with a specific gravity about 2.4 g/cm^3. It is usually used as a partial replacement for Portland cement. Since the use of metakaolin material reduces the emissions of CO_2during cement manufacture, it was considered in the industry for the production of precast concrete [2–4].

Expanded clay is an artificial lightweight aggregate which is considered as an essential approach to reduce the demand on the natural aggregate. Concrete mixes producing by this approach providing several advantages such as (1) Reduction of dead load of building that will reduce the dimensions of structural members giving a reduction in the quantity of reinforcement. (2) Lighter and smaller pre-cast elements which lead to less expansive casting, handling and transportation operations. (3) Providing large space due to reductions in the sizes of columns, beams and slabs dimensions. (4) High thermal insulation and increased fire resistance [5].

A newly modified lightweight concrete containing combination of expanded clay, waste glass and metakaolin materials was investigated, and, herein, the mechanical behaviour of highly metakaolin ratio is concerned.

2 Experimental program

The experimental program of this research aims to investigate the mechanical behaviour of high metakaolin lightweight concrete which would be suitable for use as a load bearing concrete walls.

A medium grade expanded clay type Techni Clay was used as coarse aggregate in the concrete mixes. This type of expanded clay was produced by the Plasmor Concrete Product Company. It has a typical moisture content and particle density of 20% ww and 550 kg/m^3 respectively. Figure 1 shows the grading of expanded clay.

Natural sand for building proposes was used as a fine aggregate to produce concrete mixes. Grading of the used sand is shown in Figure 1.

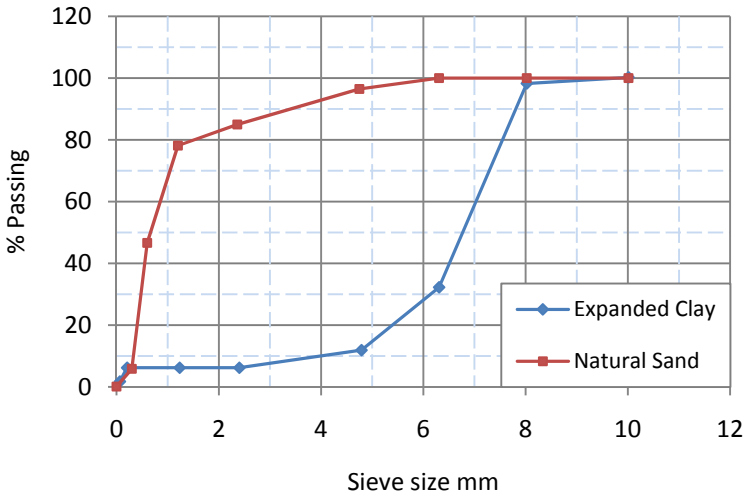


Figure 1: Grading of expanded clay and natural sand.

Waste glass with particles size of 0.5-1 and 1-2 mm was used as a partial replacement to natural sand with ratio of 15% by volume. The waste glass was provided by the Specialist Aggregate Ltd Company with a specific gravity of 2.52.

Metakaolin material was used as a partial replacement to the ordinary Portland cement with ratio of 15% by weight.

The mix proportions were 1: 0.76: 1.5 by volume which was equivalent to 1: 1.27: 0.63 by weight with 50 mm slump. Cement content and W/C ratio of controlled concrete (0% glass + 0% metakaolin) were 392 kg/m^3 and 0.45 respectively. Figure 2 shows the materials which were used in producing the lightweight concrete mixes.

The mixing operation was carried out according to BS EN12390-2 [6] using 0.1 m^3 vertical portable mixer. The fresh concrete was casted in moulds by

means of three layers. Each layer was completely compacted using a vibrating table until there was no further appearance of large air bubbles on the surface of the concrete. The excessive concrete above the upper edge was removed to achieve relatively smooth surface by using steel trowel.

Figure 2: The used materials in the experimental program: A- Expanded clay, B- Waste glass 1-2 mm, C- Waste glass 0.5-1 mm and D- metakaolin material.

After casting, the samples were kept out in the laboratory condition and coated by nylon sheet to ensure a humid air around the specimens. After 24 hours, the samples were demoulded, marked and immersed in a basin of water at a temperature 20 ± 2˚C until the date of test.

Before performing the experimental tests, the apparent moisture was removed from the specimens, and the surfaces were cleaned from any loose grit or any other materials that could be in contact with the loading plate.

The mechanical behaviour tests included unit weight, compressive and splitting tensile strengths. All these tests were conducted according to the relevant British /EN standards [7–11].

An average value of three specimens was adopted for each test result. Short and long term behaviours were investigated. The tests results were compared with the obtained results of controlled and lower metakaolin contents concrete samples which were previously studied.

3 Experimental results and discussion

3.1 Workability of concrete mixes

The approach of workability test was maintaining a constant amount of mixing water (water/cementitious ratio) for all metakaolin replacements. The worst workability was recorded for a mix containing 15% metakaolin. It required more effort using vibration process to cast. This may be related to the higher specific surface area of metakaolin particles compared to ordinary Portland cement. This caused a higher tendency to absorb the mixing water, sequentially accelerates the cement hydration and increases the micro-hardness of the paste-aggregate interfacial zones [3]. The lower metakaolin contents seem to be adequate in term of workability.

3.2 Unit weight

The unit weight behaviour of controlled and modified concrete mixes at different test ages is shown in Figure 3. Modified concrete mixes showed slightly lower density than the controlled mix. This is expected as the specific gravity of waste glass at 2.52 is less than that for natural sand at 2.65. Since the glass replacement was by volume of natural sand, the resulting unit weight of concrete will be reduced. Among the used metakaolin replacements, the ratio of 10% showed the higher decreasing in value of unit weight, while 15% metakaolin replacement produced the heaviness modified concrete mix. These results confirm with the previous studies [4].

In general, all lightweight aggregate concrete mixes exhibited a continuous decrease in density with increasing age of test except at 28 days for the modified

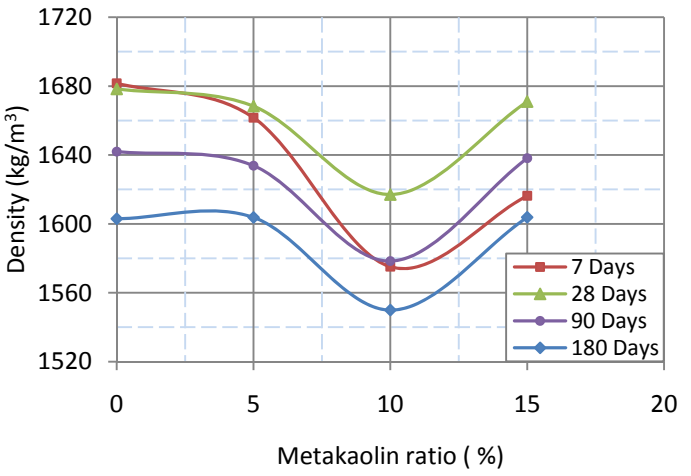


Figure 3: Effect of metakaolin ratio on the density of lightweight concretes.

concrete mixes (Figure 4). This behaviour may be attributed to two phenomena: the first is the consumption of the mix water by hydration processes, the second is reduction of free water from the gel pores by evaporation. Increasing density of modified concrete mixes at age of 28 days compared with that at 7 days could be explained by the higher hardness of these mixes due to pozzolanic reaction of metakaolin material associated with their capability to hold the mixing water.

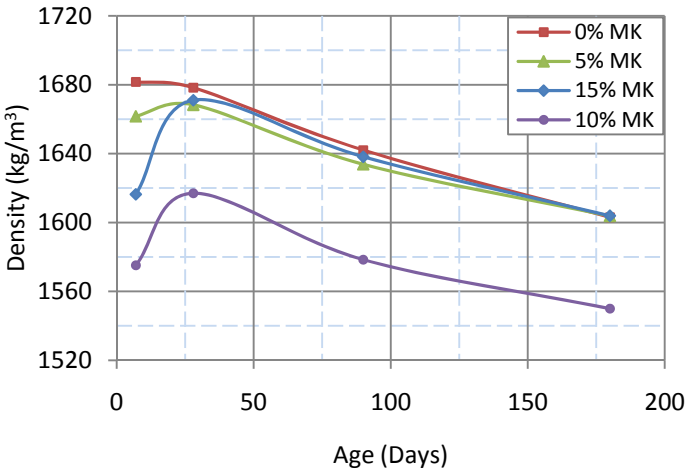


Figure 4: Density-age relationship of different metakaolin ratios.

3.3 Compressive strength

The tests results of compressive strength are presented in Figure 5. This figure indicates that the modified concrete mixes possess higher values of compressive strength than the controlled mix. This performance could be related to the filler and pozzolanic reaction of metakaolin material which improves growth of compressive strength, in addition to the acceleration of Portland cement hydration [3, 4]. Due to the high silica content of the pozzolanic material, it has capability to react with the calcium hydroxide which liberated from hydration process to produce gel of calcium silicate hydrate (CSH). The CSH component is the essential responsible of the concrete strength [12–14].

$$SiO_2 + Ca(OH)_2 \rightarrow CSH \quad (1)$$

For all concrete mixes, the compressive strength increase with the increasing of time as shown in Figure 6. Percentage increases in compressive strength for concrete mixes containing 5%, 10% and 15% metakaolin material at age of 90 days compared with the controlled concrete at the same age were 4.45%, 10.71% and 21.6% respectively.

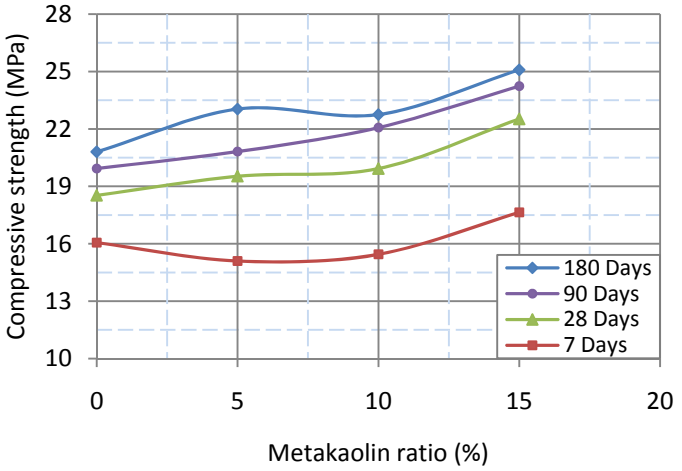


Figure 5: Effect of metakaolin ratio on the compressive strength of lightweight concretes.

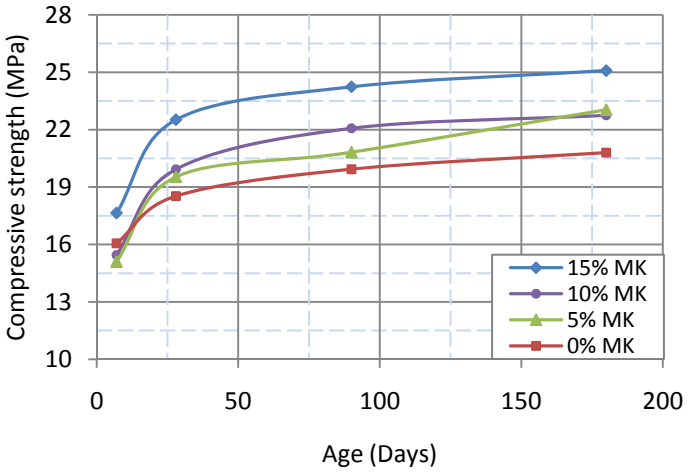


Figure 6: Compressive strength-age relationship for various metakaolin ratios.

The expression suggested by BS EN 1992-1-1 [15] to estimate the compressive strength at time (t) has been checked using SPSS statistic software with a nonlinear regression scheme to investigate its consistency with the results of high metakaolin lightweight aggregate concrete, the estimated expression is:

$$f_{cm}(t) = \beta_{cc}(t) f_{cm} \tag{2}$$

$$\beta_{cc}(t) = exp\left\{s\left[1 - \left(\frac{28}{t}\right)^{1/2}\right]\right\} \quad (3)$$

where

(s) depends on the type of cement and curing temperature, it is in range of (0.2-0.38). However, (s)value of 15% metakaolin concrete mix was 0.29 with R-squared of (≈80%).

The pozzolanic reactivity of metakaolin material which is described in BS EN 196-5 [16] was measured according to values of compressive strength as in [17]. The specific strength ratio R which is an indicator of contribution of mineral admixture to strength is defined as

$$R = f_c/p \quad (4)$$

where f_c is the compressive strength and p is the hydraulic cement or mineral admixture percentage. By eliminating the reduction effect of the glass aggregate on the values of compressive strength, the contribution of the pozzolanic effect of metakaolin R_p to concrete strength is given by Eq. (5).

$$R_p = R_M - R_C \quad (3)$$

where R_M the contribution of unit hydraulic cement when metakaolin is used and R_C the contribution of unit hydraulic cement to the concrete strength without using metakaolin material.

The index specific strength K is the ratio of R_M to R_C . The contribution of pozzolanic effect P to concrete strength can be expressed as:

$$P = (R_P/R_M) \times 100 \quad (4)$$

The values of R, R_p, K and P for controlled and modified concrete mixes at ages of 7, 28, 90 and 180 days were calculated and presented in Table 1.

The benefit of metakaolin on the strength of concrete can be clearly seen. This benefit increases with an increase of metakaolin content, where up to 30% of the concrete strength resulted from the pozzolanic action of metakaolin material. For all modified concrete mixes, the improvement of compressive strength continues for long term duration (180 days), this is in agreement with the results of [4, 17]. On the other hand, at earlier ages (7 days) the compressive strength decreased for concrete mixes containing 5% and 10% metakaolin. This may be due to the negative effect of glass aggregate on the strength which reduces the amount of hydration products at this time.

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)

Table 1: Calculated values of R, R_p, K and P for controlled and modified concrete mixes.

Pozzolanic parameter	Controlled mix				15% G +5% MK				15% G +10% MK				15% G +15% MK			
	7 days	28 days	90 days	180 days	7 days	28 days	90 days	180 days	7 days	28 days	90 days	180 days	7 days	28 days	90 days	180 days
f_c	16.06	18.53	19.93	20.8	15.1	19.53	20.82	23.04	15.45	19.93	22.07	22.75	17.64	22.52	24.24	25.09
R	0.160	0.185	0.199	0.208	0.158	0.205	0.219	0.242	0.171	0.221	0.245	0.252	0.207	0.264	0.285	0.295
R_p	0.000	0.000	0.000	0.000	-0.001	0.020	0.019	0.034	0.011	0.036	0.045	0.044	0.046	0.079	0.085	0.087
K	1.000	1.000	1.000	1.000	0.940	1.053	1.044	1.107	0.962	1.075	1.107	1.093	1.098	1.215	1.215	1.206
P(%)	0.000	0.000	0.000	0.000	-1.03	9.86	9.04	14.23	6.446	16.32	18.70	17.71	22.61	30.05	30.09	29.53

G = glass aggregate, MK = metakaolin material

3.4 Splitting tensile strength

Improvements in splitting tensile strength were observed for modified concrete mixes relative to controlled mix as shown in Figure 7. These may be related to an increase in compressive strength of theses mixes resulting from the pozzolanic action of metakaolin material. The splitting tensile strength increases whenever metakaolin content increased. However, a slight decrease was recorded for the mix containing 10% metakaolin at age of 180 days relative to concrete mix of 5% metakaolin. This behaviour could be due to the surface effect of specimens. Percentage increases in splitting tensile strength of concrete mixes containing 5%, 10% and 15% metakaolin material at age of 120 day compared with the controlled concrete mix were 11%, 12.5% and 26% respectively.

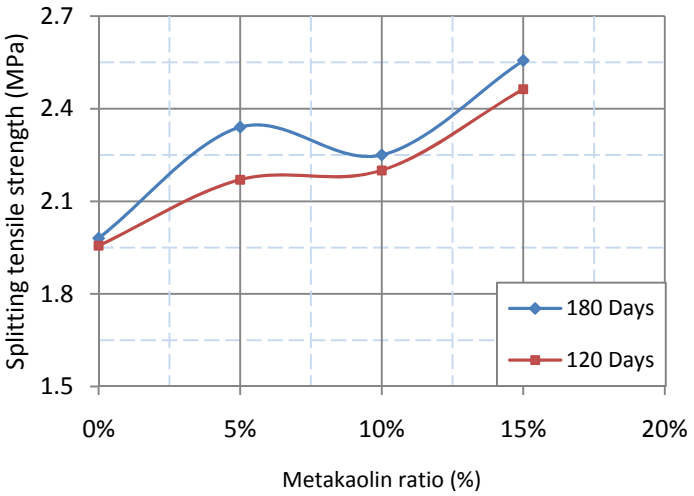


Figure 7: Effect of metakaolin ratio on the splitting tensile strength of lightweight concretes.

4 Conclusions

The experimental measurements of high metakaolin lightweight aggregate concrete were reported, discussed and then compared with the mechanical behaviour of lower metakaolin contents. The main results can be summarized as follow:

- The high metakaolin lightweight aggregate mix exhibited inferior workability due to its role in acceleration the hydration of cement, while an adequate workability was recorded to the lower metakaolin contents.
- A slightly decreasing in density values of modified concrete mixes compared with the controlled mix and the higher reduction was at 10% metakaolin content for all test ages.

- A vast improvements in value of compressive strength were achieved when metakaolin material is used as a partially substitute to ordinary Portland cement.
- The pozzolanic activity of the metakaolin material increases with an increase of its content, and it reached up to 30% from the contribution aspect of compressive strength.
- The pozzolanic action of metakaolin material continues improved the long term behaviour of lightweight aggregate concrete.
- Consistent positive improvements in tensile strength to that in compression feature were also recorded at short and long term behaviours.

Acknowledgements

The authors would like to express their gratitude to the Plasmor Concrete Product Company, staff of the Civil Engineering Laboratory of the University of Manchester and the Ministry of Higher Education and Scientific Research in Iraq.

References

[1] Miao, L., *Incorporating ground glass in self-compacting concrete*, Construction and Building Material, 25, 919-925, 2011.
[2] Eva, V., et al., *High performance concrete with Czech metakaolin: Experimental analysis of strength, toughness and durability characteristics*, Construction and Building Material, 24, 1404-1411, 2010.
[3] Parande, A., et al., *Metakaolin: a versatile material to enhance the durability of concrete – an overview*, Institute of Civil Engineering, doi: 10.1680/stco, 2009.
[4] Wild, S., et al., *A relative strength, pozzolanic activity and cement hydration in superplasticised metakaolin concrete*, Cement and Construction research, 26, No.10, 1537-1544, 1996.
[5] Kayali, O., *Fly ash lightweight aggregates in high performance concrete*, Construction and Building Materials, 22 (12): 2393-2399, 2008.
[6] BS EN 12390-1, *Making and curing specimens for strength test*, British Standards: 1-12, 2009.
[7] BS EN 206-1, *Concrete: Specification, performance, production and Conformity, British* Standards: 1-74, 2000.
[8] BS EN 12390-7, *Testing hardened concrete, Part 7: Density of hardened concrete*, British Standards: 1-14, 2009.
[9] BS EN12390-3, *Testing hardened concrete, Part 3: Compressive strength of test specimens'* British Standards: 1-22, 2009.
[10] BS EN 12390-6, *Testing hardened concrete, Part 6: Splitting tensile strength of test specimens*, British Standards: 1-14, 2009.
[11] BS ISO 1920-10, *Testing of concrete, Part 10: Determination of static modulus of elasticity*, British Standards: 1-12, 2009.

[12] Roy, D., et al., *Effect of silica fume , metakaolin, and low calcium fly ash on chemical resistance of concrete*, Cement and Concrete Researches, 31, 1809-1813, 2001.
[13] Batis G, P., *The effect of metakaolin on the corrosion behaviour of cement mortar*, Cement and Concrete Composition, 27, 125-130, 2005.
[14] Kim, H., et al., *Strength properties and durability aspects of high strength concrete using Korean metakaolin*, Construction and Building Materials, 21: 1229-1237, 2007.
[15] BS EN 1992-1-1, *Eurocode 2: Design of concrete structures: Part 1-1: General rules and rules for buildings*, British Standards: 1-230, 2004.
[16] BS EN 196-5, *Method of testing cement: Part 5: Pozzolanicity test for pozzolanic cement*, British Standards: 1-16, 2005.
[17] Kiruthia, J., et al., *Self-compensating autogenous shrinkage in PC-MK-fly ash pastes*, Advance in Cement Research, 12, No.1, 35-43, 2001.
[18] Paulo, C., et al., *Effect of Portuguese metakaolin on hydraulic lime concrete using different curing conditions*, Construction and Building Materials, 24: 71-78, 2010.

Experimental validation of numerical modelling of the bridge–track–moving train system

P. Szurgott[1], M. Klasztorny[1], J. Grosel[2] & Z. Wojcicki[2]
[1]*Department of Mechanics and Applied Computer Science, Military University of Technology, Poland*
[2]*Institute of Civil Engineering, Wroclaw University of Technology, Poland*

Abstract

A new methodology of physical and FE modelling and simulation of bridge–track –moving train (BTT) systems has been developed with the use of commercial CAE systems. A methodology is related to composite (steel-concrete) bridges, ballasted tracks and high-speed trains. In the methodology, Altair HyperMesh, LS-DYNA, LS-PrePost and HyperView software was applied. The methodology is based on homogenization of reinforced concrete (RC) platform slab, RAIL_TRACK and RAIL_TRAIN LS-Dyna's modules for simulating the moving train–track interaction, non-linear modelling of rail fastenings and crushed stone ballast, application of cylindrical and revolute constrained joints and discrete springs and dampers for modelling suspensions in rail-vehicles. For experimental validation of numerical modelling and simulation of BTT systems, the KNI 140070 composite viaduct and the EuroCity EC 114 train moving at 160 km/h have been selected. The experimental setup contained Keyence LK-G 157 system (CCD laser displacement sensors), PULSE system (acceleration sensors), and PHANTOM v12 high-speed camera. According to the experiment plan, selected vertical displacements and vertical and horizontal accelerations vs. time were measured. The simulated time-histories of displacements and accelerations have been compared to respective experimental diagrams. The results have proved that the validation is positive.

Keywords: railway bridge, ballasted track, high-speed train, numerical modelling, simulation, experimental tests, validation.

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110101

1 Introduction

Nowadays, serious problems with durability protection of bridge superstructures, tracks and approach zones loaded by high-speed trains are observed. First of all, it results from complexity of bridge-track–moving train (BTT) systems, for which nonlinear models are described by a huge number of parameters. Many of these parameters, describing fasteners, ballast, subsoil layers, rail-vehicles' suspensions, track irregularities, settlements etc., are only estimated and difficult for identification. Producers and research institutions involved in modern high-speed trains do not bring to light structural details, values of parameters or their research results. These circumstances make exact prediction of dynamic response of bridges to moving trains very difficult.

In the 2nd half of the 20th century scientists mostly developed analytic–numerical methods in dynamics of railway bridges, summarized in monographs (Klasztorny [1, 2]). Simple problem-oriented computer codes were created and used for simulations. At present, one may observe various numerical approaches to dynamics of railway bridges but commercial CAE systems based on FEM are, in general, not used in this field by Klasztorny [2], Yang *et al.* [3], Cheng *et al.* [4], Au *et al.* [5], Zhang *et al.* [6], Song *and Choi* [7]. Summing up, assuming vibrations of BTT systems may be considered as 3D but symmetric with respect to the vertical longitudinal plane of symmetry. Applications of advanced CAE systems in dynamics of bridges are at early stage of its development.

2 Description of the KNI 140070 composite viaduct, the ballasted track and the EC 114 PRAHA train

The composite (steel–concrete) viaduct No. KNI 140070, located at 200.794 km on the Polish Central Main (PCM) Line No. 4-E 65, has been selected for numerical modelling, simulation and validation [8]. There is considered the bridge span under track No. 1 (fig. 1).

After recent modernization, the viaduct has the $k = +2$ rail-line classification coefficient. A track spacing equals 4.57 m, the spans are of 14.40 m theoretical length and 15.34 m total length. Four main beams are 0.77 m high and made of

Figure 1: The modernized viaduct No. KNI 140 070 composed of two separate spans, located at 200.794 km on the PCM Line, Poland: bottom view on the main beams and bracing.

St3M steel. Bottom flanges have been enforced with additional cover plates. The thickness of a new RC platform ranges from 0.29 m in the track axis to 0.25 m at the side wall. The platform is made of C35 concrete reinforced with AII/18G2-b steel rebars. The side wall is made of C30 concrete and has vertical dilatations at ¼, ½, and ¾ of the span length. The RESTON rolling bearings (on the left support) can shift up to ±50 mm in the longitudinal direction. Bearings under the left inside main beam are unmovable in the lateral direction; the remaining bearings can displace in the lateral direction up to ±20 mm.

A scheme of the longitudinal section of the KNI 140070 viaduct is depicted in fig. 2 where all elements taken in the FE modelling are marked, i.e. the homogenized platform (the slab and the walls), the main beams set, the vertical ribs welded to webs of the main beams, the horizontal bearing plates welded to the bottom flanges of the main beams (over the bearings).

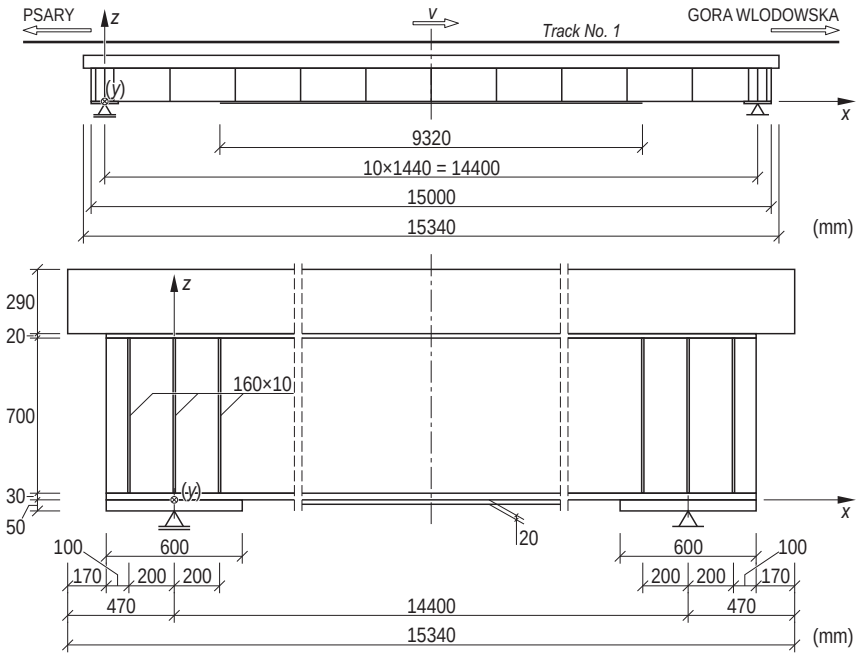


Figure 2: The longitudinal section of the KNI 140070 viaduct in the *xz* plane.

The track structure before modernization consists of UIC 60 rails, PS-94/SB/ /UIC-60 sleepers with SB3 fasteners, and crushed stone ballast of the first class. The ballast layer is 0.35 m thick under sleepers. The embankment in the approach zones contains cement-stabilized subsoil while outside the approach zones a 0.2 m thick sand-gravel mix top layer has been applied. The tracks over the viaduct are quasi-rectilinear; there exists a horizontal arc of $R = 4000$ m radius. Schemes of the longitudinal and cross sections of the ballasted track in the viaduct area are shown in fig. 3.

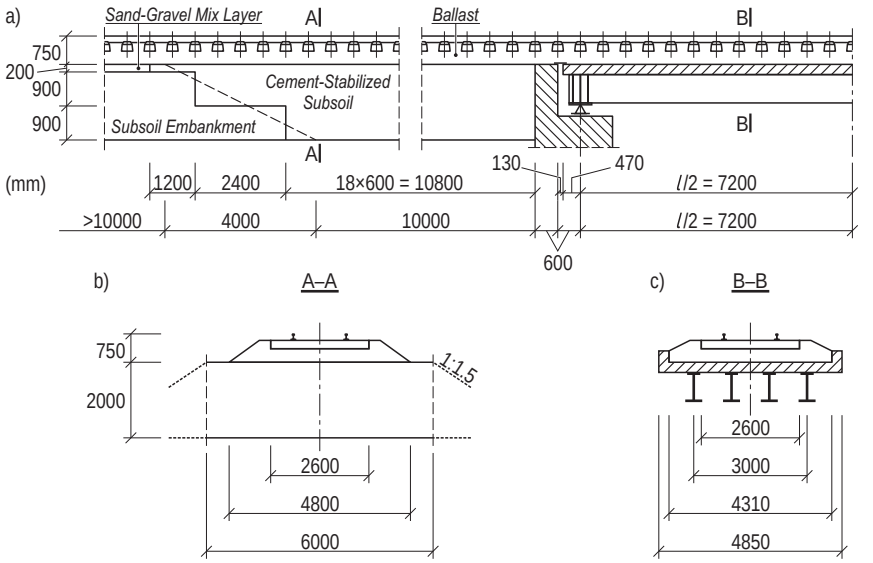


Figure 3: The ballasted track in the KNI 140070 bridge zone: the longitudinal section (a); the cross-section in the approach zone (b); the cross-section over the bridge span (c).

The EC 114 PRAHA EuroCity train, moving at 160 km/h over the bridge, has been taken into consideration. The trainset consists of 6 units. Car lengths, centre pins' distances and wheel sets' distances are reflected in fig. 4. All units are equipped with two independent two-axle bogies.

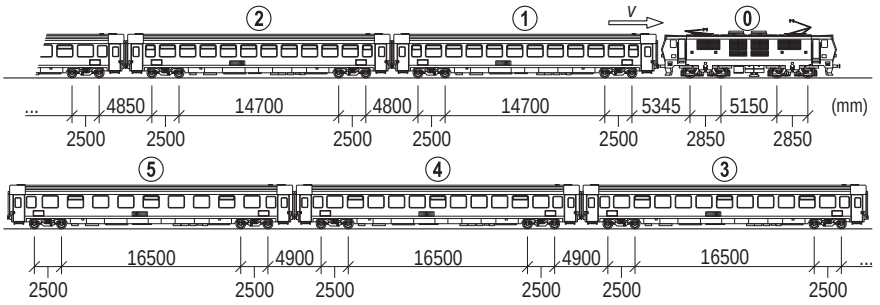


Figure 4: A scheme of EC 114 PRAHA Euro City trainset.

3 Physical and numerical modelling of the BTT system

The following concept has been developed in physic modelling of the viaduct. Since the reinforcement of the RC platform is distributed quasi-uniformly in the specified platform sectors, the slab and the wall of the platform can be homogenized. After homogenization the slab and the wall are reflected by linear

viscoelastic orthotropic materials described by 3 Young's moduli, 3 Poisson's ratios, and 3 shear moduli in each subarea. A cross-section of the platform slab is approximated by a rectangle of 0.27 m height. The platform is symmetrised via replacing a single dilated wall with two smaller dilated walls on both sides of the platform slab. The original and approximate structures have the same values of mass and cross-section area. The material of steel main beams is linearly viscoelastic and isotropic. The vertical and horizontal bracing in the main beams' set is neglected. All bearings on the abutments are assumed to be unmovable in the lateral direction. The temperature changes are not taken into consideration.

Figure 5 depicts the original (dashed lines) and symmetrised (continuous lines) cross-sections, the longitudinal section of the viaduct and the symmetrised reinforcement close to the original one. Four homogenised subareas in the cross-section have been distinguished.

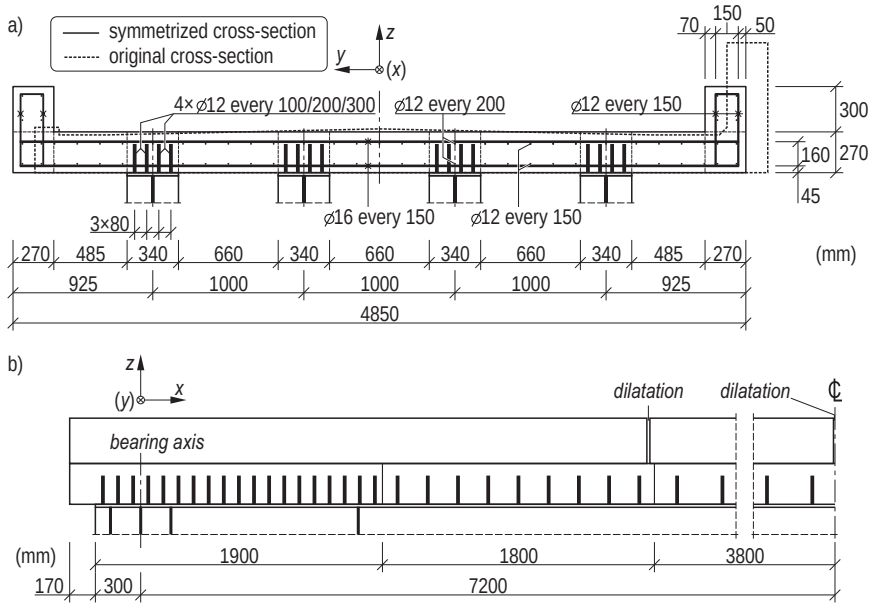


Figure 5: The original and symmetrised cross-sections (a), and the longitudinal section (b) of the KNI 140070 viaduct and the symmetrised reinforcement.

The FE model of the bridge superstructure has been created in Altair HyperMesh software (fig. 6). The numerical model of the bridge superstructure consists of 3896 4-node shell elements (steel main beams) and 5568 8-node 48 DOF solid elements (the homogenized RC platform divided into orthotropic parts). Cover plates of the beam bottom flanges are modelled as an additional layer of finite shell elements. Roller bearings on the left bridge abutment and pivot bearings on the right bridge abutment have been reflected by respective constraints.

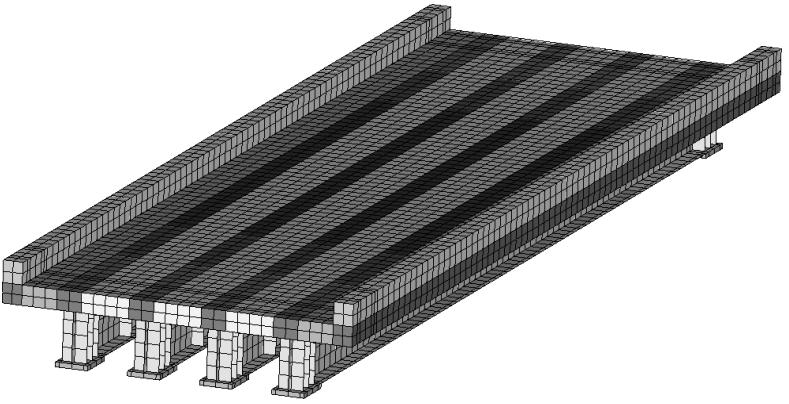


Figure 6: The FE model of the KNI 140070 viaduct superstructure.

The following assumptions have been made in physic modelling of the track. The rail-line axis is rectilinear, and — in the unloading state — the rails are rectilinear. No rail surface irregularities appear. Vibrations of the track are small and symmetric with respect to the vertical *xz* plane. The rails are prismatic beams deformable in flexure and shear, made of linearly viscoelastic material. Layers of the embankment are considered as a linearly viscoelastic material continuum.

Rail fasteners were simulated using massless one-dimensional discrete non-linear spring and damper elements oriented vertically. The embankment has been reflected approximately by a rectangular prism with unmovable side and bottom boundary surfaces and meshed using 8-node 24 DOF solid elements. Sleepers are modelled as rigid beams vibrating only vertically, using finite beam elements and respective constraints. The ballast layer has been divided into cubicoid columns in coincidence with FE mesh of the parts under the ballast (9 ballast columns under each sleeper). Each ballast column was reflected by a vertical set of nonlinear spring and damper elements. The lumped mass distribution for the ballast has been put into the bottom set of the nodal points contacting the platform slab and the top subsoil layers.

Values of geometrical and mechanical parameters of the ballasted track parts are extracted from Klasztorny [1, 2], Niemierko *et al.* [11], and refs. [10, 12, 13].

RAIL_TRACK and RAIL_TRAIN modules available in LS-DYNA [15] were applied for approximate modelling the train-track interaction (without simulation of wheels' rotation). The wheel-rail contact stiffness amounts to 2 MN/mm as suggested in ref. [15]. Hughes-Liu beam elements (2-node elements with 12 DOFs [15]) were used for FE modelling of rails bent in the vertical planes. In order to declare a set of integration points for the rail cross-section, the INTEGRATION card has been applied. For each rail a substitute double-tee asymmetric cross-section was assumed, denoted in ref. [15] as Type 10: I-Shape 1. The actual values of the centre-of-gravity location, the area and the geometrical moment of inertia have been saved with respect to the horizontal principal axis of the cross-section.

The FE numerical model of the ballasted track has been created in HyperMesh and LS-PrePost software. The main dimensions of the track, the abutments and the embankment are depicted in fig. 3, whereas one of the FE mesh schemes of the track are partly reflected in fig. 7.

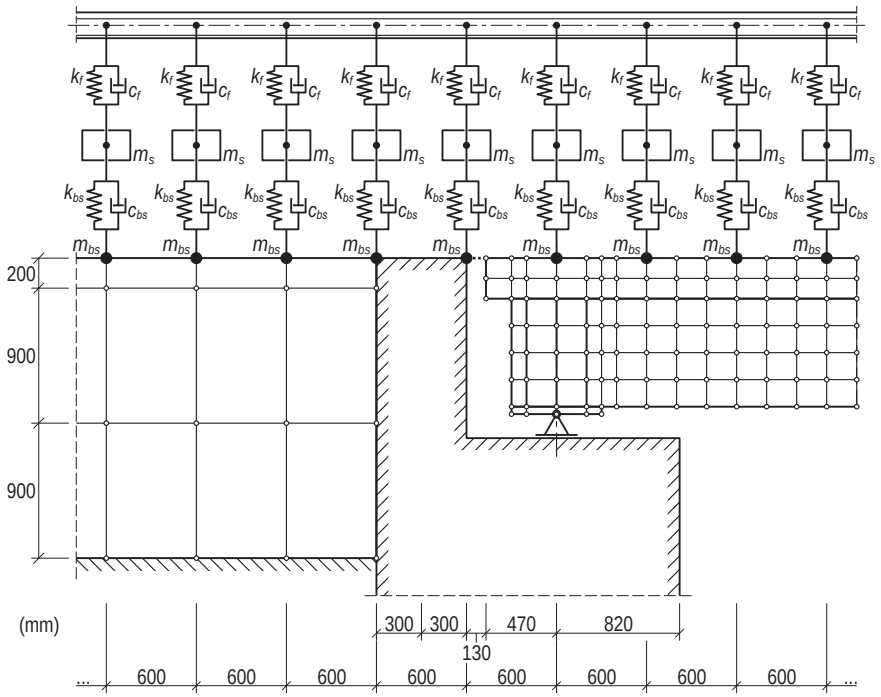


Figure 7: The side view on the physical and FE model of the track in the left abutment zone of the KNI 140070 viaduct.

The total length of the track section modelled numerically was equal to 810 m, and contains the following sections: the initial train position (with zero static wheel pressures), the zone of increasing the static wheel pressures up to the full values, the train-track vibration stabilization zone (lasting 1 sec), 60 m long main zone (including the approach zones and the bridge), the zone of bridge free vibrations (lasting 1 sec), the final train position zone. In total, the FE track model contains ~141,800 beam, shell, brick and discrete elements and ~21,700 point mass elements.

Modelling of the EC 114 trainset has been performed in LS-PrePost software [15] under assumption that vibrations of the train units are symmetric with respect to the main longitudinal vertical plane of symmetry. A numerical model of the EC 114 trainset consists of the following components: carbodies, bogie frames, wheel sets, and vertical massless discrete linear viscoelastic elements reflecting the primary and secondary suspension systems. All mass components were modelled using shell and beam elements treated as rigid bodies. Wheel sets have been reflected

by vertical rigid beam elements. Respective constraints have been put into the system via incorporating translational CONSTRAINED_JOINT_CYLINDRICAL and rotational CONSTRAINED_JOINT_REVOLUTE elements [15]. A side-view scheme of the 3D model of the train units is shown in fig. 8. The values of mechanical parameters of the EC 114 train units have been determined based on Klasztorny [1] and Bachmann [14].

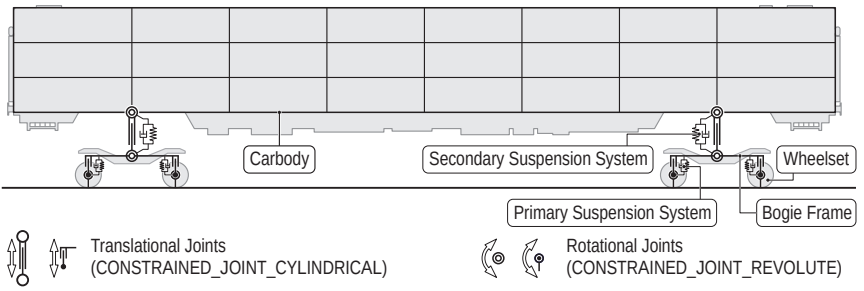


Figure 8: A side-view scheme of the 3D model of the EC 114 train units.

The FE models of carbodies, bogie frames and wheel sets were created at full conformity between actual vehicles and the numerical models with respect to their masses and principal mass moments of inertia. In total, the FE model of the 6-unit EC 114 train contains ~950 beam, shell and discrete finite elements and ~50 point masses. In the simulations, the DYNAMIC_RELAXATION option [15] has been replaced with loading the system by a set of vertical forces put in the moving vehicle-rail contact points according to the formula:

$$P(t) = \frac{P_0}{2}\left(1 - \cos\frac{\pi t}{t_0}\right), \tag{1}$$

where P_0 is the static pressure of a single wheel on the rail head, $t_0 = 2$ sec is a time of increasing of the static pressures up to the full values ($0 \le t \le 2$ sec).

A constant service velocity of the vehicle FE model was declared in two steps, with options INITIAL_VELOCITY for $t = 0$ PRESCRIBED_MOTION_RIGID for $t > 0$ applied for all carbodies and bogies FE models [15].

Selected output quantities were registered using HISTORY_NODE_SET and HISTORY_NODE_SHELL options [15]. The sampling frequency amounts to 20 Hz in the zone of increasing the static wheel pressures, 200 Hz in the train-track vibration stabilization zone and 1,000 Hz in the 60 m long main zone. The computations have been made using the 120-P supercomputer. At the service velocity 160 km/h the real time equals 14.4 sec, while the CPU time amounts to ~90 hrs. Selected time-histories for displacements and accelerations were created using the LS-PrePost software.

4 A plan of the dynamic experimental test

The KNI 140070 composite viaduct located on the Polish Central Main Line has been selected for validation tests. Spatial vibrations of the viaduct induced by EC 114 EuroCity express train moving at 160 km/h on November 9, 2009 over the viaduct were both registered experimentally and simulated numerically.

The following types of measurements have been executed:

- measurements of selected vertical displacements vs. time using laser sensors,
- measurements of selected vertical and horizontal accelerations vs. time using accelerometers,
- measurements of the train service velocity using a high-speed camera.

Locations of the measurement points are marked in fig. 9. The vertical vibrations in displacements were measured at the midspan under webs of the main beams constituting a pair under the internal rail. KEYENCE CCD Laser Displacement Sensors of LK-G 157 type have been applied, at distance of 150 mm from the beam surfaces, with sampling frequency of 2000 Hz and measuring accuracy of ±0.03%. The vertical and horizontal vibrations in accelerations were measured at ¼, ½, and ¾ of the bridge span under webs of selected main beams (fig. 9). The acceleration measurements were carried out with Brüel & Kjær PULSE platform. Sampling frequency was equal to 4096 Hz. Following types of accelerometers were used: miniature triaxial charge accelerometer 4326 A, miniature triaxial DeltaTron (CCLD) accelerometers 4504 A, high-shock DeltaTron (CCLD) accelerometers 8339-001, miniature uniaxial CCLD piezoelectric accelerometers 4507 B 005 and 751-10. PHANTOM v12 (1280×800 px, 6242 to 1 mln fps dependent on resolution) high-speed camera has been used for measurement of the train service velocity.

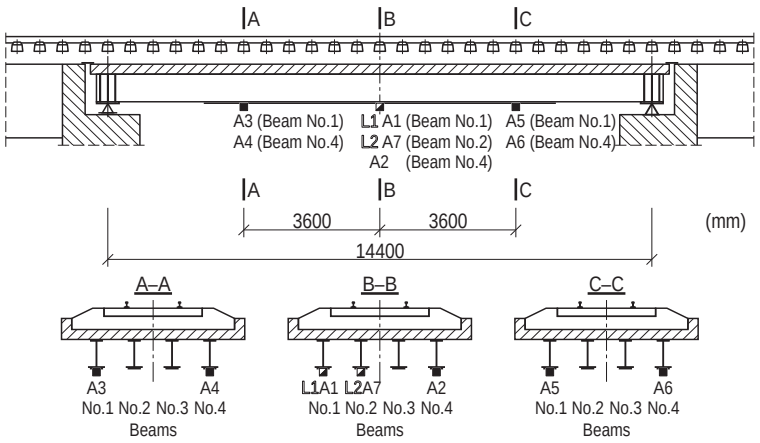


Figure 9: Dynamic experimental tests of KNI 140070 viaduct. Locations of the measurement points on the main beams: L1 and L2 — laser sensors; A1–A6 — vertical accelerometers; A7 — horizontal (lateral) accelerometer.

5 Results and analysis of the numerical and experimental tests

Time-histories of the vertical deflections $w_1(t)$, $w_2(t)$, corresponding to the main beams No. 1 and 2 are shown in fig. 10 (registered experimental time-histories) and fig. 11 (simulated time-histories). The experimental and numerical maximum deflections are set up in table 1. These results confirm good coincidence between the simulation and the experiment. Figures 12 and 13 present the experimental and simulated acceleration time-histories of the vertical accelerator A1 (see fig. 9).

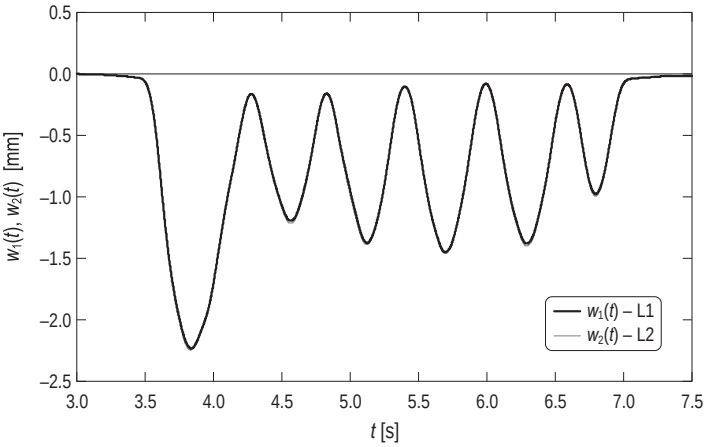


Figure 10: The validation test. Registered experimental time-histories of vertical displacements vs. time at the midspan for beams No. 1 and No. 2.

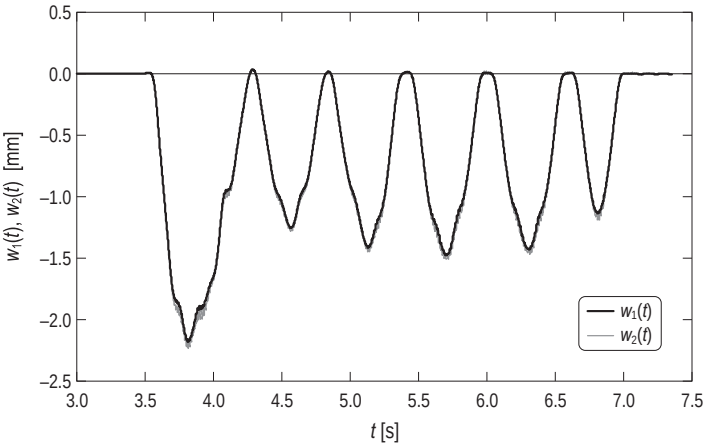


Figure 11: The validation test. Simulated time-histories of vertical displacements vs. time at the midspan for beams No. 1 and No. 2.

Table 1: The experimental and simulated maximum vertical deflections of beams No. 1 and 2 during the validation test.

Deflection	Experiment	FE Analysis
max $w_1(t)$ [mm]	2.23	2.18
max $w_2(t)$ [mm]	2.25	2.23

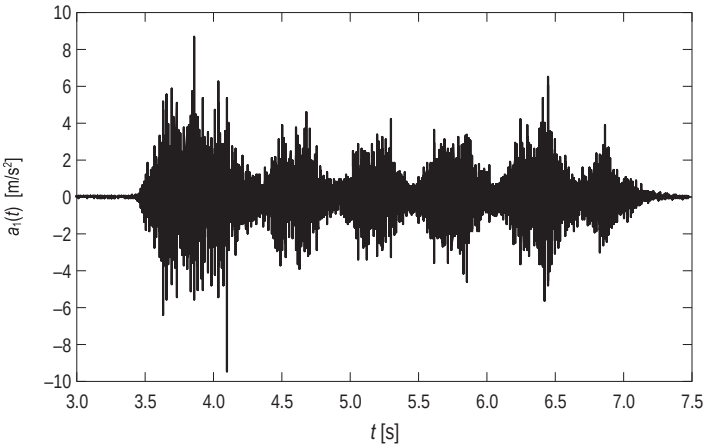


Figure 12: The experimental vertical acceleration time-history registered from accelerator A1.

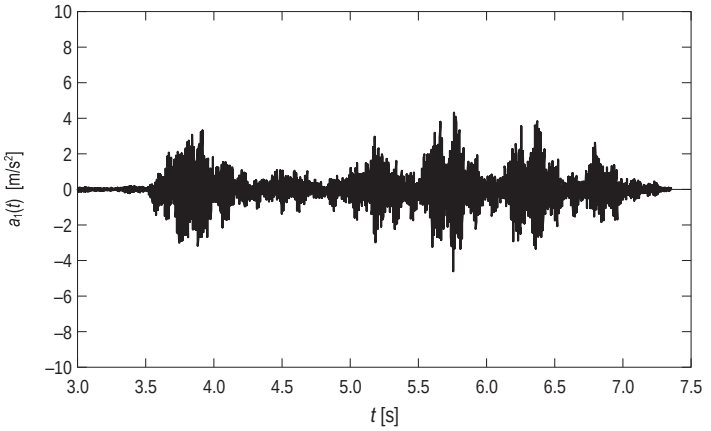


Figure 13: The simulated vertical acceleration time-history corresponding to the accelerator A1.

First of all, qualitative conformity of the experimental and simulated diagrams is visible but, as expected, the experimental acceleration levels are greater than

the numerical ones owing to neglecting the effects resulting from rolling of wheelsets.

6 Conclusions

Based on the results of the validation test, the following main conclusions have been formulated. The experimental validation of the numerical modelling and simulation of the BTT systems, examined on the KNI 140070 composite viaduct, located on the Polish Central Main Line, loaded by the EC 114 PRAHA express train moving at 160 km/h, is positive. The simulated (numerical) and experimental dynamic responses in displacements of the bridge superstructure are in good conformity, both qualitatively and quantitatively. Simplifications assumed in nonlinear physic modelling of the BTT systems are acceptable for one-track simply-supported bridge spans loaded by high-speed passenger trains. The simulated and experimental dynamic responses in accelerations exhibit longer durability of fully symmetric bridge superstructures compared to slightly asymmetric real superstructures. The examined modernized composite viaduct is insensitive dynamically at 160 km/h service speed of the EC 114 train.

Note that the viaduct has been redesigned in order to adopt it to service speeds of 300–350 km/h. Further investigations should be aimed at the control and validation tests at higher service velocities of trains.

Acknowledgements

This paper is a part of a research project No. N N509 2923 35, realized by Military University of Technology, Poland in the period 2008–2011. Financial support of Ministry of Science and Higher Education, Poland is gratefully acknowledged.

References

[1] Klasztorny, M., *Vibrations of single-track railway bridges under high-speed trains* [in Polish], Wroclaw University of Technology Press: Wroclaw, 1987.
[2] Klasztorny, M., *Dynamics of beam bridges loaded by high-speed trains* [in Polish], WNT Press: Warsaw, 2005.
[3] Yang, Y.-B., et al., Vibrations of simple beams due to trains moving at high speeds, *Eng. Struct.*, **19(11)**, pp. 936–944, 1997.
[4] Cheng, Y.S., et al., Vibration of railway bridges under a moving train by using bridge-track-vehicle element, *Eng. Struct.*, **23(12)**, pp. 1597–1606, 2001.
[5] Au, F.T.K., et al., Impact study of cable-stayed bridge under railway traffic using various models, *J. Sound & Vibration*, **240(3)**, pp. 447–465, 2001.
[6] Zhang, Q.-L., et al., Numerical simulation of train-bridge interaction dynamics, *Computers & Structures*, **79**, pp. 1059–1075, 2001.

[7] Song, M.K., Choi, C.K., Analysis of high-speed vehicle-bridge interactions by a simplified 3-D model, *Struct. Eng. & Mech.*, **13(5)**, pp. 505–532, 2002.
[8] Modernization of the rail-line 4 –E 65. Viaduct No. KNI 140070. *Building Project* [in Polish], Warsaw, 2007.
[9] Jones, R.M., *Mechanics of composite materials*, T. & F.: London, 1999.
[10] Polish Standard: PN-H-84027-07:1984. *Steel for rail-engineering. Normal rails. Sorts* [in Polish], PKN Press: Warsaw, 1984.
[11] Niemierko, A., et al., *Reinforcing the track and the subgrade in the approach zones to engineering objects* [in Polish], Project No. M1-123, Research Institute for Roads and Bridges: Warsaw, 2009.
[12] EC114 PRAHA *trainset on November 9, 2011 – Relation: Warsaw East – Praha hl*. N, PKP-PLK SA Report, Warsaw, 2009.
[13] http://kolej.krb.com.pl/dt/zal1.htm
[14] Bachmann, H., *Vibration problems in structures*, Birkhäuser: Berlin, 1995.
[15] Hallquist, J.O., *LS-DYNA V971 R4 Beta. Keyword User's Manual*, LSTC Co.: Livermore (CA, USA), 2009.

This page intentionally left blank

Estimation of drying parameters including moisture diffusivity by using temperature measurements

G. H. Kanevce[1], L. P. Kanevce[2], V. B. Mitrevski[2]
& G. S. Dulikravich[3]
[1]*Macedonian Academy of Sciences and Arts, Macedonia*
[2]*Faculty of Technical Sciences,*
St. Kliment Ohridski University, Macedonia
[3]*Department of Mechanical and Materials Engineering, Florida International University, USA*

Abstract

This paper deals with the estimation of the moisture diffusivity, together with other thermophysical properties, as well as the heat and mass transfer coefficients of a convective drying body, on the basis of single thermocouple temperature measurements by using an inverse approach. Potato and apple slices have been chosen as representative drying bodies with significant shrinkage effects. A mathematical model of the drying process of shrinking bodies has been applied. The Levenberg-Marquardt method and a hybrid optimization method of minimization of a resulting least-squares norm were used to solve the present inverse problem. The experiments have been conducted on the experimental setup that is designed to simulate an industrial convective dryer. An analysis of the influence of the drying air speed, temperature and relative humidity, drying body dimensions, and drying time on the estimation of the unknown parameters enables the design of appropriate experiments that have been conducted as well. The estimated moisture diffusivities are compared with the results published by other authors. The experimental transient temperature and moisture content changes during the drying are compared with numerical solutions.

Keywords: inverse approach, thermophysical properties, drying.

WIT Transactions on Modelling and Simulation, Vol 51,
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110111

1 Introduction

An inverse approach to parameter estimation in the last few decades has become widely used in various scientific disciplines. In this paper, application of inverse concepts in drying is analyzed. A mathematical model of the drying process of shrinking bodies has been applied where the moisture content and temperature field in the drying body are expressed by a system of two coupled partial differential equations. The system of equations incorporates several coefficients that are functions of temperature and moisture content and must be determined experimentally. All the coefficients, except for the moisture diffusivity, can be relatively easily determined by experiments. The main problem in the moisture diffusivity determination by classical or inverse methods is the difficulty of moisture content measurements. We have recently analyzed a method for moisture diffusivity estimation by the temperature response of a drying body [1–6]. The main idea of this method is to make use of the interrelation between the heat and mass transport processes within the convective drying body and from its surface to the surroundings. Then, the moisture diffusivity, together with other thermophysical properties of the body, as well as the heat and mass transfer coefficients can be estimated on the basis of an accurate and easy to perform single thermocouple temperature measurement by using an inverse approach.

2 A mathematical model of drying

Drying of a single slice of thickness $2L$ initially at uniform temperature and uniform moisture content is analysed in the paper, fig. 1. The surfaces of the drying body are in contact with the drying air. The problem is symmetrical relative to the mid-plane of the slice. The thickness of the body changes during the drying from $2L_0$ to $2L_f$.

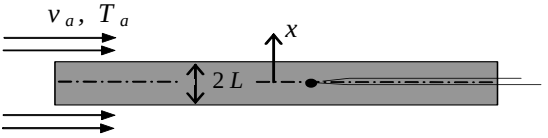


Figure 1: Scheme of the drying experiment.

In the case of an infinite flat plate the unsteady temperature, $T(x, t)$, and moisture content, $X(x, t)$, fields in the drying body are expressed by the following system of coupled nonlinear partial differential equations for energy and moisture transport

$$c\rho_s \frac{\partial T}{\partial t} = \frac{\partial}{\partial x}\left(k \frac{\partial T}{\partial x}\right) + \varepsilon \Delta H \frac{\partial (\rho_s X)}{\partial t} \tag{1}$$

$$\frac{\partial(\rho_s X)}{\partial t} = \frac{\partial}{\partial x}\left(D\rho_s \frac{\partial X}{\partial x}\right) \tag{2}$$

Here, t, x, c, k, ΔH, ε, D, ρ_s are: time, normal distance from the mid-plane of the plate, specific heat, thermal conductivity, latent heat of vaporization, ratio of water evaporation rate to the reduction rate of the moisture content, moisture diffusivity, and density of dry solid, respectively.

The shrinkage effect of the drying body was incorporated through the changes of the specific volume of the drying body. There are several models for describing the changes of the specific volume of the body during drying. In this paper, linear relationship between the specific volume, v_s, and the moisture content, X, has been used

$$v_s = \frac{1}{\rho_s} = \frac{V}{m_s} = \frac{1+\beta' X}{\rho_{b0}} \tag{3}$$

where m_s is the mass of the dry material of the drying body, V is the volume of the drying body, ρ_{b0} is the density of a fully dried body and β' is the shrinkage coefficient.

The problem of the moving boundaries due to the changes of the dimensions of the body during the drying was resolved by introducing the dimensionless coordinate

$$\psi = \frac{x}{L(t)} \tag{4}$$

Substituting the above expression for ρ_s (=$1/v_s$) and ψ into eqns. (1) and (2) and rearranging, the resulting system of equations for the temperature and moisture content prediction becomes

$$\frac{\partial T}{\partial t} = \frac{k}{\rho_s c}\frac{1}{L^2}\frac{\partial^2 T}{\partial \psi^2} + \frac{\psi}{L}\frac{dL}{dt}\frac{\partial T}{\partial \psi} + \frac{\varepsilon \Delta H}{c}\frac{\rho_s}{\rho_{b0}}\left(\frac{\partial X}{\partial t} - \frac{\psi}{L}\frac{dL}{dt}\frac{\partial X}{\partial \psi}\right) \tag{5}$$

$$\frac{\partial X}{\partial t} = D\frac{\rho_{b0}}{\rho_s}\frac{1}{L^2}\frac{\partial^2 X}{\partial \psi^2} + \left[\frac{\rho_{b0}}{\rho_s^2}\frac{1}{L^2}\frac{\partial(D\rho_s)}{\partial \psi} + \frac{\psi}{L}\frac{dL}{dt}\right]\frac{\partial X}{\partial \psi} \tag{6}$$

The initial conditions are

$$t = 0: \quad T(\psi,0) = T_0, \; X(\psi,0) = X_0. \tag{7}$$

The temperature and the moisture content boundary conditions on the surfaces of the drying body in contact with the drying air are

$$-k\frac{1}{L}\left(\frac{\partial T}{\partial \psi}\right)_{\psi=1}+j_q-\Delta H(1-\varepsilon)j_m=0 \tag{8}$$

$$D\rho_s\frac{1}{L}\left(\frac{\partial X}{\partial \psi}\right)_{\psi=1}+j_m=0 \tag{9}$$

The convective heat flux, $j_q(t)$, and mass flux, $j_m(t)$, on these surfaces are

$$j_q=h\left(T_a-T_{\psi=1}\right) \tag{10}$$

$$j_m=h_D\left(C_{\psi=1}-C_a\right) \tag{11}$$

where h is the heat transfer coefficient, and h_D is the mass transfer coefficient, T_a is the temperature of the drying air, and $T_{\psi=1}$ is the temperature on the surfaces of the drying body.

The water vapour concentration in the drying air, C_a, is calculated from

$$C_a=\frac{\varphi\ p_s(T_a)}{R_w\ T_{k,a}} \tag{12}$$

where φ is the relative humidity of the drying air and p_s is the saturation pressure. The water vapor concentration of the air in equilibrium with the surface of the body exposed to convection is calculated from

$$C_{\psi=1}=\frac{a(T_{\psi=1},X_{\psi=1})\ p_s(T_{\psi=1})}{R_w\ T_{k,\psi=1}} \tag{13}$$

The water activity, a, or the equilibrium relative humidity of the air in contact with the convection surface at temperature $T_{\psi=1}$ and moisture content $X_{\psi=1}$ are calculated from experimental water sorption isotherms.

The boundary conditions on the mid-plane of the drying slice are

$$\left(\frac{\partial T}{\partial \psi}\right)_{\psi=0}=0,\quad \left(\frac{\partial X}{\partial \psi}\right)_{\psi=0}=0\,. \tag{14}$$

Problem defined by eqns. (5)–(12) is referred to as a direct problem when initial and boundary conditions as well as all the parameters appearing in the formulation are known. The objective of the direct problem is to determine the temperature and moisture content fields in the drying body.

In order to approximate the solution of eqns. (5) and (6), an explicit numerical procedure has been used.

3 Inverse approach

For the inverse problem of interest here, the thermophysical properties and the boundary conditions parameters of a drying body are regarded as unknown parameters.

The estimation methodology used is based on the minimization of the ordinary least square norm

$$E(\mathbf{P}) = [\mathbf{Y} - \mathbf{T}(\mathbf{P})]^{\mathrm{T}}[\mathbf{Y} - \mathbf{T}(\mathbf{P})] \tag{15}$$

Here, $\mathbf{Y}^{\mathrm{T}} = [Y_1, Y_2, \ldots, Y_{\mathrm{imax}}]$ is the vector of measured temperatures, $\mathbf{T}^{\mathrm{T}} = [T_1(\mathbf{P}), T_2(\mathbf{P}), \ldots T_{\mathrm{imax}}(\mathbf{P})]$ is the vector of estimated temperatures at time t_{i} (i = 1, 2, …, imax), $\mathbf{P}^{\mathrm{T}} = [P_1, P_2, \ldots P_{\mathrm{N}}]$ is the vector of unknown parameters, imax is the total number of measurements, and N is the total number of unknown parameters (imax ≥ N).

A hybrid optimization algorithm OPTRAN [7] and the Levenberg-Marquardt method [8] have been utilized for the minimization of $E(\mathbf{P})$ representing the solution of the present parameter estimation problem.

The Levenberg-Marquardt method is a stable and straightforward gradient search minimization algorithm that has been applied to a variety of inverse problems. It belongs to a general class of damped least squares methods.

An alternative to the Levenberg-Marquardt algorithm, especially when searching for a global optimum of a function with possible multiple minima, is the hybrid optimization program OPTRAN. OPTRAN incorporates six of the most popular optimization algorithms: the Davidon-Fletcher-Powell gradient search, sequential quadratic programming algorithm, Pshenichny-Danilin quasi-Newtonian algorithm, a modified Nelder-Mead simplex algorithm, a genetic algorithm, and a differential evolution algorithm. Each algorithm provides a unique approach to optimization with varying degrees of convergence, reliability and robustness at different stages during the iterative optimization procedure. The hybrid optimizer OPTRAN includes a set of rules and switching criteria to automatically switch back and forth among the different algorithms as the iterative process proceeds in order to avoid local minima and accelerate convergence towards a global minimum.

4 Experimental

Real experiments have been conducted to investigate the applicability of the method to food processing, when involving drying of thin flat samples. The experiments have been conducted on the experimental setup that is designed to simulate an industrial convective dryer.

Drying of approximately three millimeter thick potato or apple slices have been examined. The slices have been in contact with the drying air from the top and the bottom surfaces. Two shelves, fig. 2, each holding three moist slices

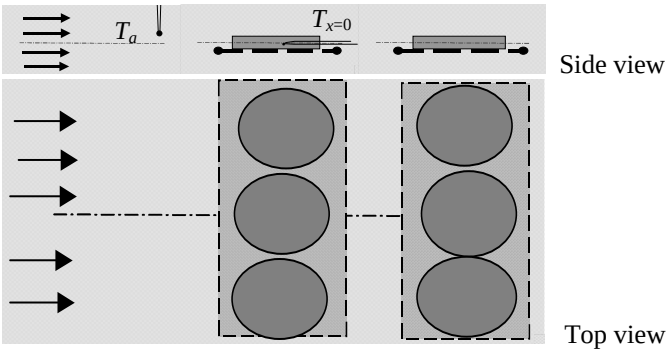


Figure 2: Scheme of the experimental setup.

have been introduced into the rectangular experimental channel of dimensions 25 x 200 mm.

A micro-thermocouple was inserted in the midplane of each of the three slices on the first shelf. An arithmetical mean of the readings from the three thermocouples was used as a transient temperature reading, for the estimation of the unknown parameters. The slices on the second shelf were weighed every ten minutes in order to obtain the volume-averaged moisture content change during drying. The temperature of the drying air, T_a, has been recorded as well. The initial moisture content, X_0, and the initial slices thickness, $2L_0$, were measured for each of the experiments.

The experiments were designed by using the so-called D-optimum criterion. From the relative sensitivity coefficients analysis, the drying air velocity, drying air temperature and drying body dimension were defined. Determinant of the information matrix with normalized elements has been calculated in order to define drying time. The duration of the drying experiment corresponding to the maximum determinant value was used.

A number of drying experiments with similar experimental conditions, (T_a = 56.6-59.5 ^{O}C, $2L_0$ = 2.36-3.14 mm, X_0 = 3.70-4.83 kg/kg and T_0 = 14.9-17.7 ^{O}C), have been carried out.

5 Results and discussion

In this paper, application of the proposed method for simultaneous estimation of the moisture diffusivity, together with other thermophysical properties of vegetables and fruits, as well as the heat and mass transfer coefficients has been analyzed.

From the conducted sensitivity coefficients analysis it was concluded that in the convective drying experiment it is possible, based on a single thermocouple temperature response, to estimate simultaneously the moisture diffusivity, the convection heat and mass transfer coefficients, and the relative humidity of the

drying air. All other quantities appearing in the direct problem formulation were taken from published data of other authors.

Moisture diffusivity of foods is very often considered as an Arrhenius-type temperature function [9, 10]

$$D = D_0 \exp[-E_0/(R\,T_k)] \tag{17}$$

with constant values of the Arrhenius factor, D_0, and the activation energy for moisture diffusion, E_0.

Table 1 shows the computationally obtained parameters and RMS-error for potato, for experiment P1: T_a = 58.13 ^{O}C, $2L_0$ = 3.14 mm, X_0 = 4.80 kg/kg and T_0 = 17.53 ^{O}C and for apple, for experiment A1: T_a = 60.17 ^{O}C, $2L_0$ = 2.96 mm, X_0 = 6.46 kg/kg and T_0 = 16.91 ^{O}C.

In fig. 3 the estimated moisture diffusivities are compared with the results published by other authors [11].

In fig. 4 the experimental transient temperature reading, $T_{x=0}$, and the experimental volume-averaged moisture content change during drying of potato are compared with numerical solutions for the estimated parameters. A similar, very good agreement was obtained for apple slice as well.

Table 1: Estimated parameters and RMS-error.

	Potato	Apple
$D_0 \cdot 10^3$ [m^2 s^{-1}]	7.985	63.905
E_0 [kJ mol^{-1}]	43.3	50.215
h [W m^{-2} K^{-1}]	31.08	24.137
$h_D \cdot 10^2$ [m s^{-1}]	3.48	2.70
φ [-]	0.0899	0.1011
RMS [^{O}C]	0.55	0.64

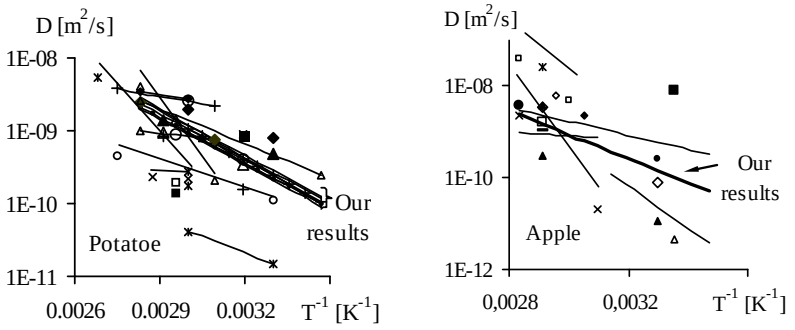


Figure 3: Moisture diffusivity of potatoes and apples.

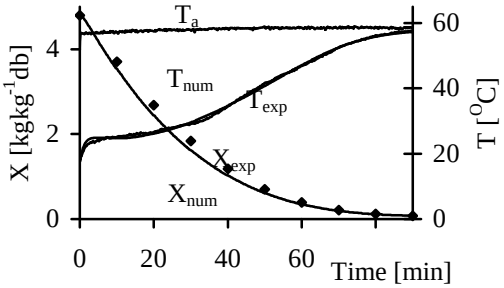


Figure 4: The mid-plane temperature, T, the temperature of the drying air, T_a, and the volume-averaged moisture content, X, changes during the drying of a potato slice.

6 Conclusions

It can be concluded that in the convective drying experiments of apples and potatoes it is possible, based on a single thermocouple temperature response, to estimate simultaneously the moisture diffusivity, the convection heat and mass transfer coefficients, and the relative humidity of the drying air.

Estimated moisture diffusivities compare well with the values obtained by other authors who utilized different methods.

Very good agreement between the experimental and numerical temperature and volume-averaged moisture content changes during drying has been obtained.

References

[1] Kanevce, G.H. Kanevce, L.P. & Dulikravich, G.S., Moisture diffusivity estimation by temperature response of a drying body, *Proc. of the 2nd Int. Conf. On Inverse Problems in Engineering Mechanics*, eds. M. Tanaka & G. S. Dulikravich, Elsevier: Amsterdam, pp. 43-52, 2000.

[2] Kanevce, G.H. Kanevce, L.P. & Dulikravich, G.S., An inverse method for drying at high mass transfer Biot number, *Proc. of the HT03 ASME Summer Heat Transfer Conference*, Las Vegas, Nevada, USA, ASME paper HT20003-40146, 2003.

[3] Kanevce, G.H. Kanevce, L.P., Dulikravich, G.S., & Orlande, H.R.B., Estimation of thermophysical properties of moist materials under different drying conditions, *Inverse Problems in Science and Engineering*, **13(4)**, pp. 341-354, 2005.

[4] Kanevce, G.H. Kanevce, L.P., & Dulikravich, G.S., Application of inverse concepts to drying, *Thermal Science*, **9(2)**, pp. 31-44, 2005.

[5] Kanevce, G.H. Kanevce, L.P., Mitrevski, V.B., Dulikravich, G.S., & Orlande, H.R.B., Inverse approaches to drying with and without shrinkage, *Proc. of the 15th Int. Drying Symposium (IDS'2006)*, ed. I. Farkas, Budapest, Hungary, Vol. A, p. 576, 2006.
[6] Kanevce, G.H. Kanevce, L.P., Mitrevski, V.B., Dulikravich, G.S., & Orlande, H.R.B., Inverse approaches to drying of thin bodies with significant shrinkage effects, *Journal of Heat Transfer*, **129(3)**, pp. 379-386, 2007.
[7] Dulikravich, G.S., Martin, T.J., Dennis, B.H. & Foster, N.F. Multidisciplinary hybrid constrained GA optimization, *Evolutionary Algorithms in Engineering and Computer Science: Recent Advances and Industrial Applications (EUROGEN'99)*, eds. K. Miettinen, M. M. Makela, P. Neittaanmaki & J. Periaux, John Wiley & Sons, Ltd., Jyvaskyla, Finland, pp. 231-260, 1999.
[8] Marquardt, D.W., An algorithm for least squares estimation of nonlinear parameters, *J. Soc. Ind. Appl. Math.*, **11**, pp. 431-441, 1963.
[9] Rovedo, C., Suarez C. & Viollaz P., Analysis of moisture profiles, mass Biot number and driving forces during drying of potato slabs, *J. of Food Engineering*, **36**, pp. 211-231, 1998.
[10] Zogzas N.P. & Maroulis Z.B., Effective moisture diffusivity estimation from drying data: A comparison between various methods of analysis, *Drying Technology*, **14(7&8)**, pp. 1543-1573, 1996.
[11] Mitrevski, V.B., Investigation of the drying processes by inverse methods, *PhD. Thesis*, University of Bitola, Macedonia, 2005

This page intentionally left blank

Investigating granular milling in a hammer mill: experiments and simulation

S. Naik and B. Chaudhuri
Department of Pharmaceutical Sciences,
University of Connecticut, USA

Abstract

Particle size reduction of dry material by milling is a key unit operation for the pharmaceutical, agricultural, food and paper industries. Knowledge of particle flow and size reduction in a hammer mill is thus critical to optimize the design and operation of such equipment. Milling experiments are performed using lactose non pareils in a laboratory scale Hammer Mill. The size and shape of the resultant progeny of particles are analyzed by sieves/light scattering and microscope/image analysis techniques respectively. Discrete Element Method (DEM) based computational methods are developed to perform a quantitative examination of granular flow, fracturing and subsequently fragmentation patterns for the same hammer mill. A parametric study was performed to understand the effect of hammer speed (rotational), feed rate, hammer-wall tolerance on size reduction process. Simulations were carried out to study the effect of mill speed on kinetic energy of particles.
Keywords: discrete element method, granular flow, fragmentation, hammer mill.

1 Introduction

Particle size reduction of dry granular material by mechanical means, also known as milling or communition, is undoubtedly a very important unit operation in pharmaceutical, agricultural, food, mineral and paper industries. For example, particle size reduction has a significant impact on pharmaceutical product performance and stability as it affects the solubility and bioavailability of many poorly soluble BCS Class II drugs [1]. The most commonly used mills are the rotary cutter, hammer mill, roller mill, ball mills and fluid energy mills, used in various stages of manufacturing. Size reduction is generally achieved by particle

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110121

fracturing and subsequent fragmentation under the action of externally applied energy Surface characteristic and fracture pattern under an applied energy is not only influenced by material properties but also by operational conditions of mill. Knowledge of particle flow, granular fracturing and size reduction in a mill is thus very critical to optimize and to scale up this operation. Substantial effort has been expended in last three decades to understand the milling performance as a function of material properties, operational conditions and machine configuration. Several experimental and mathematical approaches to understand the milling behavior in Ball Mill have been performed [2–6]. Experiments and Discrete Element Method (DEM) based modeling of milling of pharmaceutical excipients in a very small scale oscillatory single was performed by Kwan *et al.* [7]. Campbell *et al.* [8] developed the relationship between the inlet and outlet particle size distribution in a roller milling operation. The impact mills have been investigated by Austin [9], Gotsis and Austin [10], Vogel and Peukert [11] and Djordjevic *et al.* [12] for different materials. Austin developed simplified analytical model of high speed milling of limestone, cement in hammer mill with parameters chosen to match the model prediction to the experimental findings, whereas PMMA/Glass and rock was the material of interest for Vogel and Djordjevic. A comprehensive understanding of the fragmentation mechanism from experiments helped the development of a series of mechanistic and population balance based models [13–15]. However, these models completely ignore the inter-particle and particle boundary interaction which dictates granular flow behavior. DEM explicitly considers inter-particle and particle wall interactions, providing an effective tool to simulate particles kinematics and granular fragmentation. Comprehensive modeling of milling based on first principles was attempted by Djordjevic *et al.* [12] using discrete element method (DEM), where granular flow properties are considered. This DEM based direct modeling of the milling process is limited to a small number (a few hundred) of spherical particles, and derived essentially no conclusions regarding optimization, control or scale-up of milling performance. To the best of our knowledge, no previous work has used 3D DEM to study granular milling of pharmaceutical granules in hammer mill. In this article, experiment and 3D DEM modeling based method, to better understand milling behavior of pharmaceutical powders in hammer mill is presented. The effect of material properties and operational conditions on milling performance is also determined.

2 Experimental

The milling equipment (Thomas Wiley Mill, Thomas Scientific, Swedesboro, NJ) used in our study is a variable speed, digitally controlled, direct drive mill; that provides continuous variation of cutting speeds from 650 to 1140 rpm with constant torque maintained throughout the speed range. Once the material is loaded through the feed hopper in to the grinding chamber, it drops by gravity and fragments after colliding with the rotating hammers. Parametric studies are conducted to study the effect of speed, load, and impeller wall distance and feed rate on particle size reduction. The experiments were performed using lactose

non pareils. In our milling experiments the material is fed at the top center, thrown out centrifugally at first impact with the hammers and ground by impact against the wall or cut due to presence of hammers at the periphery. The material is retained until it is small enough to fall through the screen that forms the lower portion of the casing. Particles fine enough to pass through the screen is discharged almost as fast as they are formed in a glass container. To obtain rigorous information for the parametric studies, samples are collected in the glass container at regular intervals for example, 5 seconds for the first four samples, 10 seconds for the next four samples and 20 seconds for the next samples. All experiments are conducted at least thrice at room temperature of 28C and relative humidity of 35% to ensure repeatability. The effect of experimental factor was studied at two-three levels. Average particle sizes for the entire sample are calculated from sieve analysis using Rotap Sieve analyzer determination was done by sieve analysis. Particle shape analysis was performed by Optical Microscope-Camera (Olympus SZ61) and Image Analysis software (Image Pro Plus). The conditions examined using the Wiley Mill are enlisted in Table 1.

Table 1: Conditions examined for the milling experiments with lactose.

Parameters	
Speed	600rpm - 1140rpm
Clearance	2.9mm-3.7mm
Feed rate	60g/min-100g/min

3 Computational method

Numerical simulations were performed using the Discrete Element Method (DEM) originally developed by Cundall and Strack [13, 14]. In this model the granular material is considered as a collection of frictional inelastic spherical particles. Each particle interacts with its neighbors or with the boundary only at contact points through normal and tangential forces. The forces and torques acting on each of the particles are calculated in the following way:

$$\sum \mathrm{F_i} = \mathrm{m_i g} + F_N + F_T + F_{\mathrm{cohes}} \tag{1}$$

$$\sum \mathrm{T_i} = \mathrm{r_i} \times F_T \tag{2}$$

The force on each particle is given by the sum of gravitational, inter-particle (normal and tangential F_N and F_T) and cohesive forces (F_{cohes}) as indicated in eq. (1) and eq (2). The corresponding torque on each particle is the sum of the moment of the tangential forces (F_T) arising from inter-particle contacts (eq. (2)). The normal forces (F_N) and the tangential forces (F_T) in inter-particle or particle-wall collision were calculated with the "latching spring model" and "incrementally slipping model" respectively, developed by Walton [15].

Cohesion (F_{cohes}) is modeled using Equation (3) following the square-well type forces by Chaudhuri *et al.* [16] in which the granular bond number (K) and frictional parameters are varied to simulate the cohesive behavior of various pharmaceutical excipients (Microcrystalline cellulose (Avicel 101), lactose). Materials varying in cohesion were selected for this study which includes Fast flow lactose, Avicel-101 and Regular lactose. Granular cohesion was introduced by simulating a cohesive force

$$F_{cohesion} = K\,mg \tag{3}$$

where eq. (3) is applied between contacting particles or particle and wall where K is the bond number and a measure of cohesion and mg is the weight of the particle. Different K values used for this study are enlisted in table 2.

Table 2: K values of material used in study.

Material	Fast Flow lactose	Avicel 101	Regular lactose
K	45	60	75

Incremental impact breakage of the particles is modeled, as breakage due to single impact is certainly rare. Repeated impacts are required before a favorable orientation relative to pre- existing flaws is available and also for the growth of a pre-existing defect to reach a critical size prior to fracture or size reduction [17]. At each simulation time step, for each spherical particle the conditions for fragmentation were examined using the Grady's Algorithm [18] shown in eq. (4).

$$D = \left(\frac{4.472 \times Kl_c}{\rho V_p S_r} \right)^{2/3} \tag{4}$$

where D is the diameter of the resultant progeny particle, Kl_c is fracture toughness (Pa $m^{0.5}$), ρ is particle density V_p is propagation velocity (m/s) of longitudinal elastic waves in the material, Sr is the induced strain rate. The strain rate is calculated from the rotational velocity of the impeller and the radial distance between the center of the rotation and the impact point [18]. Following, fragmentation the resultant particles are considered as spheres, obeying the laws of conservation of total mass and momentum [19]. Powder characteristics responsible for breakage of solid particles such as, Young's Modulus, hardness and critical intensity factor were incorporated in the model. The material properties and DEM parameters used in our simulation are shown in Table 3.

We simulated the flow and fragmentation of 4000 particles of size 5mm radius in a hammer mill at different rotating conditions. The four hammers are simulated by rigidly grouping particles of 2mm size which perform solid body rotation in unison.

Table 3: DEM parameters for simulation.

DEM parameters	Values
Total number of particles	4000
Radius of the particles	5mm
Hopper angle	72.83 degrees
Coefficient of restitution	
Particle/particle	0.7
Particle/wall	0.5
Normal stiffness coefficient	
Particle/particle	600N/m
Particle/wall	600N/m
Friction coefficient particle/particle	0.7
Friction coefficient particle/wall	0.3
Time step	$2.0 * 10^{-5}$ seconds
Young Modulus	22.13 – 9.8 GPa
Critical Stress Intensity	0.5 – 0.8 MPa $m^{0.5}$

4 Results and discussion

4.1 Experiments

4.1.1 Effect of feed rate

It is important to study the effect of feed rate on size reduction since it determines the hold up of material in sizing chamber and hence energy required for size reduction. Milling experiments are performed at two different feed rates of 60g/min and 100g/min while the hammers are rotated at 600rpm and 1140 rpm. The hammer-wall clearance is kept constant at 2.9mm. Mean particle size for all samples at different time intervals were measured. As depicted in Figure 1a and Figure 1b a lower feed rate produced a smaller size distribution at the given impeller speeds. Besides achieving a smaller particle size, the feed rate at 60g/sec reduced the polydispersity of the sample at 1140 rpm, producing almost a constant particle size of around 450um. Thus the right choice of the feed rate and mill speed can also enable us to produce not only a small size but also control the variability in size distribution. A lower size distribution is obtained at lower feed rate due to longer path lengths of particles resulting in higher impact velocity and hence a finer size distribution. At higher hold up the number of collisions is high, but the kinetic energy per particle may be low leading to poor breakage probability.

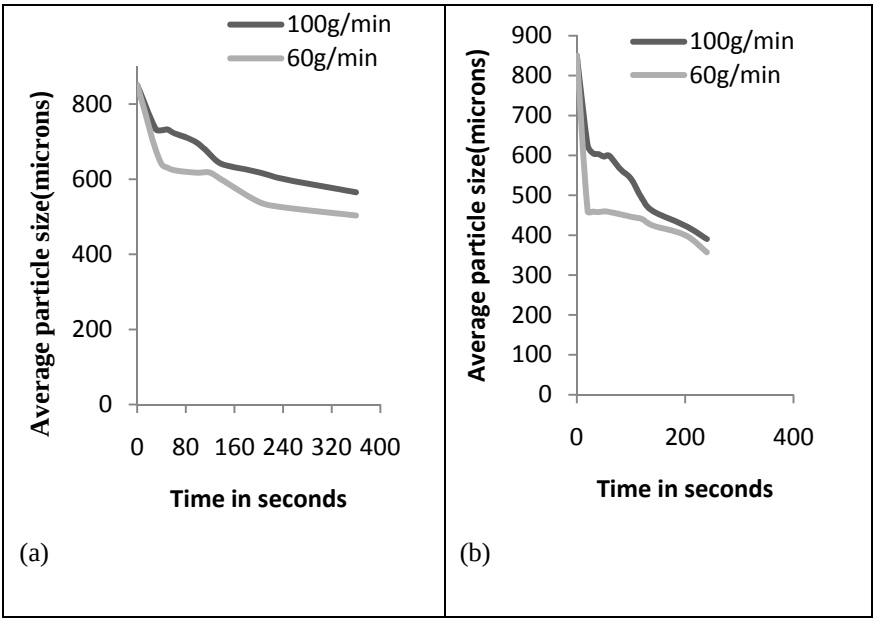


Figure 1: (a) Effect of feed rate on particle size reduction at hammer tip speed of 600rpm. (b) Effect of feed rate on particle size reductions at hammer tip speed of 1140rpm.

4.1.2 Effect of speed

The speed of the impeller determines the mechanism of size reduction. In most cases lower speeds result in attrition whereas as higher speeds cause size reduction by impact. Figure 2 depicts the behavior of size reduction upon changing the speed at a given feed rate. A higher speed of the mill produced smaller particle size. A high feed rate with lower speed can result in choking of the mill and can slow the process to a great extent. This is because under such conditions the power consumption of the motor increases and rotation speed decreases.

4.1.2.1 Shape analysis Particle shape analysis was performed using Optical Microscope-Camera (Olympus SZ61) and Image Analysis software (Image Pro Plus). Analysis of samples at an impeller speed of 600rpm, feed rate of 60g/min and a clearance of 2.9mm revealed a mechanism of chipping and attrition as shown in Figure 3a. For samples at a speed of 1140 rpm as shown in Figure 3b, the particle shape revealed that most probably, size reduction by fragmentation dominates. This s because of the greater centrifugal force experienced by the particle at higher impeller speeds.

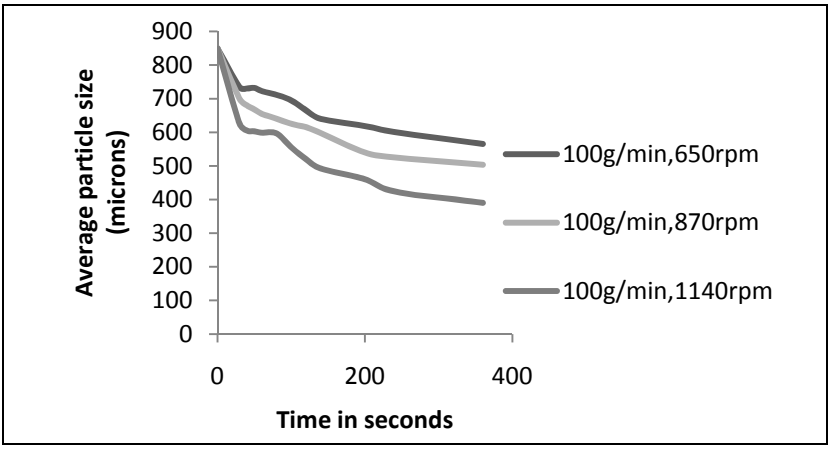


Figure 2: Change of average particle size with time as a function of Impeller speed.

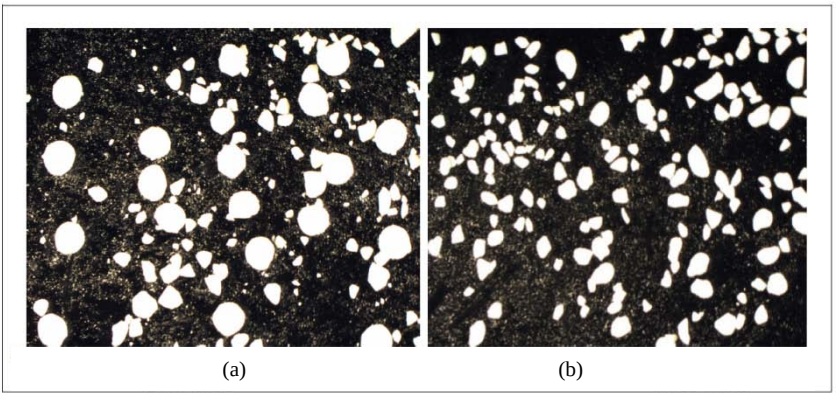


Figure 3: (a) Shape analysis of particles at clearance of 2.9mm, 600rpm. (b) Shape analysis of particles at clearance of 2.9mm, 1140rpm.

4.1.3 Effect of clearance on size reduction

The distance between the hammer and the wall is an important factor that controls size reduction. A larger clearance can result in continuous rolling of particles and hence significantly reduce the rate of size reduction. A very small tolerance on the other hand increases fine formation resulting in increased energy consumption by the mill. The hammer wall clearance is determined by the intrusion of four knives from the inner wall of mill. The knives are situated at 1.5" a part of the circular inner wall of mill. Figure 4a and Figure 4b depict the behavior of size reduction upon changing the clearance at a given feed rate and speed. This change in clearance significantly influenced the mechanism of size reduction, changing it from impact to a combination of rolling and attrition

irrespective of hammer speeds. This is because much of the particle shape was retained at hammer tolerance of 3.7mm whereas fragmentation was observed at hammer wall tolerance of 2.9mm As seen in figure 4a and figure 4b increasing the clearance decreased the rate of size reduction of the mill. This resulted in accumulation of particles in chamber. As a consequence, the particle size distribution turned coarser and wider.

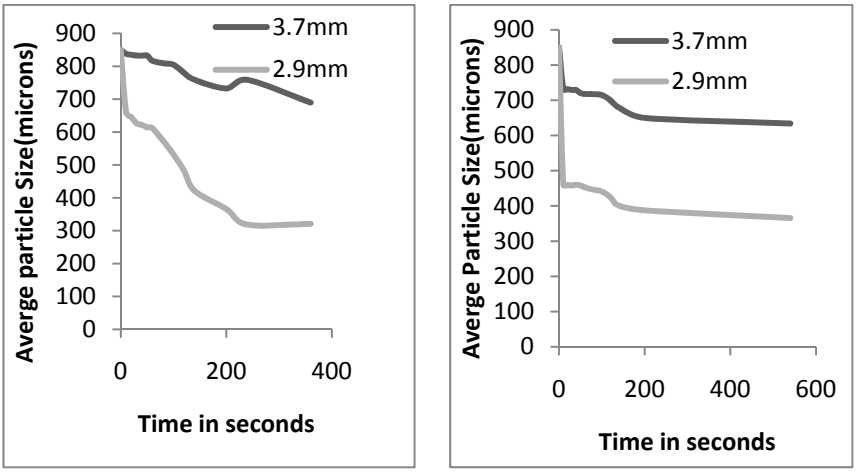


Figure 4: (a) Change of average particle size with time as a function of clearance at 600rpm. (b) Change of average particle size with time as a function of clearance at 1140rpm.

4.2 Numerical simulations of granular flow and fragmentation

4.2.1 Effect of speed

The flow and fragmentation of particles in the hammer mill is simulated using DEM as described in section 3. The hammer mill geometry is similar to the one used in our experiments and is implemented in the particle dynamics simulation. We use 4000 spherical particles with material properties described in table 1. Initially all particles are deposited in the hopper. The hopper outlet is chosen to be closed during the deposition of particles. After deposition particles are discharged from hopper into the hammer mill. The hammer is also set into motion at prescribed velocity of 650–1140 rpm. The DEM predicted velocity data is used to calculate the temporal function of total kinetic energy of the particles at different impeller speeds as illustrated in Figure 5.

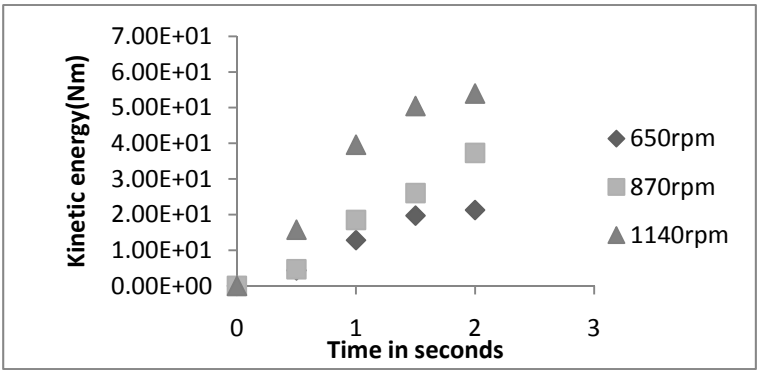
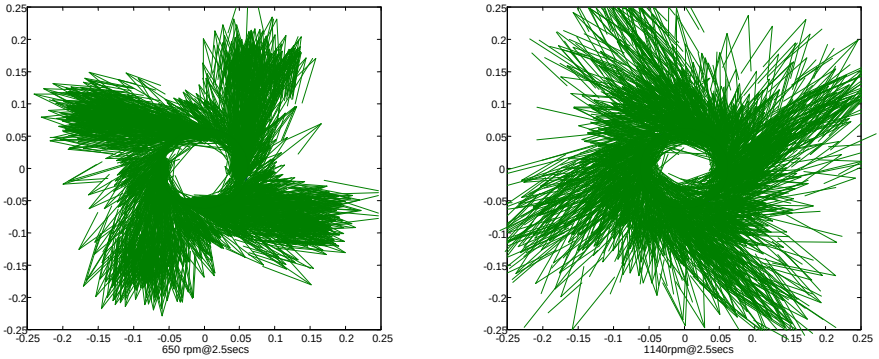


Figure 5: Temporal variation of kinetic energy of particles as function of speed.

The increase in kinetic energy of particles with increase in impeller speed is due to greater centrifugal forces, is well predicted by the simulation model. The velocity vector profiles of all particles illustrated in figure 6a and 6b further corroborates the greater impact experienced by the particle at higher speed of 1140rpm as compared to lower speed of 650 rpm.



(a) (b)

Figure 6: (a) Velocity profile of particles at 650 rpm. (b) Velocity profile of particles at 650 rpm.

4.2.2 Effect of material properties

The effect of granular cohesion on fragmentation is examined while keeping the operational condition (650rpm, 2.9mm clearance) constant. As described in section 3, the materials of different levels of cohesion is simulated by varying K value were used. Figures 7a and 7b shows the time series of axial snapshots from simulation of fragmentation of Avicel-101 (MCC) and Regular Lactose granules.

Regular lactose with a higher cohesion (K value) are more difficult to mill because of their poor flow properties, is evident from the two different sizes of fragments (fig 7a) as compared to only one size of fragment for Avicel-101 (fig 7b). This manifests into generation of more new particles for Avicel-101 as compared to Regular lactose as shown in figure 8a, which was also observed by Kwan [7] in oscillatory ball mill experiments. In fig 8a, maximum number of new particles is formed for Fast-Flo lactose which is the least cohesive material.

$$E_{nt} = \sum_{j=1}^{n} \tfrac{1}{2} m^* v_j^2 \tag{5}$$

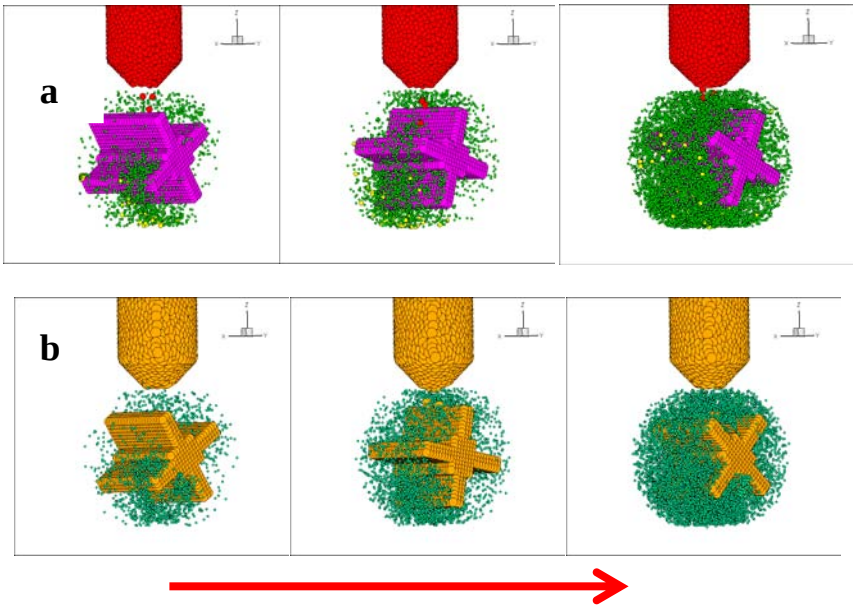
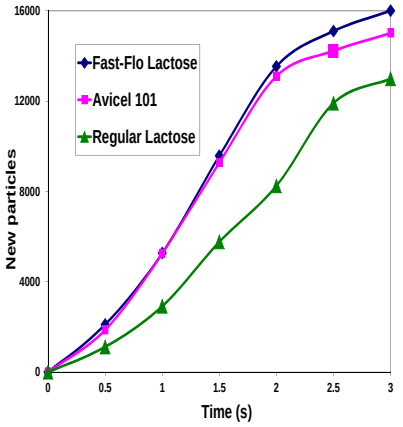
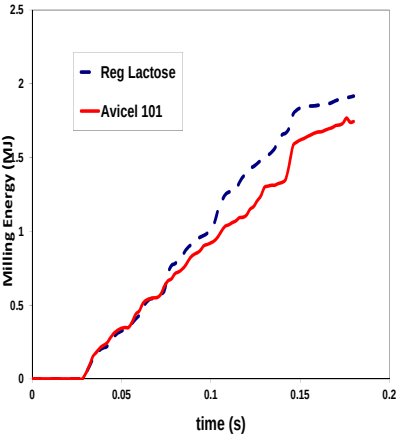


Figure 7: Evolution of particle fragmentation as function of cohesion over time a: Regular lactose, b: Avicel-101.

We can further corroborate the behavior of fragmentation of cohesive particles by computing the energy of fragmentation using eq (5), where m* is the geometric mass of the colliding particles and v_j is the relative velocity of the colliding particles. Figure 8b depicts that a greater energy of fragmentation is required for more cohesive material, regular lactose, as compared with same used for fragmenting Avicel 101.



(a)



(b)

Figure 8: (b) Effect of cohesion on fragmentation of particles. (b) Effect of cohesion on fragmentation energy.

5 Conclusion

DEM simulation and experiment based parametric studies is performed to study the effect of different operating conditions on size reduction in a hammer mill. In experiments, greater size reduction is observed at higher speeds and low feed rates owing to the greater centrifugal force experienced by the particles and longer mean free path lengths respectively. Particle shape analysis reveals fragmentation as the dominant process of size reduction in Hammer mill under the investigated conditions. In simulation, we find the kinetic energy of particles increases with impeller speed contributing greater fragmentation. Materials with higher cohesion show lower fragmentation.

References

[1] Jounela A.J., Pentikainen P.J., Sothmann A., Effect of particle size on the bioavailability of digoxin. European Journal of Clinical Pharmacology 8, pp. 365-370, 1975.

[2] Clearly P.W., Predicting charge motion, power draw, segregation, wear and particle breakage in ball mills using discrete element method, Miner. Eng., 11, pp. 1061-1080, 1998.

[3] Watanbe, H., Critical rotation speed for ball-milling, Powder Technology, 104, pp. 95-99, 1999.

[4] Misra B.K., Rajamani R.K., Simulation of charge motion in ball mills, Part 2: numerical simulations, Int. J. Miner. Processing, 40, pp. 187-197,1994.

[5] Hlungwani O., Rikhotso J., Dong H.,. Moys H., Further validation of DEM modeling of milling: effects of linear profile and mill speed., Minerals Engineering, 16, pp. 993-998, 2003.
[6] Austin L.G., Lucker P.T., A simulation model for air-swept ball mill grinding coal, Powder Technology, 38, pp. 255-266, 1984.
[7] Kwan C.C., Mio H., Chen Y.C., Ding Y.L., Saito F., Papadopoulos D.G., Benthem A.C., Ghadiri, M., Analysis of the milling rate of pharmaceutical powders using distinct element method, Chemical Engineering Science, 60, pp. 1441-1448 2005.
[8] Campbell G.M., Bunn P.J., Webb C., Hook S.C.W., On predicting roller mill performance, Part II. The breakage function, Powder Technology, 115, pp 243-255, 2001.
[9] Austin L., A preliminary simulation model for fine grinding in high speed hammer mills, Powder Technology, 143-144, pp 240-252, 2004.
[10] Gotsis C., Austin L.G., Batch grinding kinetics in the presence of a dead space as in a hammer mill, Powder Technology, 41, pp. 91-98, 1985.
[11] Vogel L., Peukert W., From single particle impact behavior to modeling of impact mills, Chemical Engineering Science, 60, pp. 5164-5176, 2005.
[12] Djordjevic N., Shi F.N., Morrison R.D., Applying discrete element modeling to vertical and horizontal shaft impact crushers, Minerals Engineering, 16, pp. 983-991, 2003.
[13] Cundall P.A., A computer model for simulating progressive large-scale movements in blocky rock systems. Proceedings of Symposium International Society of Rock Mechanics, 2, pp. 129, 1971.
[14] Cundall P.A., O. D. L. Strack, A discrete numerical model for granular assemblies. Geotechnique, 29, pp. 47-65, 1979
[15] Walton O. R., Numerical simulation of inclined chute flows of mono disperse, inelastic, frictional spheres. Mechanics of Materials, 16, 239-247, 1993.
[16] Chaudhuri B., Alexander A.W.A., Faqih A, Muzzio, F.J., Davies C., Tomassone M.S., Avalanching flow of cohesive powders, Powder Technology, 164, pp. 13-21, 2006.
[17] Morrison R.D., Shi F., Whyte R., Modeling of incremental rock breakage by impact – For use in DEM models, Mineral Engineering, 20, pp. 303-309, 2007.
[18] Grady D.E., Fragmentation under impulsive stress loading, In: W.L. Fourney et.al. (Eds), Fragmentation by blasting, Society for Experimental Mechanics, Connecticut, USA, 63-72, 1985.
[19] Poschel T., Schwager T., *Computational Granular Dynamics: Model and Algorithms*, Springer Verlag, Berlin, Germany, 210, 2005

Study of the influence of asphaltenes on modeling of viscosity of Nigerian crude mixtures

A. Miadonye[1], F. Dan-Ali[2], R. Onwude[2] & O. O. Osirim[2]
[1]*Department of Chemistry, Cape Breton University, NS, Canada*
[2]*Department of Petroleum Chemistry,*
American University of Nigeria, Yola, Nigeria

Abstract

In this study, the effect of asphaltenes on the viscosity of Nigerian light crude oils at different temperatures was studied. Different weight per cent of bitumen containing 14 per cent asphaltenes and deasphalted bitumen samples from Alberta, Canada were mixed with Nigerian crude oils. The mixtures were prepared from five light crude oil samples and two bitumen samples. The experimental viscosity data obtained at different temperatures and weight per cent of the bitumen samples were correlated with the viscosity equation of Singh et al. The viscosity of pure crude oils (for instance, Forcados crude oil) containing asphalted and deasphalted bitumen (Rush Lake and Plover Lake bitumen) increased by as much as 160% and 60% respectively.

However, the results obtained from correlating the viscosity data with the equation indicated that the discrepancies in measured data could be mainly attributed to the presence of asphaltenes, but the percent errors between measured and predicted viscosity were consistent for the entire samples. The results were identified the light crude oils as better alternative diluents for the production and transportation of heavy crude petroleum.
Keywords: Nigerian crude oil, asphaltenes, bitumen, hydrocarbon condensates, solvents, asphaltenes solubility, light crude oil, viscosity modeling.

1 Introduction

Pipeline transportation of heavy oil has been among the major challenges faced by the upstream petroleum industry. The high viscous nature of heavy oil,

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110131

viscosity in the range of $1x10^4$cP, requires that the viscosity be reduced to approximately 400cP for its effective and efficient production and transportation. Researchers in this field have reported several techniques employed by industry to achieve reduction in heavy oil/bitumen viscosity such as the use of heating system in the formation and/or pipeline [1, 2], addition of liquid diluents and injection of liquids or gaseous diluents into the formation, and the use of water and emulsion [3–5]. Although these methods, to an extent have been effective in reducing the oil viscosity the high costs of energy and diluents restrict their frequent applications. The diluents used by producers with greater success in large quantities are aromatic solvents, as in the Vapex process to reduce viscosity of heavy oil under reservoir conditions. Other classical diluents which have been used are naphthas, condensates, kerosene and light paraffins. These diluents, although environmentally friendlier and more cost effective than the aromatic solvents, have been identified with the problem of precipitation of asphaltenes from the heavy oils/bitumen which can cause severe clogging of wells and pipelines and totally halt production leading to loss in revenue [4].

Viscosity is an important physical parameter for crude oil that characterizes not only one of the rheological property, but also the quality of most crude oil products such as lubricating oils and lubricants [6]. Asphaltenes impact high viscosity to crude oils, negatively impacting production. Their precipitation and deposition are undesirable situation that can create a myriad of production problems. The choice of appropriate diluent is essential in maintaining the asphaltenes and other solids in solution during production operations. This study aims at the influence of light crude from different Nigerian oilfields in reducing the viscosity of pure bitumen and deasphalted bitumen. The effects of weight percent of asphaltenes on the viscosity of the bitumen-light crude oil mixtures are examined. The influence of composition and temperatures on the viscosities of the mixtures is correlated with independent and generalized viscosity model of Singh et al [7].

2 Experimental methods

The light crude oil samples used as diluents were obtained from oil wells in the following locations in the Niger Delta region of Nigeria; Umutu Flow station, Beryboye Flow station, Beniboye Flow station, Forcados field, Warri Terminal and Kaduna Refinery. Two bitumen samples, Rush Lake and Plover Lake, (from Alberta, Canada) were supplied by Alberta Research Council. The solvent n-pentane is analytical grade chemical (Sigma-Aldrich, HPLC grade, 99.9%) commercially available and was used as received.

Asphaltene precipitation was carried out using 1:40 ml bitumen : n-pentane ratio. The mixture was agitated using a G10 gyrotary shaker for 24 hours, left to stand for two hours and filtered using a 7.0cm whatman filter paper as described in ASTM 863-63. The procedure used to recover the deasphalted bitumen has been documented in our previous publication [8]. To prepare the mixtures of light crude oil containing 5wt% to 40wt% bitumen, a known amount of the bitumen was added quantitatively to the light crude oil in increment of 5wt%.

The mixtures, contained in tightly closed 200ml sample bottles, were then placed in the shaker for 7 days. This was to ensure for complete dispersion of the bitumen and the homogeneity of the mixtures. The kinematic viscosity of the samples was determined with Canon-Finske opaque reversible capillary viscometer in accordance with the procedure in ASTM D-445-83-1986.

3 Discussions of results

Figures 1 and 2 illustrate the viscosity-temperature relationship for the pure light crude oils, and the mixtures that contained different wt% of pure and deasphalted Rush Lake and Plover Lake bitumen. As expected, the results clearly show that the presence of asphaltenes, even in a small amount, in crude oil has a significant impact on the viscosity of the oil. The results are consistent with those obtained in our previous work with calorimetric method [8]. Figures 1 and 2 gave a true

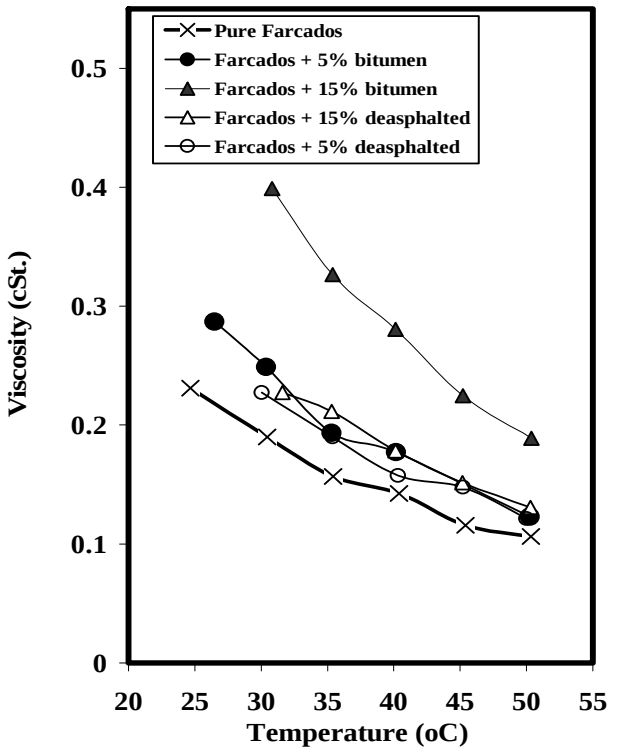


Figure 1: Viscosity-temperature relationship for Farcados crude oil with different wt% Rush Lake bitumen.

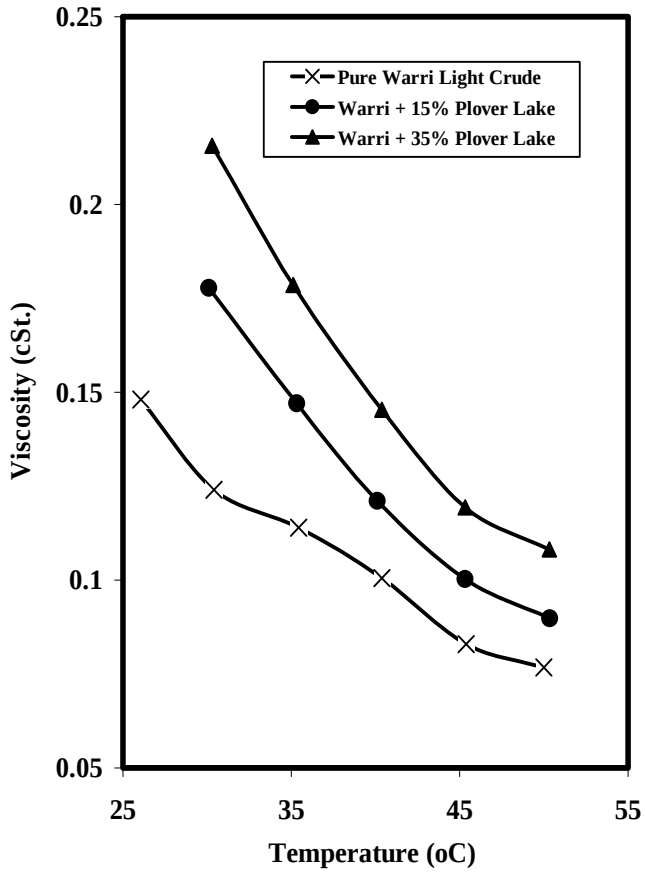


Figure 2: Viscosity-temperature relationship for Warri light crude oil with different wt% Plover Lake bitumen.

representation of the results observed with all the crude oil samples examined in this study. The light crude oils used contained no asphaltenes, and after blending with asphalted bitumen the mixtures remained homogeneous throughout the experiment. Mixtures examined contained as much as 40wt% asphalted bitumen and 25wt% deasphalted bitumen. For the light crude oil samples used in this study, the viscosity increases exponentially with higher weight percent of asphalted bitumen. Similar upward trend was observed with deasphalted bitumen, but with significant reduction in the percent increment as shown in Figures 3 and 4. The impact of asphaltenes on viscosity is clearly evident in Figure 3. With 20wt% addition of asphalted bitumen the oil viscosity at 30ºC increased by approximately 160% and by only 60% for the same amount of deasphalted bitumen. This trend was observed over the studied temperature range for the entire mixtures, as shown in Figure 4.

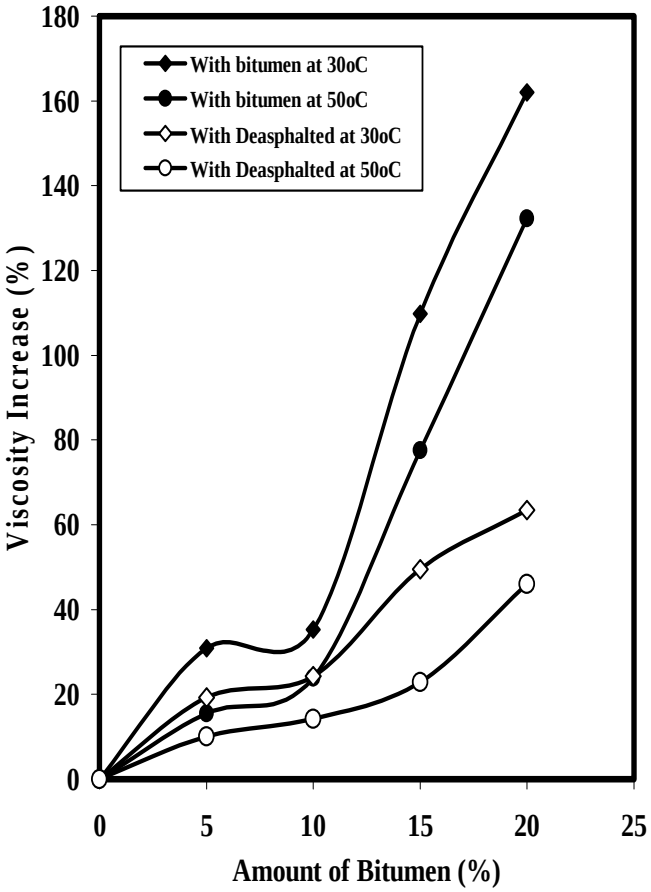


Figure 3: Percent increase in viscosity of Farcados light crude with different wt% of asphalted and deasphalted bitumen.

It is evident from the results that the presence of asphaltenes is mainly responsible for high viscosity characteristics of crude petroleum. The results show also that at higher temperature there is no significant difference in the percent reductions in viscosity between the samples (light crude oils, asphalted mixtures and deasphalted mixtures). These observations can be attributed to the fact that asphaltenes structures are hardly affected at these temperatures to have significant impact on viscosity, as a result of heat transfer to diluents. It is well known that temperature has a strong effect on viscosity of heavy petroleum. At high temperatures the high molecular weight hydrocarbon components of the heavy petroleum disentangles, breaking the physical bonds and in most cases the chemical structures of the asphaltenenes and other solids, and by so doing the oil viscosity is reduced [9]. However, our results illustrate that at the temperature range studied, viscosity reduction of heavy petroleum with diluents is mostly

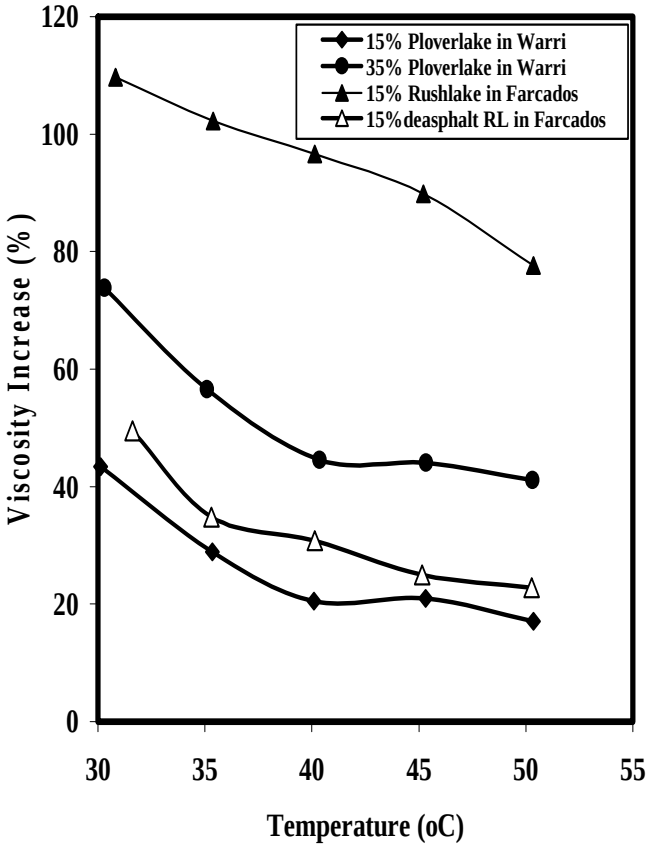


Figure 4: Effect of temperature on percent increase in Viscosity of light crude oils with asphalted and deasphalted bitumen.

governed by molecular disentanglement through diffusion phenomenon rather than temperature effects on the chemical structures of asphaltenes and other heavy components in the mixtures.

The viscosity-temperature relationship for different Nigerian light crude oil samples are shown in Figure 5, and in Figures 1 and 2 for those containing different weight percent of asphalted and deasphalted Rush Lake and Plover Lake bitumen. Singh et al [7] proposed a modified version of the Roelands Equation to correlate the viscosity of light oils as follows:

$$\log \eta = \frac{b}{1+\left(\frac{T-37.78}{310.93}\right)^{S}} + C \qquad (1)$$

η is kinematic viscosity (cSt), T is temperature (ºC), C is given as -0.8690, b is the characterization parameter defined as: $b = log_{10}\eta_{37.78°C} - C$, and S is the shape

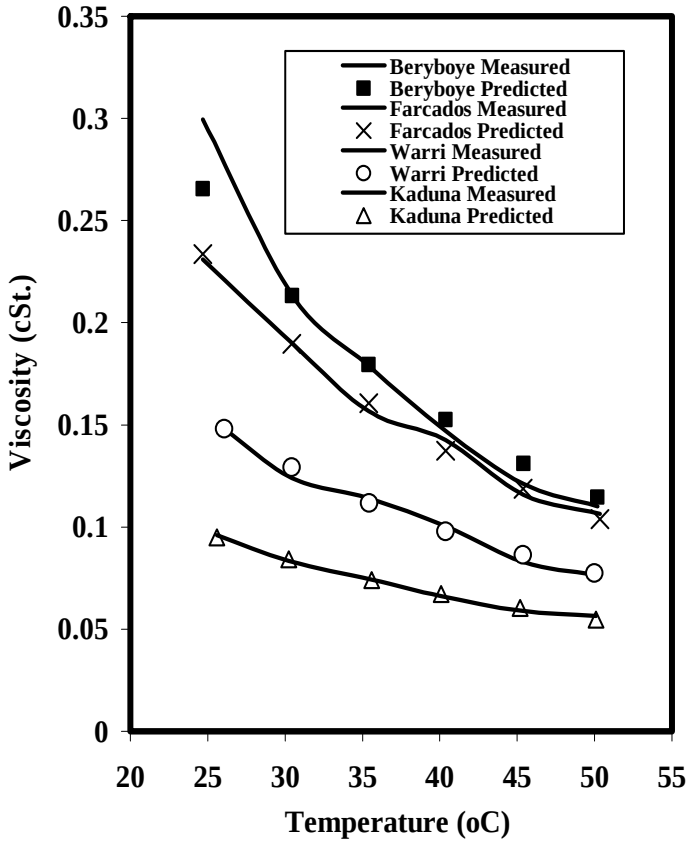


Figure 5: Comparison of measured and predicted viscosity for different Nigerian Light crude oils.

factor which relates to the characterization parameter as follows: *S= 0.28008*b – 1.6180*. The viscosity data were fitted with Eq.(1) to obtain new parameters, for *C=-2.1042* and *S =.2800*3.0166*, with regression coefficient, R^2, of 0.998. With the optimized characterization parameters, the equation was able to better estimate the viscosities of the light crude oil-bitumen mixtures. Figure 5 shows the comparison between measured and predicted viscosities at different temperatures for different samples. Except for Beryboye light crude at 25ºC there is a good match between measured and predicted viscosities for all the examined light crude oil samples. The correlation percent errors at different temperatures for the light crude samples, and mixtures containing different weight percent of either asphalted or deasphalted bitumen are compared in Figures 6–8. The percent errors were within 4% for pure Farcados light oil (Figure 6) and Warri

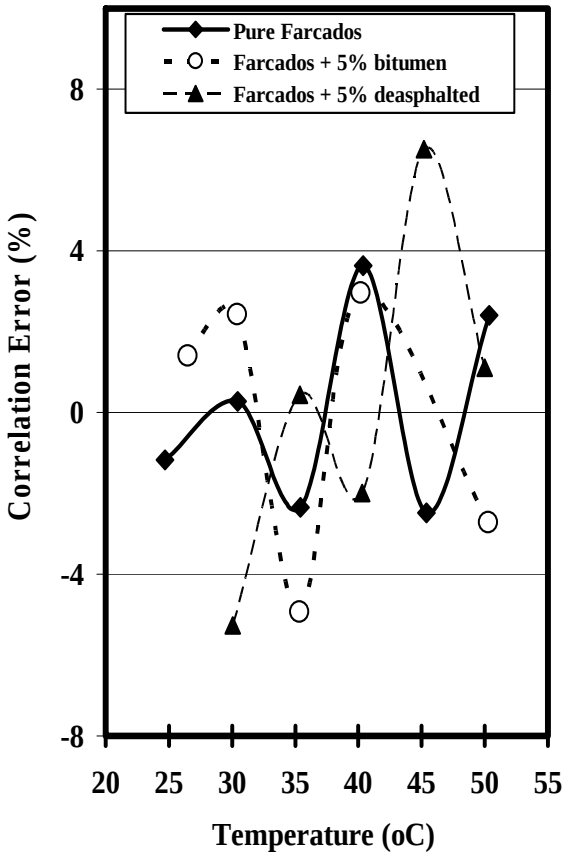


Figure 6: Deviations between measured and predicted viscosity at different temperatures for 5wt% Rush Lake in Farcados crude.

light oil (Figure 8). Similar results were obtained for the entire pure light oil sample used in this study. The percent average deviations for the samples are given in Table 1. The results show that the light oil viscosity characteristics were not affected by the addition of asphalted and deasphalted bitumen up to 40wt%. As mentioned earlier, the addition of bitumen increased the viscosity of the light oils (Figures 1 and 2), with asphalted bitumen nearly tripled the increase in viscosity obtained with deasphalted bitumen.

Besides this phenomenon, the overall viscosity behavior of the mixtures remained similar to that of the pure light crude oils even at high temperatures. The percent errors obtained with the viscosity equation (Eq.(1)) for the mixtures illustrate that the mixtures examined were homogeneous, as the predicted viscosity closely matched the measured viscosity. In many cases, better match between measured and predicted viscosities were obtained for the mixtures than for pure light crude oil samples. The percent average deviations for these

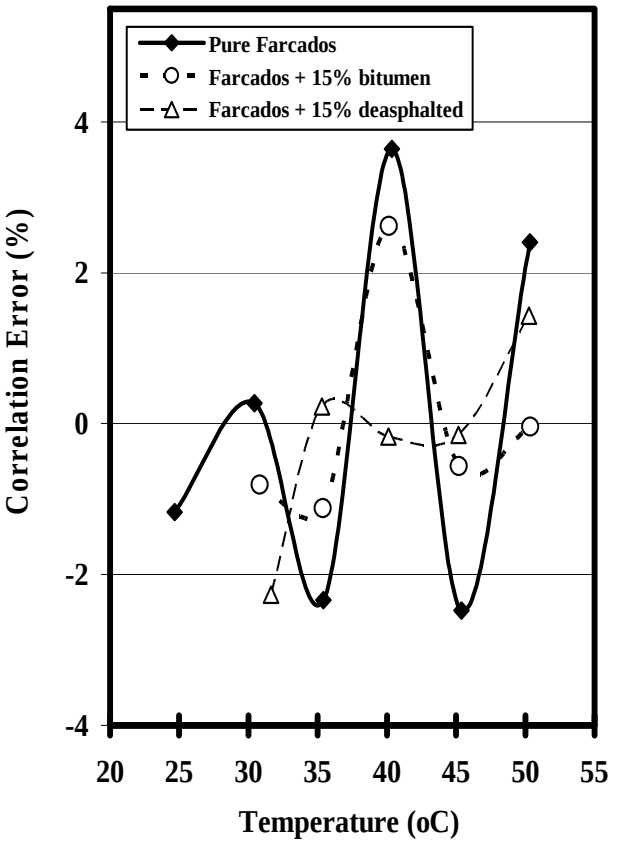


Figure 7: Deviations between measured and predicted Viscosity at different temperature for 15wt% Rush Lake in Farcados Crude.

samples are compared in Table 1. The results illustrate that light oil is arguably a good solvent for bitumen, as the asphaltenes and other solids (resins, maltenes, heavy saturates) appeared to be soluble in the oils. It is worthy to note that the light crude oils used in this study contain no asphaltenes, and the compositions are mainly low molecular weight paraffins. Aromatic solvents, such as toluene and xylene, have been long identified as the best solvent for reducing viscosity of heavy oil/bitumen due to their ability to maintain asphaltene molecules in solution. For the light crude oils used here, molecular diffusivity, intermolecular attraction and low hydrogen bonding rather than molecular polarity might be the crucial factors that characterize the oil as good solvent to reduce the viscosity of bitumen.

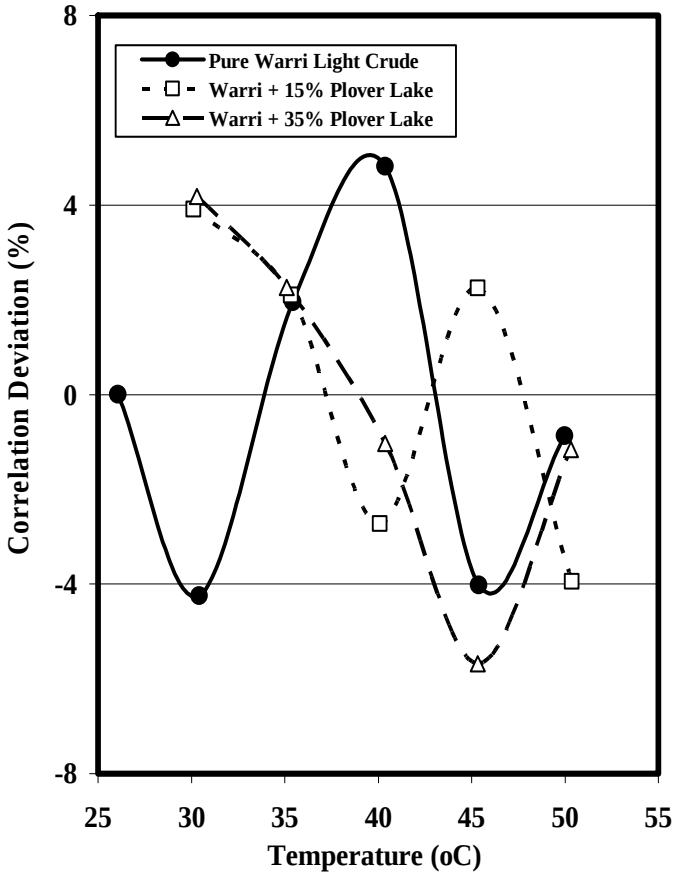


Figure 8: Deviations between measured and predicted viscosity at different temperatures for Plover Lake in Warri crude oil

4 Conclusions

The most important benefit of this study is that knowing the relatively huge reserve of non-asphaltic light crude oils and the untapped bitumen and heavy oil resources in Nigeria, we have shown that non-asphaltic light crude oils can be successfully used in production and pipeline transportation of bitumen. The results show that for up to 40wt% bitumen the viscosity characteristics of the mixtures remain identical to that of the light crude oil, an indication that the asphaltenes and other resinous solids remain soluble in the mixtures. The viscosity correlations for both pure light oil samples and the mixtures gave similar percent deviations with an overall absolute deviation of 2.6%; making light crude oils a good diluent for bitumen.

Table 1: Percent average deviations between measured and predicted viscosity values for pure and mixed samples.

Oil Sample	Data Points	Mean
Beryboye light crude	5	4.28
Farcados light crude	5	2.25
Beniboye light crude	5	2.93
Wari light crude	5	2.98
Kaduna light crude	5	1.35
Farcados-5%-Rushlake bitumen	5	2.60
Farcados-10%-Rushlake bitumen	5	1.74
Farcados-15%-Rushlake bitumen	5	1.20
Farcados-20%-Rushlake bitumen	5	2.17
Farcados-5%-Deasphalted bitumen	5	3.31
Farcados-10%-Deasphalted bitumen	5	1.95
Farcados-15%-Deasphalted bitumen	5	2.85
Farcados-20%-Deasphalted bitumen	5	1.82
Farcados-25%-Deasphalted bitumen	5	3.13
Warri-15%-Plover Lake Bitumen	5	3.90
Warri-35%-Plover Lake Bitumen	5	2.86
Average Absolute Deviation (%)	**85**	**2.60**

Acknowledgements

This work was funded in part by NSERC, Canada, and PTF, Nigeria. The supports of American University of Nigeria (AUN), and NNPC Research Lab (Port Harcourt), are highly appreciated. Ours thanks to Harrigan A. Pepple (NNPC, Warri), MacDonald I. Bara-Hart (NNPC, Benin), and Samuel O. Robinson (AGIP Oil, Umoku) for their various contributions in the acquisition of field samples and related data.

References

[1] Urquhart, R.D. Heavy Oil Transportation: Present and Future. J. Can. Petrol. Technol., **25** (2), pp.68-71, 1986.

[2] Chang, C., Nguyen, Q.D., & Ronningsen, H.P. Isothermal start-up of pipeline transporting waxy crude oils. J. Non-Newt. Fluid Mech., **87**, pp.127-54, 1999.

[3] Mokrys, I.J., & Butler, K.M. In-Situ Upgrading of Heavy Oils and Bitumen by Propane De-Asphalting:The VAPEX Process. Paper SPE 25452 presented at SPE Symposium, Oklahoma City, Oklahoma, March 21-23, 1993.
[4] Das, S.K., & Bulter, R.M., Mechanism of the Vapour Extraction Pocess for Havy Oil and Bitumen. *J. Petrol. Sci.*, **12**, pp. 219-231, 1995.
[5] Byran, J.,& Kantzas, A. Enhanced Heavy-Oil Recovery by Alkaline-Surfactant Flooding, Paper SPE 110738 presented at SPE Tech. Conf. & Exhibition, Anhein, California, Nov. 11-14, 2007.
[6] Sanchez, J.H. & Mansoori, G.A. Prediction of the Phase Behavior of Asphaltne Micelle/Aromatic Hydrocarbon Systems. Petrol. Sci. & Technol,**16** (3&4), pp.377-394, 1998.
[7] Singh, B., Miadonye, A. & Puttagunta, V.R. Modelling the Viscosity of Middle-East Crude Oil Mixtures. Ind. & Engr. Chem. Res.,**32** (9), pp.2183-2186, 1993.
[8] Miadonye, A., Secka, K.A., Cyr, J. & Britten, A. Study of thermo-physical properties of bitumen in hydrocarbon condensates. WIT Transact. on Modelling and Simulation, Numerical Methods, **48**, pp.125-134, 2009.
[9] Khan, M.R. Rheological properties of heavy oil emulsions. Energy Source, **18**, pp.385-391, 1996.

"Optic Ped Scan": an alternative inexpensive technique to study plantar arches and foot pressure sites

A. R. Jamshidi Fard[1] & S. Jamshidi Fard[2]
[1]*Department of Physiology and Biophysics, School of Medicine, Arak University of Medical Sciences, Iran*
[2]*McTimoney College of Chiropractic, Abingdon, UK*

Abstract

Early diagnosis and management of foot arch abnormalities would reduce future complications. Conventional, mainly non-objective clinical examinations were not evidence based and somehow due to expert ideas with high inter rater differences. In new Pedscope we made the Optic Ped Scan (OPS), patient stand on a Resin made 5-10 mm Plexiglass while the image of the whole plantar surface was digitally scanned, showing the pressure sites in different colours based on a Ratio-Metric Scale. Any off-line measurement or relative pressure ratios could be easily studied. The outcome of the OPS is an image file resulting from the subject's body weight on emptying the capillaries of plantar skin which causes the colours to change. These physiological changes of plantar colour could be amplified when passing though the clear, hardly elastic form of plexiglass (Acrylic or methyl methacrylate Resin or Poly methyl 2-methylpropenoate - PMMA), we prepared in Arak Petrochemical Company. We studied 2007, school age students as a pilot study of the technique in Arak schools and measured: Foot Length, Axis Length, Heel Expanse, Mid-foot Expanse, Forefoot Expanse, Arch Length (medial), Arch Depth, Hallux Distance and relative pressures of 10 defined zones.

Students had 28.15% Flat Foot, 1.54% Pes Cavus, 11.01% Hallux Valgus, 0.64% Hallux Varus, 0.04% Convexus and 0.04% complex/various deformities. OPS worked properly for both diagnosis and measurements. The new technique could be several times cheaper than other techniques.

Keywords: foot arches, foot pressure sites, optic pod scan, foot deformity.

WIT Transactions on Modelling and Simulation, Vol 51,
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110141

1 Introduction

In 1976, Sharrard mentioned flat foot, knock knee and intoeing or metatarsus varus, as the three most common causes for parental concern when their child starts to walk [1]. Accurate diagnosis and proper management of foot arch abnormalities may reduce future musculoskeletal complications. Diagnosis, interventional treatments and estimation of the prevalence of foot deformities in screening studies need more reliable quantitative techniques than a physician's observation. Conventional, mainly non-objective clinical examinations were not evidence based and somehow due to expert ideas with high inter rater differences.

In past decades, arch exams were performed by means of a pedscope chair, a wooden chair with a thick glass on its top roof that reflects an image of plantar surface of the subject's foot to an oblique (45º) adjusted mirror set so the image could be seen from outside. With technique, the examiner was not able to measure the foot arch parameters, estimate the pressure of any particular sites or and record the findings for later comparisons but it was still helpful to show the plantar contact and weight bearing surface area of the foot with the brighter sites implying more contact pressures. The traditional solution to recording these sorts of findings, as still remembered by many middle aged colleagues, was to puff the patient's bare feet with Talc powder before asking him to stand again on a black cardboard as a printout or less durable evidence of the study.

On the other hand, new techniques such as LASER, 3D Scanners, Multi cameras and pressure platforms or sensitive mats (Pedobarographs) are much more accurate and advanced methods but they are mostly expensive and still not easily accessible in a physician's office.

This study was designed to test the hypothesis that the conventional clinical examination of the foot just by looking at a child's feet as they stand up, stand on tiptoe or dangle the foot in the air as they sit on an exam table, is insufficient for diagnosing of foot deformities.

Optic Ped Scan (OPS) or similar quantitative measurement can lead to a much better views of foot plantar surface and its actual pressure bearing areas than the conventional approach and judging based on the soft tissue mass of the medial arch.

In this paper, OPS is introduced as an alternative technique and the results of its first pilot application to detect foot deformities are presented.

2 Methods

In new Pedscope we made (OPS), patients stands on a Resin made of 5-10 mm Plexiglass while the image of the whole plantar surface was digitally scanned by an ordinary office type colour scanner *(hp - 3770)* fixed beneath the plexiglass sheet to a computer. A 300 dpi resolution of scanning gave reasonably sharp images, showing the arches and pressure sites of the subject's foot floor in different colours based on a Ratio-Metric Scale with bright yellow indicating highly pressured areas and dark red for non pressured areas.

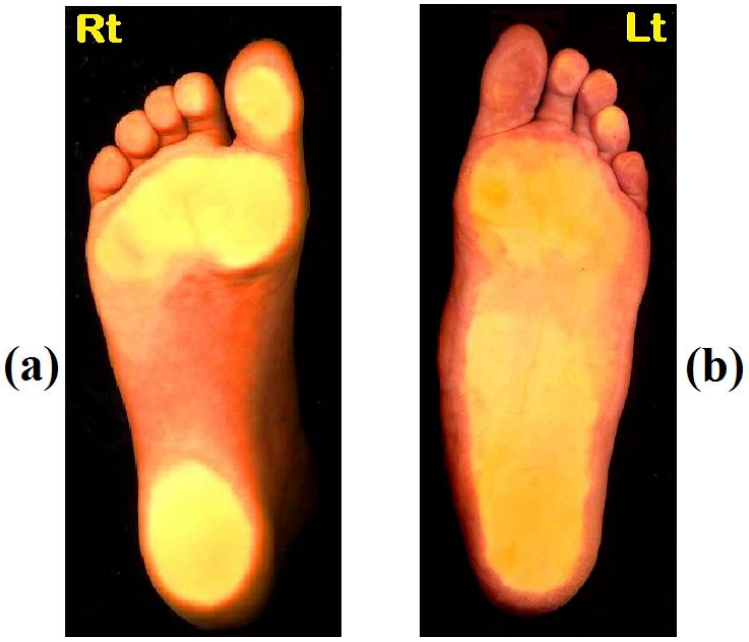


Figure 1: Examples of OPS layouts. (a) Pes Cavus and (b) Flat foot.

The outcome of the OPS is an image file (JPG/TIFF/ BMP) resulting from the subject's body weight on emptying the capillaries of plantar skin which turns the pressured sites into a yellowish colour while the other areas with lower weight bearing ratios are pink-reddish.

These physiological changes of plantar colour could be amplified when passing though the clear, hardly elastic form of plexiglass (Acrylic or methyl methacrylate Resin or Poly methyl 2-methylpropenoate – PMMA) which was prepared after several pilot studies (Figures 1 and 2). We optimized the aimed optical quality of this thermoplastic PMMA by reducing its stage of "Rubber Toughening" at Arak Petrochemical Co., R&D Unit *(www.arpc.ir)* in Arak, Iran.

The colour density of any particular area of the image could still be a rational representative of the pressure applied or a percentage of 100% of subject's body weight if standing on one foot. In this way, all three anatomic and functional arches of the foot [2] could be scanned and sent to a computer. Any off-line measurements or calculating relative pressure ratios of each part of the achieved footprint could be easily studied.

To examine the application of the new technique and optimize the proper thickness of the PMMA sheet for a different range of subject's weight, we studied 2007, 12-13 year old Students in Level 5 of primary schools in Arak. 874 girls and 1133 boys participated in a pilot study of the technique. Each subject was scanned three times while standing on one foot (right and then left foot) and a normal double feet standing. All scanned images of each subject were saved for later analysis.

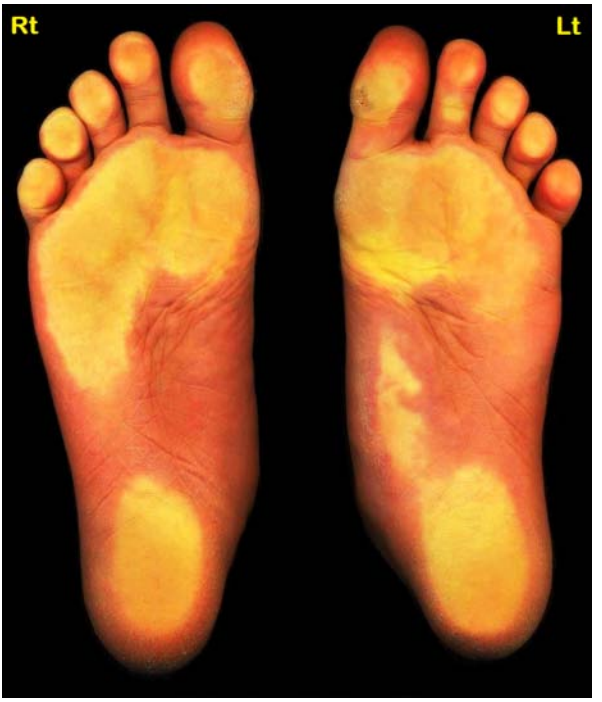


Figure 2: Examples of OPS layouts. Complex deformities in a 12 y.o. girl.

Recorded files of subjects of this technique can be opened by any Windows compatible painting or graphics software and all intended 2 dimensional measurements could be available, using the ruler of the screen in Inches, Centimetres, Millimetres or even Pixels. We used Corel Photo-Paint (CorelDraw 5, Corel Corporation, Canada) facilities for all measurements.

Most of the commercial graphics software also has an 'Equalize' filter which can redistribute shades of colours. 'Equalize' makes the darkest colour black and the lightest colour white and stretches the rest of the colours in between. We used this software to equalize the outcome image of OPS first to improve its appearance before measuring the foot plantar parameters. The Equalize filter was not used before the relative pressure measurements were taken.

The following parameters of footprints were measured: Foot Length (from heel to toe tip), Axis Length (from heel to second toe tip), Heel Expanse, Mid-foot Expanse (Foot expanse at mid-arch level), Forefoot Expanse (Foot expanse at the Metatarsal heads level), Arch Length (medial bow or first ray), Arch Depth (from soft tissue line to weight borne medial border), Hallux Distance (from toe soft tissue border to main Axis) (Figure 3).

For the evaluation of the pressure distribution of the foot, each footprint obtained from OPS was divided into ten areas as functional columns of the weight bearing (Figure 4). These were the first metatarsal head, second and third

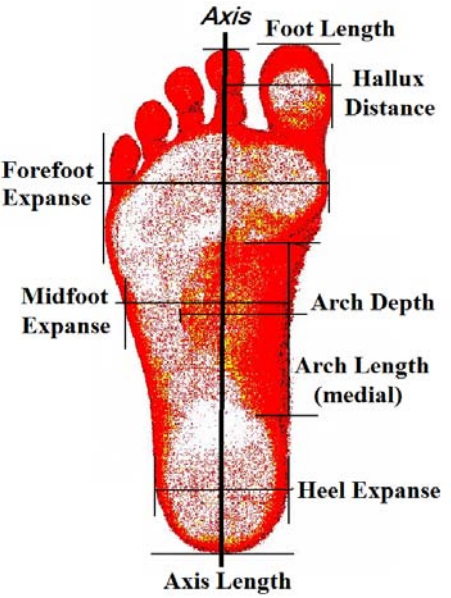


Figure 3: Foot parameters which can be measured exactly in each OPS layout.

metatarsal heads, fourth and fifth metatarsal heads and their corresponding toes in forefoot. The other areas of loading were the heel with two medial and lateral parts, the mid-foot, and the area beneath the navicular bone as the origin of the medial arch first ray [3, 4].

Almost all graphics software has the ability to measure a grey scale in its edit/ make colour toolbar (RGB indicator flyout) which can scale a grey colour between white (255) and black (0) relatively. To measure the relative pressure of 10 divided zones, each footprint image was first converted from colour to grey scale (B&W), then the mean relative whiteness (load) of each zone was measured, assuming the total borne weight as 100% (Figure 4).

The relative pressure measurements, even for each particular zone, could not be pooled to calculate normative data. These measurements could vary greatly because of multiple variables and the relative pressures should be studied in each subject separately.

3 Results

28.15% of students (boys 34.8% and girls 22.9%) had Flat feet based on clinical observation while based on OPS results and Denis classification only 19.8% did (boys 21.7% and girls 17.5%). In this type of foot deformity clinical observation over estimated by 8.3% over the actual prevalence. OPS files also showed 1.54% Pes Cavus (boys 1.05% and girls 2.17%), 11.01% Hallux Valgus (boys 5.82% and girls 17.73%), 0.64% Hallux Varus (boys 0.52% and girls 0.8%),

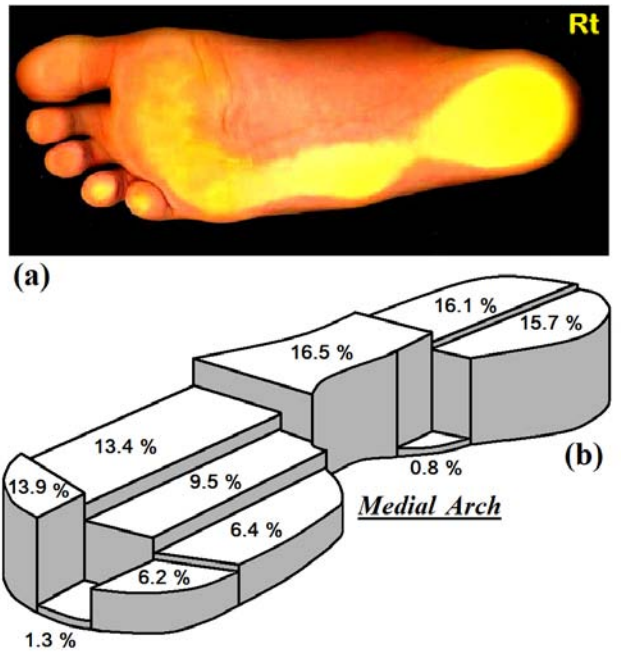


Figure 4: (a) OPS of right foot in a 12 y.o. female student with Bow legs and its abnormal relative pressure distributions on lateral border as a second heel and (b) shows its relative pressure of ten plantar zones, in schematic bars as percentage of 100% borne weight when standing on her right foot.

0.04% Convexus (1 girl) and 0.04% complex/various deformities (1 girl). 222 subjects had two types of deformities, therefore a sum of 28.3% of studied students were diagnosed with abnormal feet.

Foot Length, Axis Length, Heel Expanse, Mid-foot Expanse, Forefoot Expanse, Arch Length, Arch Depth, and Hallux Distance measures of each foot are summarized in Table 1. Between the two ganders, some measured parameters such as Foot Length, Axis Length, Heel Expanse, were significantly different in girls ($P<0.05$).

As an example of relative pressure measurements, the results of an abnormal weight distribution in right foot of one subject are presented in Figure 4. She was one of the participant students with bow legs and normal feet apparently.

Comparing to LASER, 3D Scanners, Multi cameras and Pedobarographs price lists of several companies with OPS set, the new technique coasts significantly less, at almost twice the price of a simple A4 colour scanner. The installed scanner software produces the image in any desired Windows format and so additional software is not necessary for this technique.

Furthermore other multi purpose graphics software available on an office computer could also be used for OPS.

Table 1: The mean of measured parameters of all subjects.

Foot Parameters *n=2007 (874 girls, 1133 boys)*	**Right Foot** *n=2007 (Mean ± SD) - Cm*	**Left Foot** *n=2007 (Mean ± SD) – Cm*
Foot Length	22.56 ± 1.02	22.46 ± 1.06
Axis Length	21.85 ± 1.13	21.76 ± 1.03
Heel Expanse	5.3 ± 0.09	5.29 ± 0.07
Mid-foot Expanse	6.66 ± 0.1	6.66 ± 0.08
Forefoot Expanse	8.13 ± 1.01	8.15 ± 0.93
Arch Length	5.85 ± 0.89	5.85 ± 0.96
Arch Depth	2.42 ± 0.66	2.8 ± 1.03
Hallux Distance	3.24 ± 0.99	3.28 ± 1.12

4 Discussion

An OPS footprint with flatfoot could be classified according to Denis [5] into three grades of severity: grade 1 in which the support of the lateral edge of the foot is half of that of the metatarsal contact area (almost Mid-foot Expanse = 1/2 of Forefoot Expanse); grade 2 in which the weight bearing areas of the central zone and forefoot are equal; and grade 3 in which the support in Mid- foot or central zone is greater than the width of the metatarsal support area. This classification could be described by OPS measurement methods.

Pfeiffer et al. [6], reported a prevalence of 44% in a group of 835 (411 girls and 424 boys) children of 3 to 6 years old (52% in boys and 36% in girls). They used a three-dimensional laser surface scanner with the subjects in a standing position. García-Rodríguez et al. [9] measured the prevalence of flat feet in 1181 school age children of Malaga in Spain as 2.7% which was the prevalence of stage 2 and 3 in Denis classification. 2.7% of students were diagnosed to have flat feet based on plantar footprints.

Children with first degree flat foot could be considered abnormal based on clinical observation while looking at their footprints usually shows no enhancement of contact area at the central zone. These incomplete arches could be considered as evolutionary foot problems without pathological significance [7, 8]. Prevalence of flat foot in OPS results in our subjects were closer to 12.3%

Although this was mentioned by Denis, other reported a prevalence several times higher than this figure [9, 10].

The critical age for development of the plantar arch is 6 years [11]. Consequently, if the sampled population of such studies cover this age, the prevalence of flat feet will overestimate the problem. Infants are born with flat feet, and the main medial arch develops gradually by weight bearing during the first decade of life [12, 13]. Most flat feet conditions correct themselves

spontaneously at the age of 6 or more and a retarded development of the foot medial arch is reported in boys [6].

Significant differences in prevalence of flat foot is also found in overweight, obese, and normal-weight children and boys showed a significantly greater tendency for flat foot than girls [6]. It seems that different variables such as age, weight and gender may influence the prevalence rates when the subject's feet are examined in static, mainly standing positions. In addition to different parameters affecting the results of prevalence of flat foot in population screening studies, data based on clinical observations may be over estimated unless an objective technique such as OPS is adopted.

The technique of OPS shows beneath the foot and its weight borne contact surface and may reduce the examiner's bias in evaluation of children's feet.

It is worth mentioning that, in our normal subjects pressure distribution patterns of feet did not differ significantly between males or females and the right or left feet when standing on one foot alone but this pattern varied greatly between subjects when the OPS was performed in double feet weight bearing. Similar results were also reported with different techniques [3].

5 Conclusions

We conclude that OPS can be used sufficiently for both screening studies, diagnosis and measurements of the foot plantar parameters and relative pressure sites, but due to different types, sizes and contact areas, different weights, reciprocal weight distribution and complex structure of the foot, it is difficult to draw the line between normal and abnormal feet. The pattern of pressure distribution in each particular area may be studied before and after different interventions or compared for the clinical diagnosis of the abnormal cases but it is not possible to justify a normative data even as a percentage of borne weight.

The actual prevalence rate of foot deformities are higher than appears in literature in the region and society / school authorities involved with child health care should reconsider the foot plantar deformities of students before they start their more active future life.

References

[1] Sharrard, WJW., Intoeing and flat foot. *British Medical Journal*, **1,** pp. 888–889, 1976.

[2] Soames, RW., Appendicular skeleton. *Gray's anatomy: the anatomical basis of medicine and surgery. 38th ed.*, ed. Williams, PL., Bannister, LH., Berry, MM., Collins, P., Dyson, M., Dussek, JE. *et al*. Churchill Livingstone: New York, pp. 733–4, 1995.

[3] Kanatli, U., Yetkin, H., Simsek, A., Ozturk, AM., Esen, E., Besli, K., Pressure distribution patterns under the metatarsal heads in healthy individuals. *Acta Orthop Traumatol Turc*, **42(1),** pp. 6–30, 2008.

[4] Myerson, MS., The diagnosis and treatment of injury to the tarsometatarsal joint complex. *J. Bone Joint Surg [Br]*, **81**, pp. 756–63, 1999.

[5] Denis, A. Pied plat valgus statique. *Encyclopedie Medico-Chirurgicale Appareil Locomoteur*. Paris, France: *Editions Techniques*; 1974.
[6] Pfeiffer, M., Kotz, R., Ledl, T., Hauser, G. & Sluga, M., Prevalence of Flat Foot in Preschool-Aged Children. *Pediatrics*, **118**, pp. 634–639, 2006.
[7] Meary, R., Symposium sur le pied plant. *Ann Orthop Ouest*. **1**, pp. 55–71, 1969.
[8] Stewart, SF., Human gait the human foot: an ethological study of flatfoot. *Clinical Orthop Res*. **70**, pp. 111–123, 1970.
[9] García-Rodríguez, A., Martín-Jiménez, F., Carnero-Varo, M., Gómez-Gracia, E., Gómez-Aracena, J., Fernández-Crehuet, J., Flexible flat feet in children: a real problem? *Pediatrics*. Jun; **103(6)**, e84. 1999.
[10] Gil, A., Estudio epidemiolo´gico del pie plano en escolares. *Acta Pediátrica Española*. **49(10)**, pp. 667–670, 1992.
[11] Rose, GK., Flat feet in children. *Br Med J*. **301**, pp. 1330–1331, 1990.
[12] Volpon, JB., Footprint analysis during the growth period. *J. Pediatr Orthop*. **14**, pp. 83–85, 1994.
[13] Forriol, F., Pascual, J., Footprint analysis between three and seventeen years of age. *Foot Ankle*. **11**, 101–104, 1990.

This page intentionally left blank

Stereovision for the determination of fireline intensity in vegetation fire

L. Rossi, T. Molinier, P. Santoni, T. Barboni & F. Morandini
UMR CNRS 6134, University of Corsica, Corsica

Abstract

Fireline intensity is one of the most relevant quantities in forest fire science. It helps to evaluate the effects of fuel treatment on fire behaviour, to establish limits for prescribed burning. It is also used as a quantitative basis to support fire suppression activities. However, its measurement at field scale for actual fire remains a challenge. Hence, it has been poorly used as a key quantity to test the new generation of models of fire spread that have been developed these last ten years. An inverse method to obtain fireline intensity is through the observation of the flame length. This geometrical information is measured using a stereovision system placed in the lateral position relative to the direction of the fire spread. Algorithms were developed in order to automatically segment the fire area of the images and estimate the 3D coordinates of salient fire points and then the flame length. The three dimensions of the information permit to obtain the flame length with metric measures. In the present work, we directly measure the fireline intensity at laboratory scale by oxygen consumption calorimetry. The results are then used to establish a relationship between fireline intensity and flame length obtained by the stereovision system.
Keywords: stereovision, 3D coordinates, fireline intensity, flame length, oxygen consumption calorimetry.

1 Introduction

Fires devastate regularly forests and shrublands as well as populated areas all over the world. Foresters and fire fighters are faced with problems such as the management of wildland/urban interfaces, the establishment of safety zones and suppression strategies. An important concept helpful in fire mitigation and fight is to scale fires in function of their potential threat. This scale is based on the

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110151

measure of the fireline intensity of fires, but such measure remains a challenge. The concept of fireline intensity was introduced by Byram [1] in the 1950s. It is the rate of heat released per unit time per unit length by the active flaming zone of the fire front. It is given by:

$$I_B = Hwr \tag{1}$$

where I_B (kW/m) is the fireline intensity, H (kJ/kg) is the heat yield of the fuel, w (kg/m^2) is the weight of fuel consumed in the active flame front and r (m/s) is the rate of spread of the fire. The fireline intensity is a widely used measure in forest fire applications: it helps to evaluate the effects of fuel treatment on fire behaviour [2], to establish limits for prescribed burning [3], to assess fire impacts on ecosystems [4]. It is also used as an indicator for the classification of fires in terms of risk [1] or as a quantitative basis to support fire suppression activities [5]. The measure of the fireline intensity according to Eqn (1) is extremely difficult because it requires determining the fuel consumption in the fire front. An alternative method consists in measuring the geometrical properties of the flame and use a correlation to derive the fire intensity. The flame length is defined as the distance from the base of the flame to the highest point of the flame [6]. For instance Byram [1] proposed the following relationships between flame length and intensity:

$$L_B = 0.0775\, I_B^{0.46} \Leftrightarrow I_B = 259 L_B^{2.17} \tag{2}$$

Several authors have proposed other correlations to estimate fireline intensity from flame length measurements [7–10]. These laws exhibit differences in the coefficients and the corresponding curves are strongly different. These differences may be attributed to the varying experimental conditions of the observed fires, to the difficulty in measuring the flame length and to the uncertainty in measuring the fireline intensity. For more than two decades, visual and infrared cameras have been used as complementary metrological instruments in fire spreading experiments [11, 12]. The image processes used only 2D information and were not able to provide metric measurements unless calibrated objects were present in the image.

The aim of this paper is to address both major sources of uncertainty. Firstly the use of a stereovision system in order to determine the flame properties is proposed. The stereovision system is built using a visual spectrum pre-calibrated stereo camera without the need of position markers for deriving scale and orthogonality from images. No a priori knowledge is necessary in order to use the proposed approach. Then, we present a methodology to measure the fireline intensity based on oxygen consumption calorimetry principle. Thanks to both methods to measure flame length and fireline intensity, a new relationship between these quantities is obtained at laboratory scale. It is then compared with the existing correlations found in the literature.

2 Stereoscopic methodology

A stereo vision system is composed by two cameras, fig. 1. Each camera, considered as pinhole, is defined by its optical center O_l (for the left one), its optical axis (O_lz_l) perpendicular to the image plane and its focal length. Each camera has intrinsic parameters (pixel coordinates of the projection center, vertical and horizontal distortions) and extrinsic parameters (characterizing the position and orientation of the camera relative to an object; three rotations and three translations parameters). Let P be a point in a 3D space; P_l and P_r are respectively the image of P in the left and right image planes. The line (e_lP_l) is the epipolar line representing the set of left image points corresponding P_l [13–16].

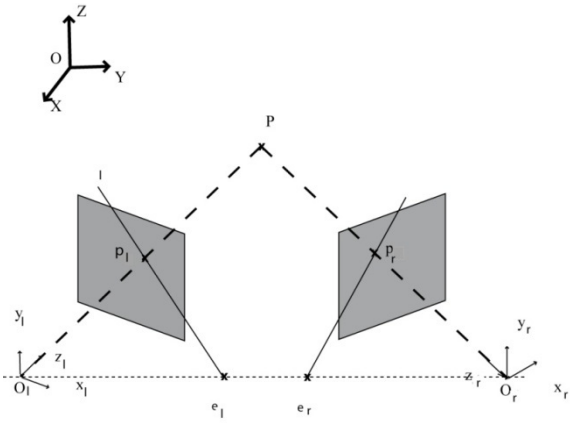


Figure 1: General case of a stereo vision system.

The fundamental matrix F is the algebraic representation of the epipolar geometry:

$$(e_l P_l) = F P_r \tag{3}$$

F incorporates intrinsic and extrinsic parameters of the cameras. The fundamental matrix is computed by calibration with a frame or right-left correspondences. The 3D position is depicted by a stereo disparity defined by the difference between the projected points in two stereo images. A 2D-search is thus needed in order to find corresponding points in both images along the epipolar lines. With the fundamental matrix, it is possible to do an image rectification. The rectification procedure transforms each image plane such that pairs of conjugate epipolar lines become collinear and parallel to one of the image axes. The rectified images can be thought of as acquired by a new stereo rig, obtained by rotating the original cameras. With rectification, computing stereo correspondences is reduced to a 1-D search problem along the horizontal raster lines of the rectified images. Once the corresponding features are extracted, a triangulation technique is used to compute their 3D coordinates. A

line l is passed through the point in the left image and the left optical center C_l and a line r is passed through the corresponding point in the right image and the right optical center C_r. A mid-point method finds the point which lies exactly at the middle of the shortest line segment joining the two projection lines. This point represents the 3D coordinate of the corresponding pixels, fig. 2.

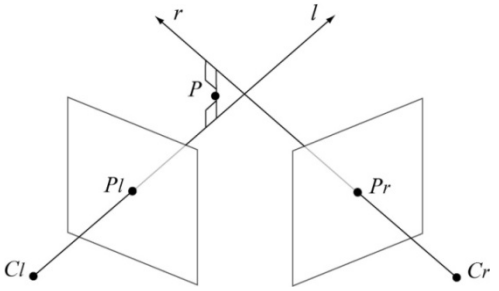


Figure 2: Triangulation method.

Feature points are salient points of the image. The Harris detector is the most commonly used operator for corner point extraction. It is based on local auto-correlation of the signal [17, 18]. This operator measures local changes of gray levels. In some cases, regions of interest detection is performed in a given color system prior to salient points extraction. This segmentation permits the extraction of connected parts of an image based on color criteria [19]. Then follows a contour detection carried out for example by using the Harris edge detector and the Corner detector based on global and local curvature properties [19–21]. Based on global Feature points are then searched along the contours. When feature points are located in the right and left images, a matching algorithm is used in order to find corresponding points in the two images. Matching algorithms can be classified as correlation-based and feature-based methods [17]. In correlation-based methods, used in this work, the elements to match are image windows of fixed size, and the similarity criterion is a measure of the correlation between windows in the two stereo images.

3 Experiments

3.1 Experimental procedure to measure the flame characteristics

The measurement of the flame length was carried out using a stereovision system placed in lateral position relative to the direction of the fire spread (circled system in fig. 3).A Point Grey pre-calibrated Bumblebee® XB3 camera was used [22, 23]. This camera is a trinocular multi-baseline stereo camera with an extended baseline of 24 cm. It has a focal length of 3.8 mm with 66° HFOV. The image sensor is a 1/3" Sony ICX445AQ CCD. This system is pre-calibrated for stereo processing. The image resolution is 1280x960 with a frame rate of 16 FPS.

Figure 3: Experimental vision device.

Image acquisition was developed using C++ with Point Grey Triclops SDK. This system permits the simultaneous acquisition of a pair of images of the scene. The captured images are stored for further processing using algorithms developed with Matlab®. The experiments were carried out with a frame rate of 0.25 Hz. Only the images within the steady phase were treated. The three-dimensional points of fire were obtained from stereoscopic images. The steps involved in this approach are [24]:

1- Segmentation of fire regions; Ko's segmentation algorithm is used in this work [25],
2- Extraction of salient features points,
3- Features selection refinement using a normalized cross-correlation matching strategy,
4- Computation of three-dimensional fire points using stereo correspondence.

Figure 4 shows the fire area segmented in the image with the matched points detected on the entire contour of the fire. Figure 5 shows the 3D points of the fire obtained from the corresponding points marked in Figure 4. The X-axis corresponds to the depth of the fire front; the Y-axis corresponds to its height, and the Z-axis to its width. The reference frame has its origin in the left image center point of the XB3 camera. In order to visualize the correspondence between the 2D and the 3D information, four points are numbered in fig. 4 and fig. 5. With 2D data, it is impossible to distinguish points that are on the ground from points located above whereas it is possible with 3D information. For example, the Y coordinate of the point number 1 shows that this point is not on the ground.

The length of the flame is defined as the distance between the ground and the highest 3D points. It is thus necessary to estimate the position of the ground. A method was developed in order to obtain automatically the ground even if the plane is inclined. The lowest 3D points of the back part of the smoldering fire

Figure 4: Segmentation of the fire region.

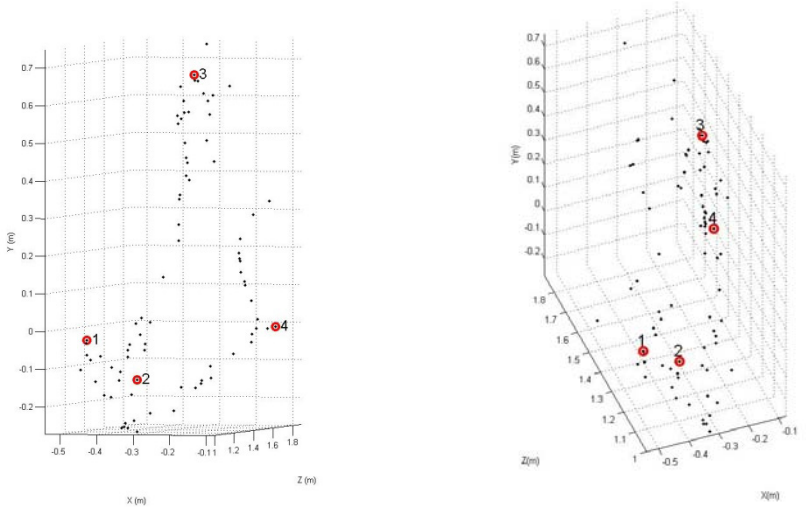


Figure 5: 3D fire points.

front are identified for all the images and a plane is estimated from these points using a least square method. Then, through homography, the coordinates of the 3D points are selected to form a base plane with $Y = 0$. The selection of the 3D lowest points is carried out as follow: the 3D points are divided in sets following the Z axis. In each set, we note Y_{max} and Z_{max} the maximum Y and Z coordinate of selected points, Y_{min} and Z_{min} the minimum Y and Z coordinate of selected points. Then we estimate $\Delta_y = Y_{max} - Y_{min}$. Due to the possible inclination of the ground, two conditions are used to select the points of the ground: $Y \leq Y_{min} + \Delta_y$ and $Z \leq Z_{min} + \Delta_z$. For each 3D set of fire points obtained at a given time, the Y coordinate of the highest point corresponds to the flame length, L. This

procedure is carried out for each image of the set of images acquired during the steady phase. From all the flame length values, the mean $\overline{L}$ and the standard deviation σ_L are computed. Then, the values of L that are within the interval $\left[\overline{L}-\sigma_L, \overline{L}+\sigma_L\right]$ are averaged in order to compute the estimated flame length of the whole steady phase.

3.2 Measurement of fireline intensity by oxygen consumption calorimetry

3.2.1 Experimental device and basic equations

Oxygen Consumption (OC) calorimetry is a common method used to measure the Heat Release Rate (HRR in kW). HRR which describes the fire size of burning materials is a fundamental parameter that is essential to estimate fire hazards [26]. It is then used to derive fireline intensity (kW/m). The oxygen consumption calorimetry principle is based on the assumption that the amount of heat released per unit mass of consumed oxygen is approximately constant for most common burning materials containing C, H and O [27]. Thus, the heat release rate can be approximated by measuring the oxygen deficit in the exhaust gas flow of burning materials. HRR measurements based on OC calorimetry were conducted using a 1 MW Large Scale Heat Release (LSHR) calorimeter (3m × 3m hood). Figure 6 shows the Open-burning HHR Calorimeter with a combustion bench inside. The HRR is given by the following relations

$$\dot{q} = \frac{E\rho_0 W_{O_2}}{W_{air}}\left(1-X^{\circ}_{H_2O}\right)X^{a^\circ}_{O_2}\dot{V}_{s,298}\left(\frac{\phi}{(1-\phi)+\alpha\phi}\right) \tag{4}$$

$$\dot{V}_{s,298} = 22.4A\frac{k_t}{k_p}\sqrt{\frac{\Delta P}{T_S}} \tag{5}$$

$$\phi = \frac{X^{a^\circ}_{O_2}\left(1-X^{a}_{CO_2}\right)-X^{a}_{O_2}\left(1-X^{a^\circ}_{CO_2}\right)}{X^{a^\circ}_{O_2}\left(1-X^{a}_{CO_2}-X^{a}_{O_2}\right)} \tag{6}$$

where $\dot{q}$ is the HRR, E is Huggett's constant [27], ρ_0 is the density of dry air at 298K and 1 atm, W_{O_2} is the molecular weight of oxygen, W_{air} is the molecular weight of air, X denotes the molar fraction, $\dot{V}_{s,298}$ is the standard flow rate in the exhaust duct. Superscript ° is for incoming air, a is for analyzers. A is the cross sectional area of the duct, k_t is a constant determined by calibration with a propane burner, k_p =1.108 for a bi-directional probe, Δp is the pressure drop across the bi-directional probe and T_s is the gas temperature in the duct.

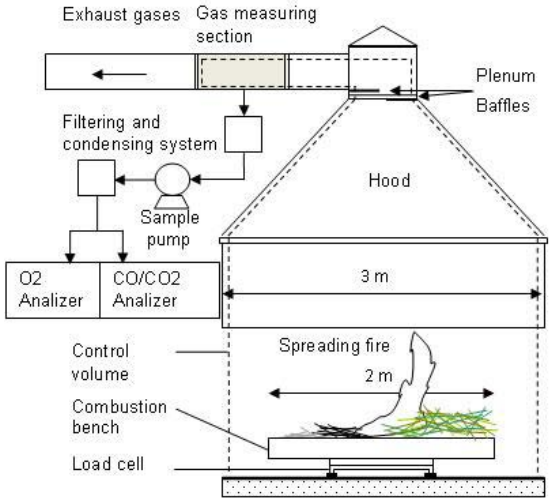


Figure 6: Large scale heat release calorimeter.

3.2.2 Experimental procedure and fuel bed properties

Three species of vegetation were considered: *Pinuspinaster* (PP) needles, *Avenafatua* (AF) straw (wild oats) and *Genistasalzmannii* (GS) spines (an endemic broom). Experimental fires were conducted under no slope and no wind for line-ignition fires (see Figure3). The fuels were scattered uniformly on the tray to obtain homogeneous beds. To ensure fast and linear ignition, a small amount of alcohol and a flame torch were used. The range of fuel bed properties for each vegetation species are provided in table 1 where w, δ, ρ, σ and β represent respectively the fuel load, the depth of the fuel bed, the density of the particles, the surface to volume ratio of the particles and the packing ratio of the fuel bed. $\Delta H_{c,net}$ is the net heat of combustion.

The fire fronts remained quasi linear during the whole spreads and exhibited a weak curvature on the flanks when reaching the end of the combustion bench.

Table 1: Range of fuel bed properties of the vegetation species.

Species	$\Delta H_{c,net}$ (kJ/kg)	σ (m^{-1})	ρ (kg/m^3)	p_M (%)	w (kg/m^2)	δ (cm)
Avena Fatua (AF)	17091	2394	287	4–7	0.6	7
Genista salzmannii (GS)	20645	3100	967	5	0.9	3.5
Pinus pinaster (PP)	20411	3057	511	3–5	0.6–1.2	3–7

The fireline intensity obtained by OC calorimetry denoted I_{OC}, corresponds directly to the measured HRR for fire front of1m width. For fuel bed of smaller width W, the fireline intensity is given by:

$$I_{OC} = \dot{q}/W \tag{7}$$

4 Results and discussion

4.1 Laboratory scale experiments

4.1.1 Geometrical characterization of the flame using stereovision

The height of the flame is estimated from visible stereoscopic images without the need of position markers for deriving scale and orthogonality present in the images. Stereovision gives metric measurements and it is then possible to distinguish easily points that are on the ground from points located above (point number 4 in Figure 5 for example). The precision of the measurement using stereovision has been tested with a fire front model (with maximum dimensions: height: 0.9; width: 1.9; depth: 0.9 m) which has been constructed from a metal frame and covered with red paper. Patterns have been added in order to have points of interest on the form that can be easily detected. The stereovision system was positioned at 4 meters of the model. The height of the estimated 3D pattern points and the maximum height of the model have been compared with the real one and the accuracy is about one millimetre.

4.1.2 Fireline intensity measured by oxygen consumption calorimetry

Figure 7 displays an example of a curve of fireline intensity for fire spreading across PP for a load of 1.2 kg/m^2. The corresponding mass loss over time is superimposed on the HRR curves.

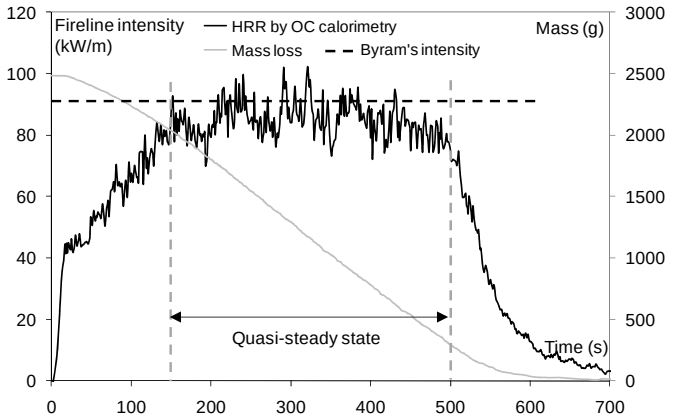


Figure 7: Fireline intensity and mass loss over time for fire spread across a fuel bed of PP for a load of 1.2 kg/m^2.

Figure 8 shows OC calorimetry intensities versus Byram's intensities for the whole set of experiments conducted. We can see that Byram's formulation I_B, overestimates I_{OC}, the fireline intensity measured by OC calorimetry. The ratio between both formulations given by the linear regression is $I_{OC}/I_B \approx 0.84$.

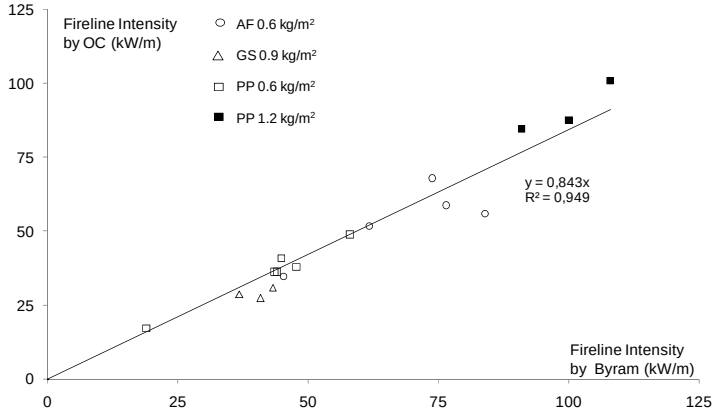


Figure 8: OC calorimetry Fireline intensity versus Byram's intensity.

4.1.3 Relationship between flame length and fireline intensity

The predictions of current correlation of flame length vs fireline intensity are plotted in Figure 9. The circle represents the fireline intensities obtained with the calorimeter. The flame length was obtained by stereovision.

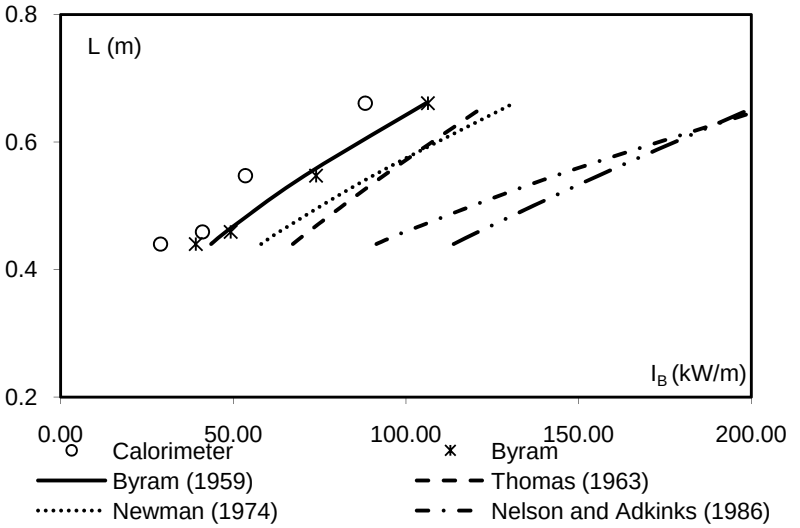


Figure 9: Flame length versus predicted fireline intensity, Byram's intensity and fireline intensity measured by OC.

We observe that all the correlation models overestimate the fireline intensity measured by OC. It should be noticed that the Byram fireline intensity computed from the measured flame length with eqn. (2) matches the one computed using eqn. (1) denoted by marker $*$ in figure 9, showing that the flame length is well measured. In order to take into account the good values (circle in figure 9) the new correlation between the intensity and the flame length is proposed as:

$$L = 0.0646\, I^{0.5281} \quad (8)$$

$$\Leftrightarrow I = 179\, L^{1.8936} \text{ or } I \approx 180\, L^{1.9} \quad (9)$$

5 Conclusion

Combined stereovision and oxygen consumption calorimetry allowed establishing a new correlation to assess fireline intensity. This first step is promising. In future, this correlation will be tested at field scale and the link between geometrical flame properties and radiant heat flux ahead of a fire front will be investigated.

Acknowledgements

This work was carried out in the scope of project PROTERINA-C supported by the EU under the Thematic 3 of the Operational Program Italia/France Maritime 2007-2013, contract: G25I08000120007.

References

[1] Byram G.M., Combustion of forest fuels (Chapter 3). *Forest fire: control and use*, ed. K.P. Davis, McGraw-Hill: New York, pp. 61–89, 1959.
[2] Fites J.A. & Henson C., Real-time evaluation of effects of fuel-treatments and other previous land management activities on fire behavior during wildfires, Report of the Joint fires science rapid response project, US Forest Service, pp. 1–13, 2004.
[3] McAthur A.G., Communication, 1962, Control burning in eucalypt forests, Forestry and Timber Bureau Leaflet, Australia, No 80.
[4] Hammil K.A. & Bradstock R.A., Remote sensing of fire severity in Blue Mountains: influence of vegetation type and inferring fire intensity. *International Journal of Wildland Fire*, **15**, pp.213–226, 2006.
[5] Palheiro P.M., Fernandes P. & Cruz M.G., A fire behaviour-based fire danger classification for maritime pine stands: comparison of two approaches. *Proc. of the Vth International Conference on Forest Fire Research*, ed. D.X. Viegas, CD-ROM, 2006.
[6] Albini F.A., Wildland fires, *American Scientist*, 72, pp.590–597, 1984.
[7] Newman M., Toward a common language for aerial delivery mechanics. *Fire Management Notes*, **35**, pp. 18–19, 1974.
[8] Thomas H., The size of flames from natural fire. In: proceedings of the 9th symposium on combustion. *Proc. of the combustion Institute*, Pittsburgh, pp. 844–859, 1963.

[9] Nelson R.M. Jr. & Adkins C.A., Flame characteristics of wind-driven surfaces fires. *Canadian Journal of Forest Research*, **16,** pp. 1293–1300, 1986.
[10] Weise D.R. & Biging G.S., Effects of wind velocity and slope on flame properties. *Canadian Journal of Forest Research*, **26**, pp. 1849–1858, 1996.
[11] Martinez-de Dios J. R., André J. C., Gonçalves J. C., Arrue B. Ch., Ollero A. and Viegas D. X. Laboratory Fire Spread Analysis Using Visual and Infrared Cameras *Inter. Journ. of Wildland Fire***15**, pp. 175-86, 2006.
[12] Adkins C.W., Fire image Analysis System version 5.0. *Proc of the 12th Conference on fire and meteo*, Jekill Island, Georgie. Poster, 1993.
[13] Hartley, R., Zissermann, A., *Multiple view geometry in computer vision*, Cambridge university press, 2nd Edition, 2003.
[14] Ayache, N, Sander, P.T., *Artificial Vision for Mobile Robots: Stereo Vision and Multisensory Perception*, MIT Press, 1991.
[15] Zhang, Z., Determining the Epipolar Geometry and its Uncertainty: A Review, *International Journal of Computer Vision*, **27(2),** pp.161-198, 1998.
[16] Hartley, R., Stereo from uncalibrated cameras, *Proc. of the Second Joint European – US Workshop on Applications of Invariance in Computer Vision*, pp. 237-256, 1994.
[17] Trucco, E., Verri, A., *Introductory Techniques for 3-D Computer Vision*, Prentice Hall, 1998.
[18] C. Harris, M. Stephens, A combined corner and edge detector, *Proc of the 4th Alvey Vision Conf*, pp. 189-192, 1998.
[19] Gonzales, R. C., Woods R. E., Eddins S. L., *Digital Image processing using Matlab*, Pearson prentice Hall, pp. 237-240, 2004.
[20] Horaud R., Monga O., *Vision par ordinateur, Outils fondamentaux*, Hermes Science Publication, 2nd Edition, 1995.
[21] He H.C. and Yung N.H.C., A corner detector based on global and local curvature properties, *Optical Engineering*; **47(5)**, pp. 057008-1 - 057008-12, 2008.
[22] http://www.ptgrey.com/products/bbxb3/index.asp
[23] Bumblebee XB3 Getting Started Manual http://www.ptgrey.com/support/downloads/documents/Bumblebee%20XB3%20Getting%20Started%20Manual.pdf
[24] L. Rossi, T. Molinier, M. Akhloufi, Y. Tison and A. Pieri, A 3D vision system for the measurement of the rate of spread and the height of fire fronts, *Meas. Sci. Technol.* **21(10)** 12p, doi:10.1088/0957-0233/21/10/105501, 2010.
[25] B.C. Ko, K.H. Cheong, J.Y. Nam, Fire detection based on vision sensor and support vector machines, *Fire Safety Journal* **44**, pp. 322-329, 2009.
[26] V. Babrauskas & R.D. Peacock, Heat release rate: The single most important variable in fire hazard. *Fire Safety Journal*, 18, pp. 255–292, 1992.
[27] C. Huggett, Estimation of the rate of heat release by means of oxygen consumption. *Journal of Fire and Flammability*, 12, pp. 61–65, 1980.

Photoelastic and numerical stress analysis of a 2D contact problem and 3D numerical solution for the case of a rigid body on a deformable one

A. Bilek & F. Djeddi
Laboratoire d'Energétique, Mécanique et Matériaux, Université Mouloud Mammeri, Tizi-Ouzou, Algérie

Abstract

A finite element analysis and a photoelastic stress analysis are conducted in order to determine the stress field developed in the pin on plan contact problem. Although this problem is relatively easy to study experimentally, the purpose here is to show the possibilities of the finite element method; after validation of the numerical procedure, problems with complicated geometry and boundary conditions can then be solved numerically. Isochromatic and isoclinic fringes, similar to the ones obtained experimentally by the photoelastic method, are obtained numerically over the whole model. The details of the finite element solution are fully given in the paper. Many studies have been achieved in order to separate the principal stresses and obtain their orientations (integration of the equilibrium equations …) in order to compare them with the simulated results. However, this requires a high precision of measurement. Here, a whole field comparison of the experimental and numerical photoelastic fringes and a local analysis using the principal stresses difference, allowed us to validate the numerical approach. Relatively good agreements were obtained. A numerical solution for a three dimensional contact problem is also developed for a rigid parallelepiped on a deformable cylinder. The mesh was refined in the neighborhood of the contact zone in order to achieve better approximation of stresses. The loading is given by the limit conditions that are simply the imposed displacement. The calculated photoelastic fringes are obtained for various sections inside the model. These simulated fringes can be compared to the experimental ones which can be obtained by slicing the model and analyzing it in a plan polariscope. The program developed allows us to calculate stresses on any

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110161

given section inside the model, particularly in the neighborhood of the contact zone.

Keywords: birefringent, isoclinic, isochromatic, fringe, photoelasticity, contact, stress field, stress freezing, slicing.

1 Introduction

Several studies have shown that failure of mechanical parts occur generally in the neighborhood of the contact zones [1–4]. Stress initiation is mainly controlled by the shear stress mechanisms, particularly for metallic materials, by displacement of the dislocations on the crystallographic plans of higher densities. It is therefore very important to determine the type and the amplitude of the imposed mechanical solicitations.

Theoretical and numerical studies of the contact stresses are in some cases very complex. Several methods have been used to analyze this type of problem. In this work two methods have been used: the photoelasticity method and the finite elements method in order to determine stresses developed on the model.

The photoelastic fringes obtained experimentally with a plan polarized light are used to determine the values of the principal stresses difference over the whole model. To obtain the individual values of the stresses, that is to separate the principal stresses, several studies have been conducted by integrating the equilibrium equations (Zenina et al. [5] and Plouzennec [6]). However, a high precision is required in the unwrapping of the isochromatic and the isoclinic fringes obtained on the analyzer to determine respectively the difference and the direction of the principal stresses.

As already done in previous papers (Bilek et al. [7, 8]), it is sufficient to make a comparison between experimental and simulated fringes. Another comparison, which is more accurate, is made between experimental and simulated values of the principal stresses difference along the vertical axis of symmetry.

2 Experimental analysis

The model, made of epoxy resin (PLM-4R) mixed with a hardener, is mounted on a loading frame (figure 1) equipped with two dynamometers. The model is loaded via a steel pin of rectangular cross section (12x12mm), the load is set to F=1300N. The loading frame with the model is then positioned on the polariscope for analysis.

Plane polarized light is used to observe the photoelastic fringes. The isochromatic and the isoclinic fringes obtained on the analyzer are used to determine the values of the principal stresses difference and the principal stresses directions, particularly in the neighborhood of the contact zone.

The model in the shape of a parallelepiped (67 x 58 x 10 mm) is cut in the birefringent material. Poisson's ratio and Young's modulus, which are necessary to implement the numerical solution, are measured with the help of electrical strain gages mounted respectively on a tensile specimen and a cantilever beam.

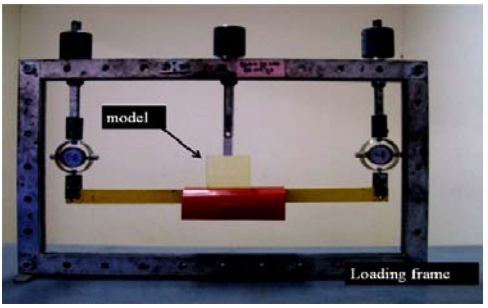


Figure 1: The model mounted on the loading frame.

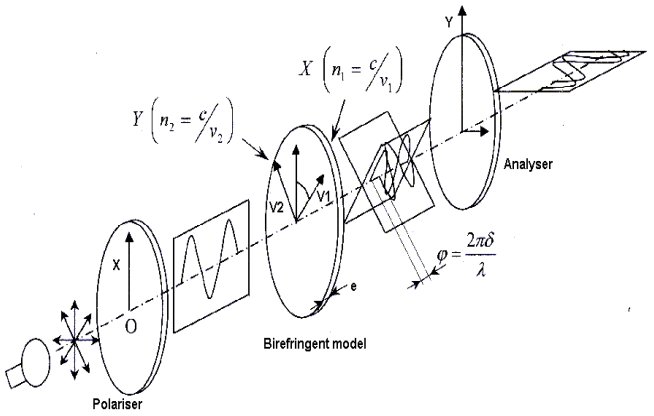


Figure 2: Light propagation through a photoelastic model.

Strains measured on the surface of the models allowed us to obtain easily these necessary values: μ= 0.37 and E=2435 MPa.

Figure 2 shows the well known photoelastic method based on the birefringent phenomenon; the refractive index n_1 and n_2 which depend on stresses in the two principal directions induce a retardation angle φ. The light intensity obtained on the analyzer after traveling through the polarizer, the model and the analyzer has two terms: $\sin^2 2\alpha$ and $\sin^2 \varphi/2$ which give respectively the isoclinic fringes and the isochromatic fringes (eq. (1)).

$$I = a^2 \sin^2 2\alpha \sin^2 \omega/2 \quad (1)$$

The isochromatic fringes allow us to obtain the values of the principal stresses difference on the model by using the well known relation (eq. (2)). This can only be done once the values of the fringe order N have been completely determined. The values of the fringe order N are determined either by the

compensation technique or, whenever possible, by starting from a non stressed region on the model were N=0. The fringe orders can then be easily deduced for the other fringes.

$$(\sigma_1 - \sigma_2). = N\,(\lambda/C)/e \tag{2}$$

The ratio f=λ/C called the fringe constant depends on the light wave used and the model material. Several solutions are available to obtain this value easily. Here, we subjected a beam (40mm x10mm cross section) to a constant bending moment (15000 N.mm) in a portion of its length (figure 3), the light wave length used is λ=546nm. We can see that the fringes are parallel to the horizontal axis of symmetry as one would expect; stresses at a same distance from the neutral axis are identical. We recall that isochromatics are loci of points having the same principal stresses difference. The isochromatic fringes are, therefore, parallel to the neutral axis of symmetry. Knowing the fringes orders and using the fact that the stress σ_2 is equal to zero (no load is applied in that direction), the value of the fringe constant can then be easily deduced by using (eq. (2)): f=11,65N/mm/fringe.

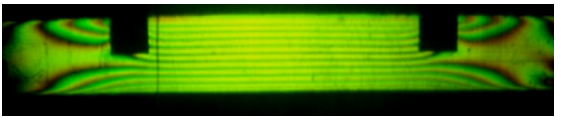


Figure 3: Model observed on a dark field polariscope.

Once the value of the fringe constant is determined, we can proceed to continue the analysis. The experimental isochromatic fringes (figure 4) on the analyzer of the polariscope are obtained with monochromatic light. Two quarter wave plates are added on the light path in order to eliminate optically the isoclinics that can hide the isochromatics which are necessary to determine the values of the principal stresses difference on the whole model, particularly in the neighborhood of the contact zone. These experimental isochromatic fringes will be compared to the simulated fringes obtained with the finite element analysis.

Another comparison is possible between the experimental isoclinics fringes and the simulated ones. The experimental isoclinics fringes (dark fringes on figure 5) are obtained for different positions of the polarizer and the analyzer. An image of the photoelastic fringes is recorded after each new position, the analyzer and the polarizer axes remain orthogonal to each other in order to always have a plan polarized light.

We recall that dark isoclinic fringes appear where the directions of the principal stresses coincide with the directions of the polarizer and the analyzer. The angle θ gives the position of the polarizer, the reference θ=0 taken as the vertical axis. The isoclinic fringes can then be easily used to obtain the principal stress trajectories called also isostatics. Several numerical methods have been developed to obtain the isostatics rapidly and efficiently.

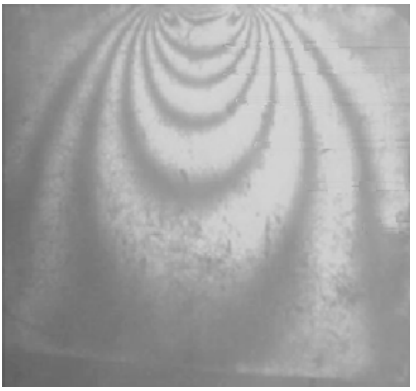


Figure 4: Isochromatic fringes obtained experimentally.

In this paper we are interested mainly in validating the finite element solution; it is therefore sufficient to compare the experimental and the calculated fringes. Another comparison, which is more accurate, is made between stresses obtained experimentally by analyzing the experimental isochromatic fringes and stresses obtained directly with the finites elements simulation, along the vertical axis of symmetry of the model.

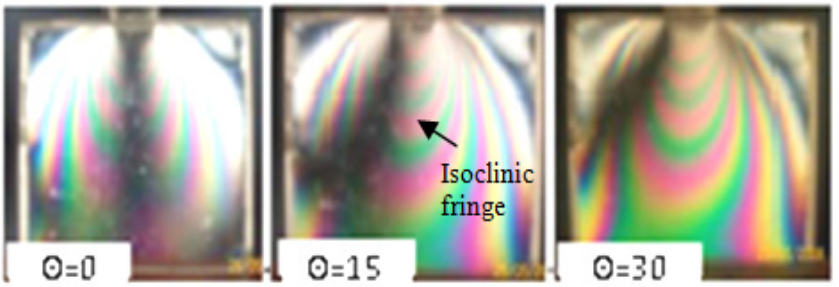


Figure 5: Isoclinic fringes at different angles.

3 Numerical analysis

In the finite element calculations, we considered that the material behaves everywhere as a purely elastic isotropic material. Young's modulus (E_1=210000MPa, E_2=2435 MPa) and Poisson's ratio (μ_1=0.3, μ_2=0.37) for the two bodies in contact were introduced in the program. The mesh was refined in the neighborhood of the contact zone (figure 6) in order to achieve better simulation of stresses.

To achieve a better simulation of the applied load, an imposed displacement is applied to the model at the contact surface between the pin and the plan. The equivalent applied load is calculated then as the sum of the elementary vertical

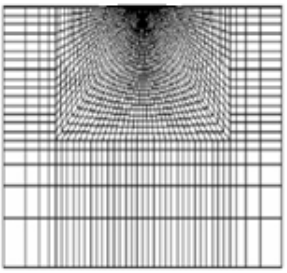


Figure 6: Finite element meshing.

load components at the nodes located at the lower surface of the model which is in contact with the loading frame.

Iterations on displacements at the contact nodes are stopped when the calculated corresponding load is equal to the value of the applied load within an acceptable error (0.1%) set in the program. The isochromatic fringes represented by $\sin^2\varphi/2$ (eq. 1) are calculated then easily over the whole model. The details of the calculation are shown hereafter.

3.1 Numerical calculation of the isochromatic fringes

The following relation (eq. 3) which can be obtained readily from Mohr's circle for stresses allows us to calculate the principal stresses difference at any point of a stressed model.

$$((\sigma_x - \sigma_y)^2 + 4\tau^2_{xy})^{0.5} = \sigma_1 - \sigma_2 = Nf/e \quad (3$$

The different values of the retardation angle φ (eq. 4) can be calculated at any point on the model using the following relation:

$$\varphi = 2\pi N = ((\sigma_x - \sigma_y)^2 + 4\tau^2_{xy})^{0.5}\ 2\pi\ e/f \quad (4)$$

The different values of $\sin^2\varphi/2$ which represents the isochromatic fringes (figure 7, left) can then be easily calculated over the whole model. A comparison can then be made with the isochromatic fringes obtained experimentally (figure 7, right). We can see relatively good agreement; however in the neighborhood of the contact zone we can see some discrepancies.

Figure 8 shows the variation of the principal stresses difference along the vertical axis. The value increases to approximately 6 MPa and then decreases away from the contact zone. For the experimental solution it is difficult to determine the stresses close to the contact zone.

The following hypotheses can be considered to explain these differences. For the finite elements procedure, the constitutive equations (mechanical behavior) taken for the material, the mesh sizes especially close to the contact zone and the boundary condition are not exactly the same as the real ones. Also, for the experimental procedure the loading is not perfectly symmetrical and the spatial resolution in the neighborhood of the contact zone is limited.

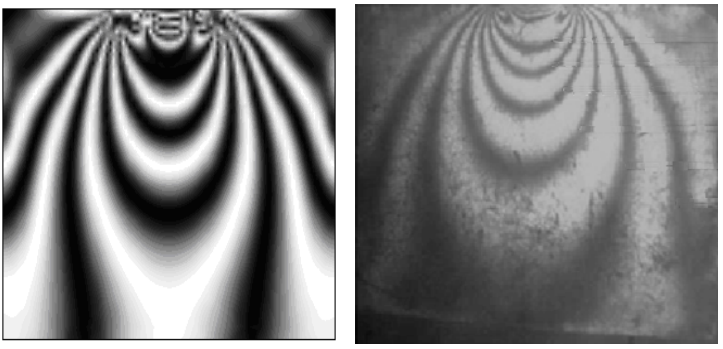


Figure 7: Simulated isochromatic fringes (left), experimental ones (right).

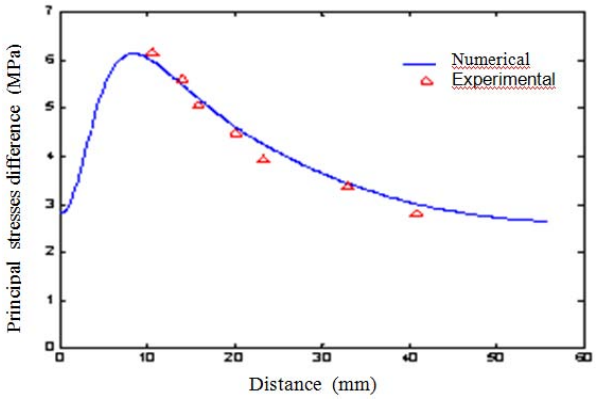


Figure 8: Principal stresses difference along the vertical axis.

The term $\sin^2 2\alpha$ represents the isoclinic fringes which are loci of points where the principal stresses directions are parallel to the polarizer and the analyzer. In the simulation program, the different values of the isoclinic parameter α can be calculated with the following relation (eq. 5) which can be obtained readily from Mohr's circle for stresses:

$$\alpha = \text{arct}\,(2\tau_{xy}/(\sigma_x-\sigma_y)) \tag{5}$$

The program calculates the different values of the parameter α. The image corresponding to the isoclinic fringes ($\sin^2 2\alpha$) can then be calculated and displayed (figure 9). The comparison is then possible with the experimental isoclinic fringes which are the dark fringes obtained experimentally (figure 5). Experimentally it is not possible, of course, to observe the isoclinics alone whereas for the finite element solution this is possible. Good agreement was obtained between the experimental and the simulated isoclinic fringes.

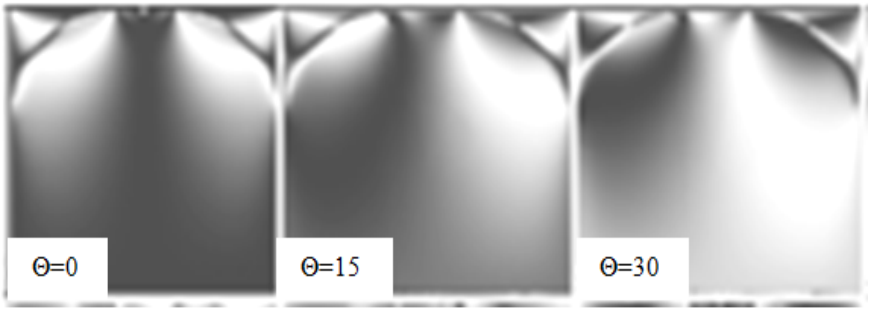


Figure 9: Simulated isoclinic fringes.

4 Numerical solution for the case of a rigid parallelepiped on a deformable cylinder

The birefringent cylindrical specimen (35mm diameter and 50mm length) is subjected to a compressive simulated load by a steel rigid parallelepiped (cross section 10x10 mm). The purpose here is to calculate stresses on any chosen plan perpendicular to the longitudinal axis of the cylinder (figure 10), particularly in the neighborhood of the contact zone. Simulated fringes similar to the photoelastic fringes obtained experimentally can then be calculated and displayed.

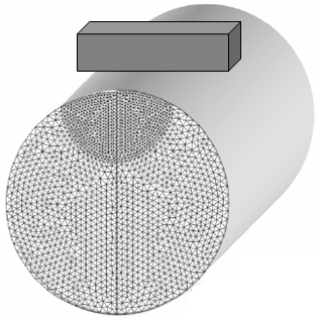


Figure 10: Experimental model.

The same as for the previous two dimensional case, displacements are imposed on the upper surface of the model at the nodes that will come into contact after the load is actually applied. The imposed displacement is calculated separately at each node; the imposed displacement decreases from a set value for the first point of contact and decreases as we move away from this point. The corresponding applied load is then calculated as the sum of the elementary vertical loads on the nodes at the lower surface of the model which is in contact with the loading frame.

4.1 Simulation for the case of mechanical slicing

If one uses the stress frizzing experimental technique to analyze the stress field by mechanically slicing the model and then analyzing stresses on a regular polariscope with circularly polarized light (quarter wave plate are added to the polariscope in the light path) to obtain the photoelastic fringes, the light intensity on the analyzer is given in by the following relation (eq. (6)):

$$I = a^2 \sin^2(\varphi/2) \tag{6}$$

In order to analyze stresses in three dimensional models we generally use the stress freezing technique which consists on locking stresses in the model. The model is then mechanically sliced with chosen plans in order to allow a two dimensional analysis on a regular polariscope.

Since for the stress freezing technique stresses should be locked inside the model at the stress freezing temperature, fringe constant and Young's modulus of the model material should be determined at this temperature. These characteristics, f=0.44 N/mm/fringe and E=15.9 MPa, are taken from reference (Bilek et al. [4]). These values should then be introduced in the finite elements program in order to implement the solution.

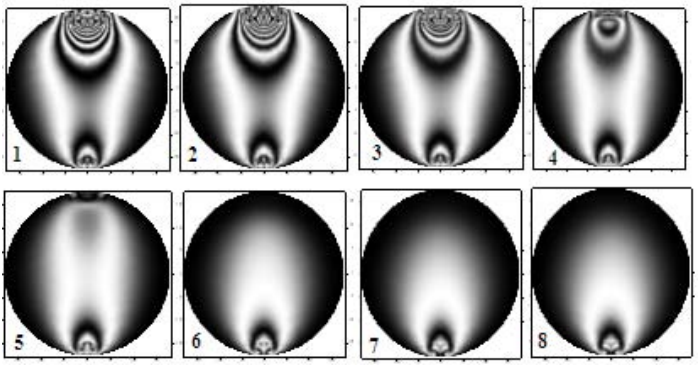


Figure 11: Simulated isochromatic fringes.

In the finite element solution, it is necessary to select the thickness of the slice to be isolated. The slice thickness should be small enough so that stresses remain relatively constant across the thickness. Here, we choose 10 mm which corresponds approximately to the generally used thickness for a two dimensional model. This process is repeated along the length of the cylinder in order to determine the variation of stresses in the whole volume. Figure 11 shows the simulated isochromatic fringes obtained with software package *castem.*

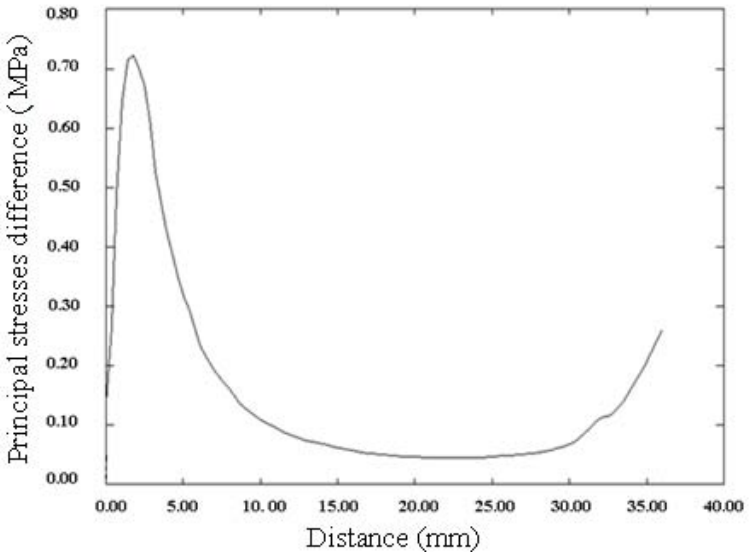


Figure 12: Principal stresses difference along the vertical axis of symmetry.

For a same slice (image 1, figure 11), we can see that stresses decrease as we move away from the contact zone. Also as we move along the cylinder (image 1 through 8) we see less isochromatic fringes which show clearly that stresses decrease. The stresses at the lower of the cylinder remain relatively constant; the load is uniformly distributed over the surface of contact with the loading frame.

The principal stresses difference along the vertical axis of symmetry increases to a maximum value of about 0.72 MPa and then decreases as we move away from the contact zone. The value of the principal stresses difference increases again, as we move close to the contact zone of the cylinder with the loading frame, to reach a value of 0.27 MPa. This graph can be obtained along the vertical axis for any plan along the length of the cylinder.

5 Conclusion

We have shown through the study of a two dimensional model that the simulation of stresses developed on a plan loaded with a pin gives relatively good agreements with the experimental ones. The isochromatic and the isoclinic fringes are comparable to the photoelastic fringes obtained on a regular polariscope. A solution for a three dimensional problem is developed. The isochromatic fringes are obtained for various sections along the cylinder. The principal stresses difference can be easily calculated in the volume of the cylinder. This allows us to locate the zones of stress concentration which is of great importance in the design of mechanical components. An experimental solution either by the stress freezing method or the optical slicing method can be used for comparison purposes.

References

[1] Mihailidis, A., Bakolas, V., & Drivakovs, N., Subsurface stress field of a dry line Contact. *Wear* V. 249, pp 546-556, 2001.
[2] Burguete, R. L. & Patterson, E. A., A photoelastic study of contact between a cylinder and a half-space. *Experimental Mechanics* V.7, No. 3, 1997.
[3] Kogut, L. & Etsion I., Elastic-Plastic contact analysis of a sphere and a rigid flat. *Journal of Applied Mechanics*, V.69, pp.657- 662, 2002.
[4] Bilek, A., Dupré, J. C., Ouibrahim, A. & Bremand, F., 3D Photoelasticity and numerical analysis of a cylinder/half-space contact problem, *Computer methods and experimental measurements for surface effects and contact mechanics* VII, pp 173 -182, 2000.
[5] Zenina, A., Dupré, J.C. & Lagarde, A., Separation of isochromatic and isoclinic patterns of a slice optically isolated in a 3-D photoelastic medium. Eur. J. Mech. A/Solids 18, pp. 633-640, 1999.
[6] Plouzennec, N., Développement de Processus d'analyse d'Images en Photoélasticimétrie par un feuillet plan obtenue par découpage mécanique et optique. *Thèse de l'Université de Poitiers*, 1996.
[7] Bilek, A., Dupré, J. C., Brémand, F. & Ouibrahim, A., Studies of contact problems by 3D photoelasticity, comparison with finite element analysis, *International conference on experimental mechanics*, Italy, 2004.
[8] Bilek, A., Ouibrahim, A., Brémand, F. & Dupré, J. C., Experimental and numerical analysis of a cylinder/cylinder contact problem. *ETDCM8, Experimental techniques and design in composite materials*, Italy 2007.

This page intentionally left blank

Section 2
Fluid flow

This page intentionally left blank

Large eddy simulation of a transition process in separated-reattached flows

Z. Yang
Aeronautical and Automotive Engineering Department, Loughborough University, UK

Abstract

Transition from laminar to turbulence in separated-reattached flow occurs frequently and plays a very important role in engineering. Hence, accurately predicting transition is crucial since the transition location has a significant impact on aerodynamics performance and a thorough understanding of the transition process can greatly help to control it, e.g. to delay the turbulent phase where laminar flow characteristics are desirable (low friction drag) or to accelerate it where high mixing of turbulent flow are of interest (in a combustor). However, it is very difficult to predict transition using conventional Reynolds-Averaged-Navier-Stokes (RANS) approach and the transition process is not fully understood. Nevertheless significant progress has been made with the simulation tools such as Large Eddy Simulation (LES) which has shown improved predictive capabilities over RANS and can predict transition process accurately. This paper presents briefly LES formalism and followed by its applications to predict/understand the transition process and unsteady behaviour of the free shear layer in separated-reattached flow.
Keywords: transition, separated-reattached flow, LES, RANS, shear layer, unsteady, turbulence.

1 Introduction

Separated flows are common and play an important role in many engineering applications from cooling of small electronic devices to airfoil and turbo-machinery design. If a separated flow reattaches downstream a separation bubble is formed and its characteristics are a crucial aspect of the engineering design process. Three types of separation bubble are possible depending on the state of

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110171

the boundary layer at separation and reattachment: laminar, transitional and turbulent. In a laminar separation bubble the flow at both separation and reattachment is laminar. For a transitional separation bubble it is laminar flow at separation while at reattachment the flow becomes turbulent, and a turbulent separation bubble is formed over an already turbulent boundary layer. Laminar boundary layer separation occurs in many engineering problems such as low Re number flow over aerofoils and turbo-machinery flow. Laminar separated flow has a tendency to become unstable even at relatively low Reynolds numbers and therefore is likely to undergo a transition to turbulence. The location where transition starts and the spatial extent within which transition takes place are of crucial interest in engineering design and performance prediction applications.

Laminar-to-turbulence transition has been under intensive research for many decades. Experimental studies have provided fundamental knowledge of parameters influencing transition, along with indications for related physical mechanisms. However, such data can only provide limited temporal and spatial resolution of flow parameters and hence a thorough description of the transition process is lacking. Theoretical studies on the other hand, suffer from the limitation imposed by nonlinearity of the transition process at later stages.

Conventional RANS approach, based on solving the time- or ensemble-averaged governing equations and hence the effect of all the scales of instantaneous turbulent motion is modelled, is most commonly applied to the solution of engineering turbulent flow problems but is not adequate to predict transition since it only predicts the time- or ensemble-averaged structure and behaviour of transitional bubbles. Other approaches such as the semi-empirical e^n method and correlations are also of limited accuracy and non universal [1].

The alternative approach is LES which was proposed as early as 1963 by Smagorinsky [2]. LES does not adopt the conventional time- or ensemble-averaging RANS approach with additional modelled transport equations being solved to obtain the so called Reynolds stresses resulting from the averaging process. In LES the large scale motions (large eddies) of turbulent flow are computed directly and only small scale (sub-grid scale) motions are modelled.LES can be more accurate than the RANS approach since the larger eddies contain most of the turbulent energy and are responsible for most of the turbulent mixing, and LES captures these eddies in full detail directly whereas they are modelled in the RANS approach. Furthermore the small scales tend to be more isotropic and homogeneous than the large ones, and thus modelling the sub-grid scale motions should be easier than modelling all scales within a single model as in the RANS approach. However, LES has received increased attention in the engineering community only since 1990's although it was proposed nearly half a century ago, mainly due to the lack of sufficient computational power since LES requires 3D time-dependent calculations with small time-steps and reasonably fine meshes.

The current paper presents briefly LES formalism first followed by its applications to study transitional separated-reattached flows, focusing on the current understanding of physics of the transition process, and concludes with

possible future trends in several important areas in LES and transitional bubble study.

2 Mathematical formulation

2.1 LES governing equations

The governing equations for any fluid flow, called the Navier-Stokes equations, are derived according to the fundamental conservation laws for mass, momentum and energy. In LES only large eddies (large scale motions) are computed directly and hence a low-pass spatial filter is applied to the instantaneous conservation equations to formulate the 3D unsteady governing LES equations. When the finite volume method is employed to solve the LES equations numerically the equations are integrated over control volumes, equivalent to convolution with a top-hat filter, therefore there is no need to apply a filter to the instantaneous equation explicitly and in this case it is called implicit filtering.

The filtered equation expressing conservation of mass and momentum in a Newtonian incompressible flow can be written in conservative form as:

$$\partial_i \overline{u_i} = 0 \tag{1}$$

$$\partial_t(\overline{u}_i) + \partial_j(\rho\overline{u_i}\overline{u_j}) = -\partial_i \overline{p} + 2\partial_j(\mu\overline{S_{ij}}) - \partial_j(\tau_{ij}) \tag{2}$$

where the bar over the variables denotes the filtered, or resolved scale quantity and:

$$\overline{S_{ij}} = \frac{1}{2}(\partial_i \overline{u_j} + \partial_j \overline{u_i}) \tag{3}$$

$$\tau_{ij} = \rho(\overline{u_i u}_j - \overline{u_i}\overline{u_j}) \tag{4}$$

$\overline{S_{ij}}$ is the resolved scale strain rate tensor and τ_{ij} is the unknown sub-grid scale or residual stress tensor, representing the effects of the sub-grid scale motions on the resolved fields of the LES, which must be modelled or approximated using a so called sub-grid scale model.

2.2 Sub-grid scale modelling

Many different kinds of sub-grid scale models have been developed [3–5]and most of them make an eddy-viscosity assumption (Boussinesq's hypothesis) to model the sub-grid scale stress tensor as follows:

$$\tau_{ij} = -2\mu_t \overline{S_{ij}} + \frac{1}{3}\delta_{ij}\tau_{ll} \tag{5}$$

μ_t is called sub-grid scale eddy viscosity and eqn. (2) then becomes:

$$\partial_t(\rho\overline{u}_i) + \partial_j(\rho\overline{u_i u_j}) = -\partial_i \overline{P} + 2\partial_j[(\mu + \mu_t)\overline{S_{ij}}] \tag{6}$$

It should be noted that a modified pressure, $\overline{P} = \overline{p} + \frac{1}{3}\tau_{ll}$, has been introduced and hence when the above equation is solved the pressure obtained is not just the static pressure only. The question now is how to determine the sub-grid scale eddy viscosity and the most basic model is the one originally proposed by Smagorinsky [2]:

$$\mu_t = \rho(C_S\overline{\Delta})^2 S \qquad S = (2\overline{S_{ij}}\,\overline{S_{ij}})^{\frac{1}{2}} \qquad \Delta = (\Delta x \Delta y \Delta z)^{\frac{1}{3}} \tag{7}$$

C_S is the so called Smagorinsky constant and typical value used for it is 0.1.

Despite increasing interest in developing more advanced sub-grid scale models this very simple model has been used widely and proved surprisingly successful although it has clear shortcomings such as that it is too dissipative (not good for transition simulation) and the Smagorinsky constant needs to be adjusted for different flows. An improvement on this simple SGS model was suggested by Germano *et al.* [6] – a dynamic sub-grid scale model, which allows the model constants C_S to be determined locally in space and in time during the simulation.

2.3 Numerical methods

The finite volume method is the most popular numerical method used in fluid flow simulation and most of LES studies have been carried out using this method. A brief discussion on many important numerical issues will be presented in this section.

2.3.1 Filtering

When the finite volume method is used there is no need to explicitly filter the instantaneous Navier-Stokes equations since the governing equations can be regarded as implicitly filtered as mentioned in section 2.1.The velocity components at the corresponding grid points are interpreted as the volume average. Any small scale (smaller than the mesh or control volume) motions are averaged out and have to be accounted for by a sub-grid scale model. However, note that it is impossible in this case to discuss the convergence properties (grid independent solution) of the LES equations because with every mesh refinement, more small scale eddies are resolved and strict convergence is only achieved in the limit of the so called Direct Numerical Simulation (DNS).

2.3.2 Spatial and temporal discretization

The most popular spatial discretization scheme used in LES is the second-order central differencing duo to its non-dissipative and conservative properties (not only mass and momentum but also kinetic energy conserving), which are

essential for LES. This is the reason why usually first- and second-order upwind schemes or any upwind-biased schemes are not used in LES since they produce too much numerical dissipation. While higher-order numerical schemes, generally speaking, are desirable and can be applied fairly easily in simple geometries, their use in complex configurations is rather difficult. In addition, it is difficult, at least for incompressible flows, to construct high-order energy conserving schemes. Hence it is likely that with increasing applications of LES to flows of engineering interest in complex geometries the second-order central differencing scheme is still the most popular choice.

As for the temporal discretization (time advancement), implicit schemes allow larger time steps to be used. However, they are more expensive because at each time step non-linear equations have to be solved. Furthermore, large time steps are unlikely to be used in LES in order to resolve certain time scales for accurate simulations of turbulence. Hence, explicit schemes seem to be more suitable for LES than implicit schemes and most researchers in LES use explicit schemes such as the second-order Adams–Bashforth scheme. Since the time steps are usually small in LES so that it is not essential to use higher-order schemes either.

2.3.3 Inflow boundary conditions

Most boundary conditions used in LES are fairly standard and similar to those used in the RANS approach but specifying inflow boundary conditions accurately for LES proves to be very difficult. This is because in LES of turbulent flow at inflow boundary, unlike the RANS computations where only time-averaged information is required that can be usually specified according to experimental data, three components of instantaneous velocity need to be specified at each time step, which are almost impossible to be obtained from experimental data. Hence normally boundary conditions in LES at inflow boundary have to be generated numerically which usually lack physical flow properties. For example, the simplest way is to specify the mean flow velocity profile (usually obtained experimentally) plus some random perturbations. However, random disturbances are nothing like real turbulence since they have no correlations; neither in space nor in time. Therefore, they decay rapidly and it takes usually a long distance downstream from the inflow boundary for a desired realistic turbulence to develop, and in some cases the use of random noise at the inlet does not develop turbulence at all. On the other hand one can use the so-called precursor simulation technique, which is basically to perform another simulation and store the data as the input for the required simulation. This can generate the most realistic turbulence information at inflow boundary but it is far too expensive. Many efforts have been made to develop a method which can generate numerically three instantaneous inflow velocity components in such a way that they have all the desired turbulence properties. However, so far there are methods developed which can generate inflow turbulence with certain properties but no methods available yet to generate inflow turbulence with all the desired characteristics such as intensity, shear stresses, length scales and power spectrum [7].

2.3.4 Near wall modelling

LES has been applied more and more, in recent years, to study practical engineering turbulent flows in complex geometries at higher Reynolds number. However, for high Reynolds number wall-bounded flows the cost of LES that resolves all the important eddies in the near wall region (the wall-layer structures) is far too high. Therefore methods to bypass the near wall region are required to perform high-Reynolds-number LES at a reasonable cost. Several methods have been developed to "model" the near wall region flow rather than resolve it directly using very fine mesh and more details can be found in a review paper by Piomelli and Balaras [8].

3 Applications of LES to study transitional bubble

This section presents some LES studies of transition in separated-reattached flows and tries to summarise the current understanding of the transition process, focusing on several important flow phenomena associated with the transition process.

3.1 Transition mechanism

Many studies have revealed that in the absence of any finite magnitude environmental disturbances, transition in the separated shear layer of a separation bubble is dominantly initiated through the inviscid Kelvin-Helmholtz (KH) instability mechanism. This mode of instability closely resembles that of the planar free-shear layer in mixing layers and jets [9]. The LES study of Yang and Voke [10]revealed a primary 2D instability of a separated shear layer (induced by a smooth leading edge) via the KH mechanism. A similar mechanism was also observed by Abdalla and Yang [11] in their LES studies of a separation bubble over a sharp leading edge. The LES study by Roberts and Yaras [12] demonstrated that transition of a separated shear layer through the KH instability does not eliminate the existence of a so called Tollmien-Schlichting (TS) instability(a viscous instability typically associated with attached flow boundary layer transition) in the inner part of the flow where the roll up of shear layer into vortical structures occurred at the dominant TS frequency. They emphasized the possibility of an interaction between the TS and KH instability modes. Several other studies have shown that KH instability plays a dominant role in the transition process of separation bubbles. A number of experimental studies have also suggested that the TS instability mechanism plays a significant role in a transitional separation bubble [13–15].

The next stage of the transition process after the dominant primary KH instability is less well understood. In planar free shear layers, the primary spanwise vortices generated by the KH instability are known to undergo an instability leading to the vortex pairing phenomenon [9, 16, 17]. This pairing of vortices is regarded as the governing secondary mechanism associated with growth of planar free shear layers. A similar vortex pairing phenomenon has also been reported in separated shear layer studies [18] but Abdalla and Yang [11]

demonstrated that transformation of 2D KH rolls into 3D structures occurs via a slightly different secondary instability known as helical instability associated with helical pairing of vortices. Fig. 1 shows the evolution of 2D KH rolls into 3D Lambda-shaped vortices in a transitional bubble formed on a flat plate with a sharp leading edge [19].

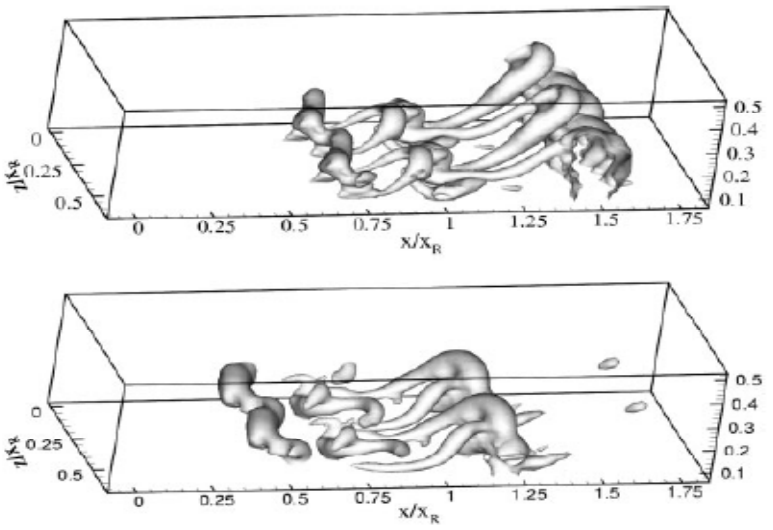


Figure 1: Low-pressure iso-surfaces showing the evolution of 2D KH rolls into 3D Lambda-shaped vortices.

In summary, the transition process in separated-reattached flow generally consists of the following stages:
1) a primary 2D instability (mostly KH instability),
2) a secondary instability leading to significant 3D motions and,
3) a breakdown stage where fully turbulent flow emerges.

Another key parameter influencing the transition process of a separated boundary layer and its following reattachment is free-stream turbulence (FST). Experimental studies have shown that FST increases the shear-layer entrainment rates, decreases the mean reattachment length and results in an earlier transition to turbulence in separated boundary layer. Yang and Abdalla [19, 28] performed LES studies of separated boundary layer transition under 2% FST. They reported a 14% reduction of the mean bubble length and an earlier breakdown of the free shear layer compared with the zero FST case. At 2% FST, 2D KH rolls were not as apparent as in the case with zero FST, but still coherent 2D structures in the early part of the bubble were observable. Lambda-shaped vortices could hardly be identified and streamwise structures were enlarged in the spanwise direction and shortened in the streamwise direction compared with the no FST case. It was concluded that in the presence of 2% FST the primary instability of the free shear layer was still the same as in the zero FST case (KH instability

mechanism) but secondary instability was different and needed to be further investigated.

3.2 Shedding phenomenon

A key feature of separated-reattached flows is vortex shedding associated with different unsteady flow phenomena of the free shear layer at different frequencies. In a steady laminar separation bubble one can define a reattachment point or line where the skin friction is zero. In transitional and turbulent separation bubbles however, the instantaneous flow field is highly unsteady around the 'mean' reattachment point and the notion of a reattachment 'point' is misleading as it continuously varies with the time. In this case, it is possible that several small bubbles or vortices are formed and then shed afterwards, leading to a vortex shedding phenomenon.

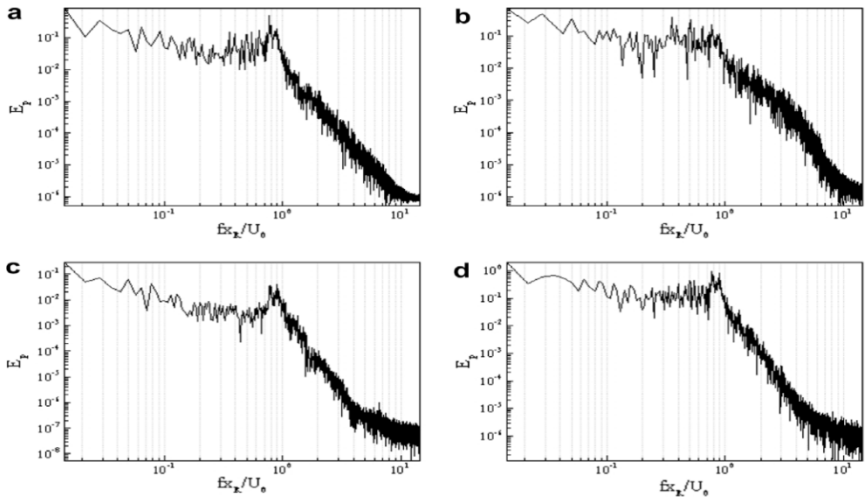


Figure 2: Pressure spectra at x/l=0.75 and four vertical locations: y/l=0.01 (a), y/l=0.05 (b), y/l=0.13 (c), y/l=0.2 (d).

Fig. 2 shows pressure spectra at several different locations in a separated boundary layer transition [19] and a peak frequency band at about 0.8–0.9 U_0/l can be clearly seen (U_0 is the free stream velocity and l is the mean bubble length). This peak frequency band was also observed in several experimental studies of separated-reattached flow over a plate with a sharp leading edge at high Reynolds number [20–22]. This peak frequency band was stated to be the characteristic frequency of the large vortices shedding from the free shear layer of the bubble. Furthermore, a low frequency peak (0.12 U_0/l) was also reported in those experimental studies near the separation line. This low frequency peak was not clearly understood and was suggested as related to the large scale shrinkage and enlargement of the bubble. A low frequency peak (0.125-0.2 U_0/l) was also observed in the LES study by Yang and Voke [10] and they suggested

that this was associated with large shrinkage of the bubble caused by a big vortex shedding at a lower frequency as shown in Fig. 3. However, this low frequency peak was not observed in some other separated boundary layer transition studies [11, 19]. Abdalla and Yang [11], in their LES of a transitional bubble over a flat plate with a sharp leading edge, showed a characteristic frequency in the range 0.7-0.875 U_0/l along with some less dominant modes between 0.3-0.6 U_0/l. They inferred that this slightly lower frequency content may be related to pairing of vortices as a similar range of frequency had been reported for the pairing phenomenon behind a backward facing step but no low frequency peak as mentioned above was observed. Yang and Abdalla [19] studied the same problem with 2% free-stream turbulence and reported a peak frequency band at about 0.8–0.9 U_0/l, in close agreement with the characteristic frequencies already measured in previous studies but again no low frequency peak was observed. Those results indicate that this low frequency mode in separated–reattached flows may only appear in the case of turbulent separation as suggested earlier by Cherry *et al*. [21] but further study is needed to clarify this.

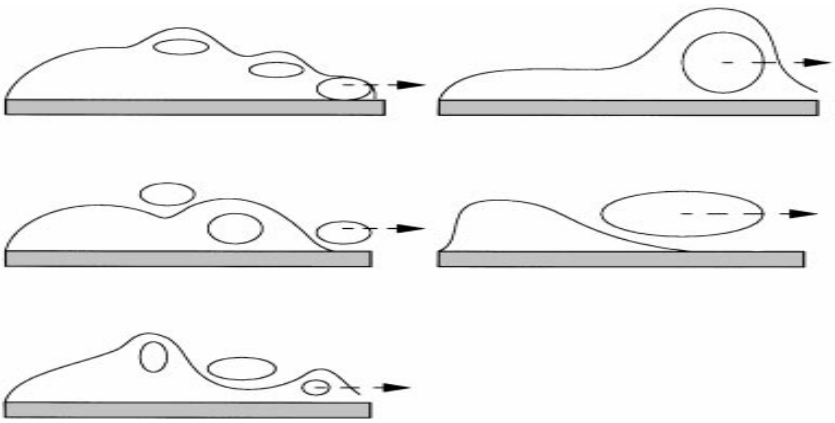


Figure 3: Two different topological structures of a separation bubble associated with the onset of vortex shedding: left, normal shedding; right, low frequency shedding.

3.3 Coherent structures

Large-scale structures (large-scale organised motions), usually called coherent structures (CS), have been revealed in many experimental studies to dominate the entrainment and mixing phenomena in free shear flows [23]. It is important to understand the physics of coherent structures so that a better insight into turbulence phenomena can be obtained (such as entrainment and mixing, heat and mass transfer, drag and aerodynamic noise generation etc.). However, despite considerable usage in the literature it seems that an approved definition for coherent structures does not yet exist. Cantwell [24] describes coherent structures as spatially coherent and temporally evolving vortical structures.

Kelvin-Helmholtz (KH) rolls, *Streaks*, *Hairpin vortices (also called Lambda-shaped vortices) and Ribs* are some of the common large-scale flow structures which are referred to as coherent structures in the literature and as shown previously in Fig. 1, the evolution of KH rolls into Lambda-shaped vortices in a separated boundary layer transition. Streaky structures are characterized with narrow regions of low velocity fluid stretched in the streamwise direction [25, 26]. Streamwise vortices are vortical structures which are predominantly oriented in the streamwise direction, although they may be bent and make an angle with the streamwise direction. Spanwise vortices are referred to as those primarily oriented in the spanwise direction such as KH rolls. Hairpin vortices (Lambda-shaped vortices) are those with two legs of quasi-streamwise vortex pairs with opposite signs and a tip of spanwise vorticity.

Coherent structures usually depend on flow geometry, flow condition, and location with respect to solid surfaces. Large-scale spanwise vortices in plane mixing layers, Lambda-shaped vortices and low-speed streaks in transitional and turbulent boundary layers and counter-rotating vortices in wakes are the dominant structures controlling the flow dynamics. Vortical structures in separated shear layers grow, merge and shed periodically from the reattachment region. KH rolls and Lambda-shaped vortices have been observed in separated layer transition and the transition process is better understood by studying the evolution of KH rolls into Lambda-shaped vortices [10, 11, 27, 28]. It is believed that reorientation of vorticity in the streamwise direction is a key mechanism for the reattachment process as it provides enhanced momentum exchange in the wall-normal direction. Abdalla *et al.* [29], in a LES study of transitional separated-reattached flow over a surface mounted obstacle and a forward-facing step, demonstrated that the coherent structures such as the Lambda-shaped and rib-like vortices, which are often associated with a flat plate boundary layer and also found in the separated-reattached flow, are not common in the separated-reattached flow over obstacles and forward-facing steps.

4 Conclusions

The present paper has presented briefly LES formalism and reviewed some of its applications to study transition process in separated-reattached flow, focusing on the current understanding of physics of the transition process. Several important issues associated with LES have been discussed. Although significant progress has been made towards a better understanding of the transition process in separated-reattached flows our current understanding is far from complete, and there are still many areas where further investigation is needed. According to the author the following issues/areas are particularly important and future research should be focused on:

- numerical methods to generate realistic turbulence at inflow for LES.
- advanced sub-grid scale models for LES of high Reynolds number engineering flow in complex geometry.

- secondary instability and later stage breakdown to turbulence in a transitional bubble.
- effect of high free-stream turbulence on transition in separated-reattached flow.
- transition control, crucial to practical engineering applications.

References

[1] Langtry, R.B. & Menter, F.R., Transition modelling for general CFD applications in aeronautics. *AIAA 2005-522*, Reno, Nevada, 2005.

[2] Smagorinsky, J., General circulation experiments with the primitive equations: I – the basic experiment. *Monthly Weather Review*, **91**, pp. 99-164, 1963.

[3] Lesieur, M. & Metais, O., New trends in large eddy simulations of turbulence. *Annual Review of Fluid Mechanic*,**28**, pp. 45-82, 1996.

[4] Sagaut, P., *Large Eddy Simulation for Incompressible Flows, an Introduction, Springer*, 2nd edition, 2003.

[5] Kajishima, T. & Nomachi, T., One-equation sub-grid scale model using dynamic procedure for the energy production. *Transaction of ASME*, **73**, pp. 368-373, 2006.

[6] Germano, P., Piomelli, U., Moin, P. & Cabot, W.H., A dynamic sub-grid scale eddy viscosity model. *Physics of Fluids*,**3(7)**, pp. 1760-1765, 1991.

[7] Veloudis, I., Yang, Z., McGuirk, J.J., Page, G.J. & Spencer, A., Novel implementation and assessment of a digital filter based approach for the generation of large-eddy simulation inlet conditions. *Journal of Flow, Turbulence and Combustion*, **79**, pp. 1-24, 2007.

[8] Piomelli, U. & Balaras, E., Wall-layer models for large-eddy simulation. *Annual Review of Fluid Mechanics*,**34**, pp. 349–374, 2002.

[9] Ho, C.M. & Huerre, P., Perturbed free shear layers, *Annual Review of Fluid Mechanics*, **16**, pp. 365-424, 1984.

[10] Yang, Z. & Voke, P.R., Large-eddy simulation of boundary layer separation and transition at a change of surface curvature. *J. Fluid Mech*., **439**, pp. 305–333, 2001.

[11] Abdalla, I.E. & Yang, Z., Numerical study of the instability mechanism in transitional separating-reattaching flow. *International Journal of Heat and Fluid Flow*, **25**, pp. 593-605, 2004.

[12] Roberts, S.K. & Yaras, M.I., Large-eddy simulation of transition in a separation bubble. *ASME J. Fluids Eng*., **128**, pp. 232-238, 2006.

[13] Lang, M., Rist, U. & Wagner, S., Investigations on controlled transition development in a laminar separation bubble by means of LDA and PIV. *Exp. Fluids*, **36**, pp. 43–52, 2004.

[14] Roberts, S.K. & Yaras, M.I., Effects of periodic unsteadiness, free-stream turbulence and flow Reynolds number on separation-bubble transition.*ASME-GT2003-38626*, 2003.

[15] R.J. Volino, R.J. & Bohl, D.G., Separated flow transition mechanism and prediction with high and low free stream turbulence under low pressure turbine conditions.*ASME-GT2004-53360*, 2004.
[16] Huang, L.S. & Ho, C.M., Small-scale transition in a plane mixing layer. *J. Fluid Mech.*, **210**, pp. 475–500, 1990.
[17] Winant, C.D. & Browand, F.K., Vortex pairing: the mechanism of turbulent mixing-layer growth at moderate Reynolds number. *J. Fluid Mech.*, **63**, pp. 237-255, 1974.
[18] Malkiel, E & Mayle, R.E., Transition in a separation bubble. *ASME J. Turbomachinery*, **118**, pp. 752–759, 1996.
[19] Yang, Z & Abdalla, I.E., Effects of free-stream turbulence on a transitional separated-reattached flow over a flat plate with a sharp leading edge. *Int. J. Heat Fluid Flow*, **30**,pp. 1026-1035, 2009.
[20] Kiya, M. and Sasaki, K., Structure of a turbulent separation bubble. *J. Fluid Mech.*, **137**, pp. 83-113, 1983.
[21] Cherry, N.J., Hillier, R. & Latour, M.E.M.P., Unsteady measurements in a separating and reattaching flow. *J. Fluid Mech.*,**144**, pp. 13–46, 1984.
[22] Kiya, M. & Sasaki, K., Structure of large-scale vortices and unsteady reverse flow in the reattaching zone of a turbulent separation bubble. *J. Fluid Mech.*, **154**, pp. 463-491, 1985.
[23] Hussain, A.K.M.F., Coherent structures and turbulence. *J Fluid Mech.*, **173**, pp. 303-356, 1986.
[24] Cantwell, B.J., Organised motion in turbulent flow. *Annual Review of Fluid Mechanics*, **13**, pp. 457–515, 1981.
[25] Kim, H.T., Kline, S.J. & Reynolds, W.C., The production of turbulence near a smooth wall in a turbulent boundary layer. *J. Fluid Mech.*, **50**, pp. 133–160, 1971.
[26] Smith, C.R. & Metzler, S.P., The characteristics of low-speed streaks in the near-wall region of a turbulent boundary layer. *J. Fluid Mech.*, **129**, pp. 27-54, 1983.
[27] Yang, Z., Large-scale structures at various stages of separated boundary layer transition. *Int. J. Numer. Meth. Fluid*, **40**, pp. 723-733, 2002.
[28] Yang, Z. & Abdalla, I.E., Effects of free-stream turbulence on large-scale coherent structures of separated boundary layer transition. *Int. J. Numer. Meth. Fluid*, **49**, pp. 331-348, 2005.
[29] Abdalla, I.E., Yang, Z. & Cook, M., Computational analysis and flow structure of a transitional separated-reattached flow over a surface mounted obstacle and a forward-facing step. *International Journal of Computational Fluid Dynamics*, **23**, pp. 25-57, 2009.

A numerical investigation of deployable drag surfaces used for recovery systems

S. Abdolahi & M. Ebrahimi
Aerospace Research Institute, Iran

Abstract

In this research, an investigation into the aerodynamic characteristics of a body with deployable drag surfaces for recovery system has been carried out using computational fluid dynamics. Two models of the body with retracted position of drag surfaces and deployed position of drag surfaces has been considered for studying the influence of drag surfaces on the flow structure and aerodynamic forces. For this purpose force measurement and flow visualization for each case has been carried out in Mach numbers 0.4 and 1.5. Validation of the results has been done through comparing aerodynamic coefficients with results of a semi-experimental method. A general study of the main aerodynamic coefficients shows that at all angles of attack, the coefficient of lift decreases and the coefficient of drag increases. Visualization of the flow structure shows a region of separated flow upstream and a dead flow region with large vortices downstream of the drag surfaces.
Keywords: numerical simulation, aerodynamic characteristics, recovery system, drags coefficient, pressure distribution, shock wave, vortex.

1 Introduction

Several methods have been employed for recovery of flying objects with various degrees of success. But the most prominent method, especially for heavier bodies is parachute recovery [1], in which a number of parachutes are deployed in a pre-defined sequence to reduce the body's rate of descent. Physics of supersonic flow around parachutes has its own complexities such as area oscillation and shock wave formation [2]. The controlled deceleration of a flying body is a vital part of many aerospace missions that involves significant technological challenges and creative engineering solutions [3]. For many flying objects,

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110181

parachutes cannot be relied on as the only means of deceleration and recovery because of their weight and altitude. Numerous studies in the past fifty years demonstrate the benefit of developing a new decelerator technology capable of operating at higher Mach numbers and higher dynamic pressures than existing decelerators allow [4].

An effective recovery system must meet three principal criteria: It must be able to reduce body's rate of descent down to the acceptable level, it must not impose unacceptable impact and shock loads to the body when its components are deployed, and finally it must have minimum sensitivity to crosswinds to prevent excessive drift.

For a flying object, which weighs approximately 100 kg, and is expected to reach a maximum altitude of 100 km, trajectory simulation shows that deploying parachutes shortly after re-entry could cause the flying object to drift beyond limits of the test range, due to strong crosswinds in higher layers of atmosphere. Trajectory simulations also show that deploying parachutes at lower levels, where the flying object reaches supersonic speeds, would expose it to impact loads beyond its established structural limit of 18g [5].

In this case, use of rigid aerodynamic decelerators such as drag plates and disk gap bands is a solution for reducing body velocity to low-subsonic before deploying parachutes [6]. A drag plate, is a circular plate with an area less than or equal to the area of the base, which is normally connected to the body by a number of belts and deploys after the body starts to descend from maximum altitude [7].

Typical drag plates have a number of inherent shortcomings that have prevented them so far from acquiring prominence among different techniques of aerodynamic deceleration. Although a drag plate can have relatively large drag coefficients in certain flight regimes, its drag generation capability is limited because its effective area cannot be larger than the body's base area. The other problem of the drag plate is that its connection to the flying object is flexible, and the number of connecting belts in this flexible connection is a concern. If a small number of belts are used, they would allow the plate, which is subject to the body's wake, to oscillate, and even become unstable. On the other hand, larger numbers of belts would add to the system's weight and complexity. In some applications, design of hinges which connect the belts to the plate has also been a challenge for the designers [7].

The mentioned shortcomings of drag plates and enhancing effectiveness and performance of the recovery system led to development of alternative method to use of rigid aerodynamic decelerator. The main concept for this method is taken from grid fins [8] and other local protuberances [9] that can be used as either aerodynamic stabilizer or a control surface. The rigid aerodynamic deceleration system presented herein is deployable drag surface folded on the body and released on flight computer's command. The presented design is the first application of this method to reduce body's speed to subsonic range before deploying parachute by increasing drag.

In this paper to study performance of using deployable drag surfaces and the influences of them on body aerodynamic characteristics, three dimensional

supersonic and subsonic turbulent flows around the body were numerically investigated. The computation of the flow field around such complex configuration, as drag surfaces adjacent fins or wings, is of considerable interest for an accurate calculation and prediction of the shock structures because there is an interaction between drag surfaces shock waves with fins or wings. Accurate determination of forces is required for the calculation of trajectory and structural analysis of body in design and operation.

In these simulations the main focus has been on aerodynamic characteristics. To demonstrate the efficiency and accuracy of the present methodology the aerodynamic characteristics of this configuration including drag and lift forces obtained by commercial CFD code for subsonic and supersonic free stream Mach numbers were compared with the semi-experimental measurements at various angles of attack.

2 Computational methodologies

This study is undertaken using Fluent software as a tool to predict the flow field around the body. To investigate the flow field, the Navier-Stokes equations are modelled by using density-based method. The flow considered here is three-dimensional, stationary, viscous, and turbulent in which the RNG K-epsilon model has been used for turbulence modelling. To accommodate accurate turbulence modelling, the standard wall functions are selected. A first order accurate method is computed to establish the flow. A second order accurate method is computed to achieve the convergence of the solution.

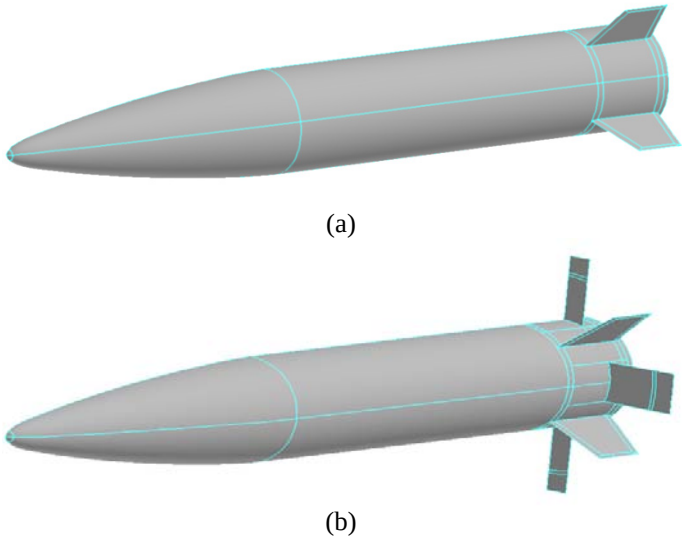
The verification of the CFD method included quantitative comparisons provided by Missile DATCOM as well as qualitative study of flow structure. Missile DATCOM is a semi-experimental aerodynamic prediction code that calculates aerodynamic forces, moments and stability derivatives as function of angle of attack and Mach number for a variety of axisymmetric and nonaxisymmetric missile configurations.

2.1 Geometry and boundary conditions

In this study to find out the influence of the drag surfaces on the aerodynamic characteristic of the body two cases have been considered. Figure 1 shows both the models.

In first case, a body-fins configuration is examined with 2350 in length and 338 in diameter. The configuration has an ogive nose with four tail fins arranged in a cruciform pattern. The fins have a supersonic airfoil cross section. The set of the fins increases the lift force and stability of the body.

In the second case, four drag surfaces deployed in between of each pair of the fins is modelled which placed across the centerline of the vehicle partially block the flow. The drag surfaces are conformed to the outer surface of the body when are retracted. In other words the deployable drag surfaces folded on the body. When the deployment command is issued by the flight computer, the drag



(a)

(b)

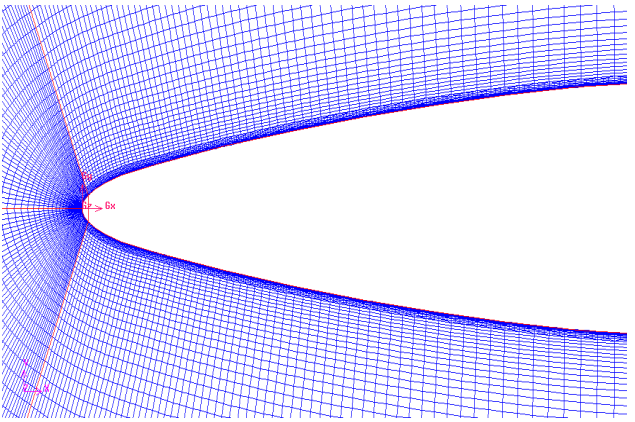
Figure 1: Geometry of body a) retracted position of drag surfaces b) deployed position of drag surfaces.

surfaces are released, and travel from the retracted position (figure 1a) to the deployed position (figure 1b). This unconventional method has been applied to enhance the drag forces. In this research a half of the model has been considered because of the symmetric flow and geometry to reduce calculation time. Three types of boundary conditions were used: wall, pressure far-field, symmetry.

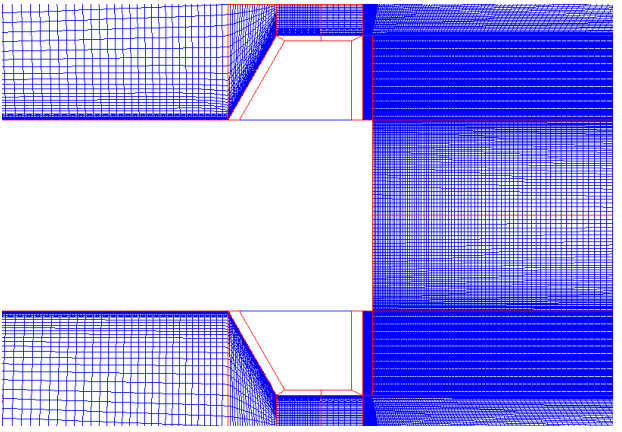
2.2 Grid generation

The accuracy of the CFD results greatly depends on how accurately the configuration is modelled and meshed. The entire solution domain is discretized by a structured grid. Several grid studies were performed to ensure that the grid was fine enough to capture the physical characteristics of the flow. It had been seen that the calculated CA approaches the constant value by increasing the grid qualification. The limitation in computer memory storage is a major factor in preventing further increase in grid size.

The total number of cells is about 4,000,000 for body-fins configuration and 4,500,000 cells for body-fins with deployable drag surfaces. Figure 2 shows the quality of the structured grid adjacent to the surface of the body, nose, fins and drag surfaces. For evaluating the quality of the structured grid adjacent to the body surface, the values of Y+ at the first of wall nodes were calculated. It was shown that most of them have the values more than 30 except for a negligible number of the grids. A trade off between computation time and quality of result led to a grid with finer mesh near to solid surface and coarser mesh adjacent to far field boundaries.



(a)



(b)

Figure 2: The quality of the structured grid (a) adjacent to the surface of the nose (b) adjacent to the surface of the body, fins and drag surfaces.

3 Results and discussion

A discussion is mainly presented herein to find out the influence of the drag surfaces on the aerodynamic characteristic of the body. A secondary interest in the investigation is the interaction of a drag plates with a stabilizing fins and the effects on the efficiency of stabilizing fins on vehicle in supersonic and subsonic flow. For this purpose force measurement and flow visualization of the numerical flow simulation for each case has been done. Validation of the results has been done through comparing forces with results of semi-experimental method.

3.1 Force coefficients

Body-fins configuration has been considered to study as the first case. The study has been done for subsonic Mach number 0.4 and supersonic Mach number 1.5 at different angles of attack. Semi-experimental method (MD code) is used as aerodynamic prediction code to determine the accuracy of numerical simulation in this paper. Figure 3 shows a comparison between predictions for drag force coefficients for numerical code and semi-experimental method for body before releasing drag surfaces. It is evident that especially at these low angles of attack there is an excellent consistency between solvers. There is a little difference in Mach number 1.5.

As the second case, the configuration of the body-fins with drag surfaces as figure 1b has been considered. In this case semi-experimental method is used as aerodynamic prediction codes to determine the accuracy of numerical simulation too with this difference that the drag surfaces are modelled as local protuberance and their effects has been considered on axial force. Therefore the axial force coefficient is only reported for this case in figure 4. This figure shows that the results from numerical modelling agree well with the trend and almost the magnitude of semi-experimental data.

3.2 Flow visualization and path line analysis

Figure 5 is a picture of contour of Mach number from side view of body after releasing drag surfaces at 0 degree angle of attack and Mach number 1.5.The nose oblique shock and expansion waves are clearly visible in this figure. There is a flow separation upstream of the drag surfaces, which causes a thickening of the boundary layer. The thickening leads to compression of the flow and eventually a shock wave. Supersonic flow faces the drag surfaces causing a second shock wave. The secondary shock structure is bowed in nature. Decreasing velocity of cross flow results region of recirculation in flow which produces a separation upstream

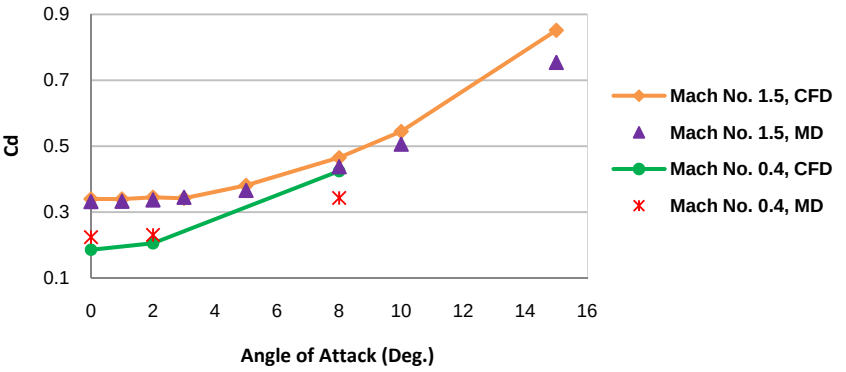


Figure 3: Comparison numerical and semi-experimental results of the drag coefficient against angles of attack before releasing drag surfaces.

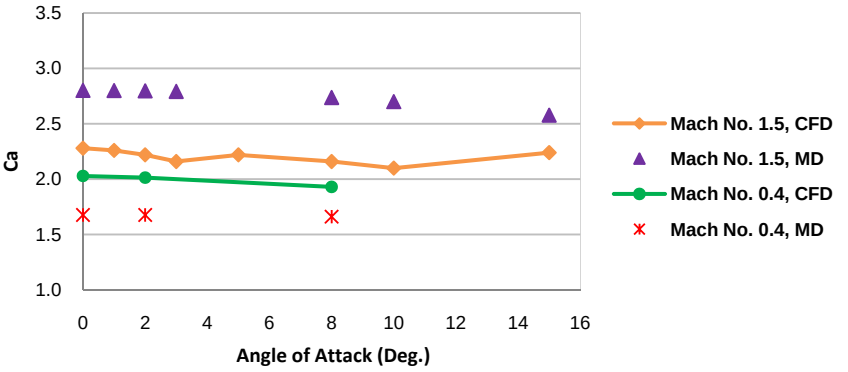


Figure 4: Comparison numerical and semi-experimental results of the axial force coefficient against angles of attack after releasing drag surfaces.

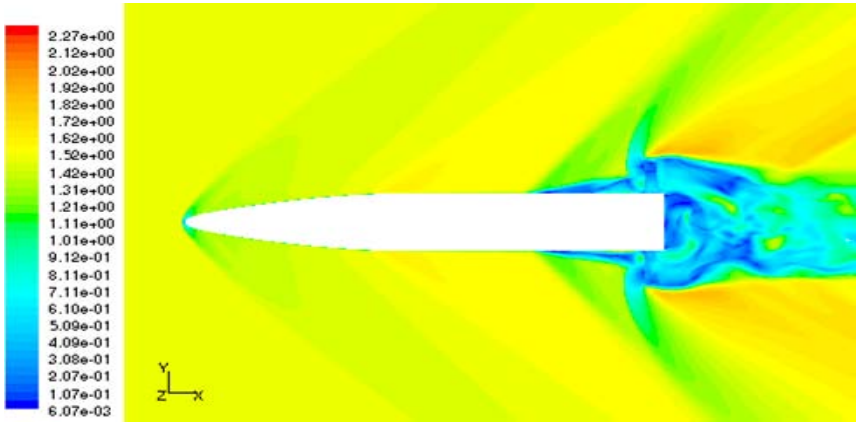


Figure 5: Mach number Contours after releasing drag surfaces, M=1.5 at $\alpha=0^{o}$.

of the drag surfaces which has not been seen on body before releasing drag surfaces. The region of influence of the separation is seen to encompass a larger area of the body at supersonic flow as compared with the subsonic flow in figure 6. Downstream of the drag surfaces, contra rotating vortices have been predicted which is shown in figure 7 in Mach number 0.4 and angle of attack 0 degree.

Since the surface area of body base is increased by releasing drag surfaces, it is expected the pressure base drag increased significantly. Eventually, the impingement upon the surface leads to changes in pressure distribution on the body surface which is shown in figure 8. This pressure distribution change results in different forces and moment on fins and body. Upstream of the drag surfaces, there is a pressure increase, which produces an existence of shock waves.

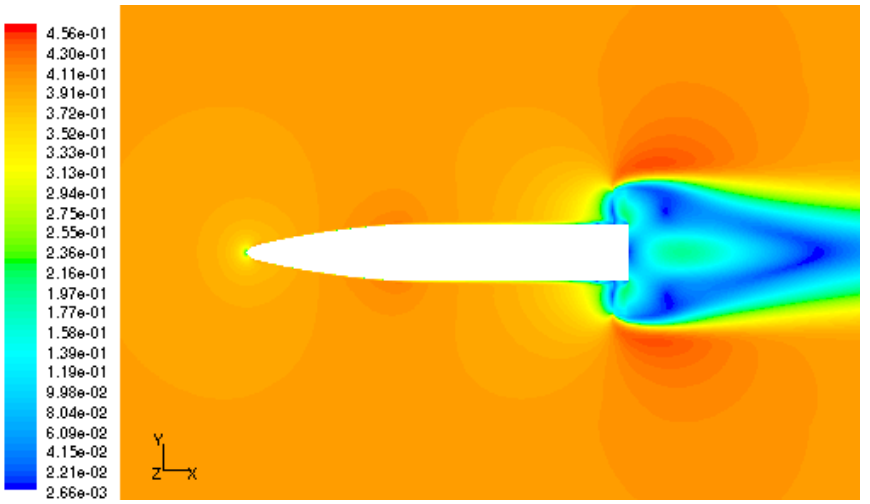


Figure 6: Mach number contours after releasing drag plates, M=0.4 at α=0^o.

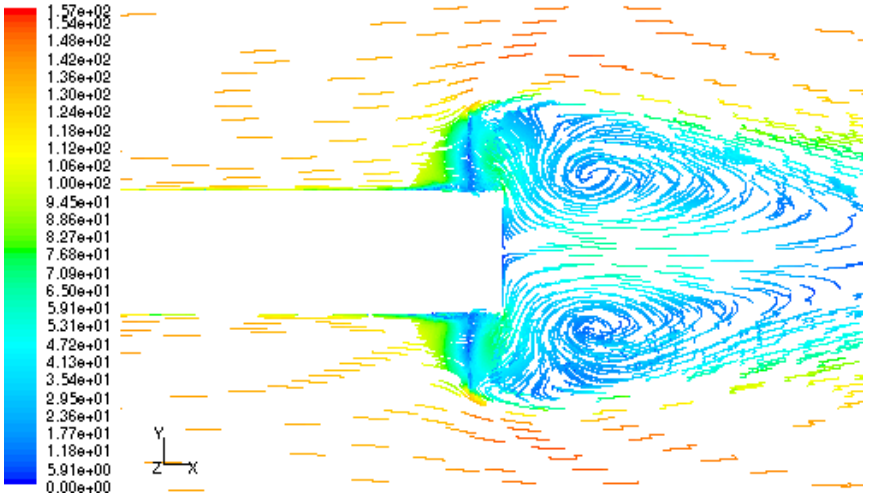


Figure 7: Side view of path line colored by velocity magnitude, Contra rotating vortices, M=0.4 atα=0^o.

A comparison between flow patterns on the body at 8 degrees angle of attack with 0 degree shows much more differences on flow. The first major difference between these cases and the zero incidence cases previously discussed is the surface area of the body affected by the drag surfaces. The boundary layer thickening occurs much earlier and the region of recirculation moves forward with increasing angle of attack on upper surface of body. The other regions of recirculation are those of the contra rotating vortices downstream of the drag surfaces. Whilst they appeared symmetric at zero incidences, they are asymmetric at an angle of attack.

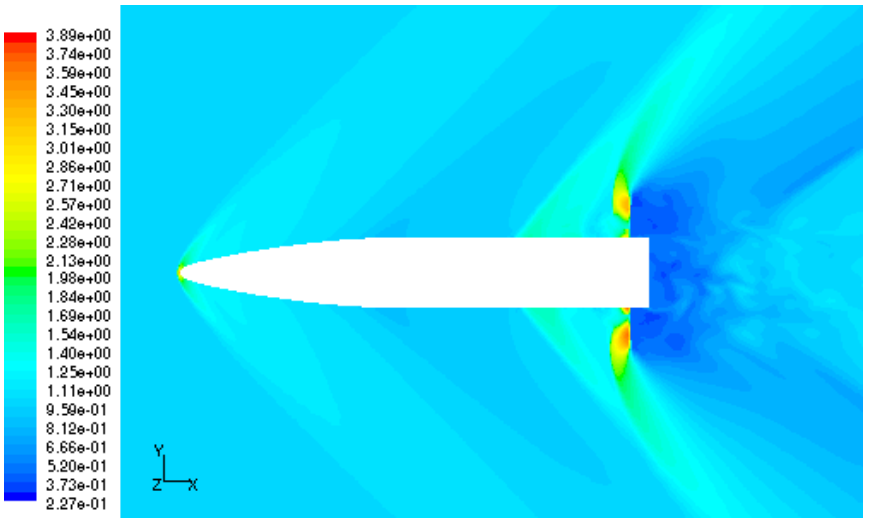


Figure 8: Static pressure contours after releasing drag plates, M=1.5 at $\alpha=0^o$ (atm).

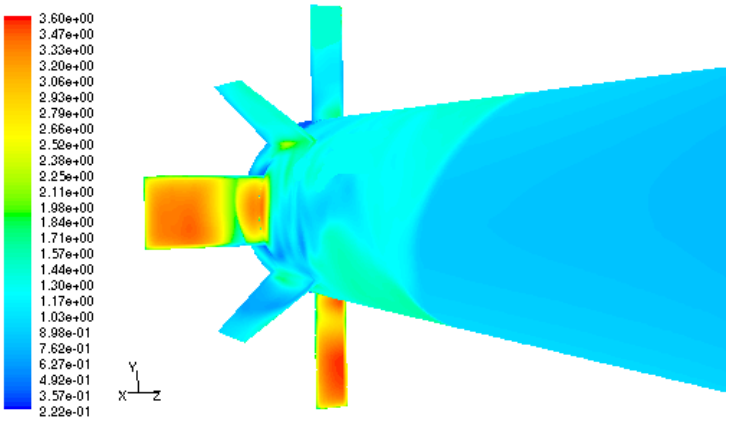


Figure 9: Static pressure contours after releasing drag plates, M=1.5 at $\alpha=8^o$ (atm).

In figure 9, the surface static pressure contours on one side of typical fin and drag surfaces are presented at Mach number 1.5 and 8 degrees of angle of attack. As can be seen, the difference flow patterns and difference pressure distribution on the opposing fin and drag surfaces produce a different load distribution.

3.3 Discussion

A series of numerical models have been computed for a range of angles of attack in Mach numbers 1.5 and 0.4. Analysis of the aerodynamic coefficients provides understanding of how the drag plates affect the body aerodynamics. Figure 10 to 13 shows the change in aerodynamic coefficient versus angle of the attack for two models, before and after releasing drag surfaces.

Figures 10 and 11 show the change in drag coefficient versus angle of the attack for two models, before and after releasing drag surfaces. A comparison between coefficients shows that the significant change in drag force. In fact with releasing drag surfaces and placing them on deployed position increase complexity in flow and produce large vortices behind the base area. This complicated structure around the drag surfaces may consist of a strong secondary shock on body before facing fins and drag surfaces and interaction boundary layer of the fins and drag surfaces with shock waves. In addition second shock wave in supersonic region has a decrease in flow velocity and reverse flow on body. Consequence of this configuration of drag surfaces, the region of the dead flow behind the base of body increased which produce more pressure base drag. These phenomena cause the drag coefficient increased up to 10 times in M=0.4 and 3 to 6 times in M=1.5 at different angle of attack.

As shown in figures 12 and 13 the lift coefficient with increasing angle of attack decreased between 25 to 45 percent in M=1.5 with different angles of attack and about 80 percent in M=0.4. This is due to the negative lift force on body which is produced by drag plates. In addition drag plates disturb the flow over its neighbouring fins which change pressure distribution on body and fins and reduce efficiency of the fins performance.

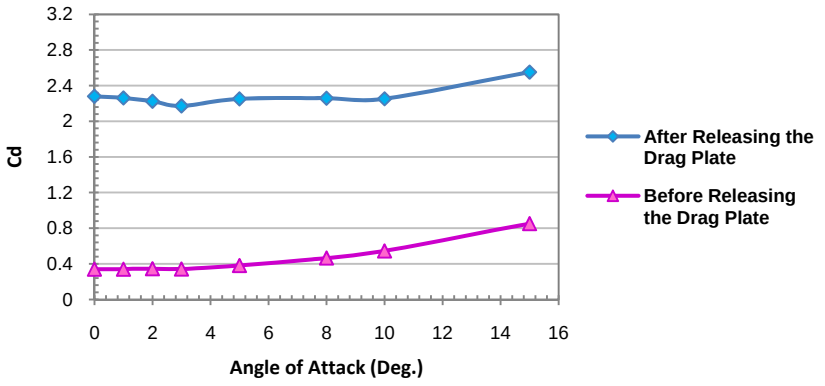


Figure 10: Comparison of the drag coefficient against angle of attack between two models, before and after releasing drag plates at M=1.5.

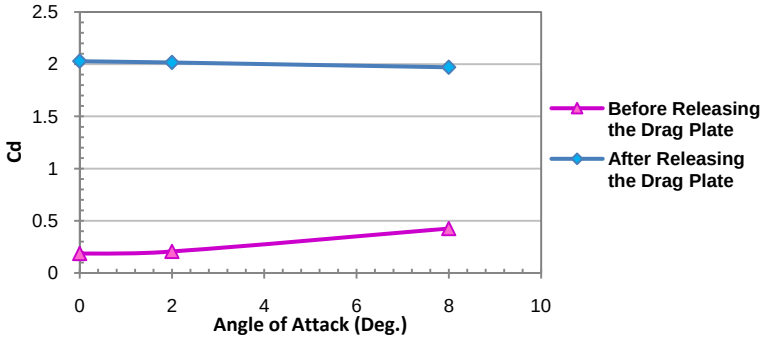


Figure 11: Comparison of the drag coefficient against angle of attack between two models, before and after releasing drag plates at M=0.4.

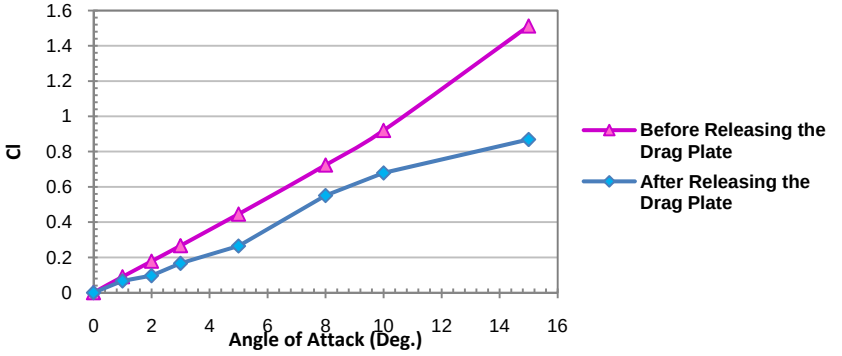


Figure 12: Comparison of the lift coefficient against angle of attack between two models, before and after releasing drag plates at M=1.5.

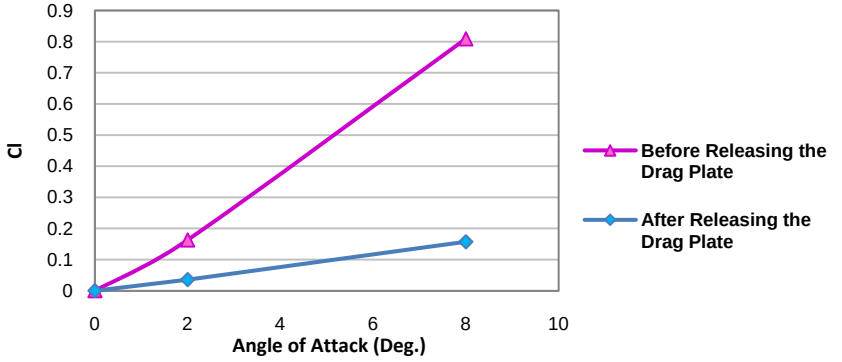


Figure 13: Comparison of the lift coefficient against angle of attack between two models, before and after releasing drag plates at M=0.4.

4 Conclusion

An investigation into the aerodynamic characteristics of a body with deployable drag surfaces is carried out. A series of numerical modelling is done for a range of angles of attack in Mach numbers 0.4 and 1.5 to include both subsonic and supersonic flow regimes. The study included two models of body with retracted position of drag surfaces and deployed position of drag surfaces. The drag and lift force coefficients were calculated from the flow field solutions. The results for retracted and deployed models were compared with each other in order to study the influence of drag surfaces on the aerodynamic force on the body.

A general study of the main aerodynamic coefficients shows that at all angles of attack, the coefficient of lift decreases and the coefficient of drag increases. This is due to the disturbing effect of the plates on the flow structure around the nearby fins which in its own turn decreases the performance of the set of fins significantly. The presence of the flat plate as local protuberance produces more drag force in different angles of attack as expected. This additional drag force brings body velocity and body rate of descent down to an acceptable level of speed for parachute deployment.

References

[1] Available Online: www.info-central.org/recovery_techniques.shtml
[2] Anita Sengupta et al, Supersonic Performance of Disk-Gap-Band Parachutes Constrained to a 0-Degree Trim Angle, *Journal of Spacecraft and Rockets*, Vol. 46, No. 6, November–December 2009
[3] Richard Benneyet al, Aerodynamic Decelerator Systems-Diverse Challenges and Recent Advances, *Journal of Aircraft*, Vol. 38, No. 5, September–October 2001
[4] Brandon P. Smith et al., A Historical Review of Inflatable Aerodynamic Decelerator Technology Development, *IEEE Aerospace Conference*, March, 2010
[5] Arash N. Lahouti et al., Design, Development and Testing of a Rigid Aerodynamic Decelerator System for Recovery Of a High-Altitude Sounding Rocket Payload, *The 1st International ARA Days: Atmospheric Re-entry Systems, Missions and Vehicles*, France, 2006
[6] Anon., Sounding Rocket Program Handbook, NASA Goddard Space Flight Center, Wallops Flight Facility, Wallops Island, Virginia, USA, June. 2005
[7] SRP-4 Design Team, "SRP-4 Design Document", University of Alaska Fairbanks, Alaska, USA, 2001
[8] Wm. David Washington et al., Experimental Investigation of Grid Fin Aerodynamics, Symposium on Missile Aerodynamics, Italy, 1998
[9] Mark Bell et al., A Numerical Study into a Local Protuberance Interaction with a Fin on Supersonic Projectile, *47th AIAA Aerospace Sciences Meeting Including the New Horizons Forum and Aerospace Exposition*, Florida, 2009

Fluid dynamics of a post-combustion chamber in electric arc steelmaking plants

L. Labiscsak[1], G. Straffelini[1], C. Corbetta[2] & M. Bodino[2]
[1]*Department of Materials Engineering and Industrial Technologies, University of Trento, Italy*
[2]*Stg Group SpA, Italy*

Abstract

In the present work, a complete three dimensional fluid dynamics simulation of reacting fluid flow in post-combustion chamber of an electric arc furnace is presented. The chemical composition of the simulated off-gases was representing typical crucial load. The gap (where oxygen enters to combust hydrogen and carbon monoxide) size was an independent variable. The optimal gap size is desirable: if the gap size is too large, the thermal efficiency diminishes. On the other hand, if the chosen gap size is too small, oxygen deficiency occurs, which lead to incomplete combustion of carbon monoxide and hydrogen. Herein established, that by means of CFD calculation, the optimal gap size can be evaluated for a particular steelmaking furnace.

Keywords: steelmaking, post-combustion, CFD.

1 Introduction

The steel production by electric arc furnace (EAF) from scrap metal is widely used technique. During the production of steel considerable amount of combustible gases are formed such as carbon monoxide and hydrogen, which are extracted directly through the fourth hole, which is placed on the water-cooled furnace roof. As a result, the furnace inner is depressurized, which helps to minimize the harmful gas emissions, and air enters the furnace from the factory's ambient. The flow rate of the penetrating air (usually called false air), is defined by: the suction flow-rate of the direct evacuation system (DES),the design of the furnace inner and the amount of generated gases in the furnace. Consequently, the operation of the DES influences the mass and energy balance of the EAF.

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110191

The oxygen in the false air oxidizes the carbon monoxide and hydrogen in the furnace. However, up to 10% (occasionally 30%) of carbon monoxide and 10% of hydrogen [1, 2] remain unburned and leave the free board via the fourth hole. Since, these reactions are highly exothermic; the complete oxidation of the abovementioned gases is desired in order to maximize the thermal efficiency of the EAF.

For this reasons, the primary line is equipped with a gap, where fresh air may enter from the factory's ambient and mix with off-gases. The oxidation takes place in the post-combustion chamber. The size of the gap can be changed and is an important operation parameter. If the gap size is too large, the temperature will drop, leading to lower thermal efficiency and lower reaction rates. By contrast, the smaller the gap, the higher the temperature, this enhances the decomposition of carbon dioxide. A small gap size also lead to oxygen deficiency, hence the complete oxidation is inhibited. The walls of the post-combustion chamber are water-cooled, which enhances efficiency and desirable for protecting filtering devices from high temperature as well. As a result, the direct evacuation system plays a dominant role in the steelmaking process.

CFD calculations are increasingly used in development and improvement of steelmaking processes [3]. Li *et al.* [2] presented a three-dimensional (3D) CFD simulation about the furnace freeboard. Chattopadhyay *et al.* [4, 5] recently published review about the application of CFD simulation in the field of iron- and steelmaking. Furthermore, Mazudmar and Evans [6] dedicated a complete section in their book about CFD relevance in the field of steelmaking. Kirschen *et al*. [1] published an analytical model about the dedusting system of the furnace and compared its results to experimental measurements. Due to the huge temperature difference between the flame and surroundings, the radiation modeling has a great importance. Habibi *et al.* [7] interpreted CFD simulation about CH_4/H_2/air flames with different radiation heat exchange models and compared the results to adiabatic flame.

In the present investigation, a complete 3D numerical simulation of reacting flow, energy and concentration fields of a post-combustion chamber were carried out. Many simulation routines were used to find the suitable models for the best accuracy. In order to optimize the operation of the post-combustion chamber, the size of the gap was changed and conversions along with thermal efficiencies of the chamber were calculated simultaneously.

2 Modeling of post-combustion chamber

2.1 Geometry and computational grid

A newly designed post-combustion chamber was simulated. The geometry of the simulated reactor is depicted in Figure 1 and the relevant chamber specifications are listed in Table 1.

The off-gas with hydrogen and carbon monoxide enters via the inlet and mixes with fresh air coming from the ambient through the gap. The size of the gap is an independent variable, i.e., a design parameter (Figure 2).

Table 1: Chamber specification.

Total volume	388,7 m^3
Inlet area	2,9 m^2
Outlet area	6,6 m^2
Water-cooled surface	430,0 m^2

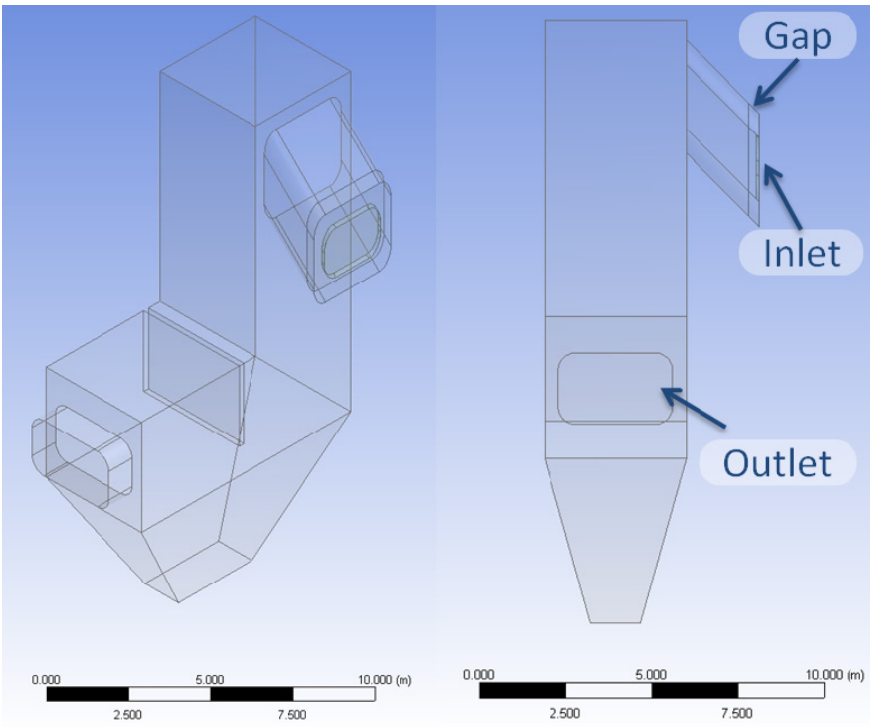


Figure 1: Geometry of the post-combustion chamber.

Consequently, the selected operating parameters define the mixing ratio, i.e. the ratio between the airflow through the gap ($\dot{V}_{air}$) and the off-gases from the EAF($\dot{V}_{EAF}$).

The geometry was discretized by tetrahedral elements, Figure 3. The computational mesh was unstructured and conformal. Local mesh refinements were applied in regions where steep gradients of the dependent variables (temperature, velocity, pressure) were expected.

The mesh was also refined in reacting zone (where carbon monoxide, hydrogen and oxygen mix) where high gradients of concentrations are expected. Steep temperature gradient was likewise expected in reacting zone due to the

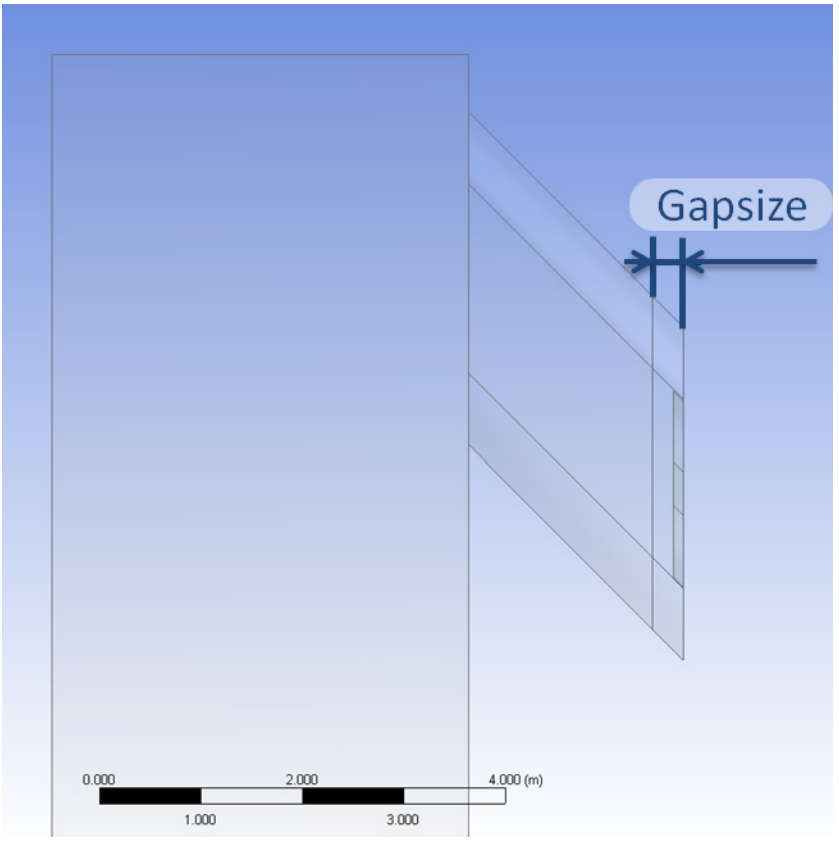


Figure 2: The gap size controls the mixing ratio (see text for definition).

exothermic reactions. The number of the cells was typically between 500.000 and 650.000, depending on the gap size. Since the present investigation is a preliminary study, the mesh independent study was not presented herein nevertheless it is currently in progress.

The convergence criteria of CFD calculations were achieved and the adopted residuals for each equation are listed in Table 2.

2.2 Kinetic mechanism

The following skeleton reactions were taken into account within the reacting flow:

$$2H_2 + O_2 \rightarrow 2H_2O \quad (1)$$

$$2CO + O_2 \leftrightarrow 2CO_2 \quad (2)$$

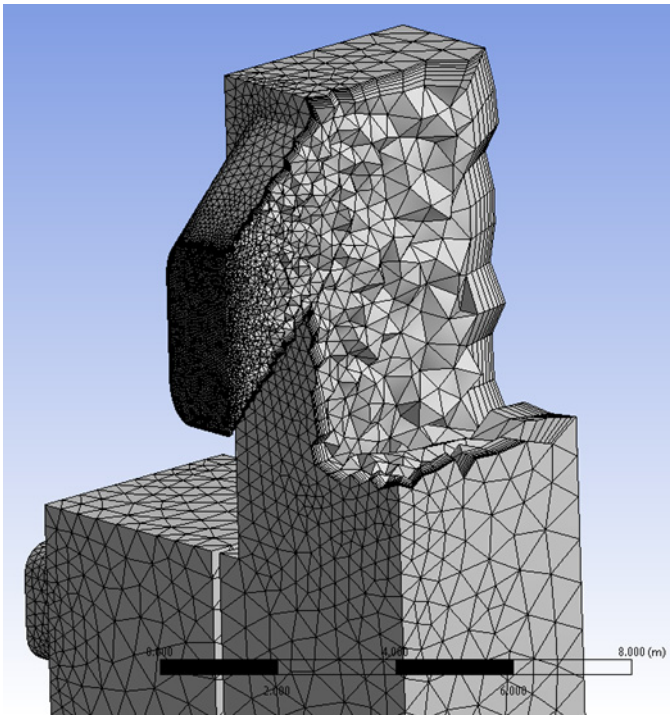


Figure 3: The discretized 3D domain.

Table 2: Adopted residuals for each equation.

Equations	Residuals
Continuity	1E-05
Velocity (x,y,z)	1E-06
Energy	1E-06
Turbulent kinetic energy (κ)	3E-06
Turbulent dissipation rate (ε)	5E-06
DO intensity	1E-06
Species	1E-06

Simplified, one-step reaction simulations give poor prediction about carbon monoxide and hydrogen content of the exiting gas and significantly over predict the carbon monoxide content at the outlet. Therefore, detailed kinetic mechanism was implemented and comprises 12 species and 37 reactions and based on the studies of Cuoci *et al.* [8]. The used rate constants are in good match with experimental measurements in wide range of temperature. The reacting flows were simulated by the Eddy dissipation concept (EDC) model [9], which allows for integrating the detailed chemical mechanisms into turbulent reacting flows.

Non-Premixed Combustion (NPC) approach [10], and the Finite Rate/Eddy Dissipation approach were also tested, however EDC model provided the most reliable results [11].

2.3 Radiation modeling

Thanks to the exothermic reactions in the chamber, the difference between fume's temperature and water-cooled walls is very large; hence, the heat transfer by radiation plays a dominant role. The approximation adopted in this study for radiation modeling is the Discrete-Ordinates approach, which is an accurate method [7]. It can handle wide range of optical thickness, at the price of a high computational cost. The absorbencies of the materials have a strong influence. The off-gas is mixture of gases, which have different absorption coefficients. Furthermore, the composition of the mixture changes rapidly along the chamber. Therefore, absorption coefficients must be handled carefully; the weighted-sum-of-gray-gases model (WSGGM) was used, since it is a reasonable compromise between the oversimplified gray gas model and the complete model, which takes into consideration the specific absorption bands [12].

2.4 Numerical scheme

The governing equations for the mass, momentum, energy, chemical species and radiation were solved segregated and the SIMPLE method was used for pressure-velocity coupling. The non-linear equations were discretized implicitly by second-order upwind scheme. The standard $\kappa-\varepsilon$ model was used for turbulence modeling. As known, this model is widely used in CFD applications in the field of steelmaking, since it is a reasonable compromise between accuracy and computational cost [5]. The appropriate under-relaxation factors were set to find the convergence.

2.5 Boundary conditions

The selected boundary conditions are shown in Table 3. The listed data are input parameters, which were taken from literature data and experimental measurements. The evacuation is determined by the power of the fan in the primary line. The discharge flow rate was set to 230.000 Nm^3/h.

Table 3: Implied boundary conditions.

Inlet		Gap	
Pressure	-10 Pa	Pressure	-10 Pa
Temperature	1523 K	Temperature	473 K
Chemical composition	Table 4.	Chemical composition	Table 4.
Outlet		Walls	
Velocity	23,6 m/s	Temperature	353 K

The selected concentration of the entering gas at the fourth hole and at the gap is listed in Table 4. These values represent a typical load during the melting period, which is the most crucial period from post-combustion point of view.

Table 4: Input data for chemical compositions. At the gap, pure air was assumed.

Material	Inlet (%)	Gap (%)
CO	30	0
CO_2	0	0
H_2	10	0
H_2O	0	1
O_2	0	21
N_2	60	78

3 Results and discussion

In our CFD calculations, the oxidation of carbon monoxide and hydrogen in the post-combustion chamber was simulated with three different gap sizes: 20, 30 and 40 cm. The heat efficiency and the conversion rates were monitored at each geometry.

The heat transfer efficiency(η) was evaluated from the ratio between the transferred heat via the water-cooled wall (H_{wall}) and the sum of the entering heat (H_{in}) and reaction heat (H_{react}):

$$\eta = \frac{H_{wall}}{H_{in}+H_{react}} \tag{3}$$

The reactions occur only in case of mixing of combustible gases and oxygen. Figure 4 showing the mass fraction of carbon monoxide and proves that the reactions take place at the very beginning of the post-combustion chamber. Consequently, the mixing can assumed very efficient. The calculated conversion factors are listed in Table 5.

It can be noted, that complete combustion of carbon monoxide and hydrogen occurred only in case of 40 cm gap size. When the gap size is 30 cm, the conversion is almost satisfactory (Table 5). Nevertheless, in case of 20 cm gap size, there is not enough oxygen for the complete reaction. Therefore, the

Table 5: Results of the simulations: conversions and final contents of CO and H2 expressed in ppm.

Gap size (cm)	CO conversion	H_2 conversion	CO (ppm)	H_2 (ppm)
20	0,53	0,72	34300	6920
30	0,97	0,99	1475	163
40	0,99	1,00	64	0,34

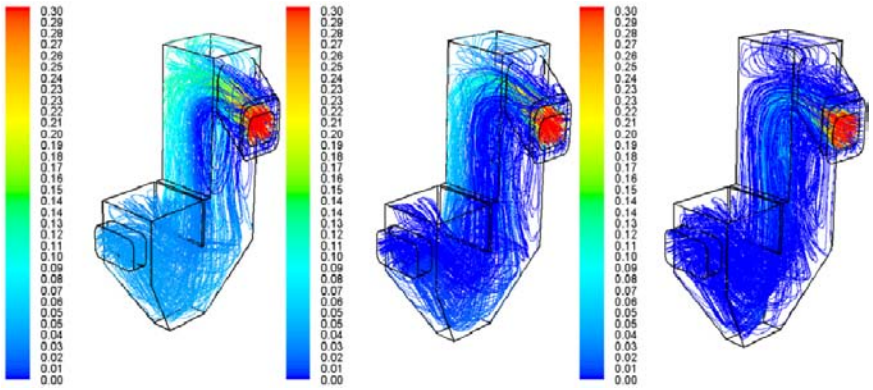


Figure 4: Carbon monoxide's mass fractions in the post-combustion chamber (color maps are scaled in mass fraction); gap sizes are 20, 30 and 40 cm, from left to right respectively.

conversion rate of H_2/CO remains low and its concentration at the exit is quite high.

In Table 6, the concentrations of the exiting gases are listed. The compositions of the leaving gases were compared to literature data [1] and experimental measurements, provided by the Stg Group Company. The calculated data match with the available data. The effect of the gap size on the oxygen content can be easily derived. The small gap size causes oxygen deficit and incomplete combustion.

Table 6: Results of the simulations: chemical composition of the exiting gas (expressed in mass percentage).

Gap size (cm)	CO (%)	CO_2 (%)	H_2 (%)	O_2 (%)	H_2O (%)
20	3,43	6,01	6,92E-01	2,90E-05	16,2
30	0,15	7,80	1,63E-02	1,18	15,9
40	6,47E-03	7,30	3,41E-05	2,67	14,8

Thermal efficiencies are lower when the gap size is large and the oxygen excess also presents (Table 7). Moreover, large gap causes lower temperature at the outlet (Figure 5), which is desirable to avoid the damage of the polymer filters.

Table 7: The mixing ratio, thermal efficiency and the temperature of the exiting gas.

Gap size (cm)	Mixing Ratio	Thermal Efficiency (η)	Outlet Temp. (K)
20	1,34	0,65	1172
30	2,09	0,62	1264
40	2,32	0,60	1252

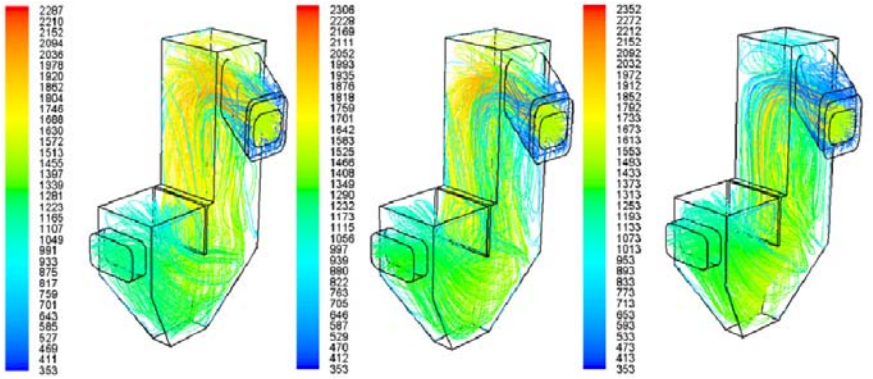


Figure 5: Gas temperature in the post-combustion chamber (color maps are scaled in Kelvin); gap sizes are 20, 30 and 40 cm, from left to right respectively.

Our results show, that the increase of the gap size reduces the emission of harmful gases due to the complete oxidation. In the meantime, the value of the thermal efficiency remains reasonably high.

From the fluid dynamic point of view, the selected geometry of the post-combustion chamber poses conflicting observations. From the one hand, the calculated path lines demonstrated a turbulent pattern of flow, with the creation of tortuous streamlines and recirculating zones. This is clearly due to the presence of sharp-edge corners, abrupt changes of direction and stagnate zones. All these factors do not help the smoothness of the flow, although help the settling of fine particles, which may present in the off-gas. On the other hand, this does not adversely affect the overall performance of the chamber, since mixing is enhanced.

4 Conclusions

In the present study, we applied CFD calculation for the simulation of post-combustion chamber. First, we obtained that the mixing in the chamber is efficient, due to the reacting zone, which is located at the beginning of the reactor. Particularly, we focused on the importance of the gap size. When the gap size was set to 30 and 40 cm, thanks to the oxygen content, the conversion rate of hydrogen and carbon monoxide was high and the oxidation is complete. In addition, we also obtained that the thermal efficiency decreases when the gap size is larger, however the reduction is not significant. Therefore, larger gap size is desirable in practice to avoid the carbon monoxide and hydrogen escape to the ambient.

References

[1] M. Kirschen, V. Velikorodov, H. Pfeifer, Mathematical modeling of heat transfer in dedusting plants and comparison to off-gas measurements at electric arc furnaces, *Energy*, **31**, pp. 2926-2939, 2006.

[2] Yun Li, Richard J. Fruehan, Computational Fluid-Dynamics Simulation of Post combustion in the Electric-Arc Furnace, *Metallurgical and Materials Transactions B*, **34B**, pp. 333-343, 2003.

[3] P. Gittler, R. Kickinger, S. Pirker, E. Fuhrmann, J. Lehner, J. Steins, Application of computational fluid dynamics in the development and improvement of steelmaking processes, *Scandinavian Journal of Metallurgy*, **29,** pp. 166-176, 2000.

[4] K. Chattopadhyay, M. Isac, R. I. L. Guthrie, Application of Computational Fluid Dynamics (CFD) in iron- and steel making: Part 1, *Iron making and Steel making*, **37(8)**, pp. 554-561, 2010.

[5] K. Chattopadhyay, M. Isac, R. I. L. Guthrie, Application of Computational Fluid Dynamics (CFD) in iron- and steel making: Part 2, *Iron making and Steel making*, **37(8)**, pp. 562-569, 2010.

[6] D. Mazudmar, J. W. Evans, *Modelling of steelmaking processes*, Boca Raton, FL, CRC Press, 2009.

[7] A. Habibi, B. Merci, G. J. Heynderickx, Impact of radiation models in CFD simulations of steam cracking furnaces, *Computers and Chemical Engineering*, **31**, pp. 1389-1406, 2007.

[8] A. Cuoci, A. Frassoldati, G. Buzzi Ferraris, T. Faravelli, E. Ranzi, The ignition, combustion and flame structure of carbon monoxide/hydrogen mixtures, Note2: Fluid dynamics and kinetic aspects of syngas combustion, *International Journal of Hydrogen Energy*, **32**, pp. 3486-3500, 2007.

[9] Magnussen B.F, On the structure of turbulence and a generalized Eddy dissipation concept for chemical reactions in turbulent flows. *19th AIAA aerospace science meeting*, St. Louis, Missouri, 1981.

[10] Y. R. Sivathanu & G. M. Faeth. Generalized State Relationships for Scalar Properties in Non-premixed Hydrocarbon/Air Flames, *Combustion and Flame*, **82**, pp. 211-230, 1990.

[11] L. Labiscsak, G. Straffelini, F. Trivellato, M. Bodino, C. Corbetta, Computational fluid dynamics simulations of post combustion chambers,*33th AIM National Congress*, Brescia, Italy, 2010.

[12] G. Krishnamoorthy, A new weighted-sum-of-gray-gases model for CO_2-H_2O gas mixtures, *International Communications in Heat and Mass Transfer*, **37**, pp. 1182-1186, 2010.

Apparent shear stress and boundary shear distribution in a compound channel flow

K. K. Khatua, K. C. Patra, S. Behera & P. K. Mohanty
Department of Civil Engineering, National Institute of Technology, India

Abstract

The flow structure in any compound channel is a complicated process due to the transfer of momentum between the deep main channel section and the adjoining shallow floodplains. The boundary shear stress distribution in the main channel and floodplain greatly affects the momentum transfer. In the present work, the shear stress distributions across an assumed interface plane originating from the junction between the main channel and flood plain using the Divided Channel Method (DCM) are analyzed and tested for different compound channels and their flow conditions using global data. An improved equation to predict the boundary shear distribution in compound channels for different width ratios is derived that gives better results than other proposed models. Analyses are also done to suitably choose an appropriate interface plane for evaluation of stage-discharge relationship for compound channels having equal roughness in the channel beds and walls. The effectiveness of predicting the stage-discharge relationship using the apparent shear stress equation and boundary shear distribution models are discussed.

Keywords: apparent shear, main channel, floodplain, compound channel, discharge estimation, interface planes.

1 Introduction

During floods, a part of the river discharge is carried by the main channel and the rest is carried by the floodplains. Momentum transfer between the flow of deep main channel and shallow floodplain takes place making the discharge prediction in compound channel more difficult. In the laboratory the mechanism of momentum transfer between the channel section and floodplain was first investigated and demonstrated by Zheleznyakov [30] and Sellin [24].

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110201

The traditional discharge predictive methods for compound channels either use the Single-Channel Method (SCM) or the Divided-Channel Method (DCM). The DCM divides a compound section into hydraulically homogeneous sub-sections generally by vertical, inclined or horizontal division lines that lead to an averaged flow velocity for each sub-section (e.g., Chow [4]). These approaches have the advantage of recognizing a particular hydraulic property in the respective compartments. Therefore, this method predicts better overall discharge as compared to SCM (Weber and Menéndez [27], and Patra and Khatua [20]) but it overestimates the flow in main channels and underestimates the flow in the floodplain due to the neglect of lateral momentum transfer.

While using the vertical interface division of DCM, Wormleaton et al. [28] proposed an apparent shear stress ratio, as the useful yardstick in selecting the best interface planes. Holden [6], Lambert and Myers [14], and Patra and Kar [10] also proposed zero shear interface plains that nullify the lateral momentum transfer. The empirical shear stress formulas to calculate the apparent shear at the shear layer between main channel and floodplain (Knight and Hamid [12]) are limited to a particular geometry and are difficult to apply to other data (Knight and Shiono [13]). Based on the published data, Ackers [1] proposed an empirical based correction to the DCM know as Coherence Method (COHM) that is recommended by the UK Environmental Agency Bristol. Shiono and Knight [25] developed a 2-dimensional (SKM) method based on the analytical solution to the depth averaged form of the Navier-Stokes equation. Lambert and Myer [14] developed the weighted divided channel method (WDCM) estimating the discharge capacity for a compound channel.

Toebes and Sooky [33] carried out laboratory experiments and showed that the horizontal interface method would be more realistic than other interface methods. The interaction phenomenon and the discharge assessment for compound sections using DCM were presented by many other researchers as well (e.g., [Bousmar and Zech [3], Knight and Demetriou [11], Knight and Shiono [13], Seckin [23], Patra et al. [19], Kejun Yang et al. [8], Khatua and Patra [10], Abril and Knight [18], Wright and Carstens [29], Yen and Overton [29], Huttof et.al. [7] etc. Failure of most subdivision methods are due to the improper accounting of the complicated interaction between the main channel and floodplain flows, more particularly for channels having wide floodplains. The present study is aimed at understanding the general nature of the interaction between the main channel and the floodplain flows. An attempt has been made to study the information on boundary shear distribution basing on which models on momentum transfer and stage-discharge relationship of compound channels are developed.

2 Experimental analyses

In the present work, a compound channel is fabricated using Perspex sheets inside a tilting flume in the Hydraulic Engineering Laboratory of the Civil Engineering Department, National Institute of Technology, Rourkela, India. The compound channel is symmetrical about the centerline of main channel making

the total width of the compound section as 440 mm (Figure 1). The main channel is rectangular in cross section having 120 mm width and 120 mm at bank full depth. Longitudinal bed slope of the channel is taken as 0.0019. The roughness of the floodplain and main channel are identical. From the experimental runs in the channel, the bed roughness coefficient (Manning *n*) is estimated to be 0.01. A re-circulating system of water supply is established with pumping of water from an underground sump to an overhead tank from where water flow under gravity to the experimental channel through stilling chamber and baffle wall. A transition zone between stilling tank and the channel helps to reduce the turbulence of the flowing water. An adjustable tailgate at the downstream end of the flume is used to achieve uniform flow over the test reach in the channel for a given discharge. Water from the channel is collected in a volumetric tank that helps to measure the discharge rate. From the volumetric tank water runs back to the underground sump.

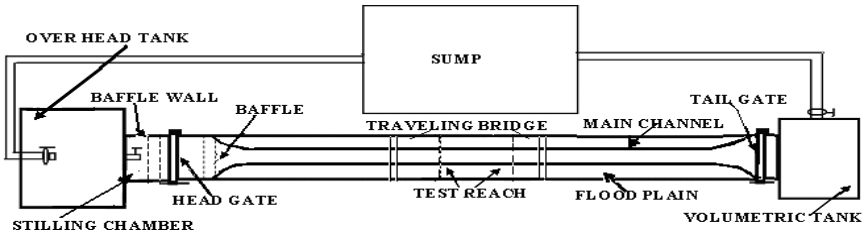


Figure 1: Plan view of experimental set up of the compound channel.

The measuring devices consist of a point gauge mounted on a traversing mechanism to measure flow depths with least count of 0.1 mm. Point velocities are measured at a number of locations across the channel section using a 16-Mhz Micro ADV (Acoustic Doppler Velocity-meter) having accuracy of 1% of the measured range. A guide rail is provided at the top of the experimental flume on which a traveling bridge is moved in the longitudinal direction of the entire channel. Readings from the micro-ADV were recorded in a computer. As the ADV (down probe) is unable to read the data up to 50 mm from free surface, a micro-Pitot tube of 4 mm external diameter in conjunction with suitable inclined manometer is used to measure velocity at other points in the flow-grid. The Pitot tube is physically rotated with respect to the main stream direction till it records the maximum deflection of the manometer reading. A flow direction finder having a least count of $0.1^{o}$is used to get the direction of maximum velocity with respect to the longitudinal flow direction. The angle of limb of Pitot tube with longitudinal direction of the channel is noted by the circular scale and pointer arrangement attached to the flow direction meter.

A total 21 numbers of stage-discharge data for both in-bank and over-bank flow conditions are observed at the test reach. Out of the 10 over-bank stage-discharge data, detailed velocity measurements for 5 stages are recorded at

Table 1: Details of geometrical parameters of the experimental and other compound channels.

Test channel	Series No.	Longitudinal slope (S)	Main channel Width (b) in mm	Main channel depth (h) in mm	Main channel side slope (s)	Ratio of *M*anning's roughness coefficients ($\gamma = n_{fp}/n_{mc}$)	Width ratio (α)	Observed discharge (Q) range in cm^3/s	Relative depth (β) ranges = $(H-h)/H$
(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
Present Channel	Type-I	0.0019	120	120	0	1	B/b = 3.667	8726-39071	0.118- 0.461
Knight and Demetriou	01	0.00096	304	76	0	1	B/b =2	5200-17100	0.108-0.409
	02	0.00096	456	76	0	1	B/b =3	5000-23400	0.131-0.491
	03	0.00096	608	76	0	1	B/b =4	4900-29400	0.106-0.506
FCF Series-A channels	01	1.027×10 - 3	1500	150	1.0	1	B/b=6.67	208200-1014500	0.056-0.400
	02	1.027×10 - 3	1500	150	1.0	1	B/b=4.2	212300-1114200	0.0414-0.479
	03	1.027×10 - 3	1500	150	1.0	1	B/b=2.2	225100-834900	0.0506-0.500
	08	1.027×10 - 3	1500	150	0	1	B/b = 4.0	185800-1103400	0.0504-0.499
	10	1.027×10 - 3	1500	150	2.0	1	B/b = 4.4	236800-1093900	0.0508-0.464

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)

a number of points at the pre defined grid points. The overall discharge obtained from integrating the longitudinal velocity plot and from volumetric tank collection is found to be within ±3% of the values. Using the velocity data, the boundary shear on the channel beds and walls are evaluated from a semi log plot of velocity distribution. Boundary shear stresses are also obtained from the manometric readings of the head differences of Preston tube techniques using Patel's [17] relationship. Error adjustments to the shear value are done by comparing the corresponding shear values obtained from the energy gradient approach. The results so obtained by the two methods are found to be consistently within ±3% values. Summary of the discharges and percentage of boundary shear in the floodplain (%S_{fp}) for different relative depths (β) observed from the experimental runs are given in Table 1.

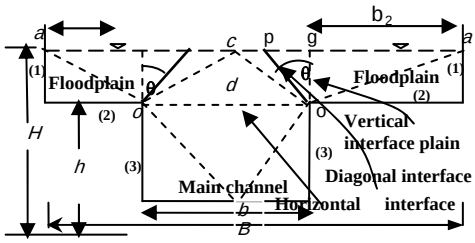


Figure 2: Interface planes dividing a compound section into subsections.

3 Methodology

3.1 Shear force on the assumed interface planes

In Figure 2, the vertical, horizontal, and diagonal plains of separation of the compound channel are represented by the interface lengths o-g, o-o, and o-c respectively. Various boundary elements comprising the wetted parameters are labeled as (1), (2), (3) and (4). Label (1) denotes the vertical wall(s) of floodplain of length [2($H - h$)], where H = total depth of flow from main channel bed, h = depth of main channel. Label (2) denotes floodplain beds of length ($B - b$), where B = total width of compound channel, and b = width or bed of main channel represented by label (4) Label (3) denotes the two main channel walls of length (2h). Experimental shear stress distributions at each point of the wetted perimeter are numerically integrated over the respective sub-lengths of each boundary element (1), (2), (3), and (4) to obtain the respective boundary shear force per unit length for each element. Sum of the boundary shear forces for all the beds and walls of the compound channel is used as a divisor to calculate the shear force percentages carried by the boundary elements (1) - (4). Percentage of shear force carried by floodplains comprising elements (1) and (2) is represented as %S_{fp} and for the main channel comprising elements (3) and (4) is represented as %S_{mc}. Following Knight and Demetriou [11], Knight and Hamed [12] proposed an equation for %Sfp for a compound channel section as

$$\% S_{fp} = 48 (\alpha - 0.8)^{0.289} (2\beta)^{m} \quad (1)$$

Equation (1) is applicable for the channels having equal surface roughness in the floodplain and main channel. For non-homogeneous roughness channels the equation is improved as

$$\%S_{fp} = 48(\alpha - 0.8)^{0.289}(2\beta)^{m}\{1+1.02\sqrt{\beta}\log\gamma\} \quad (2)$$

in which α = width ratio = B/b, β = relative depth = $(H - h)/H$, γ = the ratio of Manning's n of the floodplain to that for the main channel, h = bank full depth and H = total depth of flow. The exponent m can be evaluated from the relation

$$m = 1/\left[0.75e^{0.38\alpha}\right] \quad (3)$$

For homogeneous roughness section (γ =1), equation (2) reduces to the form of Knight and Hamed [12] i.e. equation (1). Due to complexity of the empirical equations proposed by the previous investigators, a regression analysis is made by Khatua and Patra [10] and equation for $\%S_{fp}$ is proposed as

$$\% S_{fp} = 1.23\ \beta^{0.1833}\ (38\ Ln\ \alpha + 3.6262)\ \{1 + 1.02\sqrt{\beta}\log\gamma\} \quad (4)$$

Once the shear force carried by the floodplain is known from equation (4), the momentum transfer in terms of apparent shear force acting on the imaginary interface of the compound section can be calculated. The analysis of momentum transfer helps in predicting the stage-discharge relationship of a compound channel, which is discussed in the later part of the paper. For any regular prismatic channel under uniform flow conditions, the sum of boundary shear forces acting on the main channel wall and bed, along with an ''apparent shear force'' acting on the interface plane originating from the main channel-floodplain junction must be equal to the resolved weight force along the main channel. Using the concept, Patra and Kar (2000) derived the percentage of shear force ASF_{ip} acting at the interface plane as

$$\% ASF_{ip} = 100 \frac{A_{mc}}{A} - \{100 - \% S_{fp}\} \quad (5)$$

in which $\%ASF_{ip}$ = percentage of shear force in the interface plane. Having computed $\%S_{fp}$ using equation (2) or (4), it is easy to evaluate equation (7) for the assumed interface plane. From experiments it is seen that apparent shear stress is higher than the bed-shear stress at low floodplain depths and reduces gradually as over-bank flow depth increases (Rajaratnam and Ahmadi [22], Myers and Elsawy [16], Knight and Demetriou [11], Patra and Khatua [20]). A smaller value of apparent shear stress renders the interface plane more suitable, but a large negative value of apparent shear stress at higher depths makes the interface plane unsuitable for separating the channel into hydraulically homogeneous sub-sections for calculating discharge of compound channels by Divided Channel Method (DCM). The momentum transfer from the main channel to flood-plain is considered as positive percentages of apparent shear and that from flood-plain to main channel is taken as negative.

3.2 Shear force along assumed interface planes in terms of angle θ

Evaluation of apparent shear force at the imaginary interface is helpful for DCM to choose appropriate sub-division lines for separating a compound channel into sub-sections for discharge assessment. Based on the previous works, a general expression for momentum transfer across any interface plane in terms of the angle (θ) made with the vertical line is derived. Consider an arbitrary interface (op), lying between extreme interfaces (oa) and (oe) which make an angle θ to vertical line at the junctions (Figure 2). The convention, used for θ is 0^o for vertical interface (og), positive if the interface lies in the main channel and negative, if the interface lies at the floodplain side. Two situations of locating interface plains can arise.

First, when interface (op) lies between (oa) to (oc) the ranges of angle θ can be defined as $\tan\ \theta \le \frac{b}{2(H-h)}$ and $\tan(-\theta) \le \frac{b_1}{(H-h)}$ or $\le \frac{b_2}{(H-h)}$. The area of main channel for this condition is given as $A_{mc} = bH - (H-h)^2 \tan\ \theta$, where b_1 and b_2 are the lengths of flood plain bed at both sides measured from vertical interface. For a symmetrical compound channel $b_1 = b_2 = (B\text{-}b)/2$. The total area of cross section of channel is $A = bH + (B-b)(H-h)$. Substituting the value of A_{mc} and A in (7) and simplifying, the expression for percentages of apparent shear force in the assumed interface is

$$\%\,ASF_{ip} = 100\,\frac{(\delta - \beta^2 \tan\theta)}{\delta\{1 + (\alpha - 1)\beta\}} - (100 - \%\,S_{fp}) \tag{6}$$

where δ = aspect ratio of the main channel $=b/h$, α = width ratio and β = relative depth which are defined earlier. The second case is when interface op lies between oc to oe, the ranges of angle θ for this situation can be calculated from the relations given as $\tan\ \theta \le \frac{2h}{b}$ and $\tan\ \theta \ge \frac{b}{2(H-h)}$. The area of main channel is $A_{mc} = bh + \frac{b^2}{4}\tan\theta$ and that for whole compound channel is $A = bH + (B\text{-}b)(H\text{-}h)$. Substituting the value of A_{mc} in (7) and simplifying we get

$$\%\,ASF_{ip} = 100\left\{\frac{\delta\cot\theta - 4\beta + 4}{4\{1 + (\alpha - 1)\beta\}}\right\} - (100 - \%\,S_{fp}) \tag{7}$$

For any given interfaces the angle θ is known so the equation (6) and (7) can be directly used to find the apparent shear along any interfaces.

4 Results and discussions

4.1 Boundary shear distribution results

It is seen that the magnitude of momentum transfer at an interface plains using equations (6), or (7) depend on the dimensionless parameters like α, β and $\%S_{fp}$. At a given over-bank flow depth for any compound river section both the parameters α and β are known. The third parameter $\%S_{fp}$ can be calculated using equation (2) of Knight and Hamed [12] or by equation (4) of Khatua and Patra [10]. Knight and Hamed [12] have shown the adequacy of equation (2) for the compound channels having width ratio α up to 4. However, Khatua and Patra [10] have shown the adequacy of α up to 5.25. Interestingly, it is found that when both the equations are tested for FCF data having $\alpha = 6.67$, significant error in $\%S_{fp}$ is estimated [around 90% by equation (2) and 71% by equation (4)]. Figure 3, illustrates the results for FCF data obtained using equations (2) and (4) and its comparison with the observed values. The errors are found to increase with increase in the value of α (width ratio) as well as with the increase of β (relative depth). Furthermore, equation (2) and (4) estimates unrealistic value of $\%S_{fp}$, that is $\%S_{fp}>100\%$ for a compound channel of $\alpha > 10$. Wide floodplains ($\alpha > 10$) are generally encountered during high floods in natural channels and it is essential to estimate boundary shear stress distribution for such channels.

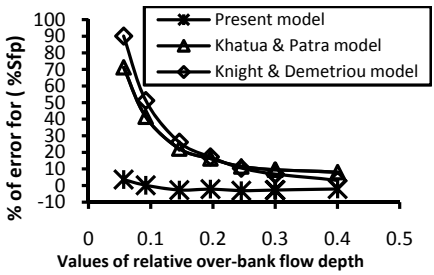


Figure 3: Variation of % error for calculating $\%S_{fp}$ with β for FCF data.

For a better understanding the boundary shear stress distribution, the authors have studied the five series of FCF phase A channels [α =2.6, 4.0, 4.2, 4.4, and 6.67], three series of compound channel data [α = 2, 3, and 4] of Knight and Demetriou [11] along with the data of present compound channel [α = 3.67]. These compound channels have homogeneous roughness both in the main channel and floodplain sub-sections. Details of the experimental arrangements relating to phase A to C of the FCF work are obtained from Greenhill and Sellin [5]. Experiments for other three channels are described by Knight and Demetriou [11]. On the basis of total 62 overbank flow depth from nine different

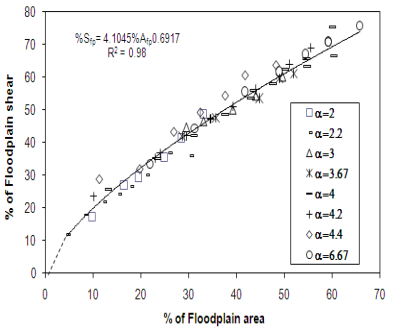


Figure 4: Variation between $\%S_{fp}$ and corresponding flow area $\%A_{fp}$.

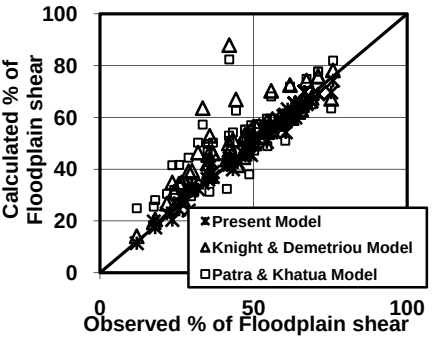


Figure 5: Variation of observed and modeled value of $\%S_{fp}$.

types of compound channels with width ratios α ranging from 2.0 to 6.67 and relative depths β ranging from 0.1 to 0.5, the authors have tried to plot the variation of percentages of floodplain area $\%A_{fp}$ verses $\%S_{fp}$ in Figure 4 given as.

$$\%\, S_{fp} = 4.1045\ \left(\%\, A_{fp}\right)^{0.6917} \tag{8}$$

By substituting $\frac{A_{fp}}{A} = \frac{\beta(\alpha-1)}{1+(\alpha-1)\beta}$ for a rectangular main channel, equation (8) is rewritten as

$$\%S_{fp} = 4.105\left[\frac{100\beta(\alpha-1)}{1+\beta(\alpha-1)}\right]^{0.6917} \tag{9}$$

Now, the equation (9) can be used for the channels having equal surface roughness in the floodplain and main channel. For non-homogeneous channels, equation (9) can further be modified as

$$\%S_{fp} = 4.105\left[\frac{100\,\beta(\alpha-1)}{1+\beta(\alpha-1)}\right]^{0.6917}\{1+1.02\sqrt{\beta}\,\log\,\gamma\} \tag{10}$$

Using the equation (10), the variation between the calculated $\%S_{fp}$ and observed values for all the ten types of compound channels are shown in Fig.5. In the same plot the variation of calculated $\%S_{fp}$ by previous investigators (i.e. equations 2 and 4) are also shown. The plot indicates high correlations (R^2 = 0.98) for(10) and R^2 = 0.68 and 0.74 respectively using equations (2) and (4).

4.2 Shear force on the assumed interface planes for an angle Θ

Further to account for the momentum transfer, the present experimental results are analysed using equation (10) and are plotted in Figure 6. The convention for momentum transfer is positive from the main channel to flood-plain and that from flood-plain to main channel is negative. As can be seen, the apparent shear

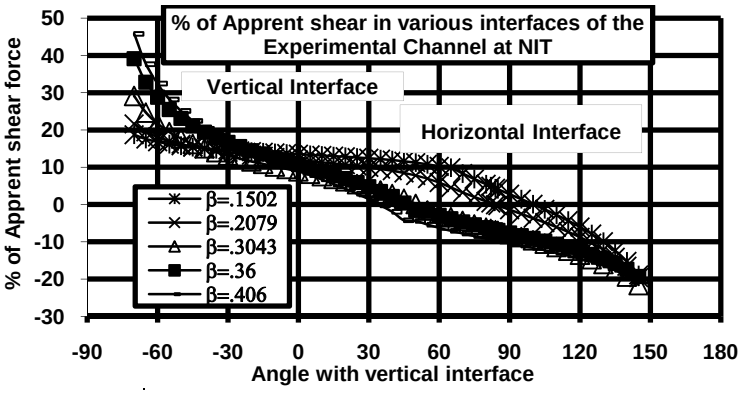


Figure 6: Variation of apparent shear along various interface planes.

in the vertical interfaces is found to be 13.5% of the total shear for the overbank flow depth of 2.12 cm ($\beta = 0.15$). It is found that the apparent shear decreases as the flow depth increases and reaches to 9.1 % for a overbank flow depth of 8.21 cm ($\beta = 0.406$). This shows that apparent shear is higher at low floodplain depths and gradually reduces as depth over floodplain increases. Similar results are also obtained for horizontal and diagonal interface plane for the present channels as well as when global data sets are tested using the concept. The interface plain of zero shear is located near the horizontal interfaces (approximately at $\theta = 99^o$) at $\beta = 0.15$ and for higher over-bank flow depths, the interface plane of zero shear is observed near a diagonal line of separation (approximately at $\theta = 40^o$).

4.3 Estimating discharge using different approaches

Let Q_c denotes the calculated discharge and Q_m the measured discharge. The percentage of error and standard error for each series of experimental runs are computed using the equation given as

$$Error\ (\%) = \frac{(Q_c - Q_m)}{Q_m} \times 100 \tag{11}$$

Using the vertical, horizontal, diagonal, and other interface planes, the error in discharge estimation for the experimental channel and one channel from the FCF (series A) are plotted in Figures 7(a) and 7(b) respectively.

In DCM, proper selection of the interface plane is required using the value of the apparent shear at the assumed interface plane. If the shear stress at this interface is zero, then there is no momentum transfer across the selected interface and therefore the length of the interface is not included to the wetted perimeter of the main channel or the floodplain for discharge calculation using divided channel method and Manning's equation. However due to the difficulty in locating this plane for all channel geometry and flow depths, investigators either include or exclude the interface lengths in calculating the wetted perimeter for the estimation of discharge. By including this interface length to the wetted

perimeter of the main channel, a shear drag of magnitude equal to the interface length times the average boundary shear is included. However, in such situations the actual interface shear is not considered as the shear along this interface is not equal to the boundary shear for all depths of flow in the channel. Similarly, by excluding these lengths, a zero shear along these interfaces is assumed.

Single channel method (curve SCM, where the whole compound section is considered as single one) is found to give higher discharge error for lower over-bank flow depths and very less error for high over-bank flow depth which is in line with the findings of Seçkin [23]. These also show that at very high overbank depths, the compound channel behaves like a single unit (Bhowmik and Demissie [2]). SCM also gives the maximum point error [e.g. for α = 6.67 discharge error is more than 45% (Figures 7b). Similarly, VDM-II (curve Vie) gives better discharge results than VDM-I (curve Vee) which is in line with the findings of Mohaghegh and Kouchakzadeh [15]. VDM-1 (curve Vee) provides higher error for compound channels of wider floodplain (e.g. α = 6.67 of Figure 7). For all the compound channels studied, the error results from HDM-I (curve Hee) is less than that from HDM-II (i.e curve Hie) which is in line with the findings of Seçkin [23]. It is again observed that HDM-I approach gives better discharge results than the corresponding vertical interface method (VDM-I) for low depths of flow over floodplain but gives large discharge error at higher depths. These findings are similar to the results of Mohaghegh and Kouchakzadeh [15]. It is also noticed that, DDM (curve Dee) gives less error (Figure 7) than all the VDM and HDM for all the compound channels studied. This finding follows the results of Wormleaton et al [28]; Knight and Hamed [12], Khatua [9] and Seçkin [23] etc. Both the area method (curve AM) and the variable inclined plain method (curve VI) gives higher standard error for

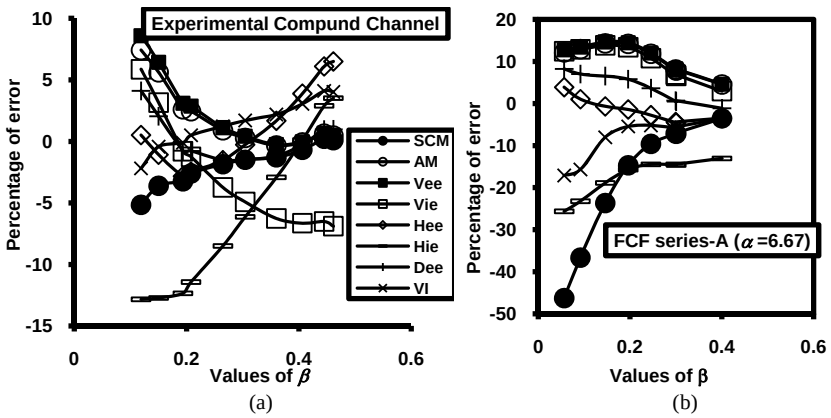


Figure 7: (a) and (b): Variation of percentage of error discharge with relative depth by various approaches for experimental channel and the FCF channel. [*SCM-Single Channel Method, Vee-*(*VDM-1*)*, Vie-*(*VDM-1I*)*, Hee-*(*HDM-1*)*, Hie-(HDM-1I), Dee-DDM, Ma-Zero Shear Interface Method (ZSIM), AM-Area Method, VI-Variable Inclined Plain Method*].

compound channels of wider floodplains (e.g FCF Series-A channels, Figure 7b). Basing on the present analysis, it can be concluded that both HDM-1 and DDM are good approaches. HDM-1 is better for low overbank depth and DDM is better for higher overbank depths.

5 Conclusions

The following conclusions can be drawn from the above investigation

- For a compound channel the important parameters affecting the boundary shear distribution are the relative depth (β), the width ratio (α), and the relative roughness (γ). These three dimensionless parameters are used to form equations to represent the total shear force percentage carried by floodplains. The present formulations for estimating the percentages of shear force carried by floodplain boundary $\%S_{fp}$ has the distinct improvement when compared to the previous investigators in the sense that it is quite adequate for all types of straight compound channel geometry (narrow as well as wide flood plain channels). Equations by the previous investigators give $\%S_{fp}$ more than 100 % when applied to a compound channel of wider floodplains (i.e. width ratio $\alpha > 10$).
- Using the boundary shear distribution results, the momentum transfer at different interfaces originating from the main channel and floodplain junction for all types of geometry are quantified. The proposed equation provides the estimation of apparent shear stress for any assumed interface in terms of an angle θ it makes with the vertical line. Furthermore the stage-discharge relationship of a compound channel using divided channel method is decided only after finding the apparent shear stress across the interface planes.
- Basing on the present analysis using DCM, it can be concluded that both HDM-1 and DDM are better than other approaches. HDM-1 is good for low overbank depths and the DDM subdivision is better for higher overbank depths. The adequacy of the developed equation for shear stress distribution along the boundary of compound channels is verified using the data from FCF-A channels.

References

[1] Ackers P, Hydraulic Design of Two Stage Channels, Proc. of the Inst. of Civil Engineers, Water Maritime and Energy, 96(4) (1992) 247–257.
[2] Bhowmik N.G., Demissie M, Carrying Capacity of Flood Plains, J Hydraul Eng, ASCE ,108(HY3) (1982) 443-453.
[3] Bousmar D, Zech Y, Momentum Transfer for Practical Flow Computation in Compound Channels, J Hydraul Eng, ASCE,125(7) (1999) 696–706.
[4] Chow V.T, Open Channel Hydraulics, McGraw Hill Book Co. Inc., New York, 1959.

[5] Greenhill R.K., Sellin R.H.J, Development of a Simple Method to Predict Discharge in Compound Meandering Channels. Proc. of the Inst. of Civil Engineers. Water, Maritime and Energy, 101 (1993) 37–44.
[6] Holden, Shear Stress and Discharge in Compound Channels, Thesis presented to the University of the Witwatersrand at Johannesburg for partial fulfillment of the requirement of the degree of M.Sc, 1986.
[7] Huthoff F.P.C, Roos D.C.M, Augustijn S.J.M, Hulscher H, Interacting Divided Channel Method for Compound Channel Flow, J Hydraul Eng, ASCE, 134(8) (2008) 1158-1165.
[8] Kejun Yang, Shuyou Cao, Xingnian Liu, Flow Resistance and Its Prediction Methods in Compound Channels. Springer Verlag, 2007.
[9] Khatua K.K, Interaction of Flow and Estimation of Discharge in Two Stage Meandering Compound Channels. Thesis Presented to the National Institute of Technology, Rourkela, in partial fulfillments of the requirements for the Degree of Doctor of Philosophy, 2008.
[10] Khatua K.K, Patra K.C., Boundary Shear Stress Distribution in Compound Open Channel Flow, J Hydraul Eng, ISH,12(3) (2007) 39-55.
[11] Knight D.W, Demetriou J.D, Floodplain and Main Channel Flow Interaction. J Hydraul Eng., ASCE, 109(8) (1983) 1073-1092.
[12] Knight D.W, Hamed M.E, Boundary Shear in Symmetrical Compound Channels. J of Hydraul Eng, ASCE, 110 (10) 1984 1412-1430.
[13] Knight D.W, Shiono K, River Channel and Floodplain Hydraulics, Floodplain Processes, by Anderson M.G., Walling D E., Paul D, 1996.
[14] Lambert M.F, Myers W.R, Estimating the Discharge Capacity in Straight Compound Channels, Proceedings of the Institute of Civil Enggs, Water, Maritime and Energy, 130 (1998) 84-94.
[15] Mohaghegh A, Kouchakzadeh S, Evaluation of Stage-Discharge Relation in Compound Channels, J Hydrodynamics, 20(1) (2008) 81-87.
[16] Myers W.R.C, Elsawy. Boundary Shear in Channel With Floodplain, J Hydraul Div, ASCE, 100(7) (1975) 933-946.
[17] Patel V.C, Calibration of the Preston tube and limitations on its use in pressure gradients. J Fluid Mech, 23(1) (1965) 185-208.
[18] Patra K. C, Kar S.K, Flow interaction of Meandering River with Flood plains. Journal of Hydraulic Engineering, ASCE, 126(8) (2000) 593-603.
[19] Patra K. C, Kar S.K, Bhattacharya A.K, Flow and Velocity Distribution in Meandering Compound Channels. J Hydraul Eng, ASCE, 130(5) (2004) 398-411.
[20] Patra K.C, Khatua K.K, Selection of Interface Plane in the Assessment of Discharge in Two Stage Meandering and Straight Compound Channels, Proc. Of Inter. Conf. on Fluvial Hydraul, Lisbon, (2006) 379-387.
[21] Prinnos P, Townsend R.D ,Comparison of Methods of Predicting Discharge in Compound Open Channels, Adv. Water Resour ,7(12) (1984)180–187.
[22] Rajaratnam N, Ahmadi R.M, Interaction between Main Channel and Flood Plain Flows. J Hydraul Div, ASCE, 105(HY5) (1979) 573-588.

[23] Seçkin G, A comparison of one-dimensional methods for estimating discharge capacity of straight compound channels. Canadian J Civil Eng, 31(4) (2004) 619–631.
[24] Sellin R.H.J, A Laboratory Investigation into the Interaction between Flow in the Channel of a River and that of its Flood Plain. La. Houilte Blanche, 7 (1964) 793-801.
[25] Shiono K, Knight, D. W. Turbulent Open Channel Flows with Variable Depth across the Channel Fluid Mech, 222 (1991) 617–646.
[26] Toebes G. H, Sooky A, Hydraulics of Meandering Rivers with Floodplains. J Wtrwy Harbor Div, ASCE, 93(2) (1967) 213–236.
[27] Weber J. F, Menéndez A. N, Performance of Lateral Velocity Distribution Models for Compound Channel Sections, Proc Int. Conf. on Fluvial Hydraul, Balkema, Rotterdam, the Netherlands, (2004) 449–457.
[28] Wormleaton P.R, Allen J, Hadjipanos, P, Discharge Assessment in Compound Channel, J Hyd. Div, ASCE, 108(HY9) (1982) 975-994.
[29] Yen, C.L, Overton D.E, Shape Effects on Resistance in Floodplain Channels, J Hydraul Div, ASCE, 99(1) (1973) 219-238.
[30] Zheleznyakov G.V, Relative deficit of mean velocity of instable river flow: Kinematic effect in river beds with floodplains. Proc. Inter. Conf. IAHR, (1965).

Wicking process in a capillary tube: a new zero-order asymptotic solution

O. Bautista[1], F. Méndez[2] & E. Bautista[1]
[1]*SEPI-ESIME Azcapotzalco. Instituto Politécnico Nacional, Mexico*
[2]*Facultad de Ingeniería, UNAM, Mexico*

Abstract

In this work, we have theoretically re-visited the capillary rise process into a circular tube for very short time scales, retaining in this manner, the physical influence of the inertial effects. We use the boundary-layer technique or matched asymptotic expansion procedure in order to treat this singular problem by identifying two appropriate time scales: one short time scale related with inertial effects, σ , and the other, τ , the large scale which is basically associated with imbibition effects. Considering that the well-known Washburn's law was derived by neglecting the inertial effects, the corresponding solution has a singular behavior for short times, which is reflected by an infinite mass flow rate. Then, for this purpose we derive a zero-order solution which is enough to avoid the singular behavior of the solution. In this manner, the Washburn's solution represents only the external solution only valid for the large time scale τ . The above analytical result is compared with a numerical solution including the case when the contact angle between the meniscus and the inner surface of the capillary tube becomes a dynamic contact angle. On the other hand, the presence of inertial effects can induce oscillations of the imbibition front which are controlled by the dynamic contact angle. Therefore, in the present work we predict a global asymptotic formula for the temporal evolution of the height of the liquid. In order to show the importance of the inertial terms, we present this evolution for different values of the dimensionless parameters involved in the analysis.

Keywords: wicking process, inertial effects, singular perturbation, matched asymptotic expansions.

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110211

1 Introduction

In recent years, the phenomenon of capillary wicking has strongly stimulated theoretical studies together with experimental evidences to show some peculiar aspects of these complex processes. In real situations, the wetting of a surface is controlled by rates of spreading of a liquid over the substrate and in general, this effect is devoted for very short times before to reach the well-known equilibrium thermodynamics Young's equation, where the surface tension force is exactly balanced with the gravity force. In addition, the movement of the contact line of the liquid front depends strongly on molecular kinetic of the dynamic contact angle. The existing theories and experimental results about the position and velocity of the contact line are not well understood yet. For instance, relevant studies of the spreading of a drop over horizontal and inclined flat plates have been developed to clarify that, in some cases; the macroscopic contact line can be preceded by a precursor film, where the van der Waals forces are not negligible. This idea was originally proposed by De Gennes [1]. Nowadays, an acceptable point of view to treat the dynamic of the contact angle is to include molecular forces, like van der Walls forces, improving in this manner, the hydrodynamics macroscopic models. In this direction, Treviño et al. [2, 3] carried out a theoretical analysis to predict the influence of the precursor film on the dynamics of an axisymmetric drop spreading over a horizontal surface. The state of the art can be found in the book of Middleman [4], where relevant topics and applications are conducted to illustrate different wicking phenomena.

Since the pioneer work of Washburn [5], several mathematical models have been proposed to analyze those cases where the capillary forces always have a predominant effect. In this direction, the classical works of Joos et al. [6] and Batten [7] show rigorously the main forces that act on a liquid rising up a capillary tube. Hamraoui et al. [8] using high-speed imaging technique and solving a particular Washburn–Rideal–Lucas equation based on a fully developed flow, showed the physical influence of the dynamic contact angle on the wicking process. In this study, the authors postulated a fundamental relationship between the dynamic contact angle and the rate of the liquid rise within the capillary tube and the mathematical solution was validated through the experimental results. A similar study was reported by Hamraoui and Nylander [9] including numerical predictions of oscillations for the imbibition front. In reality, these oscillations were previously reported by Quéré [10] if the liquid viscosity is low enough.

In this paper, following the proposed analytical models by Hamraoui et al. [8] and Duarte et al. [11], we present an asymptotic analysis of zero-order to characterize the initial step of the wicking penetration process into a capillary circular tube, for the case when the inertial terms are important. The fundamental idea is to improve the existing theoretical predictions reported by previous schemes and related with the velocity of the front, using singular perturbation techniques. We anticipate that a singular behavior prevails if, for example, inertial and gravity forces are neglected. In this case, the most simple Washburn's law shows that the penetration velocity is of order of

$U\left(=dh/dt\right)\sim t^{-1/2}$ (see Ref. [4]) and for values of $t\to 0$, the velocity U is undetermined. In order to avoid this singularity, which reflects, in addition, an infinite mass flow rate into the capillary tube, we use two time scales to identify the dominant forces. For each time scale, we derive an appropriate zero-order governing equation and the corresponding solutions are matched according a well pre-established singular perturbation technique. Recognizing the importance of the inertial terms for the initial step of the wicking process, we anticipate that is only sufficient to derive a zero-order global solution, correcting up to this order the singular behavior of the Washburn's law. Therefore, in this first part of the present work the associated corrections with higher-order terms are omitted for simplicity.

2 Theoretical analysis

The present model corresponds to eqn. (1a) from Hamraoui et al. [8], and the main forces that act on a liquid rising up a capillary tube are due to surface tension, gravity, viscosity and inertia, respectively:

$$2\pi R\gamma\left(1-\frac{\beta}{\gamma}\frac{dh}{dt}\right)=\pi R^2\rho gh+8\pi\mu h\frac{dh}{dt}+\pi R^2\rho\frac{d}{dt}\left(h\frac{dh}{dt}\right) \tag{1}$$

where R is the radius of the capillary, β is a constant related with the dynamic contact angle, $h\left(t\right)$ is the height of the liquid at time t. γ, ρ, μ are the surface tension, density and dynamic viscosity of the liquid, respectively, and g is the acceleration due to gravity force. In order to derive the above equation were basically neglected the entrance hydrodynamic effects and the liquid was assumed Newtonian, then we can write the average velocity of liquid rising at the capillary from Poiseuille's law as $U=dh/dt=R^2\Delta P/8\mu h$. In addition, we adopt the relationship between dynamic contact angle, θ_d and the rate dh/dt given by $\cos\theta_d(t)=1-\left(\beta/\gamma\right)\left(dh/dt\right)$. The details of the above considerations can be found elsewhere, [8]. In this form, the present forces included in eqn. (1) are expressed in terms of the unknown height $h(t)$. This scheme has widely been used in lubrication theory to analyze the fluid flow in thin-liquid films (Oron et al. [12]). The above non-linear ordinary differential equation must satisfy two initial conditions. Traditionally, the majority of the published works only include the first of these initial conditions. The reason is based on neglecting the inertia terms. However, for very short times, the inertial terms must be included, even more for those cases of large radius. Therefore, we propose the following initial conditions:

$$t=0:\ h=0 \text{ and } \frac{dh}{dt}=Q. \tag{2}$$

In the last initial condition, Q is the unknown volumetric flow rate per unit area at the base of the capillary tube and must be determined as a part of the solution. However, in this manner the system of eqns. (1)–(2) is incomplete because Q is unknown. Therefore, it is necessary to add other restriction. This other condition comes from the equilibrium condition of the imbibition front. The above means that there is an equilibrium time, t_{eq}, for which we reach the equilibrium height h_{eq} given by $2\gamma/\rho gR$. This last relationship is readily derived from the balance between the surface tension and gravity forces. Obviously, the equilibrium time t_{eq} must then be determined as a part of the solution. From this point of view, therefore, the additional condition can be written as

$$t = t_{eq} : h = h_{eq} = \frac{2\gamma}{\rho gR} \tag{3}$$

Now, for solving eqn. (1), we use appropriate dimensionless variables taking into account that, in a first approximation, the characteristic time t_c of the wicking penetration process is determined by a balance between the surface tension and viscosity forces. This order relationship can be written as $2\pi R\gamma \sim 8\pi\mu h_{eq}^2 / t_c$. Introducing the equilibrium height into the above order relationship, we obtain that $t_c \sim 16\mu\gamma / \rho^2 g^2 R^3$. We assume $t_c \sim t_{eq}$. Therefore, introducing the following dimensionless variables:

$$\tau = \frac{t}{t_c},\ Y = \frac{h}{h_e} \tag{4}$$

the system of Equations (1)–(3) are transformed to:

$$\left(1-\alpha\frac{dY}{d\tau}\right) = Y + Y\frac{dY}{d\tau} + \varepsilon\frac{d}{d\tau}\left(Y\frac{dY}{d\tau}\right), \tag{5}$$

with

$$\tau = 0\ :\ Y = 0\ ,\ \frac{dY}{d\tau} = \lambda\ \ ;\ \ \tau = \tau_{eq}\ :\ Y = 1 \tag{6}$$

and the parameters α, λ, and ε are defined as:

$$\alpha = \frac{\beta\rho gR^2}{8\mu\gamma}, \lambda = \frac{8\mu Q}{\rho gR^2} \text{ and } \varepsilon = \frac{\rho gR^2}{128\gamma}\frac{gR^3\rho^2}{\mu^2} = \frac{Bo\ Ga}{128}, \tag{7}$$

where *Bo* and *Ga* represent the Bond and Galileo numbers, respectively. Therefore, the solution of the problem (5)–(6) shall provide $Y = Y\left(\tau, \tau_{eq}, \alpha, \lambda, \varepsilon\right)$. In the above relationship τ_{eq} is defined as $\rho^2 g^2 R^3 t_{eq} / 16\mu\gamma$ with t_{eq} unknown. Therefore, the set of eqns. (5)–(6) represent an eigenvalue problem because the time t_{eq} is unknown. In this manner, the solutions are constructed by taking into account that for arbitrary values of the dimensionless parameters α, λ and ε, we must obtain a unique solution of $Y = Y\left(\tau, \tau_{eq}, \alpha, \lambda, \varepsilon\right)$ for each value of the equilibrium time τ_{eq}. In the remainder of this paper we analyze and classify the solutions according to the assumed values of α, taking advantage of the fact that in general, α and ε are very small compared with unity. We anticipate that the values of λ are irrelevant for the zero-order solution of Y.

2.1 Asymptotic limit of $\alpha \ll 1$ (pre-wetting surface)

As we are concerned, a first approximation, to clarify the role of the inertial terms we have omit in the present section, the physical influence of the dynamic contact angle θ_d. This particular situation can occur, for instance, when the internal surface in contact with the liquid has been previously wetted. Thus, we have that for $\alpha \ll 1$ the dynamic contact angle, θ_d, practically coincides with of the equilibrium or static angle, θ_{eq}. For simplicity, we take $\theta_{eq} \sim 0$, which is a good approximation for wetted surfaces. However, we emphasize that this particular restriction does not significantly affect the general validity of the present analysis. Indeed, for $\theta_{eq} \neq 0$ the first term of the right-hand side of eqn. (5) can be easily modified by a constant factor given by $\cos\left(\theta_e\right)$. Thus, eqn. (5) reduces to:

$$1 = Y + Y\frac{dY}{d\tau} + \varepsilon\frac{d}{d\tau}\left(Y\frac{dY}{d\tau}\right). \tag{8}$$

The numerical values for ε are generally small and therefore, the above equation dictates that the inertial terms are only important for a time scale of the order of $\tau \sim \varepsilon$. Otherwise, for $\tau \sim 1$, the inertial term represented by the last term of the right-hand side of the above equation is negligible, in a first approximation. Thus, we can introduce two time scales to study the problem.

2.1.1 Formulation for large times ($\tau \sim 1$)

For this relevant limit, we propose the following expansion

$$Y = Y_0 + \varepsilon Y_1 + O\left(\varepsilon^2\right), \tag{9}$$

and substituting into eqn. (5), by retaining only terms of zero order, we obtain that

$$1 = Y_0 + Y_0 \frac{dY_0}{d\tau}, \tag{10}$$

and the solution is readily derived and given by (undetermined only by the constant C_0):

$$1 - Y_0 - Ln\left(1 - Y_0\right) = \tau + C_0. \tag{11}$$

In principle, this solution can not satisfy the initial condition $Y(\tau = 0) = 0$, because the inertial terms were neglected. Therefore, in order to evaluate the constant C_0, we must retain the inertial terms. We show that is necessary to use a small time scale of order $\tau \sim \varepsilon$ consider the influence of the inertial effects.

2.1.2 Formulation for short times ($\tau \sim \varepsilon$)

In this case, we accept that for short times the dominant terms of eqn. (8) are basically controlled by the surface tension and inertial forces. In this manner, the suitable scales to analyze this initial step are $\sigma = \tau / \varepsilon \sim O(1)$ and $y(\sigma) = Y / \sqrt{\varepsilon} \sim O(1)$ and eqn. (8) is transformed to:

$$\frac{d}{d\sigma}\left(y \frac{dy}{d\sigma} \right) + y \frac{dy}{d\sigma} + \varepsilon^{1/2} y = 1. \tag{12}$$

Now, in order to find the zero-order solution of this equation, we propose the following expansion:

$$y = y_0 + O(\varepsilon) \tag{13}$$

and the leading order equation is governed by a balance between the inertia, viscosity and surface tension forces, given by

$$\frac{d}{d\sigma}\left(y_0 \frac{dy_0}{d\sigma} \right) + y_0 \frac{dy_0}{d\sigma} = 1, \tag{14}$$

with the initial conditions:

$$\sigma = 0 : y_0 = 0 \text{ and } \frac{dy_0}{d\sigma} = \sqrt{\varepsilon}\lambda \sim O(1). \tag{15}$$

The non-linear differential equation (14) of second order can be easily solved because admits a first integral. The resulting first order non-linear equation is solved by using an appropriate integrating factor; for simplicity the details are

omitted. Following the above procedure and applying the initial conditions (15), we obtain:

$$y_0 = \sqrt{2\sigma + 2\left[\exp\left(-\sigma\right) - 1\right]}. \tag{16}$$

It is very important to note that the asymptotic solution for short times $\sigma \to 0$ is $y_0 \sim \sigma\left(1 - \sigma/6\right) + \ldots$. The same asymptotic formula was previously derived by Quéré [10] using other formulation for modeling the initial steps, where predominate inertial and surface tension effects. Thus, for short times the rising of the capillary wicking has a linear behavior, which avoids that the velocity displays a singular behavior. In addition, the above asymptotic relationship is independent of the dimensionless mass rate λ.

Retaining then the inertial terms in order to obtain a global solution valid for all time, we can use the well-known asymptotic matching technique, which can be found elsewhere, (Van Dyke, [13]). In this case, we construct a global solution given by,

$$Y = Y_0\left(\tau\right) + \sqrt{\varepsilon}\, y_0\left(\sigma\right) - Y_{match}, \tag{17}$$

where Y_{match} represents the intermediate solution valid in the matching region $\varepsilon < \tau < 1$. In this zone, the solutions have a similar behavior and can be calculated by using the following matching asymptotic principle [13],

$$Y_{match} \sim \lim_{\tau \to 0} Y\left(\tau\right) \sim \lim_{\sigma \to \infty} \sqrt{\varepsilon}\, y\left(\sigma\right). \tag{18}$$

Thus, applying this condition to the solutions (11) and (16) and expanding adequately both solutions, can be easily shown that the constant $C_0 = 1$. In this form, the zero-order global solution is based effectively on an intermediate region $\varepsilon < \tau < 1$, where the short and large times solutions are equivalent. This matching solution is given as $Y_{match} = \sqrt{2\tau}$ and the global solution with the aid of the above comments is:

$$Y = \sqrt{2\tau + 2\varepsilon\left[\exp\left(-\tau/\varepsilon\right) - 1\right]} + Y_0 - \sqrt{2\tau}, \tag{19}$$

where Y_0 is the implicit solution of:

$$Y_0 + Ln\left(1 - Y_0\right) = -\tau. \tag{20}$$

The above transcendental equation for $Y_0\left(\tau\right)$ has an approximate analytical solution. We follow the approach presented elsewhere (Fries and Dreyer [14]). The solution for eqn. (20) is given by

$$Y_0 = 1 + W\left(-e^{-1-\tau}\right) \tag{21}$$

where W is the Lambert's function. Finally, substituting eqn. (21) into eqn. (19), we obtain that

$$Y = \sqrt{2\tau + 2\varepsilon\left[\exp\left(-\tau/\varepsilon\right) - 1\right]} + 1 + W\left(-e^{-(1+\tau)}\right) - \sqrt{2\tau}\,, \tag{22}$$

which is a global analytic solution for $Y(\tau)$. It is really important to note that in order to calculate the numerical value of the Lambert function W that appears on eqns. (21)-(22), we use an approximated function previously developed [14], with a maximum relative error of 0.1% and given by the relationship

$$W\left(-e^{-(1+\tau)}\right) \sim -1 + \frac{\sqrt{2(1-e^{-\tau})}}{1 + \dfrac{4.13501\sqrt{2(1-e^{-\tau})}}{12.7036 + \sqrt{2(1-e^{-\tau})}}}, \tag{23}$$

valid for $-e^{-1} \le -e^{-1-\tau} \le 0$, where e is the well-known Euler's number given by $e = 2.718281...$

2.2 Numerical scheme

In this subsection, we present some details related with the numerical procedure to complete the solutions of eqns. (5)–(6) for values of $\alpha \neq 0$. We have included these numerical estimations to compare with the asymptotic solution of zero-order. The class of governing equation given by eqn. (5) can be readily integrated by the classical Runge-Kutta method of fourth-order. In our case, we can define the following variables

$$Y = \varphi_1 \quad \text{and} \quad \frac{dY}{d\tau} = \varphi_2\,, \tag{24}$$

and the second-order nonlinear differential equation (5) with the aid of the above relationships can be transformed to system of two first-order equations given by

$$\begin{aligned} \frac{d\varphi_1}{d\tau} &= \varphi_2 \\ \frac{d\varphi_2}{d\tau} &= \frac{1 - \varphi_1 - (\alpha + \varphi_1)\varphi_2 - \varepsilon\varphi_2^2}{\varepsilon\varphi_1}\,. \end{aligned} \tag{25}$$

It is well-known that this method requires initial conditions to begin the first iterations. In our case and from eqn. (6), the part that only corresponds to the initial conditions is provided by $Y(0) = 0$ and $dY(0)/d\tau = \lambda$, which can be

rewritten as $\varphi_1(0)=0$ and $\varphi_2(0)=\lambda$. The use of the above initial conditions yields a divergent behavior for both functions φ_1 and φ_2. Therefore, the numerical procedure to integrate eqns. (25) is to replace the above initial conditions by the asymptotic relationships for short times derived in section 2.1.2. In terms of the functions φ_1 and φ_2, the initial conditions are the following

$$\varphi_1(0)\sim\varphi_1(\Delta\tau\to 0)=\frac{\Delta\tau}{\varepsilon^{1/2}}\left(1-\frac{\Delta\tau}{\varepsilon}\right)+O\left[\frac{(\Delta\tau)^3}{\varepsilon^{5/2}}\right] \tag{26}$$

and

$$\varphi_2(0)\sim\varphi_2(\Delta\tau\to 0)=\frac{1}{\varepsilon^{1/2}}\left(1-\frac{\Delta\tau}{3\varepsilon}\right)+O\left[\frac{(\Delta\tau)^2}{\varepsilon^{5/2}}\right]. \tag{27}$$

We can appreciate that for finite and small values of the time increment, $\Delta\tau$, the initial condition (26) remains always different to zero and the other initial condition, given by eqn. (27), depends only the parameter ε, which assumes finite values in our numerical essays. In the present estimations, we use $\Delta\tau=0.00001$ and different values of the parameter ε. In this manner the divergence is eliminated numerically.

3 Results and conclusions

In this paper we developed a singular asymptotic analysis to describe the initial steps of the temporal evolution of wicking penetration of a fluid into a capillary tube. The use of two time scales permits us to conduct a boundary layer analysis to clarify the physical role of the inertial terms. In such case, the singularity of the front velocity is eliminated. On the other hand, the present analysis serve to emphasize that the inertia terms are also important not only for a short time scale as we comment from the analytical results, lines below. In this direction, we see that for different substances the inertia effect has a relevant influence even for large times. This occurs as a consequence of the dimensionless parameters α and ε. In Fig. 1, we have plotted the dimensionless height of the wicking front as a function of the dimensionless time for the case of $\alpha=0$ and two different values of the parameter ε; *i. e.*, $\varepsilon=0$ and $\varepsilon=0.05$, that correspond to the inertial and non-inertial cases, respectively. We see the influence of the inertial terms is to decrease the temporal growth of the liquid's front for small values of the dimensionless parameter ε. Therefore, the asymptotic analysis confirms that the inertial terms always modulate the initial steps of the wicking penetration process. This is an acceptable result always that we do not taken into account time scales smaller than ε, because in this case, capillary penetration can be controlled by hydrodynamic instabilities and more complex phenomena related with the influence of precursor film, chemical

adsorption activity in the inner surface of capillary tube, intermolecular forces, etc. In addition, the asymptotic solution dictated by eqn. (22) confirms that the usage of the singular perturbation analysis, the zero-order solution is enough to describe the main physical characteristics of the problem.

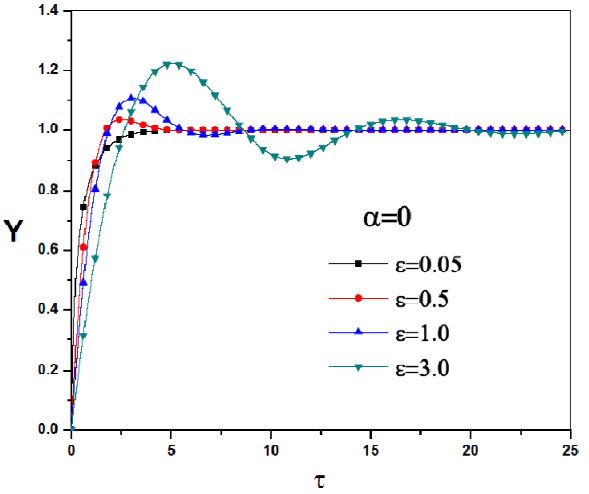


Figure 1: Evolution of capillary rise, for $\alpha = 0$.

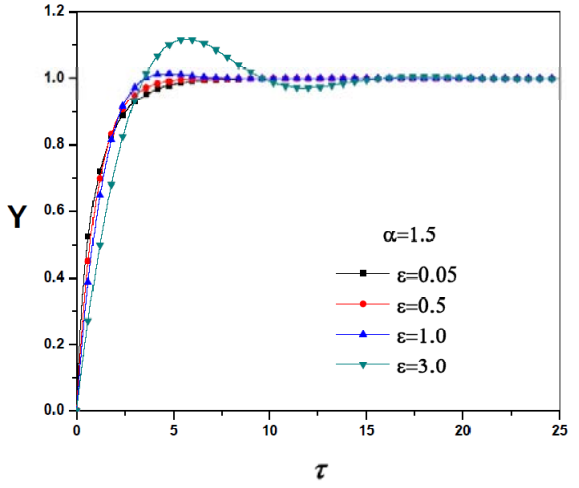


Figure 2: Evolution of capillary rise, for $\alpha = 1.5$.

In Fig. 2, we have shown the combined influence of dimensionless parameters α and ε. For instance, in this figure, we use four different values of the parameter ε $(= 0.05, 0.5, 1.0, 3.0)$ and α =1.5 to plot the dimensionless

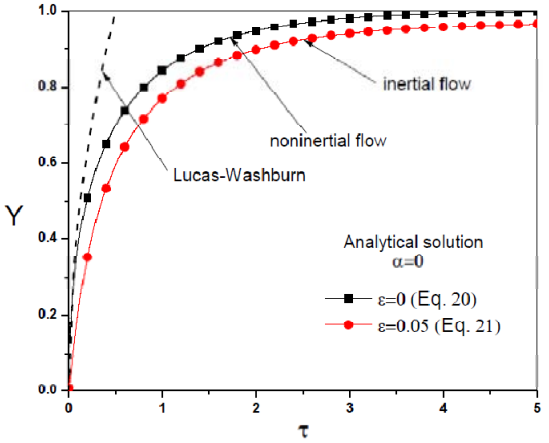


Figure 3: Comparison between the Lucas-Washburn's formula, the present asymptotic zero-order solution, Eq. (22), and the non-inertial solution $(\varepsilon = 0)$.

height of the imbibition front as a function of the dimensionless time. Clearly, parameter α works as a dissipative effect because causes a slower wicking process. Therefore, the well-known oscillating pattern previously reported by different authors [10, 11] is modulated by the influence of the parameter α. Thus, for increasing values of this parameter is easier to reach the equilibrium height of the imbibition. In this sense, the limiting factor above is controlled by the parameter ε. The Fig. 3 shows the difference between the Washburn´s law and the obtained in the present work, showing that the evolution of the imbibition front is strongly influenced by considering the inertial effects in the dynamic equation describing the capillary process.

Acknowledgements

This work has been supported by the research Grant No. 58817 by Consejo Nacional de Ciencia y Tecnología anf 201000375 of Instituto Politécnico Nacional at Mexico.

References

[1] De Gennes, P. G., Wetting: statics and dynamics**.** *Rev. Mod. Phys. Rev*., (**57**), pp. 827- 863, 1985.

[2] Treviño, C., Ferro-Fontán, C., Méndez, F., Asymptotic analysis of axisymmetric drop spreading**.** *Phys. Rev. E.*, **(58)**, pp. 4478-4484, 1998.

[3] Treviño, C., Méndez, F., Ferro-Fontán, C., Influence of the aspect ratio of a drop in the spreading process over a horizontal surface**.** *Phys Rev. E.*, **(58)**, pp. 4473-4477, 1998.

[4] Middleman, S., 1995. Capillary penetration dynamics (Chapter 8). *Modeling Axisymmetric Flows*, Academic Press, Inc., New York, pp. 211-239, 1995.
[5] Washburn, E. W., The dynamics of capillary flow. *Phys. Rev.,* **(17)**, pp. 273-283. 1921.
[6] Joos, P., van Remoortere, P., Bracke, M., The kinetics of wetting in a capillary. *J. Coll. Int. Sci.,* **(136)**, pp. 189-197, 1990.
[7] Batten Jr., G. L., Liquid imbibition in capillaries and packed beds. *J. Coll. Int. Sci.,* **(102)**, pp. 513-518, 1984.
[8] Hamraoui, A., Thuresson, K., Nylander T., Yaminsky, V., Can a dynamic contact angle be understood in terms of a friction coefficient? *J. Coll. Int. Sci.,* **(226)**, pp. 199-204, 2000.
[9] Hamraoui, A., Nylander, T., Analytical approach for the Lucas–Washburn equation. *J. Coll. Int. Sci.,* **(250)**, pp. 415-421, 2002.
[10] Quéré, D., Inertial capillarity. *Europhys. Lett.,* **(39)**, pp. 533-538, 1997.
[11] Duarte, A. A., Strier, D. E., Zanette, D. H., The rise of a liquid in a capillary tube revisited: A hydrodynamical approach**.** *Am. J. Phys.,* **(64)**, pp. 413-418, 1996.
[12] Oron, A., Davis S. H., Bankoff, S. G., Long-scale evolution of thin liquid films**.** *Rev. Mod. Phys.,* **(69)**, pp. 931-980, 1997.
[13] Van Dyke, M., *Perturbation Methods in Fluid Mechanics,* second ed. Academic Press, Inc., New York, 1964.
[14] Fries, N., Dreyer, M., An analytic solution of capillary rise restrained by gravity. *J. Coll. Int. Sci*., **(320)**, pp. 259-263, 2008.
[15] Corless, R. M., Gonnet, G. H., Hare, D. E. G., Jeffrey D. J., Knuth, D. E., On the Lambert *W* function. *Adv. Comp. Math*. (**5**), pp. 329-359, 1996.

Section 3
Heat transfer and thermal processes

This page intentionally left blank

Experimental validation of analytical solutions for a transient heat conduction problem

N. Simões[1], I. Simões[1], A. Tadeu[1], C. Vasconcellos[2] & W. Mansur[3]
[1]*CICC, Department of Civil Engineering,*
Faculty of Sciences and Technology, University of Coimbra, Portugal
[2]*Department of Fortification and Construction Engineering,*
Military Institute of Engineering, Brazil
[3]*Department of Civil Engineering,*
COPPE/Federal University of Rio de Janeiro, Brazil

Abstract

This paper analyses heat transfer across multilayer systems when boundary conditions are unsteady. The results of analytical simulations and experimental tests were compared in order to validate the analytical formulation. The formulation that is proposed to solve this problem uses Green's functions to handle the conduction phenomena. The Green's functions are established by imposing the continuity of temperatures and heat fluxes at the interfaces of the various layers. The technique used to deal with the unsteady state conditions consists of first computing the solution in the frequency domain (after the application of time and spatial Fourier transforms along the two horizontal directions), and then applying (fast) inverse Fourier transforms into space-time. The thermal properties of the multilayer system materials have been previously defined experimentally.

For the experimental measurements the multilayer system was mounted on a guarded hotplate capable of imposing a controlled heat variation at the top and bottom boundaries of the system. Temperatures were recorded using a thermocouple set connected to a data logger system. Comparison of the results showed that the analytical solutions agree with the experimental ones.
Keywords: experimental validation, transient heat conduction, Green's functions formulation, frequency domain.

WIT Transactions on Modelling and Simulation, Vol 51,
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110221

1 Introduction

A dwelling's interior comfort is a fundamental issue in building physics and it depends on the building's envelope. In order to better evaluate the thermal performance of the construction elements used throughout the building envelope, more accurate models must be developed. Thermal behaviour depends largely on unsteady state conditions, and so the formulations for studying those systems should take the transient heat phenomena into consideration.

Most schemes devised to solve transient diffusion heat problems have either been formulated in the time domain (time-marching approach) (e.g. Chang *et al.* [1]) or else use Laplace transforms (e.g. Rizzo and Shippy [2]). An alternative approach is to apply a Fourier transform to deal with the time variable of the diffusion equation, thereby establishing a frequency domain technique, and then obtain time solutions are obtained by using inverse Fourier transforms into space-time (e.g. Tadeu and Simões [3]).

In general, multilayer systems, built by overlapping different layers of materials, are used to ensure that several functional building requirements, such as hygrothermal and acoustic comfort, are met. One of the requirements is to get high thermal performance and thus reduce energy consumption and promote building sustainability. The importance of multilayer solutions has motivated some researchers to try and understand the heat transfer in those systems (e.g. Kaşka and Yumrutaş [4], Chen et al. [5] and Sami A. Al-Sanea [6]).

In this paper is presented an experimental validation of a semi-analytical Green's functions solution that simulates heat conduction through multilayer systems when they are subjected to heat generated by transient sources. The proposed semi-analytical solutions allow the heat field inside a layered medium to be computed, without having to discretize the interior domain. The problem is formulated in the frequency domain using time Fourier transforms. The technique requires knowing the Green's functions for the case of a spatially sinusoidal, harmonic heat line source placed in an unbounded medium. The Green's functions for a layered formation are formulated as the sum of the heat source terms equal to those in the full-space and the surface terms required to satisfy the boundary conditions at the interfaces, i.e. continuity of temperatures and normal fluxes between layers. The total heat field is found by adding the heat source terms equal to those in the unbounded space to the set of surface terms arising within each layer and at each interface (e.g. Tadeu and Simões [3]).

The experimental results were obtained for several systems built by overlapping different materials. These test specimens were subjected to a transient heat flow produced by cooling and heating units which established a heat flow rate that could reach a pre-programmed mean test temperature in the specimen. The temperature changes in the different specimen layers were recorded by a thermocouple data logger system. The thermal properties of the different materials, such as thermal conductivity, mass density and specific heat were obtained experimentally. The temperature variation in the top and bottom surfaces of the multilayer system was used as an input for the semi-analytical model designed using the thermal properties obtained experimentally. This

paper first formulates the three-dimensional problem and presents the Green's function in the frequency domain for a heat point source applied to a multilayer formation. A brief description of the mathematical manipulation follows, and the experimental setup is then described. Some final remarks are presented after the experimental measurements have been compared with computational results.

2 Problem formulation

Consider a system built from a set of *m* plane layers of infinite extent bounded by two flat, semi-infinite media, as shown in Figure 1. The top semi-infinite medium is called medium 0, and the bottom semi-infinite medium is assumed to be *m*+1. The thermal material properties and thickness of the various layers may differ. This system is subjected to a point heat source somewhere in the domain.

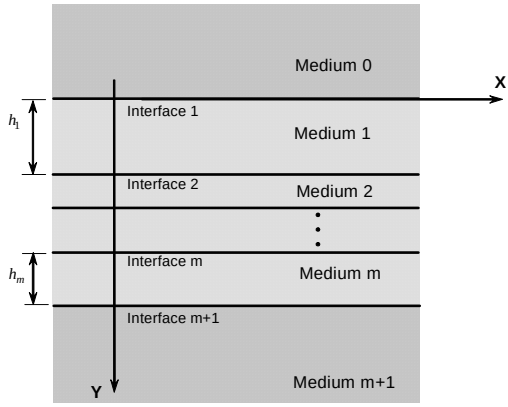


Figure 1: Multilayer system bounded by two semi-infinite media.

The transient heat transfer by conduction in each layer is expressed by the equation

$$k_j\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)T(t, x, y, z)=\rho_j\, c_j\,\frac{\partial T(t, x, y, z)}{\partial t}, \tag{1}$$

in which t is time, $T(t, x, y, z)$ is temperature, j identifies the layer number, k_j is the thermal conductivity, ρ_j is the mass density and c_j is the specific heat.

3 Semi-analytical solutions

The solution is defined in the frequency domain as the superposition of plane heat sources. This is done after applying a Fourier transform in the time domain

and a double Fourier transformation in the space domain along the x and z directions.

Applying a Fourier transformation in the time domain to eqn (1) gives equation

$$\left(\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)+\left(\sqrt{\frac{-\mathrm{i}\omega}{K_j}}\right)^2\right)\hat{T}(\omega,x,y,z)=0, \tag{2}$$

where $\mathrm{i}=\sqrt{-1}$, $K_j=k_j/\left(\rho_j\, c_j\right)$ is the thermal diffusivity of the layer j, and ω is the frequency. For a heat point source applied at $(x_0,y_0,0)$ in an unbounded medium, of the form $T_{inc}(\omega,x,y,z,t)=\delta(x-x_0)\delta(y-y_0)\delta(z)e^{\mathrm{i}(\omega t)}$, where $\delta(x-x_0)$, $\delta(y-y_0)$ and $\delta(z)$ are Dirac-delta functions, the fundamental solution of eqn (2) can be expressed as

$$\hat{T}_{inc}(\omega,x,y,z)=\frac{e^{-\sqrt{\frac{\mathrm{i}\omega}{K_j}}\sqrt{(x-x_0)^2+(y-y_0)^2+z^2}}}{2k_j\sqrt{(x-x_0)^2+(y-y_0)^2+z^2}}. \tag{3}$$

Applying a Fourier transformation in the *z* direction leads to the solution

$$\tilde{T}_{inc}(\omega,x,y,k_z)=\frac{-\mathrm{i}}{4k_j}H_0\left(\sqrt{-\frac{\mathrm{i}\omega}{K_j}-{k_z}^2}\;r_0\right), \tag{4}$$

where $H_0(\ )$ are Hankel functions of the second kind and order 0, and $r_0=\sqrt{(x-x_0)^2+(y-y_0)^2}$.

The full three-dimensional solution is then found by applying an inverse Fourier transform in the k_z domain. This inverse Fourier transformation can be expressed as a discrete summation if we assume the existence of virtual sources, equally spaced at L_z, along z, which enables the solution to be obtained by solving a limited number of two-dimensional problems,

$$\hat{T}_{inc}(\omega,x,y,z)=\frac{2\pi}{L_z}\sum_{m=-M}^{M}H_0\left(\sqrt{-\frac{\mathrm{i}\omega}{K_j}-{k_{zm}}^2}\;r_0\right)e^{-\mathrm{i}k_{zm}z}, \tag{5}$$

with k_{zm} being the axial wavenumber given by $k_{zm}=\frac{2\pi}{L_z}m$. The distance L_z chosen must be big enough to prevent spatial contamination from the virtual sources. Eqn (5) can be further manipulated and written as a continuous superposition of heat plane phenomena,

$$\tilde{T}_{inc}(\omega,x,y,k_z)=\frac{-\mathrm{i}}{4\pi\ k_j}\int_{-\infty}^{+\infty}\left(\frac{e^{-\mathrm{i}\nu_j|y-y_0|}}{\nu_j}\right)e^{-\mathrm{i}k_x(x-x_0)}dk_x \quad , \tag{6}$$

where $\nu_j=\sqrt{-\frac{\mathrm{i}\omega}{K_j}-{k_z}^2-{k_x}^2}$ and $\mathrm{Im}\left(\nu_j\right)\le 0$, and the integration is performed with respect to the horizontal wave number (k_x) along the x direction.

Assuming the existence of an infinite number of virtual sources, we can discretize these continuous integrals. The integral in the above equation can be transformed into a summation if an infinite number of such sources are distributed along the x direction, spaced at equal intervals L_x. The equation can then be written as

$$\tilde{T}_{inc}(\omega,x,y,k_z)=\frac{-\mathrm{i}}{4k_j}E_{0j}\sum_{n=-\infty}^{n=+\infty}\left(\frac{E_j}{\nu_{nj}}\right)E_d \quad , \tag{7}$$

where $E_{0j}=\frac{-\mathrm{i}}{2k_jL_x}$, $E_j=e^{-\mathrm{i}\nu_{nj}|y|}$, $E_d=e^{-\mathrm{i}k_{xn}(x)}$, $\nu_{nj}=\sqrt{-\frac{\mathrm{i}\omega}{K_j}-{k_z}^2-{k_{xn}}^2}$ and $\mathrm{Im}\left(\nu_{nj}\right)\le 0$, and $k_{xn}=\frac{2\pi}{L_x}n$, which can in turn be approximated by a finite sum of equations (N). Note that $k_z=0$ is the two-dimensional example.

The total heat field is achieved by adding the heat source terms, equal to those in the unbounded space, to the sets of surface terms arising within each layer and at each interface, that are required to satisfy the boundary conditions at the interfaces, i.e. continuity of temperatures and normal fluxes between layers.

For the layer j, the heat surface terms on the upper and lower interfaces can be expressed as

$$\tilde{T}_{j1}(\omega,x,y,k_z)=E_{0j}\sum_{n=-\infty}^{n=+\infty}\left(\frac{E_{j1}}{\nu_{nj}}A_{nj}^t\right)E_d \quad , \tag{8}$$

$$\tilde{T}_{j2}(\omega,x,y,k_z)=E_{0j}\sum_{n=-\infty}^{n=+\infty}\left(\frac{E_{j2}}{\nu_{nj}}A_{nj}^b\right)E_d \quad , \tag{9}$$

where $E_{j1}=\mathrm{e}^{-\mathrm{i}\nu_{nj}\left|y-\sum_{l=1}^{j-1}h_l\right|}$, $E_{j2}=\mathrm{e}^{-\mathrm{i}\nu_{nj}\left|y-\sum_{l=1}^{j}h_l\right|}$ and h_l is the thickness of the layer l. The heat surface terms produced at interfaces 1 and m+1, which govern the heat that propagates through the top and bottom semi-infinite media, are respectively expressed by

$$\tilde{T}_{02}(\omega,x,y,k_z)=E_{00}\sum_{n=-\infty}^{n=+\infty}\left(\frac{E_{01}}{\nu_{n0}}A_{n0}^b\right)E_d \tag{10}$$

$$\tilde{T}_{(m+1)2}(\omega,x,y,k_z)=E_{0(m+1)}\sum_{n=-\infty}^{n=+\infty}\left(\frac{E_{(m+1)2}}{\nu_{n(m+1)}}A^t_{n(m+1)}\right)E_d \qquad (11)$$

A system of 2(m+1) equations is derived, ensuring the continuity of temperatures and heat fluxes along the m+1 interfaces between layers. Each equation takes into account the contribution of the surface terms and the involvement of the incident field. All the terms are organized according to the form $\underline{\underline{F}}\,\underline{a}=\underline{b}$. When the heat source is placed in medium 1, the following system of equations is obtained

$$\begin{bmatrix}
-1 & -1 & e^{-i\nu_{n1}h_1} & \dots & 0 & 0 & 0\\
\frac{1}{k_0\nu_{n0}} & -\frac{1}{k_1\nu_{n1}} & -\frac{e^{-i\nu_{n1}h_1}}{k_1\nu_{n1}} & \dots & 0 & 0 & 0\\
0 & e^{-i\nu_{n1}h_1} & -1 & \dots & 0 & 0 & 0\\
0 & \frac{e^{-i\nu_{n1}h_1}}{k_1\nu_{n1}} & \frac{1}{k_1\nu_{n1}} & \dots & 0 & 0 & 0\\
\dots & \dots & \dots & \dots & \dots & \dots & \dots\\
0 & 0 & 0 & \dots & -1 & e^{-i\nu_{nm}h_m} & 0\\
0 & 0 & 0 & \dots & -\frac{1}{k_1\nu_{nm}} & -\frac{e^{i\nu_{nm}h_m}}{k_m\nu_{nm}} & 0\\
0 & 0 & 0 & \dots & e^{-i\nu_{nm}h_{m1}} & -1 & -1\\
0 & 0 & 0 & \dots & \frac{e^{-i\nu_{nm}h_m}}{k_m\nu_{nm}} & \frac{1}{k_m\nu_{nm}} & \frac{1}{k_{m+1}\nu_{n(m+1)}}
\end{bmatrix}
\begin{bmatrix}
A^b_{n0}\\ A^t_{n1}\\ A^b_{n1}\\ \dots\\ A^t_{nm}\\ A^b_{nm}\\ A^t_{n(m+1)}
\end{bmatrix}=
\begin{bmatrix}
-e^{-i\nu_{n1}y_0}\\
\frac{e^{-i\nu_{n1}y_0}}{k_1\nu_{n1}}\\
-e^{-i\nu_{n1}|h_1-y_0|}\\
-\frac{e^{-i\nu_{n1}|h_1-y_0|}}{k_1\nu_{n1}}\\
\dots\\ 0\\ 0\\ 0\\ 0
\end{bmatrix} \qquad (12)$$

The resolution of the system gives the amplitude of the surface terms at each interface. The temperature field for each layer formation is found by adding these surface terms to the contribution of the incident field, which leads to the following equations:

top semi-infinite medium (medium 0 *)*

$$\tilde{T}(\omega,x,y,k_z)=E_{00}\sum_{n=-\infty}^{n=+\infty}\left(\frac{E_{01}}{\nu_{n0}}A^b_{n0}\right)E_d\text{, if } y<0 \qquad (13)$$

layer 1 (source position)

$$\tilde{T}(\omega,x,y,k_z)=\frac{-i}{4k_1}H_0\left(K_{t1}\,r_0\right)+E_{01}\sum_{n=-\infty}^{n=+\infty}\left(\frac{E_{11}}{\nu_{n1}}A^t_{n1}+\frac{E_{12}}{\nu_{n1}}A^b_{n1}\right)E_d\text{, if } 0<y<h_1 \qquad (14)$$

layer j $(j\neq 1)$

$$\tilde{T}(\omega,x,y,k_z)=E_{0j}\sum_{n=-\infty}^{n=+\infty}\left(\frac{E_{j1}}{\nu_{nj}}A^t_{nj}+\frac{E_{j2}}{\nu_{nj}}A^b_{nj}\right)E_d\text{, if } \sum_{l=1}^{j-1}h_l<y<\sum_{l=1}^{j}h_l \qquad (15)$$

bottom semi-infinite medium (medium $m+1$*)*

$$\tilde{T}_{(m+1)2}(\omega,x,y,k_z) = E_{0(m+1)} \sum_{n=-\infty}^{n=+\infty} \left(\frac{E_{(m+1)2}}{\nu_{n(m+1)}} A^t_{n(m+1)} \right) E_d \qquad (16)$$

Note that when the position of the heat source is changed, the matrix $\underline{\underline{F}}$ remains the same, while the independent terms of $\underline{b}$ are different. As the equations can be easily manipulated to consider another position for the source, they are not included here.

4 Experimental validation

4.1 Specimen description

The multilayer systems were built by overlapping 500x500 mm^2layers of different current insulating materials: natural cork (NC), molded expanded polystyrene (EPS) and medium-density fibreboard (MDF).

Each multilayer system was composed of 4 layers. One homogeneous (System 1) and two heterogeneous (Systems 2-3) systems were prepared. System 1 is composed of four NC layers (20.63mm, 20.68mm, 20.52mm and 20.60mm thickness). System 2 was prepared with these four layers: EPS (18.84mm), MDF (19.65mm), NC (20.60mm) and EPS (19.86). System 3 had the following four layers: NC (20.52mm), EPS (18.84mm), MDF (19.65mm) and NC (20.60mm). Each material used was tested to determine its thermal conductivity, mass density and specific heat. The thermal conductivity was found by means of the Guarded hot-plate method (ISO 8302:1991 [7]) using the EP-500 Lambda-meter from Lambda-Mebtechnik GmbH Dresden, a single-specimen model. The test procedure defined in EN 12667:2001[8] was used. The mass density was determined using the procedure described in EN 1602:1996[9]. The specific heat was obtained using a Netzsch apparatus, model DSC200F3, following the ratio method.

Table 1 gives the averages of those properties for the three materials used in the experiments.

Table 1: Thermal properties of the materials.

Material	Conductivity, k ($W.m^{-1}.^oC^{-1}$)	Mass density, ρ ($kg.m^{-3}$)	Specific heat, c ($J.kg^{-1}.^oC^{-1}$)
Natural Cork	0.046	130.0	1638.0
Molded Expanded Polystyrene	0.041	14.3	1430.0
Medium-Density Fiberboard	0.120	712	1550.0

4.2 Experimental procedure

The experiments required imposing an unsteady heat flow rate on each multilayer system using the single-specimen EP-500 Lambda-meter apparatus. Before running any test, the specimens were conditioned in a climatic chamber, Fotoclima 300EC10 from Aralab, in a controlled environment with a set-point temperature of (23±2)ºC and (50±5)% relative humidity, until constant mass was reached.

The tests were carried out in a controlled laboratory environment (temperature (23±2)ºC and relative humidity (50±5)%). The single-specimen Lambda-meter EP-500was first programmed to reach a mean temperature of 23ºC in the test specimen, establishing a 15º temperature difference between the heating and the cooling units. So, during the test, the temperature of the top multilayer surface (in contact with the heating plate unit) increased, while the temperature of the bottom multilayer surface (in contact with the lower plate) decreased. The energy input was maintained until a permanent heat flow rate was reached, that is, when there were no temperature variations at the multilayer interfaces. The system was then allowed to cool until the initial temperatures were reached again.

The temperature variation at each interface layer was measured using type T (copper) thermocouples made of 0.2 mm diameter wire. Three thermocouples were placed at each system interface, including the top and bottom surfaces, which were in contact with the heating and cooling plates (see Figure 2). The data were recorded by a Yokogawa MW 100 data logger, with a time interval of 10 seconds.

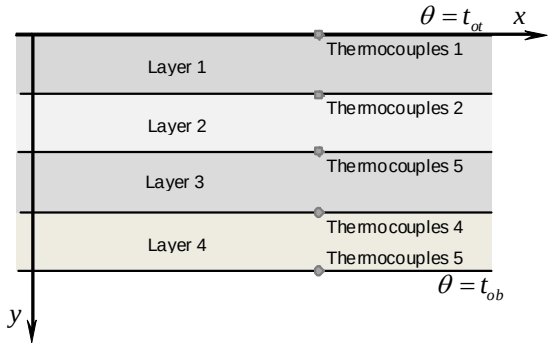


Figure 2: Thermocouple positions.

5 Results and discussion

In this section the experimental measurements are compared with the semi-analytical results. The semi-analytical solutions are obtained using the formulation presented in section 3, after they have been manipulated to simulate

the heat transfer through the multilayer systems where the temperature variations are prescribed for the top and bottom surfaces. The materials' thermal properties (see Table 1) were used in these simulations.

5.1 Semi-analytical model

Equations (8) and (9) are manipulated by removing the media 0 and m+1 and by imposing temperatures t_{0t} and t_{0b} on the external top and bottom surfaces (interfaces 1 and m+1). Temperatures t_{0t} and t_{0b} are obtained by applying a Fourier transformation in the time domain to the temperatures recorded at the external multilayer system surfaces during the guarded hot plate test.

The total heat field is achieved by adding together the sets of surface terms arising within each layer at each interface and by imposing continuity of temperatures and normal fluxes at the internal interfaces.

The following system of 2m equations is obtained:

$$\begin{bmatrix} \frac{1}{k_1\nu_{n1}} & \frac{e^{-i\nu_{n1}h_1}}{k_1\nu_{n1}} & \dots & 0 & 0 \\ e^{-i\nu_{n1}h_1} & -1 & \dots & 0 & 0 \\ \frac{e^{-i\nu_{n1}h_1}}{k_1\nu_{n1}} & \frac{1}{k_1\nu_{n1}} & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & -1 & e^{-i\nu_{nm}h_m} \\ 0 & 0 & \dots & -\frac{1}{k_1\nu_{nm}} & -\frac{e^{i\nu_{nm}h_m}}{k_m\nu_{nm}} \\ 0 & 0 & \dots & \frac{e^{-i\nu_{nm}h_m}}{k_m\nu_{nm}} & \frac{1}{k_m\nu_{nm}} \end{bmatrix} \begin{bmatrix} A_{n1}^t \\ A_{n1}^b \\ \dots \\ A_{nm}^t \\ A_{nm}^b \end{bmatrix} = \begin{bmatrix} t_{0t} \\ 0 \\ 0 \\ \dots \\ 0 \\ 0 \\ t_{0b} \end{bmatrix} \tag{17}$$

Given that the temperatures t_{0t} and t_{0b} are uniform along the interfaces, this system is solved by imposing $k_{xn} = 0$ and $k_z = 0$. The resolution of this system gives the amplitude of the surface terms at each interface, leading to the following temperature fields at layer j:

$$\tilde{T}(\omega, x, y, k_z) = E_{0j}\left(\frac{E_{j1}}{\nu_{0j}}A_{0j}^t + \frac{E_{j2}}{\nu_{0j}}A_{0j}^b\right), \text{ if } \sum_{l=1}^{j-1} h_l < y < \sum_{l=1}^{j} h_l \tag{18}$$

5.2 Results

Temperatures t_{0t} and t_{0b} were first defined by applying a direct discrete fast Fourier transform in the time domain to the temperatures recorded by the thermocouples on the external surfaces of the system and subtracting the initial temperature. Analysis of the experimental responses led to an analysis period of 16h being established. This was enough to find the energy equilibrium of the

multilayer system with the environment (temperatures at the interfaces were again restored almost to the initial test temperatures). The upper frequency of the analysis was defined such that its contribution to the global response is negligible.

The analytical computations were performed in the frequency domain for frequencies ranging from $0.0\,\text{Hz}$ to $\frac{1.0}{32\times3600}\,\text{Hz}$, with a frequency increment of $\frac{1.0}{32\times3600\times2048}\,\text{Hz}$, which determined a full analysis window of 16h.

The temperature variation imposed on the top and bottom multilayer surfaces may be of any type. To obtain the temperature in the time domain, a discrete inverse fast Fourier transform was applied in the frequency domain. The aliasing phenomena were dealt by introducing complex frequencies with a small imaginary part, taking the form $\omega_c = \omega - \mathrm{i}\eta$ (where $\omega = 0.7\Delta\omega$, and $\Delta\omega$ is the frequency increment). This shift was subsequently taken into account in the time domain by means of an exponential window, $\mathrm{e}^{\eta t}$, applied to the response.

The final temperatures were obtained by adding the initial test temperatures to these responses.

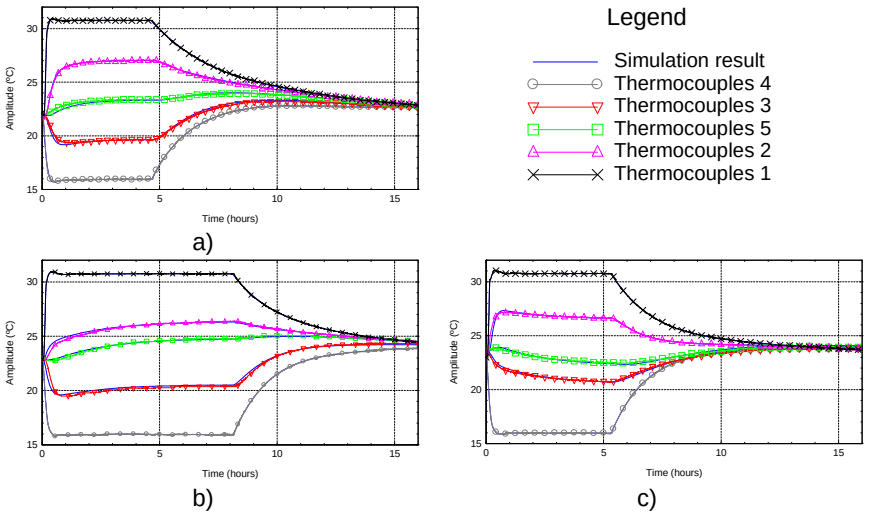


Figure 3: Semi-analytical and experimental results. Temperature change at the layer interfaces of the multilayer systems: a) System 1 – homogeneous multilayer system. b) System 2 – heterogeneous multilayer system. c) System 3 – heterogeneous multilayer system.

The results of the experimental measurements are presented below and compared with those computed analytically. In the figures the solid lines correspond to the semi-analytical responses and the experimental measurements are represented by the lines with marked points. The plotted experimental results

at each interface correspond to the arithmetic mean of three thermocouple temperatures.

Figure 3 shows the results obtained for the different multilayer systems. System 1 is the natural cork (NC) homogeneous system, while the other two are heterogeneous (see section 4.1). These responses show a good agreement between the semi-analytical responses and the experimental results. The results are similar over the full time window, i.e. the period in which the heating and cooling units are receiving energy, the period with a constant heat flow rate, and the time when apparatus has no power input. Note that at the beginning of the process all the thermocouples show a temperature similar to the ambient temperature, which conforms with the initial conditions defined for the semi-analytical simulation.

Figure 3b) and c) shows the temperature change at each interface of the heterogeneous systems. System 2 is composed of EPS sandwiching MDF and NC layers, while System 3 has NC layers sandwiching an EPS and MDF layer. Comparing the semi-analytical solutions and experimental measurements for each system, it can be seen that the results are very similar. During the heating and cooling phase, when temperatures are becoming constant, we can see that the lowest temperature gradients occur in the MDF layers, given their higher conductivity and lower thermal diffusivity.

6 Conclusions

Three-dimensional semi-analytical solutions for transient heat conduction in a multilayer system in the frequency domain have been validated experimentally. The results showed a good agreement between experimental measurements and the computed solutions, thus we can conclude that the proposed semi-analytical model formulated in the frequency domain is reliable for studying transient heat conduction in multilayer systems.

Acknowledgements

The research work presented herein was supported by the Portuguese Foundation for Science and Technology (FCT), under research project PTDC/ECM /114189/2009 and doctoral grant SFRH/BD/48138/2008, and by the Coordination of Improvement of Higher Education Personnel (CAPES), a Brazilian government agency.

References

[1] Chang, Y.P., Kang, C.S., Chen, D.J., The use of fundamental Green's functions for solution of problems of heat conduction in anisotropic media. *International Journal of Heat and Mass Transfer* **16,** pp. 1905-1918, 1973.

[2] Rizzo, F.J., Shippy, D.J., A method of solution for certain problems of transient heat conduction. *AIAA Journal* **8,** pp. 2004-2009, 1970.

[3] Tadeu, A., Simões, N., Three-dimensional fundamental solutions for transient heat transfer by conduction in an unbounded medium, half-space, slab and layered media. *Engineering Analysis with Boundary Elements*, **30(5)**, pp. 338–349, 2006.
[4] Kaşka, Ö., Yumrutaş, R., Comparison of experimental and theoretical results for transient heat flow through multilayer walls and flat roofs. *Energy*, **33**, pp. 1816-1823, 2008.
[5] Chen, Y., Zhou, J., Spitler, J.D., Verification for transient heat conduction calculation of multilayer building constructions. *Energy and Buildings*, **38**, pp.340-348, 2006.
[6] Sami A. Al-Sanea, Thermal performance of building roof elements. *Building and Environment*, **37(7)**, pp. 665-75, 2002.
[7] International Organization for Standardization, **ISO 8302**: Thermal insulation – Determination of steady-state thermal resistance and related properties – Guarded hot plate apparatus, 1991.
[8] European Standards, **EN 12667**: Thermal performance of building materials and products. Determination of thermal resistance by means of guarded hot plate and heat flow meter methods. Products of high and medium thermal resistance, 2001.
[9] European Standards, **EN 1602**: *Thermal insulating materials, Thermal insulation, Construction materials, Density measurement, Bulk density, Test specimens, Testing conditions, Buildings, 1996.*

On liquid crystal thermography for determination of the heat transfer coefficient in rectangular ducts

Z. Ghorbani-Tari[1], B. Sunden[1] & G. Tanda[2]
[1]*Department of Energy Sciences, Division of Heat Transfer, Lund University, Sweden*
[2]*Dipartimento di Macchine, Sistemi Energetici e Trasporti (DIMSET), Università degli Studi di Genova, Italy*

Abstract

Liquid crystal thermography (LCT) has been employed by researchers in heat transfer and fluid flow communities as a reliable and non-intrusive temperature measurement technique. This technique correlates the colour response to temperature for a heated surface which has been treated with thermochromic liquid crystals (TLCs). The LCT has been used extensively in convective heat transfer research in duct flows. In this paper, some experiences by LCT in thermal measurements for rectangular duct flows are provided. A few TLCs examples associated with continuous ribs for two different values of rib pitch-to-height ratio of 4 and 8 at Re=8900 and 28500 are illustrated. Important issues such as heating principles, calibration of TLCs, image acquiring and analysis, problems of treating the surface by TLCs, and expected measurement accuracy are summarized. The main purpose is to assist newcomers in the field and provide some guidelines for proper use of the LCT in heat transfer research.
Keywords: LCT, convective heat transfer, measurement accuracy.

1 Introduction

Liquid crystal thermography (LCT) has emerged as reliable measurement sensors for the visualization and determination of surface temperature distributions leading to convective heat transfer coefficients [1]. The advantages of LCT are, e.g., its flexibility to be used from micron sized electronic circuits to

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110231

large scale gas turbine blades, high spatial resolution, steady state and transient temperature measurements, temperature accuracy ± 0.1 ºC, optical system based on visible light and independent of surface emissivity, ability to use with common color video and CCD cameras and conventional recorders, cost effective compared to infrared thermography (IR) systems etc. The LCT has been used extensively in convective heat transfer research in duct flows (applications in heat exchangers, and internal cooling of gas turbines and electronics cooling). References [2–4] are examples of published papers. Descriptions of LCT measurement technique can be found in, e.g., refs. [5–8].

The LCT correlates the color response to temperature for a heated surface which has been treated with thermochromic liquid crystals (TLCs). The TLCs are complex organic substances that once illuminated with a white light; reflect visible light of a narrow frequency band depending on temperature. The dominant reflected wavelength decreases with increasing temperature of the TLCs: red color appears at the lowest temperature of the range of indicated temperatures and blue at the highest temperature [9]. This behavior is exhibited within a range of temperatures called event temperature range. The TLCs with an event temperature range within 5-30 ºC are termed broad band and those within 0.5-4 ºC are termed narrow band [5]. Bandwidth as small as 0.1 ºC has also been reported [10]. The use of narrow band TLCs is an attractive alternative for temperature and heat transfer measurements due to their higher accuracy in temperature measurements and also the fact that narrow band TLCs are less affected by variation in illumination-viewing angles and illumination disturbances [5]. For full field measurements where the surface temperature is non-uniform, the narrow band TLC cannot be used directly but instead a more tedious procedure is needed. The wide band TLCs has better ability to record the entire surface temperature distribution from a single image, if the TLC bandwidth is large enough to span all temperatures experienced on a test surface. However, the primary drawback is that wide-band TLCs require a significant number of calibration points to accurately resolve the highly non-linear variation of hue versus temperature of typical TLC [11].

In this paper, calibration of TLCs, image acquiring and analysis, some important issues like heating principles (e.g., how to ensure uniform heating), and problems of treating the surface by TLCs are summarized. Some examples of TLCs images for a straight rectangular duct equipped with arrays of continuous ribs at different Re values are illustrated. Two different values of rib pitch-to-height ratio of 4 and 8 are considered to demonstrate heat transfer coefficients distribution on the end wall. The main purpose is to assist newcomers in the field and provide some guidelines for proper use of LCT in heat transfer research.

2 Thermochromic liquid crystals – literature review

The TLC compounds are difficult to deal with under most heat transfer laboratory conditions. Several adverse effects are present in any laboratory or field setting such as chemical contamination and exposure to ultra-violet (UV)

light which primarily degrade the colour response. Several approaches have been developed to make TLCs more practical to use [12]. Micro-encapsulation is the most popular method of protection and the liquid crystal is enclosed in polymer spheres which can be as small as a few microns. Commercial micro-encapsulated TLCs may be purchased in water-based slurry and are generally applied to flat or curved surfaces using an air-brush spray. The thickness of coating must be controlled carefully and further reading can be found in ref. [13, 14]. The most convenient manner to measure the thermal field on a surface is to coat it with a pre-packaged assembly consisting of a liquid crystal layer painted on a plastic sheet having a background colour with black ink [12]. The pre-packaged assembly TLC is limited in application because they are not typically available in a sprayable medium and limited to relatively flat surfaces. Edge effects due to chemical contamination may destroy the pre-packaged surface when a cutout portion of a manufactured sheet is used. To be used in advanced implementations like research on heat transfer coefficients on duct walls one needs to heat the surface of interest, apply calibration for the color-temperature response, acquire images and analyze the images by suitable software for image processing. Hue-based image processing [15] is the most common technique used in applications of LCT to interpret the TLC signal. The color of TLCs observed by a RGB-data (red-green-blue) acquisition system is transformed to hue (H), saturation (S) and intensity (I) color space. HSI color space provides approximately the human description of color [1]. Hue is an optical attribute related to actual colour of the point of observation [12]. Hue is the only among these parameters being retained due to its monotonic relationship with surface temperature [16]. The hue value of each pixel indicates the temperature of corresponding location on the surface. This property of hue and its independence of the light intensity levels make it suitable for temperature measurements [17].

In experimental heat transfer, one demanding issue related to liquid crystals is the colour-temperature relationship, and any quantitative application of TLCs requires an accurate calibration [13]. It is essential to keep the lighting and viewing conditions of the thermal experiments identical to the calibration experiment as a colour pattern can be significantly affected by lighting and viewing angles [12]. There are several papers related to TLC calibration and the main factors being pertinent to calibration. Some of these are summarized here. The illumination source has a significant effect on the shape of the hue-temperature calibration curve and this may lead to higher or lower temperature uncertainties over the applicable range. If the hue-temperature curve is a straight line, this would be the sensitivity at every hue; however, hue-temperature curves have regions of higher and lower sensitivity [18]. The background lighting is another important factor which may influence the shape of the hue-temperature calibration curve [18]. The hysteresis is characterized by a decrease in reflectivity and a shift in temperature associated with the peak reflected intensity for each of R, G and B components during cooling. This causes a shift in the hue-temperature relation of the TLC. This results in temperature biases when cooling occurs instead of heating [19]. The hysteresis is not a limiting measurement error factor if the temperature of a measured surface remains

within or below the active range of the TLC [17]. Conversely, the magnitude of the hysteresis can be significant if the temperature exceeds the TLC active range even for short periods of time [17, 19]. The magnitude of hysteresis and aging effects is related to the bandwidth; smaller the bandwidth greater the effect [19]. Ref. [17] reported that a thickness of the TLC layer affected the calibration and vulnerability to aging, i.e., a change in the hue-temperature relationship with time due to prolonged exposure to high temperature and usage. Ref. [19] reported a similar effect; the permanent damage to TLCs during exposure to temperatures above the clearing point. The sensitivity to aging and damage is shown to be more acute for thinner film thicknesses [13]. The reading of the TLC colours depends on the angles between the light source and the TLC surface, and between the camera and the TLC surface [17]. Further details about the effects of lighting and viewing arrangements on TLCs colour can be found in, e.g., refs. [15, 17, 20]. The perceived color of a TLC depends on the optical properties of the measurement path and the reader is referred to refs. [13, 20].

2.1 Selection of calibration methods

The calibration experiment can be conducted in a separate system (on which controlled thermal conditions are imposed) or directly in a test surface (*in-situ*), using the same optical arrangement as thermal measurement section [12]. The calibration conducted in a separate system may result in some problems. This might be due to the situation imposed in the actual thermal measurements and it might be difficult to enable the similar conditions precisely for the calibration. The *in situ* calibration might be more beneficial as it is referred by refs. [11, 21]. However, the arrangement of *in situ* calibration is more complicated because several thermocouples have to be fitted into a test section. Therefore, it can be

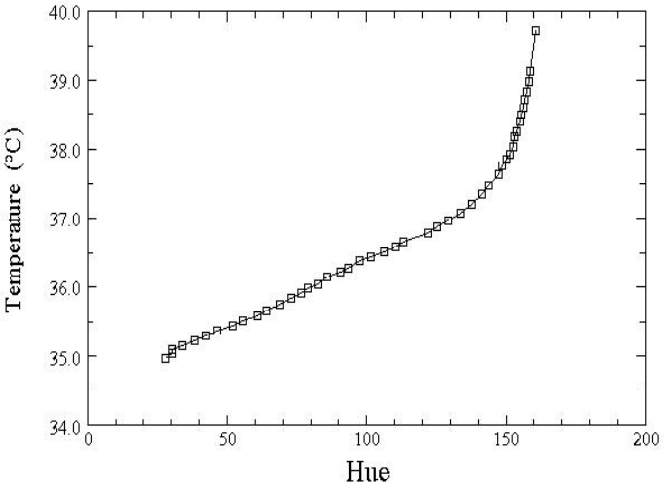


Figure 1: Hue-temperature calibration [22].

more practical to make calibration on a separate plate where several thermocouples are fitted and the thermally boundary conditions can be more easily controlled.

Fig. 1 shows an example of a typical hue-temperature calibration curve. The process of calibration is omitted here and the reader is referred to ref. [22]. The active range starts at a temperature of 35 ºC with a bandwidth of 5 ºC. The hue is fairly linear and repeatable in the range of 60-120 and thus can be adopted for the reconstruction of heat transfer coefficient distribution. The linear portion of the hue field is usually adopted due to its higher sensitivity; sensitivity at every hue value, and thus is reliable to be applied in thermal measurements.

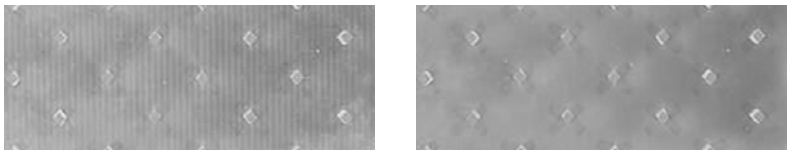
3 Selection of heat measurement method

Two commonly used methods for heat transfer measurement are steady state where experimental properties do not vary with time and transient liquid crystal method in which heat transfer coefficients are deduced from a time-temperature history record [1]. The steady state heat transfer experiment with a constant heat flux density is conceptually simple and relatively easy to analyze. The heat transfer coefficient is obtained from dividing the heat flux by the measured temperature difference between the TLC surface and the fluid. However, the steady state method requires the maintenance of a uniform surface heat flux and this attainment might be difficult even by using an electric heating foil. The maximum heat flux must not cause an appreciable temperature difference across the TLC layer either. For the transient method a surface temperature is subject to a step change in the fluid flow temperature. The surface is then considered to be the interface between the fluid and a semi-infinite body, which implies that an analytical solution for the transient heat conduction problem can be used to evaluate the heat transfer coefficient. If a surface cannot be considered as the boundary of a semi-infinite body but must be regarded as a thin layer, then the temperature is assumed to be uniform in the lateral direction. A lumped heat transfer analysis is then performed in the evaluation of the convective heat transfer coefficient.

3.1 Steady state-constant heat flux

The steady state thermal measurement is employed and TLCs are used to monitor the surface temperature. The test section coated by heating foil is electrically heated to maintain the surface temperature greater than the air temperature, i.e., $T_{LC} > T_a$, and convective heat transfer occurs between the surface and the fluid. The surface electric heater is employed such that local heat flux is known and this, together with the local surface temperature gives the local heat transfer coefficient [23]. The important issue is that the applied heating foil provides a uniform heat flux through the test section. If TLC is attached to the heating foil directly, the heat flux might be more concentrated over the electrical path and that may lead to a non-uniform heat flux distribution over the test surface, see fig. 2 (a). One way to solve this might be to apply a thin metallic

plate to which a plane heater is attached and then a more uniform heat flux is obtained, see fig 2 (b). The thermal conductivity of the selected metallic plate may not be too high because then a uniform wall temperature distribution might be achieved instead. The thermal contact resistance is another important issue. This occurs when different conducting surfaces, e.g., heating foil, metallic foil, TLC are not fit tightly together and a thin layer of air is trapped between them. As a consequence, heat is transferred by conduction through the contact points of the solid, while convection and radiation take place across the trapped interfacial air and result in a temperature drop across the interface and this cannot be ignored. The contact resistance is a function of pressure holding different surfaces and this highlights the importance of which metallic interface should be considered to increase the pressure and thus reduce the thermal contract resistance. As a result, one should always be aware of the existence of the interface resistance and the resulting temperature difference across the interface [24].



(a) (b)

Figure 2: TLC directly attached to a plane heater: the electric tracks are clearly visible, symptom of a non-uniform heat flux at surface (a); A thin stainless steel between TLC and the plane heater: the heat flux is more evenly distributed over the surface (b).

4 Thermal measurement

4.1 Thermal test section apparatus

The experimental apparatus is not shown here for the sake of brevity and the reader is referred to ref. [16]. The test section is a rectangular duct as wide and high as the entry and exit sections and delimited by a thin heated plate (0.1 m wide and 0.28 m long) lateral and frontal Plexiglas walls. The heated side of the test section consists of 0.5 mm-thick stainless steel foil to which a plane heating foil is glued. A thin TLC sheet is applied on the stainless steel foil, on the side exposed to the air flow, to measure local surface temperature. The rear side of the heated plate is thermally insulated to convey as much as possible of the electric power dissipated by the heater to the convective air flow. Electric power is supplied by a DC to measure voltage and current of the heater. Fine-gauge thermocouples were placed inside the rectangular duct, directly exposed to the airflow, and at several sites inside the duct wall material. These sensors are used to measure the air temperature at the test section inlet, to estimate conduction heat losses to the surroundings and to control the attainment of the steady state

conditions. The TLC used here is supplied by pre-packaged sheet (0.15 mm thick) to measure temperature distribution on the heated surface. The colour distribution of TLC is observed by a CCD video camera through the Plexiglas wall opposite to the heated surface and stored in a PC. The light for the illumination is switched on only for the time required for image acquisition to avoid radiant heating of the test surface [16].

4.2 Measuring procedure

The test surface is roughened by insertion of multiple ribs. The airflow is adjusted to a prescribed velocity and at the same time DC current is supplied to the heater. The test surface temperature is maintained within the linear portion of the calibration curve by controlling the input power delivered to the heater and therefore no information is gained outside that region. That necessitates repeating the tests at the same flow rate and for several different values of the input power to move the color pattern toward the regions not previously monitored and to extract local the heat transfer coefficients (h), eqn (1). One important issue is that the input power and TLC images are recorded at steady state. The heat transfer coefficient distributions from each test (at a given heat flux) are superimposed on some regions of the test surface and an average value is taken with the requirement that the discrepancy between h values for the same pixel never exceeds 2% [16].

$$h = \frac{Q_{el} - Q_{loss}}{A \cdot \left(T_{LC} - T_b\right)} \tag{1}$$

The convective heat flux is assumed to be uniformly distributed and it is obtained by measuring input power to the heater (Q_{el}) corrected by heat loss rate, divided by the difference between surface temperature detected by the TLC and air bulk temperature (T_b) at the current position along the streamwise direction. The surface area (A) is assumed as that of the smooth surface. The heat loss rate is calculated by radiation, conduction (through the insulation on the rear face of the heater) and heat dissipation from the ribs. Although, the ribs made of Plexiglas are deemed adiabatic due to low thermal conductivity, the conventional one-dimensional fin model is applied to estimate the small amount of heat transfer rate from the rib sides to the air flow [16].

4.3 Heat transfer coefficients for the ribbed duct

Continuous ribs are positioned on the surface of the duct (end wall) at periodic sites and oriented transversely to the main flow direction. The geometric parameters (heated surface length L and width W, channel height H, rib height e, rib pitch p, streamwise rib length l) are summarized in table 1. The rib pitch 20 and 40 mm are considered for the demonstrated configurations. The height and the streamwise length of the ribs were kept fixed at 5 and 3 mm, respectively.

Table 1: Geometric parameters of rib configurations [16].

config.	L (mm)	W (mm)	H (mm)	l (mm)	p (mm)	e (mm)	p/e
A	280	100	20	3	20	5	4
B	280	100	20	3	40	5	8

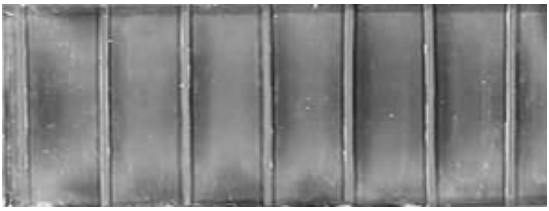


Figure 3: TLC image for configuration B [16].

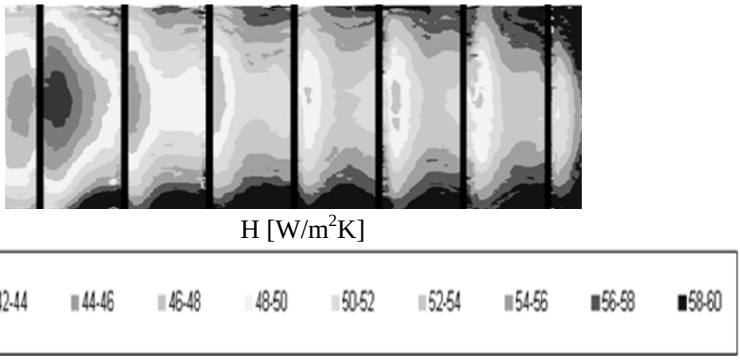


Figure 4: 2-D plot heat transfer coefficient (W/m^2K) for configuration at Re = 8900 [16].

An example of a TLC image taken by a CCD video camera for configuration B at the Reynolds number (Re = 8900) and a specific heat flux is illustrated in fig. 3. As explained previously, only points within the linear portion of the calibration curve is taken from some images like the one illustrated in fig. 3 and processed for the reconstruction of full field heat transfer coefficients, as shown in fig. 4. The inter-rib h-distribution at midline is demonstrated in fig. 5 for configurations A and B, respectively [16].

Inspection of fig. 5 shows that for the configuration A (p/e = 4), the inter-rib h-distribution is characterised by a monotonic increase that periodically repeats itself after the third-fourth rib. This probably indicates the presence of a trapped vortex between ribs without a reattachment point. Conversely, for configuration

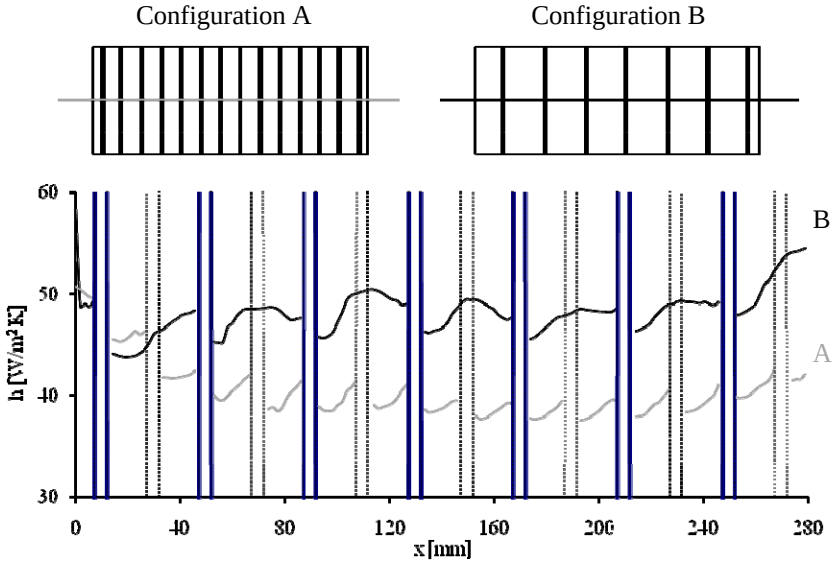


Figure 5: Distribution of local center-line heat transfer coefficient for the continuous rib configuration A and B at Re = 28500 [16].

B (p/e = 8), the heat transfer coefficient reaches a maximum value at the point of flow reattachment. Downstream of the reattachment point, the heat transfer coefficient slightly decreases up to the vicinity of the successive rib. The periodic fully developed conditions are attained approximately after the third rib [16].

5 Estimation of uncertainties

Several factors (image noise, lighting angle, TLC coating thickness and coating quality) may influence the hue-temperature relationship and the expected measurement accuracy. In addition, the uniformity of the heating procedure and various sources of losses to the surroundings may influence the accuracy of the measurement. The image noise can be reduced by using filtering techniques to erase hue spots not related to temperature and that leads to improvement in accuracy of the measurements. The lighting angle is known to have a significant effect on TLCs hue-temperature curve and thus on the accuracy of the measurements. Therefore, it is important that the lighting set up in a real measurement is the same as at the calibration stage. The coating thickness has a significant effect on a hue-temperature curve but it has been found to have a non-distinctive effect on the measurement uncertainty. The coating quality affects both the hue-temperature curve and the accuracy of the measurement. Further reading is available in ref. [6]. In applications where the overall purpose is to determine convective heat transfer coefficients, other uncertainties impact on the

final accuracy; i.e., the uncertainties in flow rate or Reynolds number and the flow temperatures. Consideration of all uncertainties is commonly taken into account applying the root-sum-square combination as described in ref. [25].

6 Concluding remarks

In this paper, some important issues related to the use of TLCs for heat transfer measurements in duct flows (treatment of the surface, TLC calibration, image acquiring and analysis, etc.) are summarized. Regardless of calibration method, each TLC has its own calibration curve. The linear portion of the calibration curve is recommended to be adopted due to higher sensitivity and thus more reliable for thermal measurements. Concerning the thermal boundary condition typically adopted in the LCT, steady-state, heat measurement method, and care should be taken in the use of an electrically heating foil to attain a uniform heat flux; often the insertion of a thin metallic plate between the heater and the TLC sheet could improve the uniform distribution of heat flux at the wall. Another concern associated with the heating issues is the thermal contact resistance and resulting temperature drop when different conducting surfaces, e.g., heating foil, metallic foil, TLC are not in perfect thermal contact.

There are several factors which impact the accuracy of thermal measurements associated with the LCT technique and some of those were outlined in this paper. The main purpose was to highlight some important issues collected from several researchers to assist newcomers and provide some guidelines for proper use of LCT in heat transfer research.

Acknowledgements

Part of this work is financially supported by the Swedish Energy Agency, Volvo Aero Corporation and Siemens Industrial Turbines through the national Swedish research program TURBO POWER, project TURB3. The work is also supported by a separate project by the Swedish Energy Agency.

Nomenclature

A	heat transfer area
e	rib height
h	heat transfer coefficient
H	channel height
HSI	hue, saturation, intensity
L, W	length and width of the heated section
l	length of the rib
p	rib pitch
Q_{el}	input power to the heater
Q_{loss}	heat loss rate by radiation and conduction
Re	Reynolds number

RGB red, green, blue
T_b air bulk temperature
T_{LC} surface temperature detected by TLCs

References

[1] Chan, T.L., Ashforth-Frost, S. & Jambunathan, K. , Calibrating for viewing angle effect during heat transfer measurements on a curved surface, *Int. J. Heat and Mass Transfer* , **44**, pp. 2209-2223, 2001.

[2] Sunden. B., On heat transfer and fluid flow in ribbed ducts using liquid crystal thermography and PIV measurements, *EXHTF-7(CD-ROM Proceedings)*, 139-152, 2009.

[3] Tanda. G., Heat transfer in rectangular channels with transverse and V-shaped broken ribs, *Int. J. Heat Mass Transfer*, **47**, 229-243, 2004.

[4] Wang, L. & Sunden, B., Experimental investigation of local heat transfer in a square duct with continuous and truncated ribs, *Experimental Heat Transfer*, **18**, 179-197, 2005.

[5] Abdullah, N., Talib, A.R.A., Jaafar, A.A., Salleh, M.A.M. & Chong, W.T, The basics and issues of thermochromatic liquid crystal calibrations, *Exp. Thermal Fluid Science*, **34**, 1089-1121, 2010.

[6] Rao, Y. & Zang, S., Calibrations and measurement uncertainty of wide-band liquid crystal thermography, *Meas. Sci. Technol.*, **21**, paper no. 015105, 2010.

[7] Ireland, P.T. & Jones, T.V, Liquid crystal measurements of heat transfer and surface shear stress, *Meas. Sci. Technol.*, **11**, 969-986, 2000.

[8] Critoph, R.E., Holland, M.K. & Fisher, M., Comparison of steady state and transient methods for measurement of local heat transfer in plate fin-tube heat exchangers using liquid crystal thermography with radiant heating, *Int. J. Heat Mass Transfer*, **42**, 1-12, 1999.

[9] Behle, M., Schulz, K., Leiner, W. & Fiebig, M., Colour–based image processing to measure local temperature distributions by wide-band liquid crystal thermography, *Applied Scientific Research*, **56**, pp. 113-143, 1996.

[10] Grodzka, P.G. & Facemire, B.R. Tracking transient temperatures with liquid crystals, *Letters Heat Mass Transfer*, **2**, pp. 169-178, 1975.

[11] Smith, C.R., Santino, D.R. & Praisner, T., Temperature sensing with thermochromic liquid crystals, *Exp. Fluids*, **30**, 190-201, 2001.

[12] Tanda, G., Application of optical methods to the study of convective heat transfers in rib-roughened channels, PhD thesis, The City University London, 1996.

[13] Kakade, V.U, Lock, G.D, Wilson, M., Owen, J.M. & Mayhew, J.E, Accurate heat transfer measurements using thermochromic liquid crystal. Part 1: Calibration and characteristic of crystals. *Int. J. Heat Fluid Flow*, **30**, No. 5, pp. 939-949, 2009.

[14] Baughn, J. W., Liquid crystal methods for studying turbulent heat transfer, *Int. J. Heat and Fluid Flow*, **16**, 365-375, 1995.

[15] Camci, C., Kim, K. & Hippensteele, S.A., A new hue capturing technique for the quantitative interpretation of liquid crystal images used in convective heat transfer studies, *ASME J. Turbomachinery*, **114**, pp. 765-775, 1992.
[16] Cavallero, D. & Tanda, G., An experimental investigation of forced convection heat transfer in channels with rib turbulators by using the liquid crystal thermography, *Exp. Thermal Fluid Sci*., **26**, pp. 115-121, 2002.
[17] Wiberg, R. & Lior, N., Error in Thermochromic Liquid Crystal Thermometry, *Review of Scientific Instruments*, **75**, No. 9, pp. 2985-2994, 2004.
[18] Anderson, M.R. & Baughn, J.W., Liquid-crystal thermography: illumination spectral effects. Part1- experiments, *ASME J. Heat Transfer*, **127**, pp. 581-587, 2005.
[19] Anderson, M.R. & Baughn, J.W., Hysteresis in liquid crystal thermography, *ASME J. Heat Transfer*, **126**, 339-346, 2004.
[20] Farina, D.J., Hacker, J.M., Moffat , R.J. & Eaton, J.K., Illuminant invariant calibration of thermochromic liquid crystals, *Exp. Therm. Fluid Sci.* **9 (1)**, pp. 1-12, 1994.
[21] Baughn, J.M., Anderson, M.R., Mayhew, J.E. & Wolf, J.D., Hysteresis of thermochromic liquid crystal temperature measurement based on hue, *ASME J. Heat Transfer*, **121**, pp. 1067-1072, 1999.
[22] Gao, X., Heat transfer and fluid flow investigations in ribbed ducts and impinging jets using liquid crystal thermography and PIV, PhD thesis, Division of Heat Transfer, Lund University, 2002.
[23] Incropera, F.P. & DeWitt, D.P., Fundamentals of Heat and Mass Transfer (Chapter 7), *External Flow*, John Wiley & Sons, Inc. pp. 347, 1996.
[24] Kreith, F. & Bohn, M.S. Principles of Heat Transfer (Chapter 1). *Basic Modes of Heat Transfer*, Harper & Row, New York. pp. 12-13, 1986.
[25] Moffat, R., Describing the uncertainties in experimental results, *Exp. Thermal Fluid Science*, **1**, 3-17, 1988.

Modelling the energy systems in a PV powered two floor desert ZEH

M. A. Serag-Eldin
American University in Cairo, Egypt

Abstract

The paper presents an integral computational model for the prediction of the thermal performance of a conceptual two-floor, zero energy house (ZEH) in the Arabian desert. The ZEH is powered by PV modules which shade the roof during the day time and retract at night to expose it the sky, thus enhancing night time cooling. The house boasts all modern comforts, including air-conditioning. Solar radiation models coupled with recently published ASHRAE environmental data and models are integrated with a time dependent heat conduction model to predict the heating and cooling loads, for given equipment and storage characteristics. The application of the computational model is demonstrated by employing it to predict the effect of various design parameters on performance and equipment sizing, for a typical desert site in the Kingdom of Saudi Arabia.
Keywords: zero-energy-house, solar energy, desert environment, sustainability, modelling.

1 Introduction

The paper is concerned with the presentation and application of an efficient computational model for the investigation of the thermal performance of the integrated energy systems in a modern two floor, zero energy house (ZEH) located in the Arabian desert. Solar energy drives the whole energy system after being converted to electrical energy employing roof mounted PV modules. Several investigations have been reported in the past for single floor, roof mounted ZEH designs, e.g. Serag-Eldin [1] and Beshr et al. [2], as they are much less challenging to meet the zero external energy requirement than for two floor houses. A two floor ZEH design was considered by Serag-Eldin [3], however synthetic data was used for solar radiation and environmental properties and only

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110241

a crude presentation of long-wave atmospheric radiation was employed. Serag-Eldin [4] demonstrated the importance of accurate site data on prediction of performance; moreover the success of the present design depends to a large extent on night-time cooling by atmospheric radiation. Thus the need arose for more precise and site specific data, particularly in view of the delicate balance between supply and demand for two floor buildings. These issues are addressed in the present work.

The paper first describes the various models comprising the computational model; these include solar radiation and environmental models, a long-wave atmospheric radiation model, and equipment characteristics models. It then demonstrates use of the computational model by analyzing the impact of each of, the percentile design factor, room temperature, cooling-air peak temperature reduction by evaporative-cooling, heat exchange between fresh-air and discharge air, and COP characteristics, on the cooling load and equipment capacities.

2 Problem specification and description of ZEH

The home under consideration lies on a secluded desert plot in Dhahran, KSA, whose elevation is 17 m above sea-level, and whose angle of latitude is 26.27^0 North. The house features two floors, each 3.3 m high, with a ground foot print of 16.5 m x 16.5m. The roof-top displays 1 m overhangs on each side, in addition to top and side-shades surrounding each window. The window heights are each 1.5 m, and the total window area is 15% of total envelope area, assumed uniformly distributed. The windows display a U-value of 3.1 W/m^2K and a solar heat gain coefficient F=0.71.

During day time, the roof-top is covered by a reflective cover and Photovoltaic modules, about 0.2 m above the roof top, so as to totally shade the roof from solar radiation, whilst simultaneously collecting solar energy; during night time both the modules and cover are retracted to expose the roof top to the sky immediately above it, thus enhancing night time cooling, which is substantial in desert environments. The roof also includes good insulation material to reduce heat transfer to the indoors.

The building side walls comprise a 0.06 m external sandstone cladding, followed by two 0.12 m red brick walls separated by a 0.01 to 0.05 m air gap. Research has shown that due to the appearance of natural convection effects for air-gaps greater than 0.01 m, the overall thermal resistance in this range is almost constant, Kreider and Rabl [5]; hence calculations are made here for conductive resistance only, with a representative air gap of 0.01 m; however in practical implementation a 0.05m air gap would be more realizable. The internal wall is covered by a 0.03 m layer of plaster. The roof surface comprises a lower 0.02 m plaster layer, above which is a 0.12 m concrete layer, covered by 0.10 m of mortar and asbestos insulation, followed by a 0.06 m thick layer of gravel. The surrounding ground short wave reflectivity is assumed to be 0.25.

The room temperature is maintained constant by an electrically driven conventional air-conditioning system, which runs straight off the output of the PV modules during day time, and extracts power from battery storage during

night time. Displacement ventilation, e.g. Serag-Eldin [6], is employed to maintain air quality with reduced fresh-air rates (0.5 air-exchanges/h). The house is built to a high level of tightness, so that a 0.35 air-exchanges/h infiltration rate is expected. The conversion efficiency of the photovoltaic cells, η_c is assumed to be 15% at source; moreover an average combined electrical storage and distribution efficiency of 70% is assumed. The nominal number of occupants is 7; the electrical light-load is estimated to be 1 kW maintained uniformly over all night-time hours, whereas the total auxiliary electrical loads (washing machines, TV sets, computers,..) are expressed by a uniform 980 W load maintained over the 24 h. The latter two numbers were derived from considerations of the experimental and theoretical work of NAHB Research Center [7], after adapting them to the present house specifications following the procedure described by Serag-Eldin [1].

3 Mathematical models

The basic models employed comprise a solar radiation model, an environmental model, a heat conduction model, and an air-conditioning COP model; they are each presented here in turn.

3.1 The Solar radiation model

The solar beam angles are determined by first deriving the declination angle δ (degrees) for any day of the year from:

$$\delta = 23.45^0 \sin(2\pi(\frac{(n+284)}{365.25}) \tag{1}$$

where n is the number of the day of the year measured from January 1st.

Next the sun beam incidence (Zenith) angle θ_s on a horizontal plane is calculated from:

$$\cos\theta_s = \cos\lambda.\cos\delta.\cos\omega + \sin\lambda.\sin\delta \tag{2}$$

where ω is the solar angle. The sunset angle ω_{ss} is given by

$$\cos\omega_{ss} = -\tan\lambda.\tan\delta \tag{3}$$

The azimuth angle from due South, φ_s, is determined uniquely from the two equations:

$$\sin\varphi_s = \frac{\cos\delta.\sin\omega}{\sin\theta_s} \tag{4.a}$$

and

$$\cos\phi_s = \frac{\cos\omega.\cos\delta.\sin\lambda - \sin\delta.\cos\lambda}{\sin\theta_s} \tag{4.b}$$

For a vertical wall, incidence angle is given by:

$$\cos\theta_i = \sin\theta_s.\cos(\varphi_s - \varphi_w) \tag{5}$$

where φ_W is the azimuth orientation of the wall from due South.

Solar radiation is composed of two components, a direct (beam) component, I_{dir}, and a diffuse component, I_{diff}. They are derived from the following two empirical equations:

$$I_{dir} = I_0 e^{-[\tau_b m^b]}$$

and (6)

$$I_{dif} = I_0 e^{-[\tau_d m^d]}$$

where τ_b and τ_d are site specific values which vary over the year and are obtained from long period meteorological measurements; values of τ_b and τ_d are reported by ASHRAE [8] for more than 4000 sites world-wide, on the 21st day of each month. The mass exponents b and d are correlated to τ_b and τ_d through the following equations:

$$b = 1.219 - 0.043\tau_b - 0.151\tau_d - 0.204\tau_b\tau_d$$

and (7)

$$d = 0.202 + 0.852\tau_b - 0.007\tau_d - 0.357\tau_b\tau_d$$

I_0 is solar irradiance outside atmosphere at normal incidence expressed by the following equation:

$$I_0 = 1367[1 + 0.033\cos(360^0(n-3)/365)] \tag{8}$$

and m is the relative air-mass ratio, representing the actual mass of atmosphere in the sun-to-earth path to that which would exist if the sun were directly overhead. It is a pure function of solar angle and is calculated from the following equation:

$$m = \frac{1}{\cos\theta_s + 0.50572(6.07995 + (90-\theta_s)^{-1.6364}} \tag{9}$$

The total incident radiation on a horizontal surface, $I_{glo,hor}$, is expressed by:

$$I_{glo,hor} = I_{dir}\cos\theta_s + I_{diff} \tag{10}$$

The total solar energy, $I_{glo,vert}$ received at a vertical wall is expressed by:

$$I_{glo,vert} = I_{dir}\cos\theta_i + I_{glo,hor}\rho_g/2 + \max(0.45, 0.55 + 0.437\cos\theta_i + 0.313\cos^2\theta_i)I_{diff} \tag{11}$$

where ρ_G is the ground reflectivity. The first term on the right hand side presents the beam radiation contribution, the second the ground reflected solar radiation, whereas the last one presents the non-isotropic diffuse radiation; the anisotropy is caused by some parts of the sky being brighter than others.

3.2 Heat conduction model

Heat is transferred across the solid medium (walls, roof and glazing) by conduction, mainly in the direction normal to the solid surface, x. Therefore, the temperature T and heat fluxes are obtained by numerical solution of the following unsteady heat conduction equation through the solid mediums of density ρ, conductivity, k, and specific heat c :

$$\frac{\partial}{\partial t}(\rho c T) = \frac{\partial}{\partial x}(k \frac{\partial T}{\partial x}) \tag{12}$$

subject to boundary conditions which express heat convection with environment and both solar and long-wave radiation exchange with the environment and nearby surfaces. Heat transfer coefficients are estimated from appropriate empirical relations adopting a nominal free wind velocity V= 3m/s.

The control-volume method is employed to derive the solution of the heat conduction differential equation, using second order differencing in both space and time. A total of 360 control-volumes were employed across the side walls and ceiling surface. A time step of 0.05 hours was adopted for the final runs. Grid-dependence tests showed no variations within at least 4 significant figures. Because of the non-linearities associated with the radiation terms at the external surfaces, air gaps and room boundaries, iteration is necessary. Moreover, to truly reach 24 hour (day) periodicity, and eliminate effect of initial (starting solution) temperatures, many repetitions are necessary. Iteration for non-linearity and cycle repetitions were handled simultaneously in the solution algorithm, by successively repeating the calculations of a full day, each time starting with the previous day results. Cycle repetitions were stopped only after maximum changes in temperature values were all within 6 significant figures.

3.3 Environmental models

These are classified here as two models, as follows:

3.3.1 Ambient temperatures model

For thousands of meteorological sites, ASHRAE [8] provides fairly detailed climatic design conditions which may be used to interpolate hourly variation of wet bulb and dry bulb temperatures. Since it is not practical to design for extreme conditions which occur very infrequently (e.g. highest dry bulb temperature in 50 years), practical HVAC design caters for meeting comfort conditions for all but a specified percentage of the year. ASHRAE provides design values for 0.4%, 2%, 5% and 10% percentiles, corresponding to 35, 175, 438 and 876 h per year, respectively. For a specified design percentile, the procedure for estimating the instantaneous temperature profiles is as follows:

i- the monthly design dry bulb temperature T_{db}, coincident wet bulb temperature T_{wb}, and their mean daily temperature ranges, $T_{db\text{-}r}$, $T_{wb\text{-}r}$, respectively, are read from the ASHRAE table for the specified location.

Since cooling load is considered, T_{db} represents the peak daily temperature $T_{db,max}$ to be employed in cooling load calculations. Like-wise, T_{wb} represents $T_{wb,max}$.

ii- the hourly values of T_{db} and T_{wb} are calculated from:

$$T_{db} = T_{db,max} - \beta\, T_{db\text{-}r} \quad \text{and} \quad T_{wb} = T_{wb,max} - \beta\, T_{wb\text{-}r} \tag{13}$$

where β is a daily temperature profile shape factor which varies between 0 and 1 [8]; Hedrick [9] and Thevenard [10] showed it to represent accurately both dry-bulb and wet bulb temperature variations on design days.

iii- instantaneous values are interpolated from hourly values using cubic splines.

iv- the instantaneous dew-point temperatures T_{dp} in ^{0}C, are derived from the instantaneous values of T_{db} and T_{wb} employing the following psychrometric relations:

$$T_{dp} = 6.54 + 14.526 \ln p_w + 0.7389 (\ln p_w)^2 + 0.09486(\ln p_w)^3 + 0.4569\, p_w^{0.1984} \qquad T_{dp} \geq 0$$
$$T_{dp} = 6.09 + 12.608 \ln p_w + 0.4959 (\ln p_w)^2 \qquad T_{dp} < 0 \tag{14}$$

where p_w is the water vapour partial pressure in the air, kPa. It is related to the humidity ratio W through the relation:

$$p_w = W\, p/(0.621945+W) \tag{15}$$

where p is the atmospheric air pressure, kPa, and W is derived from:

$$W = [(2830 - 0.24\, T_{wb})\, W_s(T_{wb}) - 1.006(T_{db}-T_{wb})]/[2830 + 1.86\, T_{db} - 2.1\, T_{wb}] \qquad T_{db} < 0$$
$$W = [(2501 - 2.326\, T_{wb}) W_s(T_{wb}) - 1.006(T_{db}-T_{wb})]/[2501 + 1.86 T_{db} - 4.186\, T_{wb}] \qquad T_{db} \geq 0 \tag{16}$$

where $W_s(T_{wb})$ is saturation humidity ratio at T_{wb}. It is related to p_{ws} at wet-bulb temperature $p_{ws}(T_{wb})$ through the relation:

$$W_s(T_{wb}) = 0.621945\, p_{ws}(T_{wb})/(p - p_{ws}(T_{wb})) \tag{17}$$

the latter is obtained from :

$$\ln[p_{ws}(T_{wb})] = C_1/T_{wb} + C_2 + C_3 T_{wb} + C_4 T_{wb}^2 + C_5 T_{wb}^3 + C_6 T_{wb}^4 + C_7 \ln T_{wb} \qquad T_{wb} < 0$$
$$\ln[p_{ws}(T_{wb})] = C_8/T_{wb} + C_9 + C_{10} T_{wb} + C_{11} T_{wb}^2 + C_{12} T_{wb}^3 + C_{13} \ln T_{wb} \qquad T_{wb} \geq 0 \tag{18}$$

where T_{wb} is expressed in K and C_1 .. C_{13} are constants whose values are displayed in Table 1.

Table 1: Constants for p_{WS} calculation.

C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	C_{11}	C_{12}	C_{13}
-5674.5	6.3925	-9.6778	6.2216×10^{-7}	2.0748×10^{-9}	-9.4840×10^{-13}	4.1635	-5800.2	1.3915	-4.8640×10^{-2}	4.1765×10^{-5}	-1.4452×10^{-8}	6.5459

3.3.2 Atmospheric radiation model

Long wave radiation exchange between roof and sky, known as atmospheric radiation, is a key factor in the success of the proposed ZEH. Therefore accurate estimate of its magnitude is of high importance. The net long-wave radiation exchange between roof and sky q_{rs} is expressed by the term:

$$q_{rs} = \varepsilon_r \sigma (T_r^4 - \varepsilon_a T_a^4) \quad (19)$$

where ε_r is the long wave emissivity of the roof surface or PV/cover material, T_r is the temperature of the roof surface or PV/cover material and T_a is the ambient temperature. Whereas ε_a is the apparent emissivity of the atmosphere, according to Bliss [11] and ASHRAE [12]; it is defined as the ratio of the atmospheric radiation on a horizontal surface per unit area to (σT_a^4), and is only a function of T_{dp} near ground. The latter is derived as shown in the previous section. Values of ε_a at given T_{dp} increments were deduced from Bliss's graphical display, then cubic splines were used to interpolate intermediate values.

3.4 Air conditioning model

In a desert environment water is scarce; hence cooling of the air conditioning system condenser must rely primarily on ambient air. The COP of the air-conditioning system suffers heavily with a rise of cooling media temperature; indeed some vapour compression systems will shut down at ambient air temperatures above 45^0C to protect themselves. However peak summer temperatures in the desert often reach 50^0C.

Thus for exceptionally hot summer days, it is proposed here to use small amounts of water for cooling of condenser coils by misting and evaporative cooling; just enough to reduce the effective cooling air temperature to an upper limit, T_{ul}. This water need not be potable, and may be the accumulated output of treatment of residential waste water, stored over a long period of time (extending to cooler seasons if need be). The model employed here allows the calculation of the daily cooling water usage for a pre-set T_{ul}. Setting T_{ul} above ambient temperature naturally deactivates this feature.

Whereas different models and makes display different COP characteristics, two characteristic profiles are considered here, one based on the average of currently operating European residential units, ECODESIGN [13], and the other a state of the art unit displaying an electronically controlled expansion valve, EEV [14]. The two characteristics are displayed simultaneously in Fig. 1.

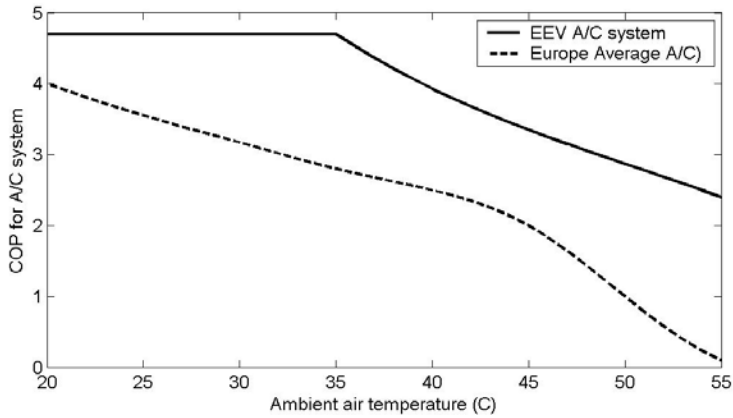


Figure 1: COP for EEV and European average.

In order to reduce cooling load, the fresh air is cooled prior to entering indoors by exhaust air leaving at room temperature, in a counter flow heat exchanger. Thus the sensible heat load is reduced, the gain increasing with increase of outdoor temperature. Since the heat capacity of the return air is at least as large as that of the fresh air(actually larger because of infiltration), theoretically it is possible in a very long double pipe heat exchanger to cool the fresh air down to room temperature and eliminate this component of the sensible heat load altogether. However, for practical reasons a relatively short heat exchanger is employed and it is assumed that the effectiveness of the heat exchanger, $\varepsilon_{HX} = 0.7$; i.e. 70% of cooling of fresh air occurs in heat exchanger.

4 Results

Figures 2-3 present results of the basic case, which employs a 2% design factor, the EEV COP characteristic, evaporative cooling on exceptionally hot days to reduce peak cooling air temperature to 43^0C , and room temperatures of 25^0C and 22^0C for June and December months, respectively; without cooling of fresh-air via a heat exchanger. They display the distribution of the cooling loads for the envelope walls, windows and auxiliary loads (fresh air, persons, infiltration, equipment, ..) as well as their sum, for June 21st and December 21st, respectively.

Table 2 displays the results of various parameters on cooling load electricity consumption (kW.h/day), auxiliary appliances and lighting electrical loads (kW.h/day), and the total daily output of the PV modules (kW.h/day). It also displays the required equipment rating for each case, namely battery storage capacity (kW.h), PV module capacity (kW) and the cooling equipment rating (kW).The first three rows display results of the basic design case for three key days representing spring/autumn, summer and winter. It is apparent that equipment rating depends only on the requirements of June 21st; only 69.8% of the roof area is required to be covered by PV modules for June 21st, which still leaves ample roof area for other equipment.

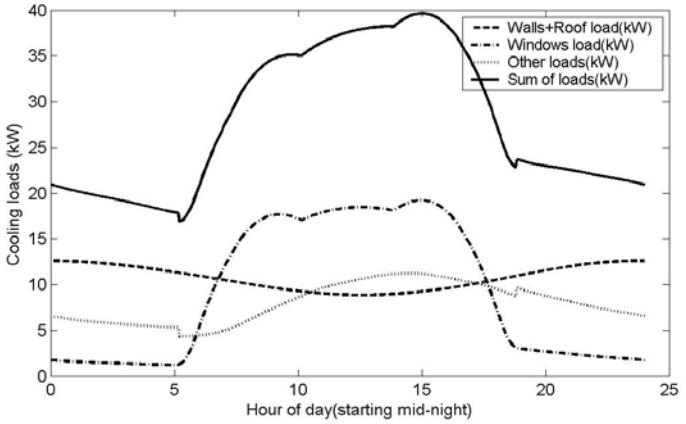


Figure 2: Cooling loads on June 21st.

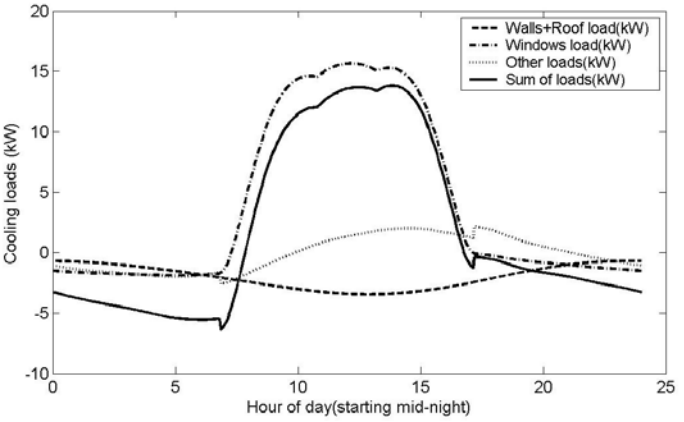


Figure 3: Cooling loads on December 21st.

Rows 4–6 display the above results for June 21st, for 3 different ASHRAE design percentile values, namely 0.4%, 2% and 5%. The lower the percentile value the higher the peak values of the ambient temperature; hence the larger the heat gain by convection at the envelope surface, sensible heat of fresh-air and infiltration, and long-wave atmospheric radiation. Thus the cooling load increases from 592 kW.h/day for 5%, to 635 kW.h/day for 2%, to 691 kW.h/day for 0.4%, corresponding to cooling equipment capacities of 36.5 kW, 38.3 kW and 40.6 kW, respectively. Although, the basic case calculations employ 2%, as this is the more common practice, using a 5% value is justified here since our calculations assume that the design day conditions are continuously repeating themselves every 24 hours; whereas in reality the previous day should be cooler since the design criteria are for the hottest day of the month. Since our building

Table 2: Effect of design parameters on loads and equipment capacities.

Row number	Case Description	Cooling load (kW.h/day)	Electric appliances + lighting (kW.h/day)	Daily PV output (kW.h/day)	Min battery-storage (kW.h)	PV modules rated capacity (kW)	Cooling equipment capacity(kW)
1	base case, March 21st	372.7	35.6	115.6	53.6	26.5	28
2	base case, June 21st	668.7	33.87	197.2	83.8	36.4	39.6
3	base case, Dec. 21st	91.9	37.2	65.3	36.3	25.7	13.8
4	design percent,0.4%	691.4	33.87	206.5	88.3	39.1	40.6
5	design percent,2.0%	634.8	33.87	189.2	80.0	35.8	38.3
6	design percent,5.0%	592.0	33.87	176.0	73.9	33.3	36.5
7	cooling-air limit 40^0	634.8	33.87	183.3	79.4	34.7	38.3
8	cooling-air limit 43^0	634.8	33.87	189.2	80.0	35.8	38.3
9	cooling-air limit 46^0	634.8	33.87	190.6	79.9	36.1	38.3
10	room temp. 23^0C	700.8	33.87	204.9	87.7	38.8	41.1
11	room temp. 25^0C	634.8	33.87	189.2	80.0	35.8	38.3
12	room temp. 27^0C	568.7	33.87	173.5	72.2	32.9	35.5
13	COP Europe average	634.8	33.87	279	113.	52.9	38.3
14	COP of EEV A/C	634.8	33.87	189.2	80.0	35.8	38.3
15	EEV + fresh-air HE	579.8	33.87	175.5	74.8	33.2	34.5

envelope displays exceptionally high thermal capacity, the effect of a cooler previous day on cooling loads is to reduce cooling load considerably.

Rows 7–9 reveal the variation of equipment capacities and loads with variation of allowable peak cooling-air temperatures, controlled by using small amounts of evaporative cooling to lower ambient temperature to the designated peak values. The lowering of peak cooling air temperature affects only the size of equipment, via raising the COP, but has no bearing on the cooling loads. The maximum ambient air temperature for this design day is 44.5^0C occurring at 14.5 hour solar time. Thus no lowering of ambient air peak temperature occurs with a peak cooling air temperature of 46^0C. The water consumption rate for the 43^0C and 40^0C peaks are 69.2 kg/day, and 369.5 kg/day, respectively. It is apparent that the gains introduced by reducing the cooling air temperature peaks are small, unless the peaks are reduced substantially, in which case the water consumption may be unacceptably high; from a cooling perspective it may then be more efficient to use it in an indoor evaporative cooler. However, it may be required to restrict the peak cooling air temperature just to protect the cooling equipment and stopping it from shutting down. Thus a peak of 43^0C is adopted for base case, which should correspond to a condenser temperature of around 48^0C.

Rows 10-12 reveal the variation of equipment capacities and cooling load with specified internal room temperature for June 21st. As the room temperature rises from 23^0C to 25^0C and then to 27^0C, the cooling load decreases from 701 kW.h/day, to 635 kW.h/day and to 569 kW.h/day, respectively; with corresponding drops in PV and cooling equipment capacities. This drop is substantial. A room temperature of 27^0C in a desert environment may be quite acceptable since the relative humidity is exceptionally low; moreover, outdoor temperatures are exceptionally high in summer and hence local summer clothing is generally made of porous cotton and loose fitting; well adapted to high temperatures with low humidity.

Rows 13,14 show the effect of COP on performance, the two COP characteristics corresponding to the ones displayed in Fig. 1. The impact of COP on capacity of equipment (PV modules, battery and air-condition) is remarkable. For the European averaged COP, the required PV rated capacity is 53 kW, whereas for the EEV COP the corresponding output is only 36 kW. Adding an exhaust-air/fresh-air heat exchanger reduces the latter to 33 kW. Since PV modules and battery storage are expensive items, it is expected to be well worth it to purchase state of the art, highly efficient air-conditioning systems. The last row presents results of using a fresh-air heat exchanger of effectiveness 0.7 on capacities; it produces an attractive 10% saving in PV capacity alone.

Further savings may be introduced if clothes are dried outdoors rather than using an indoor electric clothes drier. According to [7] the energy consumed in clothes drying is at least 25% of the total 980 W estimated here, i.e. 245 W. Indeed, a solar clothes drier [15] may even be employed to lower peak air temperatures. If the load for the internal electric clothes drier is removed, the basic case PV output will be reduced by about 6 kW.h/day, i.e. 3%.

5 Summary and conclusion

The paper presents a complete mathematical model for the prediction of the performance of a desert ZEH with roof mounted PV modules. The model makes use of ASHRAE meteorological data which is available for thousands of meteorological sites world-wide; thus reducing the need for prior site measurements.

The model is employed to investigate the effect of various design and operation parameters on the capacity of the PV modules, cooling equipment and battery storage. It is demonstrated that under optimum design conditions the capacity requirements may be reduced substantially over that of the basic design.

Acknowledgements

This work was funded by King Abdalla University for Science and Technology (KAUST) project on Integrated Desert Building Tech, grant held by AUC.

References

[1] Serag_eldin, M.A., Thermal design of a modern, air-conditioned, single-floor, solar-powered desert house, in press, *Int. J. of Sustainable energy, 2011*.
[2] Beshr, M., Khater, H. and Abdel Raouf, A., Modeling of a Residential Solar Stand-Alone Power System, *Proc. of 1st Int. Nuclear and Renewable Energy Conf.*, Amman, Jordan, March 21-24, 2010.
[3] Serag-Eldin, M.A., Thermal Design of a Modern, Two Floor, Zero Energy House in a Desert Compound, *Proc. Thermal Issues in Emerging* Technologies, Theta-3, Cairo, Egypt, Dec 19-22, 2010.
[4] Serag-Eldin, M.A., Influence of site on thermal design of a two floor ZEH in the desert, Theta-3 Conference, Cairo, December 19-21, 2010.
[5] Kreider, J. and Rabl, A. 1994, *Heating and Cooling of Buildings: Design for Efficiency*, McGraw-Hill Inc., New York, pp.257.
[6] Serag-Eldin, M.A., Displacement Ventilation for Efficient Air-conditioning and Ventilation of GYMs, *Proc. of Heat-SET-2007 Conference, Chambery, France*, 18-20 April, 2007, paper # P-128.
[7] NAHB Research Center, Inc., "Zero Energy Home Armory Park Del Sol, Tucson, Arizona", *Final report submitted to NREL*, 1617 Cole Boulevard, Golden, CO 804101-3393, June 30th, 2004.
[8] *ASHRAE Fundamentals Handbook*, 2009, American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., USA.
[9] Hedrick, R. Generation of Hourly Design-Day Weather Data (RP-1363). ASHRAE Research Project, Final Report, 2009.
[10] Thevenard, D. Updating the ASHRAE Climatic Data for Design and Standards, (RP -1453). ASHRAE Research Project Report, 2009.
[11] Bliss, R.W., Atmospheric Radiation Near the Surface of the Ground: A Summary for Engineers, *Solar Energy* 59(3), pp.103-120.
[12] ASHRAE HVAC Applications *Handbook*, 2007, American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., USA.
[13] ECODESIGN-Preparatory Study on the Environmental Performance of Residential Room Conditioning Appliances, *Draft report Task 4*, July 2008, Contract TREND/D1/40-2005/LOT10/S07.56606.
[14] Chinnaraj, C. and Govindarajan, P., Performance Analysis of Electronic Expansion Valve in 1 TR Window Air Conditioner using Various Refrigerants, *I. J. of Eng. Sc. and Tech.*, Vol. 2(9), 2010, 4020-4025.
[15] Suntivarakorn, P., Satmaromg, S., Benjapiyaporn, C. and Theerakulpisut, An experimental Study on Clothes Drying Using Waste Heat from Split Type Air Conditioner, World Academy of Sc., Eng. and Tech., 53, 2009, pp.168-173.

Numerical simulation of blowing ratio effects on film cooling on a gas turbine blade

H. A. Daud[1], Q. Li[1], O. Anwar Bég[1] & S. A. A. AbdulGhani[2]
[1]*Department of Engineering and Mathematics, Sheffield Hallam University, UK*
[2]*Mechanical and Industrial Engineering, Qatar University, Qatar*

Abstract

This article investigates the film cooling effectiveness and heat transfer in three regimes for a film-cooled gas turbine blade at the leading edge of the blade with 45^0 angle of injection. A Rolls Royce blade has been used in this study as a solid body with the blade cross section from Hub to Shroud varying with a degree of skewness. A 3-D finite-volume method has been employed (FLUENT 6.3) with a $k-\varepsilon$ turbulence model. The numerical results show the effectiveness cooling and heat transfer behavior with increasing injection blowing ratio BR (1, 1.5 and 2). In terms of the film cooling performance, high BR enhances effectiveness cooling on pressure side and extends the protected area along the spanwise direction from hub to shroud. The influence of increased blade film cooling can be assessed via the values of Nusselt number in terms of reduced heat transfer to the blade.

Keywords: turbine blade, film cooling, blowing ratio, CFD, heat transfer, effectiveness cooling.

1 Introduction

Increasing the thrust, overall efficiency and reducing the fuel consumption as much as possible are major issues in modern gas turbine engineering, and this is generally achieved via increasing the turbine inlet temperature. These higher temperatures however have detrimental effects on the integrity of high pressure turbine components and materials composing the turbine blades. Film cooling technology is justified to protect blades surfaces from incoming hot gas and for increasing life time. Numerical and experimental studies of three-dimensional

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110251

film cooling for real turbine blade models are much less common than the flat plate models which deal with injection of a coolant fluid. In addition, rotating blade studies compared with the stationary blades are rare. The film cooling method is an efficient way to protect the blades of turbines from the hot gas and is widely used in contrast with Impingement Cooling, Transpiration cooling, and Convection Cooling. Numerous of the prior studies examined aerodynamics, heat transfer and film cooling over a flat, curved plate in two dimensions. For example, Kadja and Bergeles [1] presented numerical computations of blade turbine cooling via injection of a fluid into the main stream flow. The study was conducted via creating a slot along the flat surface. Typically, the local film cooling effectiveness increased with increasing blowing ratio and the Nusselt number was decreased near the slot. Hung *et al.* [2] studied 3-D effectiveness cooling and heat transfer over concave and convex plates using the liquid crystal thermograph technique. The performance of film cooling was investigated with one row of injection holes (angle of injection 35^0) for various blowing ratios (BR=0.5, 1, 1.5 and 2). The concave model results showed that with increasing blowing ratio the heat transfer coefficient and film cooling effectiveness increase and the convex model surface results demonstrated that the compound angle (0^0, 45^0 and 90^0) increases both heat transfer and film cooling effectiveness at moderate and high blowing ratio. The thermal effect of the turbine blade has been studied for a NACA 0021 airfoil by Kassim *et al.* [3]. Good agreement was obtained between the experimental and computational results using 6.2 Fluent code for the different angles of attack, different jet angle and velocity ratio on the penetration area. Lakehal *et al.* [4], Theodoridis *et al.* [5] utilized a simple geometry in their study of a stationary blade (symmetrical turbine blade). Two dimensional and three dimensional finite volume simulations were employed to model film cooling holes at the leading edge of the blade with the FAST 3D CFD code. The computational results for different blowing ratios (BR = 0.3 to 1.1) were compared with experimental data including isentropic Mach number, velocity and pressure fields. Characterizations of the gas flow and heat transfer through a modern turbine are important engineering considerations, which represent very complex flow fields and high turbulence levels. Forest *et al.* [6] described the influence of hole shape on heat transfer coefficients and film cooling effectiveness through a series of experimental measurements compared with a heat transfer design code (HTDC). A computational method was used to simulate transonic flow in a Rolls Royce blade. Flared hole film cooling and heat transfer coefficient results were found to correlate poorly on the blade pressure side; on the suction side flared holes were found to provide increasing effectiveness cooling compared with cylindrical holes.

Improving the arrangement of holes for the film cooling in rotating turbine blade has a strong influence on aero-thermodynamics. Burdet and Abhari [7] and also Garg and Abhari [8] predicted Nusselt numbers for such regimes. Their computations were based on an explicit finite volume code with an algebraic turbulence model and implemented in excess of 2.5 million grid points to calculate the flow. CFD results were compared with experimental data to predict transonic film flow. These studies highlighted the macro-flow structure for the

interacting coolant jet process especially near the hole region, at different hole rows. On the pressure surface the results generated however were poor.

After reviewing the previous published studies, the majority of studies communicated hitherto have focused on 2-D or 3-D aerodynamic flow and heat transfer for a simple blade geometry, flat or curved plate, NACA 0021, symmetrical turbine blade and simple cross section blade from hub to tip. Consequently, this paper aims to extend these studied by focusing on:

- Using different cross section blade geometry (from hub to shroud) with angle of twist.
- The main flow (hot gas) and coolant system (cooled fluid) differing in temperature, pressure and chemical composition for the hot gas and cooled air.
- Solid body thermal properties which will be simulated by inserting the type of material type e.g. carbon steel (SAE4140) in the FLUENT software property specification pre-processor. The blade taken in the previous studies as a *shell surface* is shown in Fig. 1.
- Aerodynamic flow and heat transfer in modern gas turbine constitutes a very complex flow field with high turbulence levels; therefore we apply film cooling in this paper from the blade leading edge. This is additionally complicated due to the resulting interference between the main flow and injected coolant.

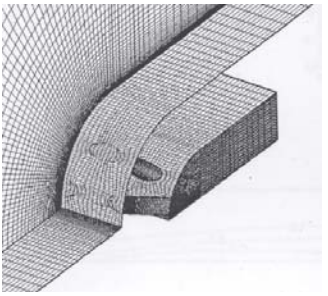


Figure 1: Previous studies defined blades as shell surface with cooling holes Azzi and Jubran [9].

The motivation of this study is to achieve robust numerical predictions of film cooling effectiveness, heat transfer, temperature distribution on a skewness solid blade (complex geometry), effect of coolant fluid property in the hub, mid and shroud area prior to further experimental tests which are currently being designed at the Sheffield Hallam University Engineering Laboratory, and which will be communicated imminently.

2 Computational flow model

Computation Fluid Dynamics (CFD) can analyze internal and external aerodynamic flow, heat and mass transfer and also multiphysical problems by solving three dimensional flows, steady state mathematical equations as

continuity equation; Reynolds averaged Navier- Stokes (RANS) flow equations and the energy equation via numerical algorithms on a computer. The transport equation is additional term utilize to simulates a non reacting gas mixture when this gases jet in to a crossflow at different density (Renze *et al*. [10]). However, the (k – ε) equation is the simplest turbulence model employed widely in engineering simulations, for which the initial boundary conditions need to be supplied. These involve algebraic approximations for turbulence kinetic energy (k) and its rate of dissipation (ε) via a semi-empirical model.

The numerical computations in this paper were completed via a series of steps, including grid checking, model solver definition (segregate implicit CFD technique) selection and turbulence simulation with the standard RNG (k – ε) model. Eghlimi, *et al*. [11] applied Renormalization group (RNG) model in Fluent code to predict the turbulent gas flow and also Lakehal [12] has elucidated that there is little advantage to using different viscosity data for (RSM) Reynolds stress turbulence model as compared with the eddy viscosity model. In consistency with these we implement herein turbulence (or eddy) viscosity constants as follows: $\mu_t = \rho C_\mu \frac{k^2}{\varepsilon}$, $C_{1\varepsilon}$: 1.42, $C_{2\varepsilon}$: 1.68, C_μ: 0.0845.

3 Turbine blade geometry and mesh

Fig. 2 illustrates the Rolls Royce blade model with a maximum chord length, C_X of 52 mm, where (C_X) represents blade axial chord. The maximum width C_Y is 27.3 mm (C_Y is the blade true chord); in addition the film cooling geometry contains, on each side, one row of holes located near the leading edge of the blade. Lateral holes have a diameter of 4 mm, within inclined angle of 9.13^o along the span and the holes diameter (d) is 1 mm with a lateral spacing of *5d*. Therefore, the number of holes on the pressure side and suction side are 42 with an angle of injection of 35^o.The incoming hot gas simulation box is represented in dimensions (4.8*2.11*3.85) Cx.

The meshing process of these complex volumes is performed using an unstructured mesh and the best way to control size mesh intervals for edges and

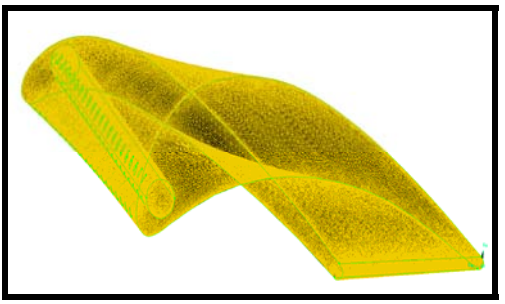


Figure 2: Local mesh of blade model as a solid body with cooling holes system.

mesh elements for face and volumes are the selection option size function from the tool mesh.

In this paper a high quality of mesh for the blade, hot gas and fluid coolant is achieved via the multi-block method with fewer cells. Therefore, a size function is employed for meshing the volumes to control the size, growth rate and (owing to the geometric complexity in the Rolls Royce blade) the holes geometries. Tetrahedral elements were used to complete the volumetric mesh and the final mesh generated for the three volumes contains in excess of 10 million cells (elements). All simulations were executed until the solution convergence factor $1*10^{-5}$and $1*10^{-7}$ were satisfied on the energy equation. The solution has been controlled by the select SIMPLIE algorithm to solve velocity-pressure coupling with an implicit procedure, such that more than 3500 iterations are sustained to convergence (solution stability).

4 Boundary conditions

The initial boundary condition details are inserted in the code the inlet main velocity at 163.4 m/sec and the mass flux (blowing) ratio values for the plenum are 1, 1.5 and 2. In addition, the ratio of the hot gas temperature to the cooled air temperature was specified as ($T_\infty/T_C = 1.6$) where; (T_C) is the coolant temperature and (T_∞) designates the incoming hot gas temperature ($T_\infty = 460$ K). In the numerical simulation the outlet flow was defined as a static pressure and the turbulence flow (T_u) was calculated as a function of the Reynolds number. The Reynolds number (R_e) is approximately $2.57*10^5$ based on the maximum axial chord of the blade model. So, T_u can be written as defined as:

$$T_u=0.16*Re^{(-1/8)} \tag{1}$$

The Rolls Royce blade is simulated as a solid body. The blade material properties are demonstrated in table 1.

Table 1: The blade material properties.

Material type (low carbon steel)	Density (ρ) kg/m^3	Thermal conductivity (k) W/m-K	specific heat (C_P) J/kg-K
SAE 4140	8030	42.7	473

5 Result and discussion

Fig. 3(a), 3(b) and 3(c) illustrate the distribution of the static temperature for the solid blade from the hub to shroud regions and from the blade leading edge to the trailing edge. The effects of blowing ratio (BR) on the distribution of blade

temperature (T) via injecting coolant fluid into main hot gas are very clear. Blowing ratio (BR) in this paper can be calculated by:

$$BR = \rho_C * V_C / \rho_\infty * V_\infty \tag{2}$$

Fig. 3(a), 3(b) and 3(c) depict contour temperature distributions for blowing ratio BR=1, 1.5 and 2, respectively. The colour temperatures have been graduated from low to high temperature (leading edge to trailing edge area). Consequently, with increasing of mass flux ratio (BR) the blade surface temperature will be reduced on both the pressure side and the suction side and also from the hub to shroud. Thereby, the protected area of the blade from incoming hot gas will be increased with increasing blowing ratio. In this study film cooling technology will be studied firstly- convection cooling though blowing coolant fluid in to the lateral hole from hub to shroud (internal cooling). Secondly, coolant fluid will be injected for both the pressure and suction side as a secondary fluid to create a blanket above the blade surface (external cooling) simultaneously. According to Fig. 3, the drops in predicted blade temperature

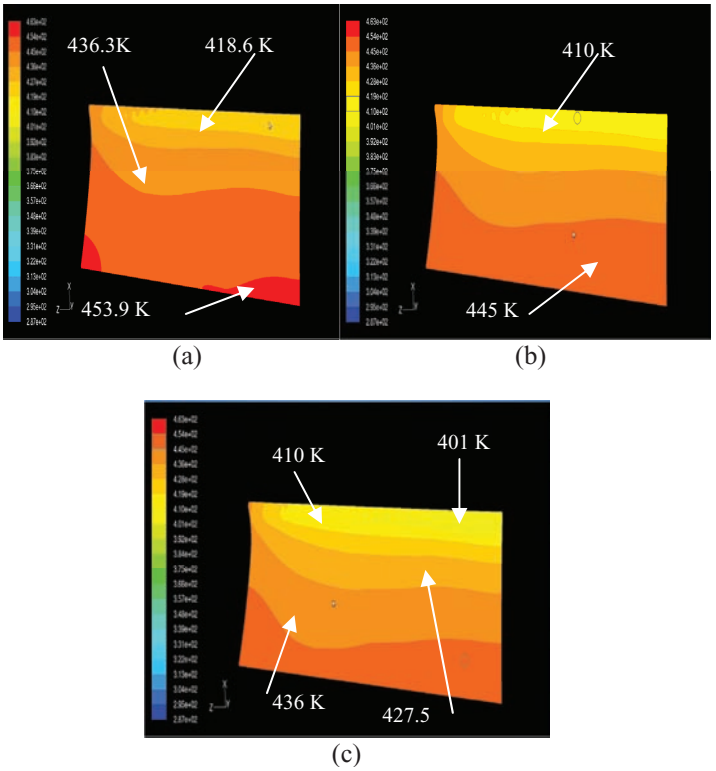


Figure 3: Predicted counters temperatures of the blade model at different blowing ratio (a) BR=1, (b) BR=1.5, (c) BR=2.

with rise of blowing ratio (BR) indicate that the effectiveness cooling will be enhanced from the hub to the shroud and from the leading to trailing edge, a result which correlates very well with the computations of Kadja and Bergeles [1].

A unique feature of our study is to demonstrate the effect of cooling on the blade as a solid body. In Fig. 4 therefore presented the trajectory of temperature and effectiveness cooling distribution for both pressure side and suction side on the solid body blade (always the effectiveness will be reversed trajectory compares with Temperature trajectory). In addition the hub, mid and shroud area temperature at blowing ratio BR=2 are also illustrated. The temperature along the span of the blade can be detected from distribution of temperature at the hub area to mid and from the mid to shroud area. Invariably the hub area temperatures will be lower than the mid and shroud area temperatures at the leading edge region. This is a result of the coolant fluid being blowing (injected) from the blade base (hub area). In the midspan section the predicted temperatures curve descends much more than the hub temperature profile at X/Cx>0.25 on the pressure surface and at X/Cx< -0.02 on the suction side Fig. 4. This pattern is due to the camber of blade and the angle of twist- the blade cross section with the span and axial cord at each section changes from hub to shroud (blade design shape).

Local film cooling effectiveness (η) is analyzed in this paper for both blade sides (pressure and suction side) as a function of (T_∞), (T_W) wall temperature and (T_C) using this equation:

$$\eta = (T_\infty - T_W)/(T_\infty - T_C) \tag{3}$$

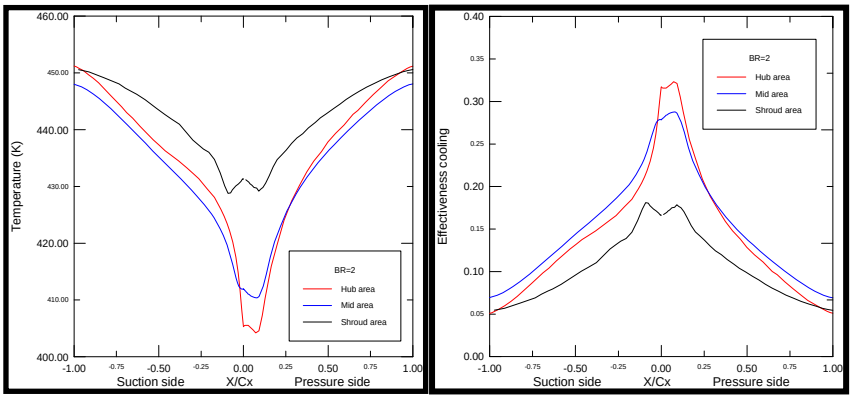


Figure 4: Temperature and the effectiveness cooling distributions difference in hub, mid and shroud area on the blade model by injecting coolant air at BR=2.

Fig. 5 shows the effectiveness cooling (η) in three regimes (hub, midspan and shroud), for different blowing ratios, BR=1, 1.5 and 2. Fig. 5(a) displays the effects of increasing blowing ratio (BR) from 1 to 2 at the hub area. Significantly, with increasing blowing ratio from BR=1 to 1.5 the effectiveness cooling will be increased by about 20.1% and 15.2% through respective increases in BR to 1.5 and 2 (near the leading edge of blade). The pressure side effectiveness cooling is higher than suction side at this regime (high BR enhanced film cooling on pressure side), a trend in agreement with Burdet and Abhari [7], and Tao *et al.* [13] attributable to the similar angle of injection (45^0) and lateral holes inclined (9.13^0) design.

In the midspan area the effectiveness cooling is illustrated in Fig. 5(b) and it also increases corresponding to bowling ratio by about 21.3% and 13.6% through BR=1 to 1.5 and from BR=1.5 to 2. At a blowing ratio BR=1 the effectiveness cooling near the injection zone (leading edge zone) is higher on the

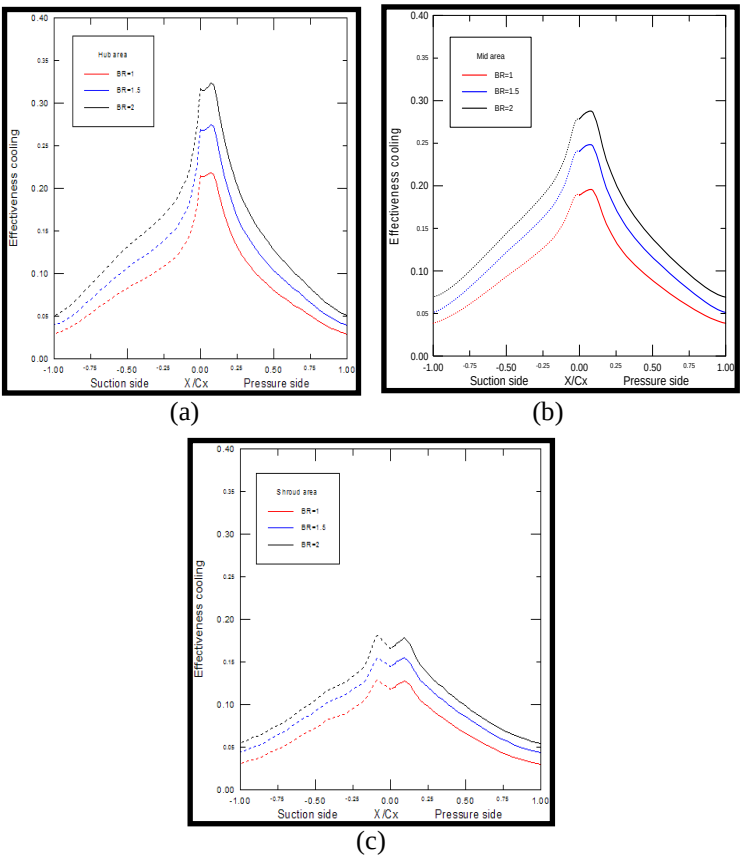


Figure 5: Effects of blowing ratio (BR) on the blade effectiveness cooling- (a) hub, (b) mid and (c) shroud.

suction side than that on the pressure side, but with increasing BR, the film cooling effectiveness on the pressure side will gradually increase compared with the suction side.

Fig. 5(c) shows variation of effectiveness cooling with blowing ratio in the shroud regime. The trajectory of the pressure and suction side is nearly analogous (matched). About 18.3% and 12.7% film cooling effectiveness will be enhanced with increase blowing ratio changing from BR=1 to 1.5 and from BR=1.5 to 2 respectively. The film cooling effectiveness contours for the pressure and suction side on the blade model at blowing ratio BR = 2 is shown in Fig.6.

The values of effectiveness cooling change gradually form the leading edge to trailing edge (colder to hottest place). Briefly, when the values of blowing ratio (BR) are reduced the hottest area will be increased on both sides and the blade will be exposed to incoming hot gas.

A good correlation has been found between previous studies e.g. Kadja and Bergeles [1] and Guangchao *et al.* [14] and other published paper results on film cooling effectiveness studies with cylindrical hole shapes. With increases in values of BR the blade surface temperature drops and the effectiveness cooling is enhanced. BR is proportional to temperature difference (ΔT) and inversely proportional to effectiveness cooling (η).

Fig.7 shows the comparison between the CFD prediction of Burdet *et al.* [15] and our computed film cooling effectiveness (η) on the blade model near hub position for BR=1 at angle of injection 35^0 . At first glance, the values of effectiveness cooling (η) at X/d < 0.3 on the pressure side are significantly higher than Burdet *et al.* [15], which is beneficial. On the suction side at X/d < 0.18 the values of (η) seems to coincide with Burdet *et al.* [15] while, the profile of the effectiveness cooling (η) descend after this region. The plummet in values of effectiveness cooling for both sides is attributed to the blade being modeled as a solid body which is much more realistic than the previous studies (where blades were analyzed as a shell) and also the effects of blade design, holes

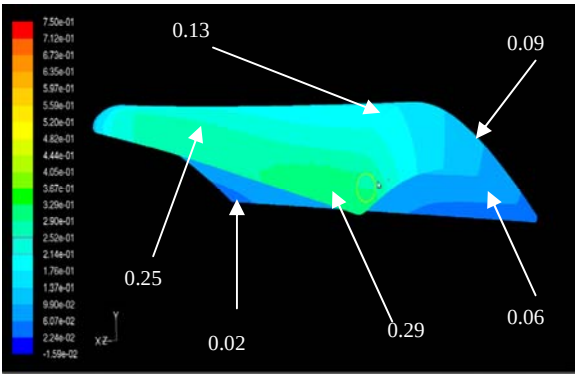


Figure 6: Film cooling effectiveness couture at blowing ratio BR= 2 on the blade model.

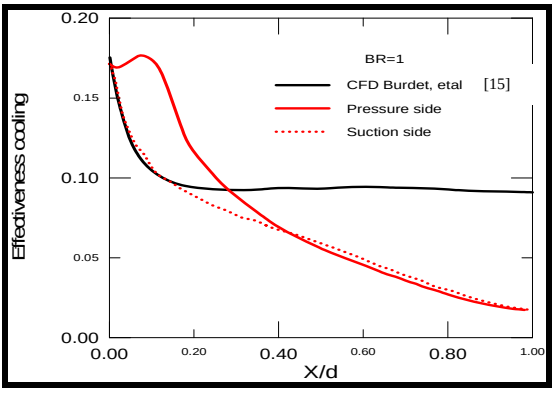


Figure 7: Distributions of film cooling effectiveness at blowing ratio BR=1 at angle of injection 35^0: comparison between CFD results of Burdet *et al.* [15], the blade model calculation.

diameter and number of holes along the span which are different from previous investigations.

Fig. 8 illustrates the effects of coolant fluid on the distribution of film cooling effectiveness for the pressure side and suction side at the midspan location with BR=1, 1.5. In this article two different coolant fluids (air at T=287.5 and T=153 K) have been injected to resolve the best way to achieve improved blade protection. High effectiveness cooling was obtained through injected air as coolant fluid at temperature 287.5K. Evidently, from equation (3) η depends on T_∞ , T_W and T_C , even with a decrease T_C there is not necessarily an increase in effectiveness cooling (η) since the blowing ratio (BR) will be also affected by the temperature property as indicated by equation (2).

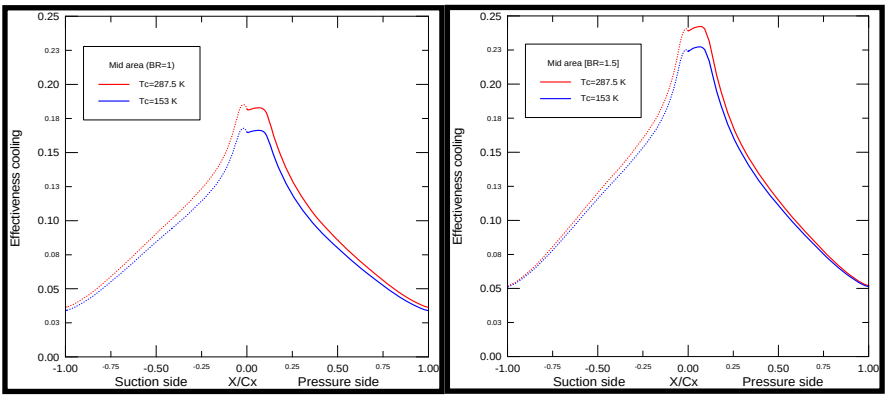


Figure 8: Effects of coolant fluid temperature on the film cooling effectiveness at BR=1, 1.5.

Prediction of heat transfer from the hot gas towards the blade can be achieved via the Nusselt number (N_U) based on the axial chord Cx and this is defined as:

$$N_U = (q*Cx)/k(T_\infty - T_W) \quad (4)$$

where, q is the blade wall heat transfer rate. Fig. 9 shows the predicted profiles of Nusselt number (N_U) at the mid span for BR=1 1.5 and 2.

Near the leading edge along the span (from hub to shroud) the heat transfer attains the maximum level and this is in excellent agreement with the results of Burdet and Abhari [7], Garg and Abhari [8] and Burdet *et al*. [15]. On the pressure side (N_U) suddenly drops after the leading edge. Certainly, the effects of film cooling are manifested via the creation of a layer of protection from the incoming hot gas on this side. At X/Cx>0.15 the BR influence is clearly observed as reducing (N_U) – again this trend concurs with the studies of Kadja and Bergeles [1] and Guangchao *et al.* [14]. On the suction side, the Nusselt number (N_U) suddenly falls after the leading edge and in the region X/Cx=-0.25 the values of (N_U) slightly increase owing to the effects of turbulence and the camber of blade and gradually falls; the better protection of blade from hot gas through increases effectiveness cooling, η and reduced blade Nusselt number (N_U).

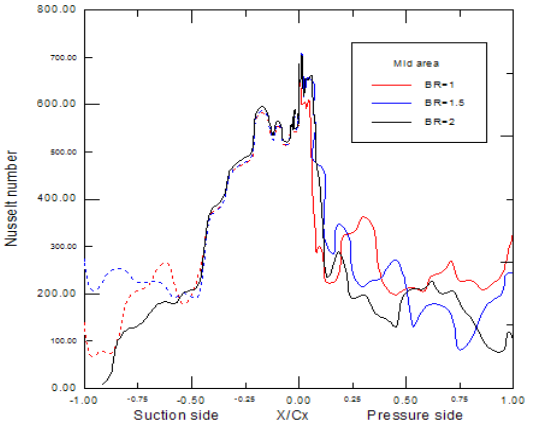


Figure 9: Predicted profile of Nusselt number (Nu) at midspan for BR=1, 1.5, 2.

Fig. 10 shows Nusselt number contours at the leading edge and film cooling regions at BR=2. The maximum value of Nusselt number (N_U) is determined at the leading edge along the blade span. This implies heat transfer from the hot gas *in the direction of the blade*. Near the holes there is no heat transfer to the blade due to coolant fluid injection. Thus, the Nusselt number (N_U) drops to negative values (heat transfer will be in the opposite direction).

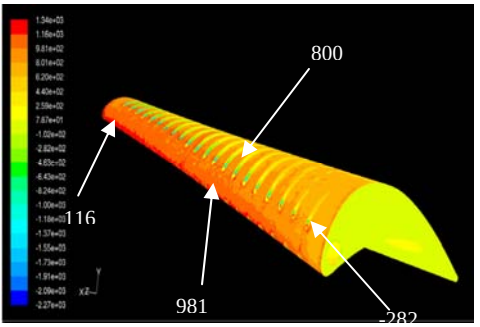


Figure 10: Nusselt number (Nu) counter on the blade model for BR=2.

6 Conclusions

The film cooling performance for a complex gas turbine blade at the leading edge with three blowing ratios has been studied numerically. The main findings of this investigation are as follows:

1) Film cooling effectiveness near the leading edge significantly increases with blowing ratio (η is proportional with BR). The influence of increasing BR on the film cooling effectiveness appeared on the pressure and suction side; the pressures side effectiveness cooling is much more enhanced than the suction side (high BR enhanced film cooling on pressure side).
2) The response of film cooling effectiveness along the span from hub to shroud section is enhanced with increasing blowing ratio. However, if there is dissimilarity between the published paper and the previous studies certainly this is due to the number of holes, holes diameter, blade section from hub to shroud and skewness.
3) There is no benefit in injection of coolant fluid at low temperature to enhance blade cooling at constant BR (see equation (5)).
4) The heat load on the blade represented by Nusselt number (N_U) is strongly influenced by increasing BR.

Acknowledgement

The authors would like to thank the Government of Iraq, Ministry of Education for their generous financial support during Harbi A. Daud's studies at Sheffield Hallam University.

References

[1] Kadja M. & Bergeles G., Computational study of turbine blade cooling by slot- injection of a gas. *Applied Thermal Engineering*, **17(12)**, pp. 1141-1149, 1997.

[2] Hung M.S, Ding P.P & Chen P.H, Effect of injection angle orientation on concave and convex surface film cooling, *Experimental Thermal and Fluid Science*, **33**, pp.292-305, 2009.

[3] Kassim, M.S, Yoosif, A.H. & Al-Khishali K.J.M., Investigation into flow interaction between jet and cross mainstream flows, *PhD thesis, University of Technology, Mechanical Engineering Department, Iraq, Baghdad*, 2007.

[4] Lakehal D., Theodoridis G.S. & Rodi, W., Three-dimensional flow and heat transfer calculations of film cooling at the leading edge of a symmetrical turbine blade model *Int. J. Heat and Fluid Flow*, 22, pp.113-122, 2001.

[5] Theodoridis, G.S., Lakehal, D., & Rodi, W., Three dimensional calculation of the flow field around a turbine Blade with film cooling injection near the leading edge, *Flow, Turbulence and Combustion*, **66**, pp. 57-83, 2001.

[6] Forest, A.E., White, A.J., Lai, C.C., Guo, S.M., Oldfield, M. L. G. & Lock, G. D., Experimentally aided development of a turbine heat transfer prediction method, *Int. J. Heat and Fluid Flow*, **25**, pp. 606-617, 2004, available online at www.ScienceDirect.com.

[7] Burdet, A. & Abhari, R.S., Three-dimensional flow prediction and Improvement of Holes Arrangement of a Film-Cooled Turbine Blade Using Feature Based Jet Model, *ASME J. Turbomachinery*, **129**, pp. 258-268, 2007.

[8] Garg, V. K. & Abhari, R. S., Comparson of predicted and experimental Nusselt number for a film cooled rotating blade, *Int. J. Heat Fluid Flow*, **18**, pp. 452-460, 1997.

[9] Azzi, A. & Jubran B.A., Influence of leading edge lateral injection angles on the film cooling effectiveness of a gas turbine blade, *Heat and Mass Transfer*, **40**, pp. 501-508, 2004, available online at www.ScienceDirect.com.

[10] Renze, P., Schroder, W. & Meinke, M., Large- eddy simulation of film cooling flow at density gradients, *Int. J. Heat Fluid Flow*, **29**, pp.18-34, 2008, available online at www.ScienceDirect.com.

[11] Eghlimi, A., Kouzoubov, A. & Fletcher, C.A.J., Anew RNG-Based tow equation model for predicting turbulent gas particles flow, *Int. Conference on CFD in Mineral & Metal Processing and Power Generation CSIRO, Sydney Australia*, 1997.

[12] Lakehal, D., Near wall modeling of turbulent convective heat transport in film cooling of turbine blades with the aid of direct numerical simulation data, *ASME J. Turbomachinery*, **124**, pp.458-498, 2002.

[13] Tao, Z., Yang, X., Ding, S., Xu, G., Wu, H., Deng, H.& Luo, X., Experimental Study of Rotation effect on Film Cooling over the Flat wall with a Single Hole, *Experimental Thermal and Fluid Science*, **32**, pp.1081-1089, 2008, available online at www.ScienceDirect.com.

[14] Guangchao, L., Huiren, Z. & Huiming, F., Influences of Hole Shape on Film Cooling Characteristics with CO_2 Injection, *Chinese J. Aeronautics*, **2**, pp.393-401, 2008, available online at www.ScienceDirect.com.
[15] Burdet, A., Abhari R.S & Martin, G.R., Modeling of Film Cooling – part II: Model for use in Three Dimensional Computational Fluid Dynamics, *ASME J. Turbomachinery*, **129**, pp.221-231. 2007.

Spray cooling by multi-walled carbon nanotubes and Fe nanoparticles

H. Bellerová & M. Pohanka
Heat Transfer and Fluid Flow Laboratory, Faculty of Mechanical Engineering, Brno University of Technology, Czech Republic

Abstract

An experimental investigation of spray cooling performed with water based nanofluid containing multi-walled carbon nanotubes and Fe nanoparticles was carried out. The concentrations of carbon nanotubes in the liquid used in the experimental program were 1 wt.%, 0.1 wt.%, 0.01 wt.%, the concentrations of Fe nanoparticles were 40 wt.%, 10 wt.%, 1 wt.%. The liquid was sprayed on the surface by a full cone nozzle from distances of 40, 100 and 160 mm with flow rates of 1 to 2 kg/min (liquid impingement densities of 1 to 40 kg/m^{2}s). A steel sensor measuring temperature history was cooled by spraying from 190 °C. The heat transfer coefficient was calculated at an interval of the surface temperature from 100 °C to 50 °C by inverse modelling, and compared with the heat transfer coefficient of water cooling. Using Fe nanoparticles showed a decrease of the heat transfer coefficient on the cooled surface. The majority of experiments with carbon nanotubes also showed a decrease of the heat transfer coefficient. However, there were some conditions during which an increase was observed.

Keywords: nanofluids, multi-walled carbon nanotubes, Fe nanoparticles, heat transfer, spray cooling, experimental.

1 Introduction

It was anticipated that some fluid heat qualities would be improved by adding metal parts, metal oxides parts, or generally those parts from materials which have suitable heat transfer characteristics. Some attempts focused on cooling were made by liquid additives in water, with particles sized in mm or μm mixed into the fluids and with nanofluids. Nanofluid is a suspension of fluid (water, ethylene glycol, oil, bio-fluids, polymer solution, etc.) and particles (metals,

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110261

metal oxides, nonmetals as carbon nanotubes, etc.) sized 1–100 nm. Not many works of research have been focused on the cooling intensity by nanofluids and very few on the cooling intensity during spray cooling (e.g. Chakraborty et al. [1], Bansal and Pyrtle [2]).

1.1 Thermal conductivity of nanofluids with carbon nanotubes

Research shows that the thermal conductivity of nanofluid is higher than that of base fluid. The first group of researchers who studied thermal conductivity of nanofluid with carbon nanotubes was Choi et al. [3]. They used oil with 1 vol.% of multi-walled nanotubes sized 25 nm in diameter and 50 μm in length and discovered an enhancement of 160%. A large number of scientists showed a keen interest in the use of carbon nanotubes, yet they did not confirm this significant enhancement of thermal conductivity by adding carbon nanotubes. The enhancement of thermal conductivity at room temperature was in the majority of the research that followed (e.g. Assael et al. [4], Amrollahi et al. [5]) 20–40% by adding less than 3 vol.% carbon nanotubes to some fluid like water, synthetic engine oil or ethylene glycol.

1.2 Thermal conductivity of nanofluids with Fe nanoparticles

The thermal conductivity of Fe_3O_4 nanoparticles in water was studied by Zhu et al. [6] and an enhancement of 38% was observed for their concentration of 4 vol.%. This enhancement was more than for other oxide aqueous nanofluids. Hong et al. [7] investigated 0.1–0.55 vol.% of Fe nanoparticles sized 10 nm in ethylene glycol. The enhancement was 18%.

1.3 Heat transfer characteristics of nanofluids with carbon nanotubes and Fe nanoparticles

According to the aforementioned information about thermal conductivity of carbon nanotubes and Fe nanoparticles, the nanofluids should be a suitable coolant and a majority of scientists really found an increase in the heat transfer coefficient by using carbon nanotubes and Fe nanoparticles. In the work of Ding et al. [8] the convective heat transfer of 0.5 wt.% of carbon nanotubes in water was enhanced by over 350%. According to Liao and Liu [9], the enhancement ratio of heat transfer coefficient was 18–25% for 2 wt.% and 49–56% for 1 wt.% at 58 °C. Park and Jung [10] found that there was an increase of up to 30% in the nucleate boiling heat transfer of 1 vol.%. Unlike the findings of Park et al. [11], concluded that the pool boiling heat transfer coefficient of aqueous suspension with 0.0001–0.5 vol.% of carbon nanotubes was lower than that of pure water. Shi et al. [12] studied the pool boiling heat transfer of Fe nanoparticles sized 26 nm in water and found that the enhancement was up to 60%.

1.4 Spray cooling by nanofluids

To date limited research on the heat transfer of nanofluid with carbon nanotubes or with Fe nanoparticles by spray cooling has been conducted. There are some papers about heat transfer by spray cooling with other nanofluids. Bansal and Pyrtle [2] examined alumina nanoparticles dispersed in water in concentrations of 0.25 to 0.505 wt.%. The initial temperature of the surface was 150 °C and the surface was cooled to 75 °C. The cooling intensity was comparable to the cooling intensity of water. Chakraborty et al. [1] used 0.1 wt.% TiO_2 in water to cool a plate heated up to 1200 °C. A significant enhancement in the cooling rate was observed.

2 Basic properties of examined nanofluids

The nanofluid with 1 wt.% carbon nanotubes was produced by the Belgian company Nanocyl S.A. The nanofluid with 40 wt.% of Fe nanoparticles was offered by prof. Miroslav Mašláň from Palacky University Olomouc. In the experiments, the concentrations of 0.1 wt.% and 0.01 wt.% of carbon nanotubes and the concentrations of 10 wt.% and 1 wt.% of Fe nanoparticles were achieved by diluting the original concentrations with distilled water. Both the density of nanofluids and the specific heat were measured. A calorimeter "Kalorymetr KL-5" was used to measure the specific heat capacity.

2.1 Properties of nanofluid with carbon nanotubes

The nanofluid with carbon nanotubes is a black and odorless aqueous dispersion with 1 wt.% of carbon nanotubes (0.5–1.5 wt.%). The length of multi-walled carbon nanotubes was 1.3 μm and the diameter 9 nm on average (fig. 1). The nanoparticles were prepared by catalytic carbon vapor deposition. The nanofluid

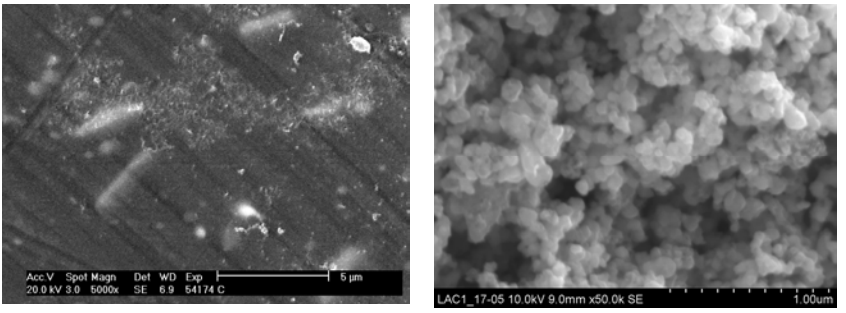


Figure 1: The left picture: cluster of carbon nanotubes (photo by Institute of Material Science, Brno University of Technology); the right picture: Fe nanoparticles (photo from prof. Miroslav Mašláň, Palacky University, Olomouc).

also contained an ionic surfactant in a concentration of 0.1–3 wt.%. The specific gravity was 1000 kg/m^3 at 20 °C according to the product list, pH 7–11, boiling point 100 °C. It is usually used for waterborne systems for antistatic coatings, transparent films, and cleaning rooms. The densities of 1 wt.%, 0.1 wt.% and 0.01 wt.% of carbon nanotubes were 1011 kg/m^3, 1004 kg/m^3 and 1002 kg/m^3 at room temperature. The specific heat was measured at 20–60 °C only for 1 wt.% of carbon nanotubes and it was 3763 J/kg K.

2.2 Properties of nanofluid with Fe nanoparticles

The Fe nanoparticles were spherical and measured 80 nm on average (fig. 1). The nanofluid did not contain any stabilizers. The densities of 40 wt.%, 10 wt.% and 1 wt.% of Fe nanoparticles were 1079 kg/m^3, 1006 kg/m^3 and 1002 kg/m^3. The specific heat was 3528 J/kg K at 20–60 °C for 40 wt.% of nanoparticles.

3 Experiment

3.1 Experimental plan

The hot surface was sprayed by the full cone nozzle Lechler 460.443.17 CA with a spray angle of 45° from three distances of 40 mm, 100 mm, and 160 mm. The flow rates of 1 kg/min, 1.5 kg/min, and 2 kg/min were steady during the experiment (the appropriate liquid impingement density is shown in Tab. 1). The initial experiments were conducted by spraying pure water and after that nanofluid. Finally, the nanofluid spray cooling intensity was compared with the pure water spray cooling intensity. The summary of experiments is shown in Tab. 1.

Table 1: Experimental plan.

Fluid	Concentration	Nozzle	Distance	Flow rate	Liquid impingement density
	wt.%		mm	kg/min	kg/m^{2}s
Pure water	100	Full cone	40	1	19.3
C nanotubes in water	1 0.1 0.01			1.5	29
				2	38.7
			100	1	3.1
				1.5	4.6
Fe nanoparticles in water	40 10 1			2	6.2
			160	1	1.2
				1.5	1.8
				2	2.4

3.2 The nozzle

The liquid impingement density in Tab. 1 was determined theoretically by the type of nozzle (full cone), distance between nozzle and surface, spray angle and flow rates. The radius of impingement area should be 41.4 mm in distance 100 mm from the nozzle. Figure 2 shows the measured pressure of water falling from a distance of 100 mm for three flow rates. Figure 2 shows that the shape of spray was not a full cone for flow rate of 1 kg/min.

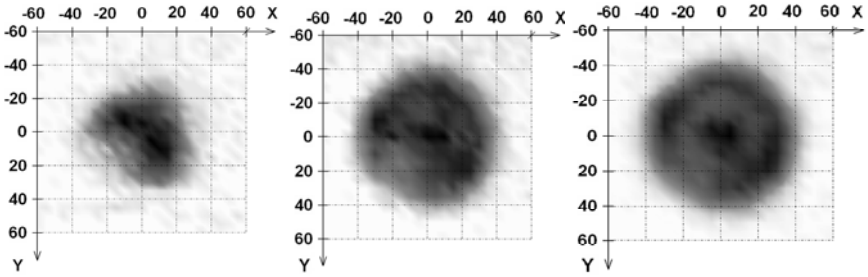


Figure 2: Measured pressure distribution of pure water spraying from full cone nozzle at a distance of 100 mm, the pictures from left to right represent the flow rates 1 kg/min, 1.5 kg/min, 2 kg/min; the axis units are mm; the darker the colour, the higher the pressure of water; the maximum value for 1 kg/min was 112 Pa, for 1.5 kg/min 186 Pa, for 2 kg/min 297 Pa.

3.3 The cooled sensor

The cooled sensor (fig. 3) was made from austenitic steel 314 AISI and built into Teflon. All surfaces excluding the cooled surface were protected against cooling. The thermocouple of type K with a diameter of 0.5 mm was soldered

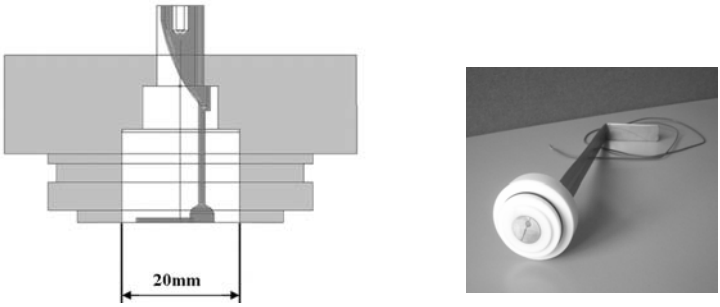


Figure 3: Cooled steel sensor; the drawing is on the left side (steel sensor body is colored white, Teflon is light grey, channel for thermocouple is dark grey); the photo is on the right side.

into a cooled cylindrical steel sensor with a diameter of 20 mm and was placed parallel to the surface 0.37 mm. The thermocouple was connected with a data logger, which is mobile equipment for recording data.

3.4 Experiment description

The steel sensor in Teflon was heated up to 190 °C by an electrical furnace and was held at that temperature for 15 minutes then put into an experimental test bench (fig. 4) with the tested surface face down. A nozzle was placed under the probe. Before the experiment, the tested fluid was located in a pressure tank pressurized by nitrogen. The manometer was located close to the nozzle. After putting the sensor into the equipment, the fluid went through the nozzle and sprayed the cooled surface. The fluid was caught in the experimental equipment after spraying. At the end of the experiment, the fluid was moved again to the pressure tank. The measurement took tenths of seconds, and the sensor surface was cooled to approximately 30 °C. During the experiment, the sensor temperature, coolant temperature and ambient temperature were measured. These data were saved by the data logger and downloaded to a computer after the test. Then they were used for an inverse calculation in a non-commercial program based on Pascal programming language. The output from the program was a physical quantity characterizing the cooling intensity.

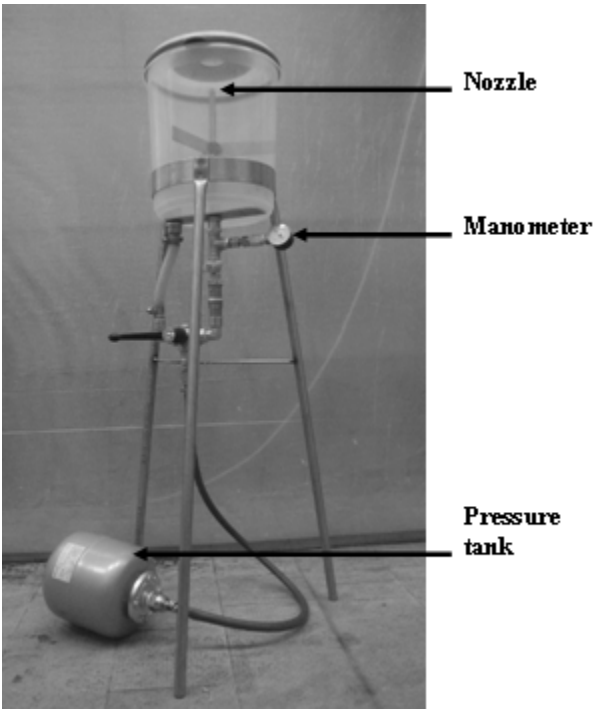


Figure 4: Experimental equipment.

4 Calculation of cooling intensity

The cooling intensity is characterized by the heat transfer coefficient HTC [W/m^{2}K]. The Beck algorithm of one-dimensional inverse task [13] was applied to find the heat transfer coefficient. The direct task of the two-dimensional thermal field was a part of the inverse solution and was solved by using the finite difference method (fig. 5). The inverse task provides the surface temperature and the heat transfer coefficient history.

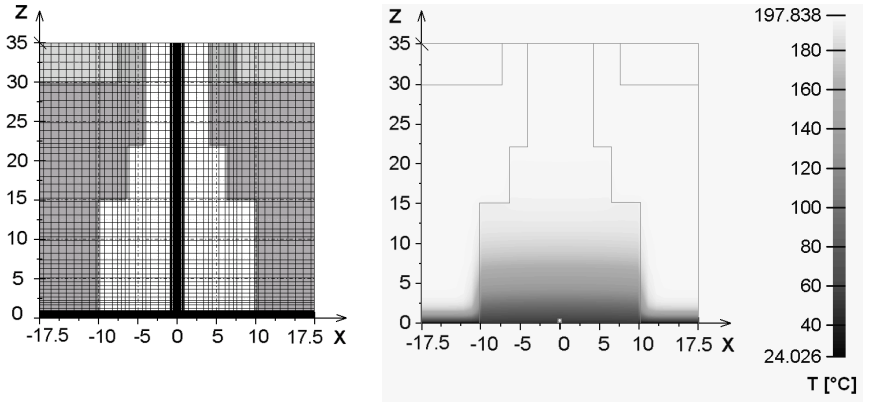


Figure 5: The picture on the left: control volumes covering model of steel body that were used for direct task (dark grey is for teflon, light grey for female screw, white for steel and for air around female screw); the picture on the right: the thermal field in cooled steel body after 5 s of cooling; the temperature of steel surface is 50 °C, the axis units are in mm.

5 Results

An example of record and results from the cooling experiment is shown in fig. 6. The measured temperature history in thermocouple position, computed surface temperature and computed HTC is shown. Some fluctuations can be seen on the HTC curves in fig. 6, it is the result of the random fluctuation of falling droplets in the impact area. The results for spraying with pure water were found to match the results published by Ciofalo et al. [14]. The temperature field in the whole steel sensor after 5 s of cooling is illustrated in fig. 5.

The average results from the experiments described in this paper are shown in figs. 7–10. The average values of the heat transfer coefficient HTC were computed for the intervals 50–100 °C of the surface temperatures. The graphs show that the HTC increases with an increase in liquid impingement density and it is possible to say that adding carbon nanotubes and Fe nanoparticles to pure water decreases HTC in most of researched cases. The HTC was surprisingly

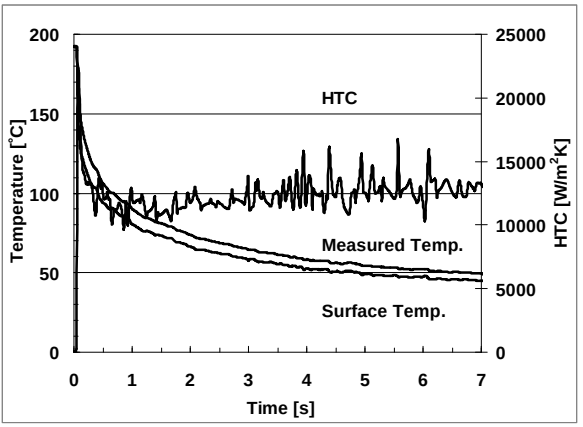


Figure 6: An example of record and results from the cooling experiment with pure water, flow rate of 1.5 kg/min and distance between surface and nozzle of 100 mm.

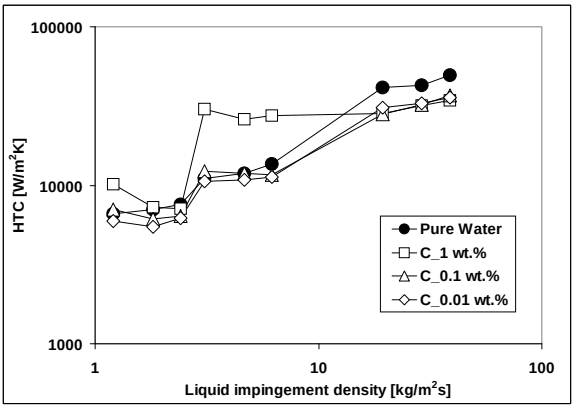


Figure 7: The heat transfer coefficient vs. liquid impingement density for nanofluids with different concentrations of carbon nanotubes.

increased by nanofluid with 1 wt.% of carbon nanotubes at a distance of 100 mm (it represents liquid impingement density 3.1 kg/m^{2}s, 4.6 kg/m^{2}s and 6.2 kg/m^{2}s). For verification of measurement, the experiments for each nanofluids at a distance of 100 mm between cooled surface and nozzle and for flow rates of 1 and 2 kg/min were provided twice. They are shown in figs. 11 and 12.

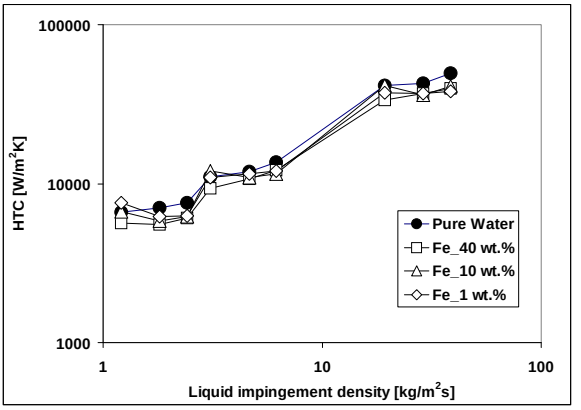


Figure 8: The heat transfer coefficient vs. liquid impingement density for nanofluids with different concentrations of Fe nanoparticles.

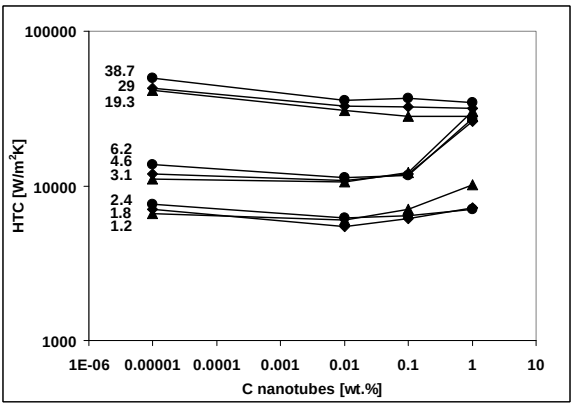


Figure 9: The heat transfer coefficient vs. concentration of carbon nanotubes for variable impingement density shown by the curves with numbers [kg/m^{2}s]; pure water is shown as nanofluid with 1e-005 wt.% of carbon nanotubes.

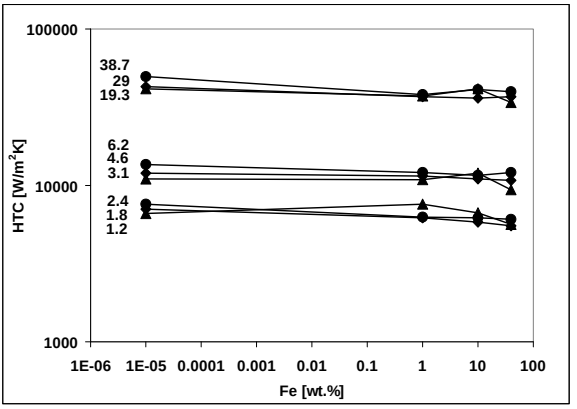


Figure 10: The heat transfer coefficient vs. concentration of Fe nanoparticles for variable impingement density shown by the curves with numbers [kg/m^{2}s]; pure water is shown as nanofluid with 1e-005 wt.% of Fe nanoparticles.

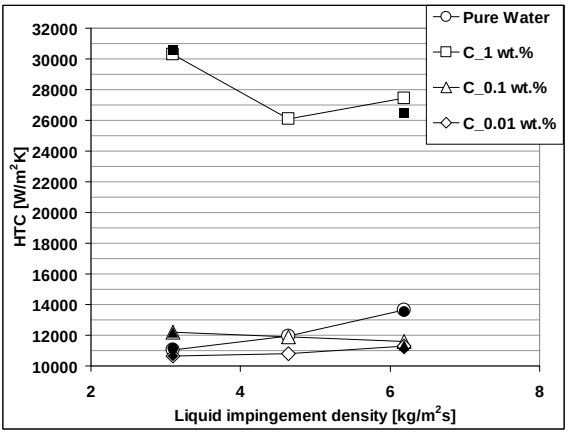


Figure 11: The repeated experiments for nanofluids with carbon nanotubes at a distance of 100 mm between cooled surface and nozzle; the empty marks represent the first measurements, the filled marks the repeated.

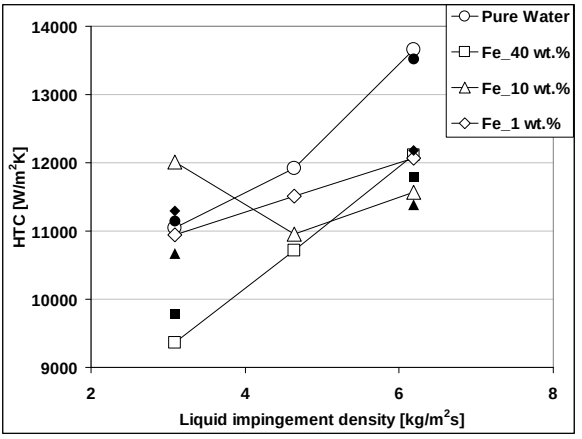


Figure 12: The repeated experiments for nanofluids with Fe nanoparticles at a distance of 100 mm between cooled surface and nozzle; the empty marks represent the first measurements, the filled marks the repeated.

6 Conclusions

Most of the research confirmed an improvement in thermal conductivity and convective heat transfer coefficient during liquid convection by adding carbon nanotubes or Fe nanoparticles to pure water. This paper investigated spray cooling by these nanofluids. It was found that by adding carbon nanotubes or Fe nanoparticles to pure water the cooling intensity during spraying of the steel surface with its temperatures of 100 to 50 °C was not increased. A high increase in cooling intensity by spraying only with 1 wt.% of carbon nanotubes at a distance of 100 mm from the nozzle was observed. For 3.1 kg/m^{2}s (flow rate 1 kg/min) the heat transfer coefficient was increased about 174% to the heat transfer coefficient of pure water, for 4.6 kg/m^{2}s (1.5 kg/min) it was 119% and for 6.2 kg/m^{2}s (2 kg/min) 101%. For other liquid impingement densities a decrease in HTC up to 32%, 32%, 26% for nanofluids with 1 wt.%, 0.1 wt.%, 0.01 wt% of carbon nanotubes, respectively, and up to 22%, 18%, 23% for nanofluids with 40 wt.%, 10 wt.%, 1 wt.% of Fe nanoparticles, respectively, in comparison with pure water was observed. The decrease in HTC was also observed in the work of Bansal and Pyrtle [2] by using alumina nanoparticles in water. Contrary to this the results in Chakraborty's work [1] show an increase in HTC by adding TiO_2 nanoparticles. The reason for this could be in the temperature range of the conducted experiments. The experiments carried out for this paper and the Bansal and Pyrtle's experiments [2] were conducted between 200 and 50 °C. However, Chakraborty's research [1] was conducted for temperatures between 1200 and 500 °C.

References

[1] Chakraborty, S., Chakraborty, A., Das, S., Mukherjee, T., Bhattacharjee, D. & Ray, R.K., Application of water base- TiO2 nano-fluid for cooling of hot steel plate. *ISIJ International*, **50**, pp. 124–127, 2010.

[2] Bansal, A. & Pyrtle, F., Alumina nanofluid for spray cooling enhancement. *ASME-JSME Thermal Engineering Summer Heat Transfer Conference*, pp. 797–803, 2007.

[3] Choi, S.U.S., Zhang, Z.G., Lockwood, F.E. & Grulke E.A., Anomalous thermal conductivity enhancement in nanotubes suspensions. *Physics Letters*, **79**, pp. 2252–2254, 2001.

[4] Assael, M.J., Chen, C.F., Metala, I. & Wakeham, W.A., Thermal conductivity of suspensions of carbon nanotubes in water. *International Journal of Thermophysics*, **25**, pp. 971–985, 2004.

[5] Amrollahi, A., Hamidi, A.A. & Rashidi, A.M., The effects of temperature, volume fraction and vibration time on the thermo-physical properties of a carbon nanotube suspension (carbon nanofluid). *Nanotechnology*, **19 (31)**, pp. 1–8, 2008.

[6] Zhu, H., Zhang, C., Liu, S., Tang, Y. & Yin, Y., Effect of nanoparticle clustering and alignment on thermal conductivities of Fe3O4 aqueous nanofluids. *Applied Physics Letters*, **89**, article number 023123, 2006.

[7] Hong, K.S., Hong, T.K. & Yang, H.S., Thermal conductivity of Fe nanofluids depending on the cluster size of nanoparticles. *Applied Physics Letters*, **88**, articles number 031901, 2006.

[8] Ding, Y., Alias, H., Wen, D. & Williams, R.A., Heat transfer of aqueous suspensions of carbon nanotubes (CNT nanofluids). *Heat and Mass Transfer*, **49**, pp. 240–250, 2006.

[9] Liao, L. & Liu, Z.H., Forced convective flow drag and heat transfer characteristics of carbon nanotube suspensions in a horizontal small tube. *Heat and Mass Transfer*, **45**, pp. 1129–1136, 2009.

[10] Park, K.J. & Jung, D.S., Enhancement of nucleate boiling heat transfer using carbon nanotubes. *Heat and Mass Transfer*, **50**, pp. 4499–4502, 2007.

[11] Park, K.J., Jung, D. & Shim, S.E., Nucleate boiling heat transfer in aqueous solutions with carbon nanotubes up to critical heat fluxes. *International Journal of Multiphase Flow*, **35**, pp. 525–532, 2009.

[12] Shi, M.H., Shuai, M.Q., Li, Q. & Xuan, Y.M.: Study on pool boiling heat transfer of nano-particle suspensions on plate surface. *Journal of Enhanced Heat Transfer*, **14**, pp. 223–231, 2007.

[13] Beck, J., Blackwell, B. & Clair, C.R., *Inverse Heat Conduction*, Wiley, 1985.

[14] Ciofalo, M., Caronia, A., Di Liberto, M. & Puleo, S. The Nukiyama curve in water spray cooling: Its derivation from temperature-time histories and its dependence on the quantities that characterize drop impact. *Heat and Mass Transfer*, **50**, pp. 4948–4966, 2007.

Section 4
Stress analysis

This page intentionally left blank

Analyses of the robustness of the mathematical model with the laboratory prototype of the drilling process during simulation and vibration control experiments

F. Abdul Majeed[1], H. Karki[1], Y. Abdel Magid[2] & M. Karkoub[3]
[1]*Department of Mechanical Engineering,*
The Petroleum Institute, U.A.E.
[2]*Department of Electrical Engineering, The Petroleum Institute, U.A.E.*
[3]*Mechanical engineering Department, Texas A&M University, Qatar*

Abstract

The drilling industry is faced with many challenges, and the sudden failure of a drill string during drilling is one of major concern. Exploration of the causes for the failures reveals vibrations as the major cause. In order to test and analyze the vibration patterns of rotary drilling, a laboratory proto type of the process is set up. The mathematical model developed to analyze the vibration presents residual error. Robustness issues pertaining to model error and modelling error is discussed. Methods to counter the errors and minimize the vibrations are also discussed.

Keywords: rotary drilling, robustness, modeling error, vibration, experimental set up, unbalanced mass, parameter uncertainty.

1 Introduction and problem formulation

There are many types and designs of drilling rigs. Drilling rigs are classified by the power used to drive the process (electric, mechanic or hydraulic), height of the derrick or the type of drill pipe used (cable, conventional and coil tubing).

The drilling rig we are concentrating on is an electric driven, conventional rotary drilling, fig. 1. This is the widely used method of drilling and drilling is achieved by the drill bit as it rotates and cuts into rock. All the major items of machinery used in the rig are driven by electric motors. Metal or plastic drill

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110271

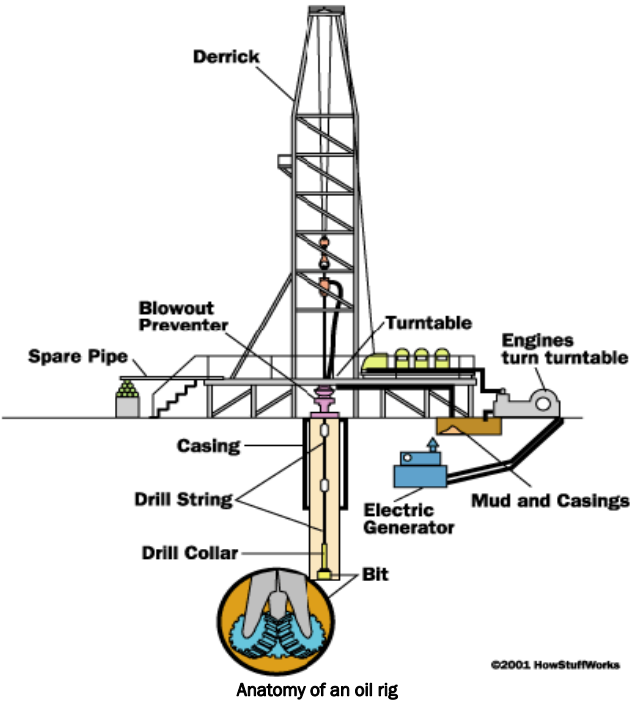


Figure 1: Rotary oil rig features.

pipe for the length of the drill string, and the rotation is achieved by turning a square or hexagonal pipe (the "Kelly") on a rotary table at drill floor level.

Before drilling, a large, heavy bit is attached to the end of a hollow drill pipe. As drilling progresses, the drill bit forces its way underground and additional sections of pipe are connected at the top of the hole. The derrick is the name for the structure which supports the rig above the surface. The taller the derrick, the longer the sections of drill pipe that the derrick can hold at a time. Although early derricks were made of wood, modern derricks are constructed of high-strength steel. Throughout the rotary drilling process, a stream of fluid called drilling mud is continuously forced to the bottom of the hole, through the bit, and back up to the surface. This special mud, which contains clay and chemicals mixed with water, lubricates the bit and keeps it from getting too hot. It also acts as a cap to keep the oil from gushing up.

The drill strings experience high vibrations and strong rocks in the path of drilling. These cause the dynamics presented by drill strings to be highly complex, non linear and unexpected. The drill string vibrations coupled together with well bore friction result in many phenomenon such as bit bounce, stick slip, forward and backward whirl. There are three main types of drill string vibration:

Axial vibration is mainly caused when drilling with roller cone bits. It leads to jumps in drilling; bouncing and can slow down the rate of penetration.

Torsional vibration results due to twisting of the drill string, and sometimes breaking. It makes the rotation of the drill bit irregular and leads to the stick slip phenomenon.

Lateral vibration occurs when the drill string is bent or when drilling in a non vertical well. The drill bit rotates with a center of rotation not coincident with the center of well, leading to hole enlargement and forward or backward whirl of the bit.

This research concentrates on lateral vibrations occurring in the drill pipe due to a bend. Ideally with zero well bore friction and assuming the drill string is a perfect straight beam rotated with an axial load, there will be no nonlinearities or vibrations during drilling. However, in the presence of curved/inclined boreholes or unbalanced WOBs the friction between the drill bit and well borehole contact is uneven and different at different contact points. This result in the drill bit centerline not being in the center of the hole, hence the centrifugal force will now act as the center of gravity causing the drill string to bend. Bend drill strings do not follow circular trajectories, causing the drill bit to hit the sides of the borehole. This will eventually lead to the stick slip phenomenon, in which large vibrations and sudden unexpected drill bit movements occur. The usual solution in oil rigs is to stall the entire drilling process, and restart. In extreme cases the drill string would break requiring a call for an entire process up haul.

2 Experimental set up and literature review

In order to understand and analyze the vibrations due to the bend drill string, experiments are conducted. The drilling process is simulated in the laboratory by a simple experimental set up, fig. 2. The set up operates by operating a motor, which rotates a large rotor connected to the motor shaft. A drill string (made of carbon steel chosen due to its proximity in properties to actual drill string material) is attached to the upper rotor by a universal joint. This joint provides two degree of rotational freedom (x and y axes).The motor rotation provides the set up with 1DOF of rotation about the z axes. In order to understand the behavior pattern of a bend drill string and accurately simulate its trajectories, a literature review is conducted. The following are the literature review results required for accurate modeling of the bend drill string features.

The performance of drill strings and their effect on drilling performance have been investigated and analyzed in a number of researches [1–8]. Many different models were set up to analyze drill string vibrations including lateral vibrations (whirl) and mode coupling [3, 9, 10]. Other researchers have focused on models which represent torsional vibration and have attempted to suggest methods to avoid stick-slip behaviour [2, 7]. Jansen [11] proposed an unbalanced mass model to represent a bend drill string section in which self excited vibrations were studied. Similar models were also studied by other researchers; for example, Melakhessou *et al.* [5] modelled the drill string as an unbalanced rotor supported by two bearings and research is concentrated on a contact zone between the drill string and the borehole wall. Dykstra *et al.* [1] explains that the

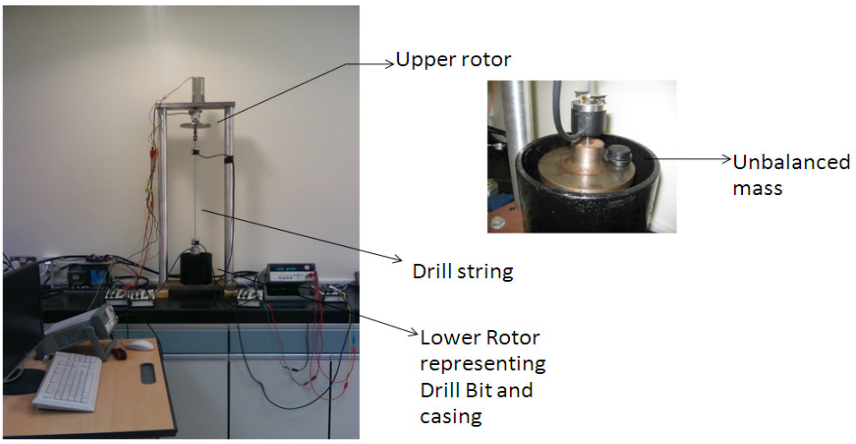


Figure 2: Laboratory drilling set up and close up of lower bit with unbalanced mass.

source of vibration is the bit and hence the centrifugal forces developed when an unbalanced drill string is rotated can be one of the major sources of vibrations.

Analyzing the literature, an unbalanced mass is placed on the lower rotor representing the drill bit to simulate the bend drill string properties. The experimental set up now has three DOFS. Apart from the rotation of the upper rotor, and lower rotor, there is tangential angular displacement for the lower rotor initiated by the new centre of rotation of the lower rotor not coinciding with the centre of rotation of the upper rotor. The lower rotor now follows an elliptical trajectory, also known popularly as bit whirl in the drilling field. This paper also analyses the behaviour of the system at low and average operating speeds of actual drilling.

3 Robustness issues

3.1 Residual error and Model error

The mathematical model for the process was identified using the system identification black box modeling approach. The experimental set up was excited with chirp input to obtain the required identification data. The chirp input has correlation function properties very similar to white noise. A Box Jenkins model was identified for the process. Box Jenkins models are especially useful when the process is affected by disturbances entering late into the system.

The laboratory operating speeds are selected to represent the rotary drilling process at its low and average operational speeds. The process is excited by command inputs of 8 RPM and 38 RPM, figs. 3 and 4. The usual rotary drilling speeds are around 35 to 45 RPM. The low speed is analyzed to understand the drill string behavior in the transient time. The upper rotary speed is the input speed of the process. The experimental responses of the process are recorded and

plotted in figures 3 and 4. The identified model is seen to have very close response to the process response. The residual signal graph between the process response and model response is seen to be very low with values around 180m RPM, figs. 5 and 6. This suggests that the model provides a good fit for analyzing the process behavior.

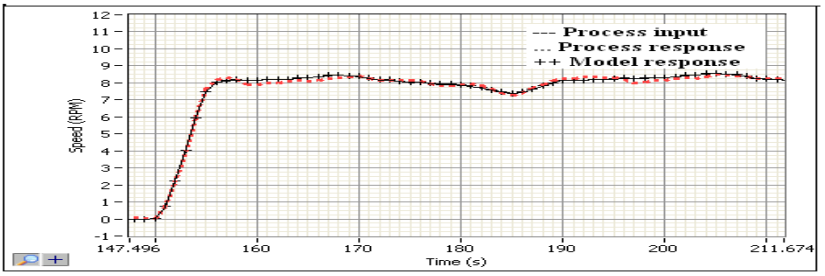


Figure 3: Process data and model response for low drilling speed.

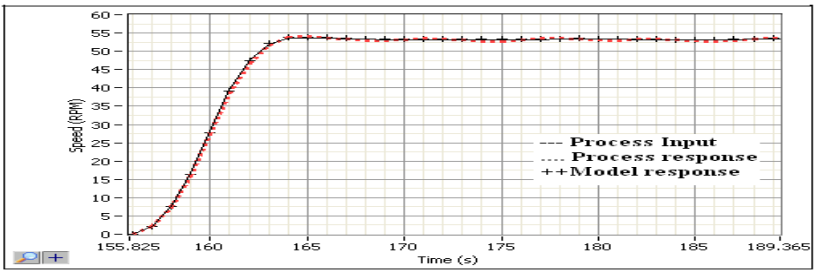


Figure 4: Process data and Model response at average drilling speed.

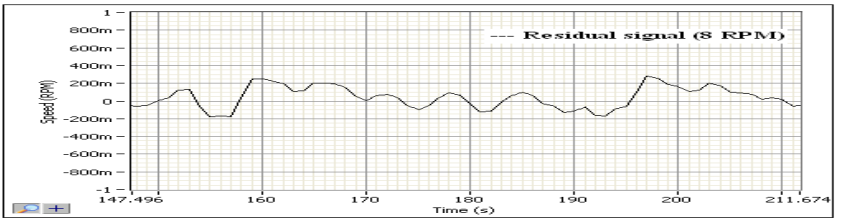


Figure 5: Residual signal for low speed analysis.

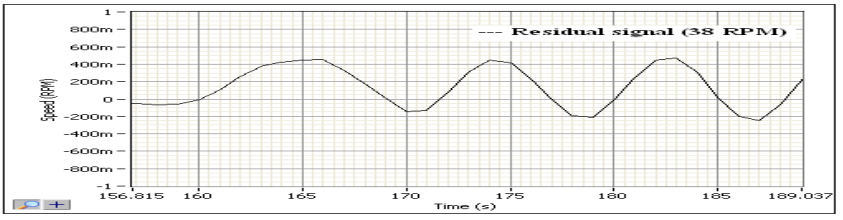


Figure 6: Residual signal for average speed analysis.

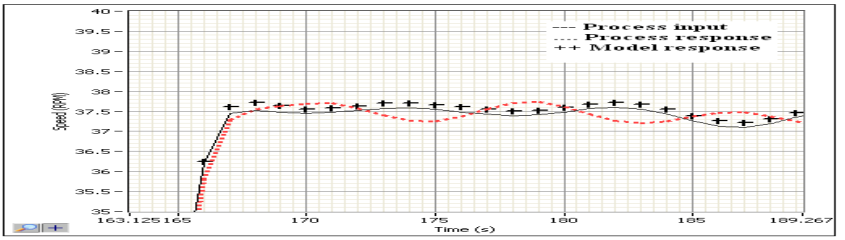


Figure 7: Larger view of vibrations in the process and model response.

However, the very small vibrations in the speed of the process at the output due to the unbalanced mass are noticeable, fig. 7. Here the unbalanced mass is very small nearly 57 gms, which is about 5% of the mass of the lower rotor representing the drill bit. This mass will represent only a very small bend in the drill string. However, in reality, drill strings when they bend slightly, present more severe vibrations due to the presence of well bore friction and higher mass of the bottom hole assembly. The black box model of the process is identified in a Box Jenkins model format specifically because the Box Jenkins models are good for processes in which disturbances enter late in to the system.

The residual error, figs. 5 and 6 presents us with a model robustness issue which needs to be dealt with. One suggestion is to combine the black box model with a separate model describing effect of the unbalanced mass using analytical principles and larger degrees of freedom Liao *et al.* [4]. The drilling system prototype concerned here can be seen to be a strictly proper system. In other words the gain tends to zero in the limit as frequency tends to infinity, fig. 8.

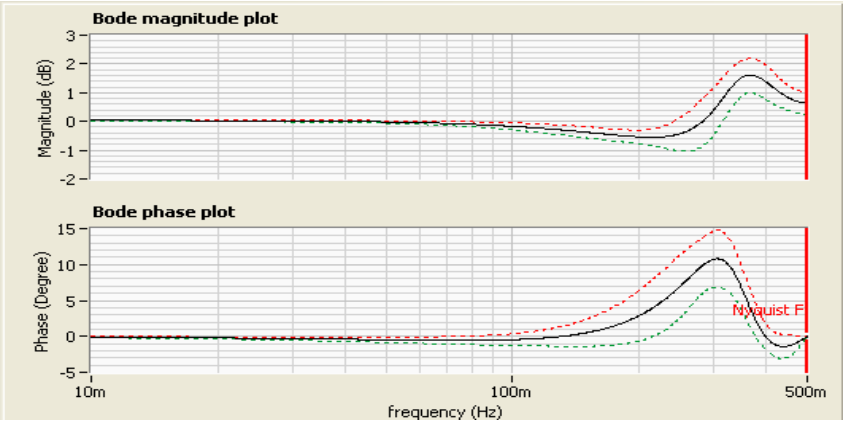


Figure 8: Bode plot of the model with lower (green) and upper (red) bounds.

This can be attributed to the presence of inertia in the system. The model itself will have robustness errors and they need to be analyzed further by looking for RHP poles and zeros, cancellations and analyzing the internal stability of the model.

3.2 Modelling error

Real time systems are nonlinear, time varying, infinite dimensional and very complicated. Modeling error is due to two major reasons.

The model obtained here assumes a third order transfer function; this approximation leads to a simpler model and lack of robustness in the model. This is one of the major sources of modeling error. Assuming that the best model approximation of the plant is $G_p(s)$;

$$G_p(s) = G(s) + \Delta G(s) \tag{1}$$

i.e. if

$$G_p(s) = \sum_{i=1}^{\infty} \frac{N(s)}{D(s)}; \tag{2}$$

$$G(s) = \sum_{i=1}^{k} \frac{N(s)}{D(s)} \tag{3}$$

and

$$\Delta G(s) = \sum_{i=k+1}^{\infty} \frac{N(s)}{D(s)} \tag{4}$$

where $\frac{N(s)}{D(s)}$ is the numerator and denominator of the plant transfer function. $\Delta G(s)$ is the modeling error, or the difference between the model and the best possible plant model.

Another source of modeling error can be deduced from analyzing the frequency response magnitude; figs. 9 and 10.The frequency response gains are plotted for two different conditions, for small mass unbalance and large mass unbalance. It can be seen that as the frequency increases the size of the resonant peaks tend to decrease after a certain point ω'. In the frequency response gain plots, the point ω' of the drilling system here can be seen to be around 150m Hz for the two cases studied. This particular frequency ω' is seen to be a constant for a particular system and does not vary with added disturbance, here the unbalanced mass. Hence we can safely assume that for frequencies higher than ω' the magnitude of the frequency response will never exceed the gain at that value, i.e;

$$20\log_{10}\left|\Delta G(jw)\right| = \begin{cases} = 0;\ 0 \le \omega \le \omega' \\ \le \rho;\ \omega' < \omega < \infty \end{cases} \tag{5}$$

where ρ is the value of gain at ω' and it represents an upper bound on the magnitude of the frequency response of the modeling error.

The second source of error in modeling is from parameter uncertainty. The parameters estimated to obtain the model will have a tolerance associated with their values. Hence at every frequency, the gain and phase response will have an

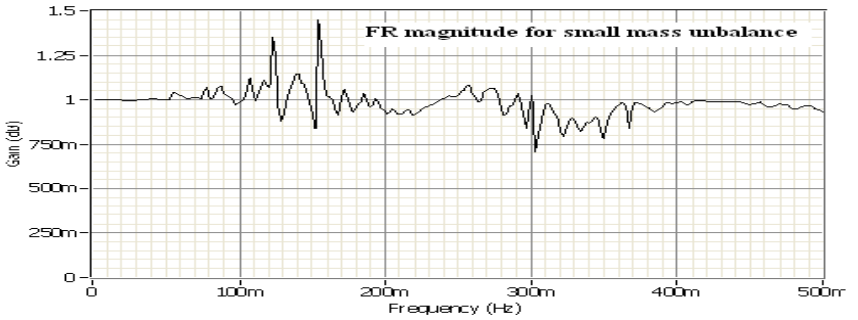


Figure 9: Frequency response magnitude plot for small mass unbalance.

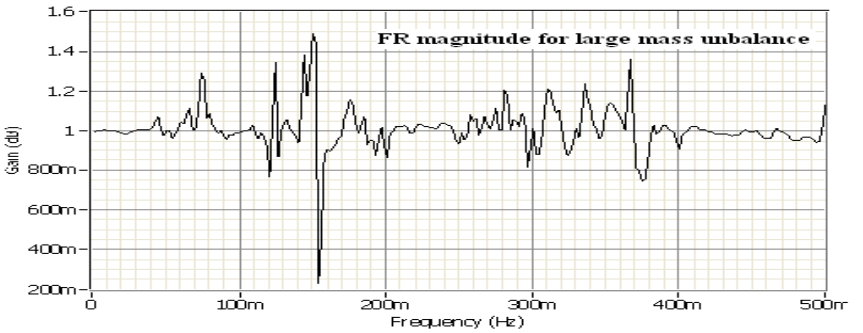


Figure 10: Frequency response magnitude plot for large mass unbalance.

uncertainty associated with its value. These are plotted in the bode plot, fig. 8, with the upper and lower bound of the magnitude and phase curves.

4 Vibration analysis

The analytical equation for the lower rotor with the unbalanced mass can be written as Inman [12]:

$$m\ddot{x} + c\dot{x} + kx = m_0 e\omega_r^2 \sin \omega_r t \tag{6}$$

where m is the lower rotor mass, m_0 is the unbalanced mass, c is the damper constant, k is the spring constant (drill string considered as a spring and damper), e is the distance of the unbalanced mass from the center axis of rotation of lower rotor, ω_r is the drilling rotational frequency and ξ is the damping ratio.

The steady state displacement of the lower rotor is

$$X = \frac{m_0 e}{m} \frac{r^2}{\sqrt{(1-r^2)^2 + (2\xi r)^2}} \tag{7}$$

where $r = \dfrac{\omega_r}{\omega_n}$; ω_n is the natural frequency.

At any particular moment t, the deflection is:

$$x_p(t) = X \sin(\omega_r t - \theta) \tag{8}$$

Analyzing the above equations it can be inferred that an increase in the deflection amplitude can be minimized by increasing the damping in the system. It can also be noted that the most vibrations occur when the set up is operated at frequencies near the natural frequency. Hence it is also advisable to operate the set up at frequencies lower or higher than the natural frequency. Sources of model error will be further expanded in forthcoming papers.

For laboratory scale assessment of vibration and related control the Box Jenkins model can be considered appropriate enough due to residual error values not exceeding 200 m RPM which is 0.4% of the rotation rate. For control purposes the controller effect is simulated and once an acceptable effect is realized the controller can be interfaced with the drilling setup to analyze experimental response. The under actuation of the lower rotor now presents another difficulty, due to any effect of the controller on the simulated response of the lower rotor will fail to achieve in the experimental set up. The unbalanced mass presents a change in the inertial mass properties of the lower rotor. The vibrations induced can only be minimized by assuring that the drilling is carried out at speeds neglecting the natural frequency, preferably low speeds so that the tangential acceleration is minimized and the drill bit does not cause too much damage to the borehole wall by enlarging it or making wedges in the bore hole and damaging the drill bit and other down hole components.

5 Summary

This paper discusses the model of a drill string system representing rotary drilling. The experimental and simulated model responses are plotted and analyzed. The residual error is discussed and the source of the error due to lack of robustness in the model is also studied. Two major reasons for the uncertainty and presence of modeling error are discussed. The source and control of vibration is also discussed. Future work involves study and discussion of robustness issues arising in the model itself, and further expansion of the modeling errors and ways to overcome the errors for a better plant model.

References

[1] Dykstra, M., Christensen, H., Warren, T., and Azar, J., Drill string component mass imbalance: A major source of drill string vibrations, *SPE Drilling and completion*,Vol.11, pp.234-241,1996.

[2] Germay, C., van de Wouw, N., Nijmeijer, H., and Sepulchre, R., Nonlinear drill string dynamics analysis, *SIAM J. Applied Dynamical systems*, Vol. 8, pp.527-553, 2009.

[3] Leine, van Campen, D.H and Keultjes, W., Stick slip whirl interaction in drill string dynamics, *ASME J. of vibration acoustics*, Vol.124, pp. 209-220, 2002.
[4] Liao. C et al., Reduced order models of drill string dynamics, *Paper presented at the Second international energy 2030 conference*, UAE, 2008.
[5] Melakhessou, H. et al., A nonlinear well drill string interaction model, *J. Vibration acoustics*, Vol.125, pp.46 -52, 2003.
[6] Mihajlovioc, N., Van Veggel, A., Van de Wouw, N. and Nijmeijer, H., Friction induced torsional vibrations in an experimental drill string system, *Paper presented at the 23rd IASTED international conference on modeling, identification and control*, pp.228-233, 2004.
[7] Mihajlovic, N., van Veggel, A., Nan de Wow and Nijmeijer, H., Analysis of friction induced limit cycling in an experimental drill string system, *J. of Dynamic systems, measurement and control*, ASME, Vol.126, pp.709-721, 2004.
[8] Navarro Lopez, E. and Suarez, R., Modeling and Analysis of stick slip behavior in a drill string under dry friction, Congresso Annual De La Amca, pp. 330-336, 2004.
[9] Christoforou, A. and Yigit, A., Dynamic modeling of rotating drill string with borehole interactions, *J. sound and vibration*, Vol.206, pp.243-260, 1997.
[10] Elsayed, M. A., and Dareing, D. W., Coupling of Longitudinal and Torsional Vibrations in a Single-Mass Model of a Drillstring, *Developments in Theoretical and Applied Mechanics, Vol. XVII, University of Arkansas*, Fayetteville, AR, pp. 128–139,1994.
[11] Jansen, J.D., Non linear rotor dynamics as applied to oil well drill string vibrations, *J. Sound and vibration*, Vol.147, pp.115 -135, 1991.
[12] Inman, D.J., Engineering Vibration, Prentice Hall, pp.120-121, 2001.

Transverse cracking of composite bridge decks

Ayman N. Ababneh
Department of Civil Engineering,
Jordan University of Science and Technology, Irbid, Jordan

Abstract

Early age cracking of concrete bridge deck is a frequent problem for bridges worldwide. This work provides a framework for a thermo-hygro-mechanical mathematical model for the analysis of early age transverse cracking of the bridge deck. The model includes the determination of the temperature and moisture gradients in the deck slab, and the prediction of the thermal and drying shrinkage strains. These strains were superimposed with the strains from creep and mechanical loads and applied to an elasto-plastic damage approach to quantify the damage and stresses in deck slab. The model was implemented in finite element computer software to accurately predict the cracking and damage evolution in concrete bridge decks. Accurate prediction of crack tendency is essential for durability design of bridge decks, thus more sustainable bridges with increased usable life span and low life-cycle costs.
Keywords: transverse cracking, bridge deck, thermo-hygro-mechanical model.

1 Introduction

Bridges usually developed early cracking of their decks [1]. Early age cracks usually develop in the transverse direction of the traffic. The cracking could initiate almost immediately after construction and sometimes appear within a few months after deck is constructed. The problem of deck cracking is still significant, even after the adoption of high performance concrete (HPC) for bridge decks. In a survey conducted by New York Department of Transportation (NYSDOT), it was observed that 48% of an 84 bridge decks built in New York State between 1995 and 1998, using HP concrete, have developed transverse cracks [2].

Figure 1 shows the mechanism of the transverse cracking of a concrete deck slab. The composite action between the deck and the girders provides restraining

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110281

to the deck. When concrete shrinks the external restraint from the girder as well as the internal restraints from the reinforcement and aggregates produce tensile stresses in the longitudinal direction of the deck. When these stresses reach the tensile strength of concrete (low at early ages), transverse cracks are developed in the deck starting from bottom and extended to the top surface. In continuous beams or in beams with fixed-end restraint, the negative moments result from mechanical loads produce tensile stress in the deck, which when combined with the shrinkage stresses aggravate the problem of deck cracking.

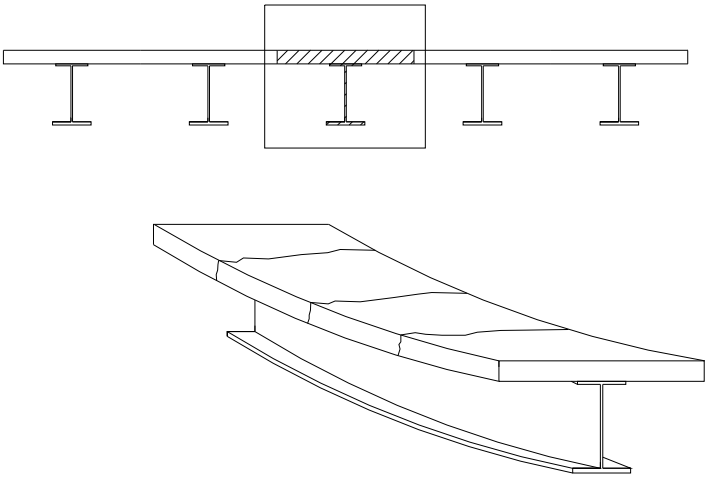


Figure 1: Transverse cracking of concrete.

Deck cracking has no immediate effect on the bridge safety, but it has detrimental effects on the long-term performance. Cracks interconnect the voids and the isolated microcracks in the concrete deck to form preferential pathways for the ingress of chlorides from deicing chemicals thus accelerate reinforcement corrosion. Fanous et al. [3] observed severe corrosion of black and epoxy coated rebars extracted from cracked location in different bridge decks. Also, leakage of water thorough cracks increases the degree of water saturation in the bridge substructure, therefore increases the risk of freeze-thaw damage. As a result, the bridge service life is reduced and maintenance and rehabilitation costs rise.

Bridge deck cracking occurs when restrained volumetric changes associated with moisture and temperature changes take place. Volumetric changes are mainly result from autogenous shrinkage, drying shrinkage, thermal shrinkage and creep. These major causes of concrete volume change with time depend primarily on the material properties and mix design, design details, construction practices, and environmental conditions. Concrete properties are the most important factors affecting transverse deck cracking since they control the shrinkage and thermal strains that cause stresses and control the relationship between strains and stresses. Understanding the concrete properties is central to

reliably modeling the mechanisms contributing to the cracking of concrete decks. Experimental and analytical models have been developed to enable bridge designers to calculate shrinkage and thermal stresses in bridge decks [1, 4], so that they can evaluate and modify designs to reduce these stresses and the risk of transverse deck cracking. Furthermore, construction can affect transverse deck cracking. Careful construction practices are often required to reduce such a risk where the first large stresses in a concrete deck develop in the first 12 to 24 hours, when temperatures change rapidly from early hydration. The exposure to environmental conditions, e.g. ambient humidity and temperature has a major effect on transverse cracks.

Previous studies have focused primarily on unveiling the extent and significance of cracking and trying to pinpoint possible causes. While there is agreement among researchers on the major causes, the relative contribution of each factor has not been completely determined and the problem of premature deck cracking still exists. The predominant reason for this is because most of the factors have simply been discussed qualitatively and there has been little quantitative analysis of these mechanisms. In the last few years, researchers have developed numerical models using finite element or finite difference methods to simulate real structures under different environmental conditions. Saadeghvaziri and Hadidi [5] studied the developed tensile stresses in bridge decks under the effect of many design factors, e.g. girder stiffness, deck thickness, girder spacing, relative stiffness of deck to girder, and amount of reinforcements; in their simulation, they assumed that the shrinkage strain is constant in value and over the deck depth.

2 Significance

The main objective of this research is to develop a mechanistic approach for the analysis of transverse cracking of composite bridge decks. This work will allow better understanding of the cracking mechanism and in turn help practical engineers to develop preventive and remedial strategies to eliminate or at least mitigate them. The FEM simulations will result in better-determined stresses and crack widths in the bridge deck structures subjected to the combined effects of hygro-thermal volume changes and load-induced cracking. Accurate prediction of crack tendency is essential for the reliability and long-term performance of newly constructed bridge decks at service load levels.

The report with recommendations that will result from the literature survey and parametric studies will provide engineers with the ability to analyze the impacts of the material properties and mix design parameters, structural design details, and construction practices on cracking of concrete bridge decks. This will enable engineers to develop materials and methods to construct bridge deck structures that can withstand a multitude of harsh environmental conditions at low life-cycle cost. These measures will result in more sustainable structures with increased usable life span and low life-cycle costs.

3 Formulation of computational framework

This section presents the framework for a mechanistic approach for the analysis of early age transverse cracking of the bridge deck and it is divided into four steps. First, mathematical models for the determination of the temperature and moisture gradients in the deck slab are descried. The temperature variation and moisture loss will be used to predict the thermal and drying shrinkage strains in bridge deck. Finally, these strains will be added to the strains from creep and mechanical loads and applied to an elasto-plastic damage approach to quantify the damage and stresses in deck slab.

3.1 Determination of the temperature and moisture gradients

The variation of the temperature over the bridge deck is governed by the heat transfer in concrete which can be described by the following equation

$$\rho c \frac{\partial T}{\partial t} = div\ [k\ \ grad(T)] + Q \tag{1}$$

where T is the temperature, ρc is the heat capacity, k is the thermal conductivity and Q is the rate of heat generation inside concrete as a result of cement hydration. A model for the heat of hydration generation of blended cement is described by Kishi and Maekawa [6], and it will used to calculate Q.

The moisture loss is governed by the moisture transport equation as follows

$$\frac{\partial w}{\partial t} = \frac{\partial w}{\partial H}\frac{\partial H}{\partial t} = div\ [D_H\, grad(H)] + G \tag{2}$$

where w is the moisture content, t is the time, H is the relative humidity, D_H is the moisture diffusion coefficient, $\partial w/\partial H$ is the moisture capacity and G is the rate of moisture loss due to hydration. The heat of hydration generation model mentioned above can be also used to determine G.

3.2 Prediction of the volumetric changes

The numerical solution of the equations above (Eq. (1) and (2)) results in temporal and spatial distribution of the temperature (T) and relative humidity (H) inside the concrete bridge deck. The volumetric changes in concrete are related to environmental factors including temperate and humidity variations. The environmental strain, ε_{ev} is the summation of the drying and thermal strains

$$\varepsilon_{ev} = \varepsilon_{sh} + \varepsilon_T \tag{3}$$

The thermal strain ε_T is a function of heating and cooling cycles and can be expressed in terms of temperature change as follows

$$\varepsilon_T = \alpha\ \Delta T \tag{4}$$

in which α is the coefficient of thermal expansion of concrete and ΔT is the temperature change.

The drying shrinkage strain is related to moisture loss and so it can be linked to the change in the relative humidity as follows

$$\varepsilon_{sh} = \beta \ \Delta H \tag{5}$$

where β is coefficient of shrinkage strain and can be calculated by a multiscale model [7].

3.3 Quantifying the early age transverse cracking

In practice, bridges at early ages are exposed to the simultaneous actions of environmental and mechanical deteriorations. The development and coalescence of early age cracks due to the combined effect of environmental factors and traffic loads progress gradually and cause gradual strain softening in the stress strain relation. These microcracks result in a reduction of both strength and stiffness at the macroscopic material level. Damage mechanics not only captures the loss of strength, but also the deterioration of material stiffness. Therefore, it is realistic to use continuum damage theory, which has the ability to describe the tensile strain softening and stiffness deterioration rather than a sudden complete cracking [8]. A fully coupled hygromechanical model was developed by Ababneh et al. [9] and Sheban et al. [10]. This model can be extended to quantify the damage due to environmental loads including the creep effect, and the mechanical loads. In this model, an elastoplastic-damage approach was used to characterize the shrinkage-induced damage. Based on the small strain theory of plasticity, the strain tensor is additively decomposed into an elastic part ε_e and a plastic part ε_p. Considering the presence of environmental strain tensor ε_{ev}, the total strain tensor can be expressed as:

$$\varepsilon = \varepsilon_e + \varepsilon_p + \varepsilon_{ev} \tag{6}$$

The constitutive relation can then be written as:

$$\sigma = E:(\varepsilon - \varepsilon_p - \varepsilon_{ev}) \tag{7}$$

in which E is the elastic stiffness tensor.

The elastoplastic and damage models are coupled by exploiting the effective stress concept known from continuum damage mechanics [11]. The effective stress represents the redistributed stress over undamaged or effective area after the damage has taken place. Based on the scalar damage model the relation between the effective stress σ_{eff} and the nominal stress σ can be expressed as:

$$\sigma_{eff} = \frac{\sigma}{1-d} \tag{8}$$

where d is the damage parameter. It is assumed that $0 < d < d_{cr}$, where d_{cr} is the critical damage in which a complete local rupture takes place. In practice, a $d_{cr} = 1$ is usually employed.

Considering the undamaged elastic stiffness E_o, we have

$$\sigma_{eff} = E_o : (\varepsilon - \varepsilon_p - \varepsilon_{ev}) \tag{9}$$

and its corresponding time derivative takes the form:

$$\dot{\sigma}_{eff} = E_o : (\dot{\varepsilon} - \dot{\varepsilon}_p - \dot{\varepsilon}_{ev}) \tag{10}$$

From Eqs. (9), (10) and (11), the constitutive relation becomes:

$$\sigma = (1-d)E_o : (\varepsilon - \varepsilon_p - \varepsilon_{ev}) = E : (\varepsilon - \varepsilon_p - \varepsilon_{ev})$$
$$E = (1-d)E_o \tag{11}$$

Following the effective stress concept, it is logical to assume that the plastic flow takes place only in undamaged area [12]. Thus, the formulae from elastoplastic theory that are dependent on stress must be modified by substituting the effective stress in place of the nominal stress. The problem can then be solved by using standard elastoplastic theory. The elastoplastic behavior of concrete will be assumed to follow the pressure-sensitive Drucker–Prager criterion. The damage initiation and evolution can be characterized by a damage model developed by Mazars and Pijaudier-Cabot [13].

4 Simulation of bridge deck cracking

Ababneh et al. [9] and Sheban et al. [10] studied the drying shrinkage-induced damage in concrete structures, and they developed a finite element simulation program capable to estimate the stress and strain variations with time in bridge decks. This program can be used to study the effects of many parameters, e.g. concrete mixing ingredients, types and proportions, ambient dry environment, and support restraints. A concrete bridge deck 4″ [10 cm] in depth was simulated by this model. The concrete has an average compressive strength of 5000 psi [34.5 MPa] and is moist cured for 28 days before drying. The deck, which is initially saturated (H_{ini}=100%), is exposed to drying on the top and bottom surfaces (H_{env}= 50 %) after curing. Figure 2 shows the geometry and boundary conditions of the investigated structure where only one quarter of the deck is modeled by four-noded plane strain elements.

The comparison results of the deck simulation are shown in Figure 3 for two cases: case I, where the effect of damage on moisture diffusivity is ignored, and case II, where the full coupling between damage and moisture diffusion is considered. It is clear that in case I, the rate of drying process is slower than case II, and case II reaches equilibrium with the environment faster. In the latter case, the drying process generates damage in the concrete, which accelerates the drying process as shown in Figure 3a. The damage parameter, d versus drying time for the bridge deck is shown in Figure 3b for both cases. The state of damage in concrete in case II accelerate and increase earlier and faster, and its value is higher than the damage in case I where no coupling between the damage and the diffusion process is assumed. Therefore, ignoring the effect of damage

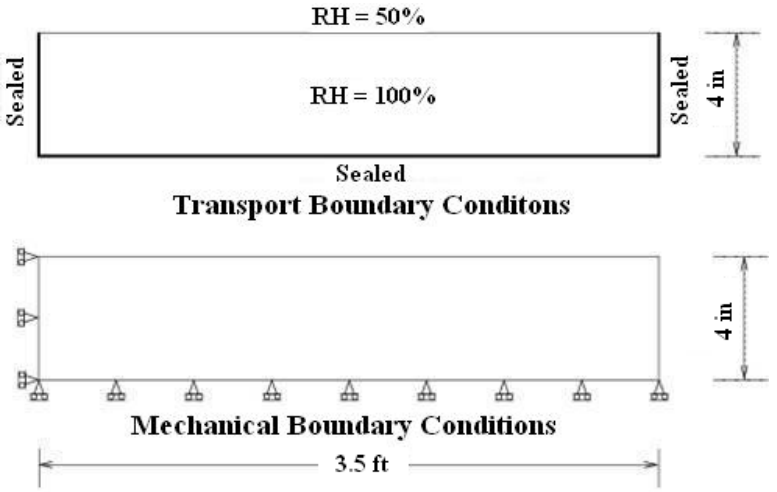
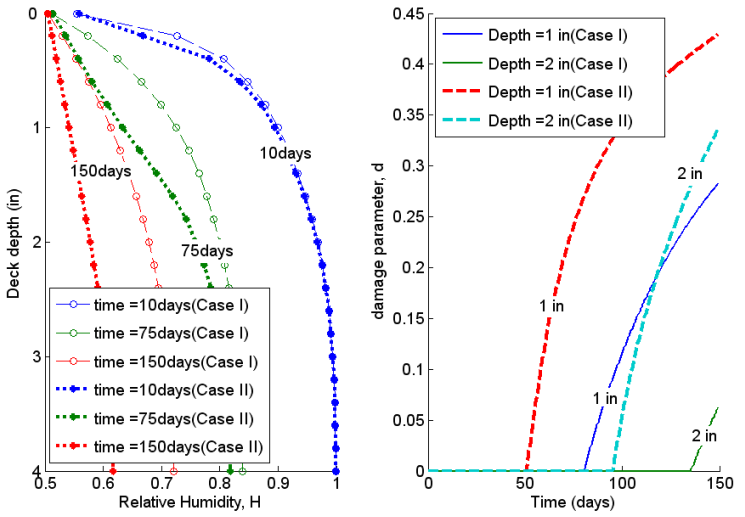


Figure 2: Boundary conditions for the concrete bridge deck.



a. Humidity profiles b. Damage evolution

Figure 3: Humidity profiles and damage versus drying time for the bridge deck.

on the moisture diffusion process yields an overestimation of the relative humidity in concrete and a lower estimation of concrete damage states.

This simulation demonstrated that the fully coupled hygromechanical model can capture the concrete structural behavior under drying conditions. This model can be used to analyze different types of concrete structures, and with some modifications, e.g. the effect of thermal stresses and stress relaxation, the computer program can be used for analysis of early age transverse cracks in composite bridge decks.

Ongoing research at Jordan University of Science and Technology seeks to develop a finite element analysis (FEA) based software to investigate the causes of transverse concrete deck cracking, quantitatively predict the tendency of a deck to crack and predict the crack sizes. To isolate the effects of individual parameters that have a potential impact on transverse deck cracking, computational software is developed, by means of which parametric studies will be conducted using finite element analysis of the complete bridge system. The information will then be used to make recommendations relative to design and construction practices to reduce the potential for transverse cracks in composite bridge decks.

5 Summary and conclusions

This paper provided a computational framework for the analysis of early age transverse cracking in composite bridge decks. A fully hygro-thermal-mechanical model for the prediction of early age transverse cracks was formulated. The model involves the formulation of the governing coupled heat and moisture transport equations, the quantification of the induced thermal and hygral damage by elasto-plastic damage model and modeling the coupling between damage and transport phenomena. The simulation of the bridge deck demonstrated that the fully coupled hygromechanical model can capture the concrete structural behavior under drying conditions. Ongoing research at Jordan University of Science and technology seeks to modify the computer software by implementing the hygro-thermal- mechanical model developed in this study.

References

[1] Krauss, P. D. & Rogalla, E. A., Transverse cracking in newly constructed bridge decks. *Rep. No. NCHRP Report 380, Transportation Research Board*, National Research Council, Washington, DC, 1997.

[2] Alampalli, S. & Owens, F.T., Improved Performance of New York State Bridge Decks. *HPC bridge Views*, Issue 7, 2000.

[3] Fanous, F., Wu, H., & Pape, J., Impact of deck cracking on durability, Center for Transportation Research and Education, Iowa State University, Ames, IA, 2000.

[4] French, C., Eppers, L., Le, Q. & Hajjar, J.F., Transverse Cracking in Concrete Bridge Decks, *Transportation Research Record*, No. 1688, TRB, National Research Council, Washington, D.C, 1999.

[5] Saadeghvaziri, M. A., & Hadidi, R., Cause and control of transverse cracking in concrete bridge decks, *Rep. No. FHWA-NJ-2002-19*, Federal

Highway Administration, U.S. Department of Transportation, Washington, D.C., 2002.
[6] Kishi, T. & Maekawa, K., Hydration heat model for blended cement including blast slag and fly ash, *Proceedings of the JCI*, **15(1)**, pp. 1211-1216, 1993.
[7] Lemaitre, J., *A course on damage mechanics*, 2nd. Ed., Springer-Verlag, NY, 1996.
[8] Bazant, Z., & Raftshol, W., Effect of cracking in drying and shrinkage specimens, *Cement and Concrete Research*, **12(2)**, pp. 209-226, 1982.
[9] Ababneh, A., Sheban, M., Suwito, A., & Xi, Y., The coupling effect of drying shrinkage and moisture diffusion in concrete, *Concreep 7*, pp. 175-180, 2005.
[10] Sheban, M. A., Ababneh, A. N., & Fulton, S. R., Numerical simulation of moisture diffusion in concrete, Joint International Conference on Computing *and Decision Making in Civil and Building Engineering*, Montréal, Canada, pp. 1015-1024, 2006.
[11] Kachanov, L. M., *Introduction to continuum damage mechanics*., Martinus Nijhoff Dordrecht, The Netherlands, 1986.
[12] Suwito, A., Ababneh, A., Xi, Y., & Willam, K., The coupling effect of drying shrinkage and moisture diffusion in concrete. *Computers and Concrete*, **3(2-3)**. pp. 103-122, 2006.
[13] Mazars, J., & Pijaudier-Cabot, G., From damage to fracture mechanics and conversely: a combined approach, *International Journal of Solids Structures*, **335(20-22)**, pp. 3327-3342, 1986.

This page intentionally left blank

Evaluation of crack growth rate and growth model of ultrafine grained copper

M. Goto[1], S. Z. Han[2], K. Euh[2], J.-H. Kang[2], K. Kamil[1],
N. Kawagoishi[3] & S. S. Kim[4]
[1]*Department of Mechanical Engineering, Oita University, Japan*
[2]*Korea Institute of Materials Science, Republic of Korea*
[3]*Department of Mechanical Engineering, Kagoshima University, Japan*
[4]*Engineering Research Institute,*
Gyeongsang National University, Republic of Korea

Abstract

High-cycle fatigue tests were carried out on smooth specimens of ultrafine grained (UFG) copper produced by equal channel angular pressing for 12 passes. The growth behavior of a small surface-crack was monitored. A major crack, which led to the final fracture of the specimen, initiated from shear bands (SBs) at an early stage of stressing. Different tendencies of growth behavior occurred depending on the ranges of crack length. To understand the changes in growth rate and fracture surface morphologies, a quantitative model describing a crack growth mechanism were developed considering the reversible plastic zone size at a crack tip. In addition, the crack growth rate of UFG copper was evaluated by applying the small-crack growth raw.
Keywords: fatigue, surface damage, fine grains, copper, crack propagation.

1 Introduction

Ultrafine grained (UFG) materials processed by equal channel angular pressing (ECAP) have many unique properties due to the unusual characteristics of the microstructure with non-equilibrium states. Regarding the fatigue of UFG materials, most studies have concentrated on cyclic deformation, fatigue life, surface damage formation and underlying microstructural mechanisms [1–6]. Since the fatigue life of machine components and structures are mainly controlled by the growth life of a fatigue crack, the crack growth behavior should

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110291

be clarified for the design of safe machine components and structures. Recently, the growth behaviors of millimeter-range cracks in UFG metals were studied for compact-tension [5, 7, 8] and single edge-notched specimens [9, 10]. On the other hand, the fatigue life of smooth specimens is approximately controlled by the growth life of a small surface crack [11]. This means that the growth behavior of small cracks must be clarified to estimate the fatigue life of plain members. However, little has been reported on the growth behavior of small surface cracks in UFG metals.

In this study, stress-controlled fatigue tests were conducted on UFG copper. The formation of surface damage and growth behavior of a surface crack were monitored to clarify the physical background of fatigue damage. In addition, fracture surface analysis was conducted by scanning electron microscopy (SEM). The growth model of a small surface-crack was discussed by considering its growth path on the specimen surface and the results of fracture surface analysis.

2 Experimental procedures

The material used was pure oxygen-free copper (OFC, 99.99 wt% Cu). Prior to performing the ECAP process, the materials were annealed at 500 ˚C for 1 hr (average grain size: 100 μm). An ECAP die used had an angle of 90˚ between intersecting channels. The angles at the inner and outer corners of the channel intersection were 90 and 45˚, respectively. Repetitive ECAP was accomplished according to the Bc-route. 12 passages of extrusion resulted in an equivalent shear strain of about 11.7, respectively. The mechanical properties before ECAP were 232 MPa tensile strength, 65% elongation and a Vickers hardness number of 63. After 12 passages of ECAP the properties changed to 402 MPa, 32%, and 131, respectively. The coarse grained copper and UFG copper processed through 12 passes are referred to hereafter as CG and UFG, respectively.

Round bar specimens of 5 mm diameter were machined from the respective processed bars. The fatigue specimens were electrolytically polished ($\approx$ 25 μm from the surface layer) prior to mechanical testing in order to remove any preparation affected surface layer.

All fatigue tests were carried out at room temperature using a rotating bending fatigue machine operating at 50 Hz. The observations of fatigue damage on the specimen surface were performed using both optical microscopy (OM) and SEM. The fracture surface analysis was performed by SEM. The crack length, l, is a length measured along the circumferential direction of the surface. The stress value referred to is that of the nominal stress amplitude, σ_a, at the minimum cross section (5 mm diameter).

For EBSD-analyses, the cross section perpendicular to press direction was observed. EBSD mappings were carried out using a Tescan Mira II SEM incorporating an EDAX-TSL Hikari EBSD detector. Each pixel was 40 nm for UFG samples and 1.0 μm for coarse grain samples and hexagonal in shape. Transverse cross sections of as-ECAPed and post-fatigued bars were cut to prepare specimens for transmission electron microscopic (TEM) observation. Specimens were mechanically polished to a thickness of 100 μm, and then

subject to twin-jet electropolishing using a solution comprising 200 ml CH_3OH and 100 ml HNO_3. The jet thinning was conducted at –30 ˚C. Microstructure and diffraction patterns were analyzed using JEOL JME-2100F at 200kV.

3 Experimental results and discussion

3.1 Microstructure

Fig. 1(a) shows inverse pole figure (IPF) maps and grain boundary (GB) maps of UFG copper. The cross section perpendicular to pressing direction was observed in each sample. The GBs in GB maps are denoted either by red lines corresponding to low angle GBs where the misorientation, θ, is between 2˚ and 15˚ or by black lines corresponding to high angle GBs with θ> 15˚. The microstructure retains large fraction of equiaxed grains with high angle GBs. The average grain/cell size was measured as 295 nm.

The characteristics of GBs plotted as a function of the misorientation angle is shown in Fig. 1(b). The solid curve in each plot represents the misorientation profile of random orientations as calculated by Mackenzie and Thomson. Fig. 1(b) shows that the distribution of misorientation angles has a weak bimodal character with peaks at low and high misorientation angles. The fraction of high angle GBs is about 65%. Although there remains an excess of low angle GBs which in inconsistent with the prediction because of the continuous introduction of dislocations in each passage of ECAP processing, the misorientation distribution appears to be nearly similar to the random orientation distribution.

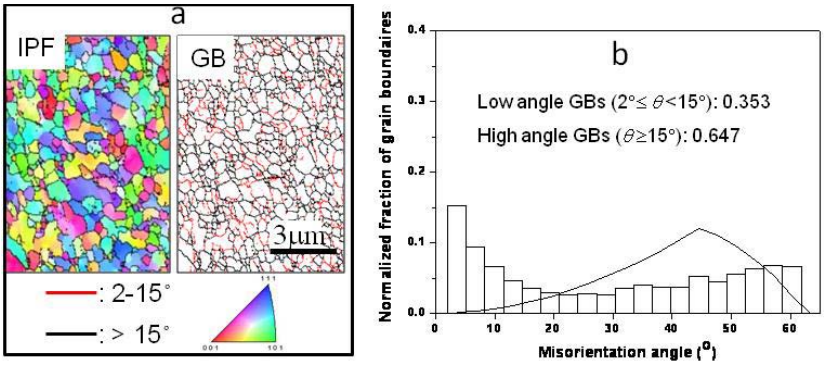


Figure 1: (a) OIM orientation and GB maps; (b) Histograms of the misorientation angles; A random distribution is indicated by solid lines superimposed on the histograms.

3.2 Fatigue characteristics and surface damage under cyclic stressing

Significant cyclic softening of UFG copper has been reported by many investigators [1, 12–14]. Agnew and Weertman [1] studied the softening property by means of hardness measurements of post-fatigued specimens. They

showed that UFG copper prepared by severe plastic deformation is susceptible to cyclic softening, where softening is found to decrease at lower strain amplitudes. In their studies, the measurements of hardness were performed on cross sections of the post-fatigued specimens. Since fatigue cracks are usually initiated at the sample surface, the surface hardness should be measured for understanding fatigue damage of UFG materials. Fig. 2 shows the change in surface Vickers hardness (load: 2.9 N) during repeated stressing of σ_a= 120 MPa. Inserted OM micrographs show a change in surface damage within the same observed area monitored at initial, medium and final stages of cycling. Here, each plot represents an average of the hardness measurements taken at six different locations. The maximum errors of the raw average hardness measurements were roughly within ±10%. Surface hardness decreases gradually with cycling in initial stages, followed by a significant drop in hardness. By investigating the change in surface hardness and formation process of surface damage together, it is found that this considerably large drop in hardness in the latter half of the fatigue life is closely related to the significant formation of damaged areas.

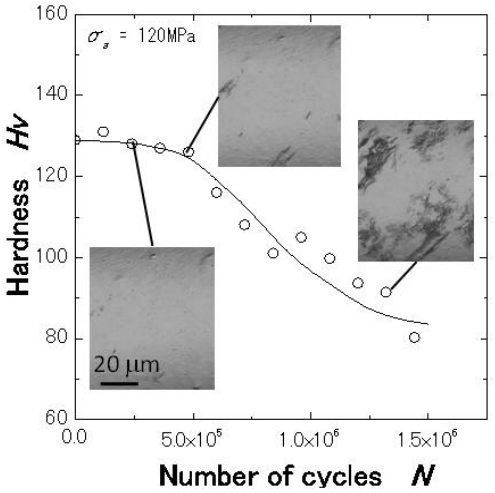


Figure 2: Change in surface hardness during stressing of σ_a = 120 MPa.

Regarding the UFG copper processed by ECAP, most studies [1, 12–14] on low-cycle fatigue indicated that the hardness measured on cross sections of post-fatigued specimens is considerably dropped but its drop is found to decrease at lower strain amplitudes. Even in the present specimens fatigued under lower stress amplitudes, the surface hardness exhibited a dramatically large drop. The hardness drop was closely related to the formation behavior of surface damage: the formation of surface damage was accelerated in the latter half of fatigue life, resulting in a significant hardness drop, whereas the mild damage formation in an initial stage of cycling brought a moderate hardness drop. It should be concluded that, thus, the initial (moderate) drop in hardness appears to result mainly from a decrease in the dislocation density inside the grains/GB-regions

and the formation of SBs. Regarding the dislocation density after fatigue, Xu et al. [15] conducted strain-controlled fatigue tests of commercial copper (99.8% Cu) processed by ECAP for 6 passes through C-route (after each pressing, the billet bar was rotated around its longitudinal axis through 180˚). From TEM observations and EBSD grain maps, they indicated that post-fatigued structures have narrower GBs and lower dislocation density in grain interiors when compared to those in virgin microstructures. For stress-controlled fatigue tests under low plastic strain amplitude (less than $5x10^{-4}$), Kunz et al. [12] indicated the formation of narrower GBs and lower dislocation densities in the grain interior for post-fatigued specimens of UFG copper (99.9% Cu).

In the latter half of the fatigue life, on the other hand, the surface hardness exhibits a significant drop tendency with a large, simultaneous extension of damaged regions (Fig. 2). As the microstructural background of hardness drop (softening) of copper processed by ECAP, the coarsening of ultrafine grains has been discussed. Höppel et al. [3] showed that pronounced grain coarsening is related to a marked cyclic softening in strain-controlled fatigue tests. Thermally activated grain coarsening must be considered as one other main reason for the

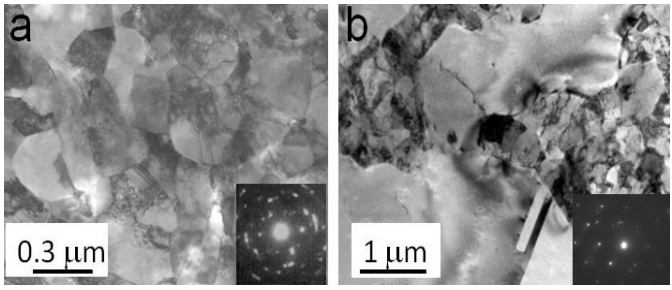


Figure 3: TEM micrographs: (a) pre-fatigue, (b) post-fatigue (σ_a = 100 MPa, $N_f = 3.1x10^6$).

cyclic softening [16, 17]. Fig. 3 shows the TEM micrographs of as-ECAPed and post-fatigued samples. In spite of very low applied stress amplitude (σ_a = 100 MPa: about 25% of tensile strength) under stress controlled fatigue, coarsened grains embedded within the original fine grain/cell regions are generated after $3.1x10^6$ repetitions. Evidently, purity and fatigue time might be important in determining the coarsening of microstructure.TEM micrograph of post fatigued specimen indicated grain coarsening and decrease in dislocation density inside coarsened grains. A cyclic softening in the latter half of the fatigue life results from the decreased dislocation density, shear banding and cell/grain-coarsening. The primary factor of significant hardness drop appears to be the grain coarsening. It has been suggested that the plastic strain localization during cyclic deformation induces the dynamic grain growth and causes the development of SBs [18]. The heavily extended surface damage in later fatigue stages indicates the formation of coarse grains, leading to the significant drop in surface hardness.

3.3 Crack growth behavior

Fig. 4 shows an initial growth behavior of a major crack. After the initiation of a crack, the crack tended to grow along the shear direction of the final pressing. When the crack length reached specific crack length (l = 0.1 – 0.2 mm at σ_a = 100 MPa), however, the change in crack growth direction occurred.

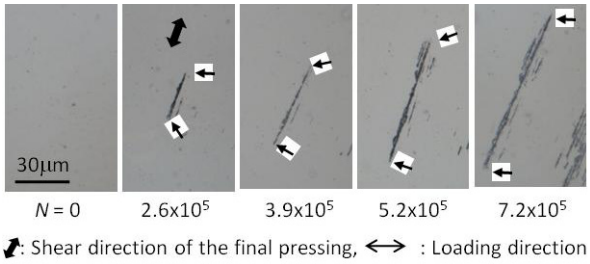


Figure 4: Changes in the surface state during stressing (σ_a = 100 MPa).

Fig. 5(a) shows the growth curve (ln l vs. N) of major cracks. Like conventional grain-sized materials, the crack growth life from an initial size (e.g. 20 μm) to fracture accounts for about 60–90% of the fatigue life of the specimens. The growth curves at higher stress amplitudes tend to approximated by straight lines, whereas the crack growth curves at σ_a= 100 and 120 MPa are roughly divided into three stages. In the first stage the crack length increases sharply with stressing and this was followed by a change in the slope of the growth curve. In the second stage the actual crack length after the slope change is smaller than the length expected from an extension of the crack growth curve from the first stage.

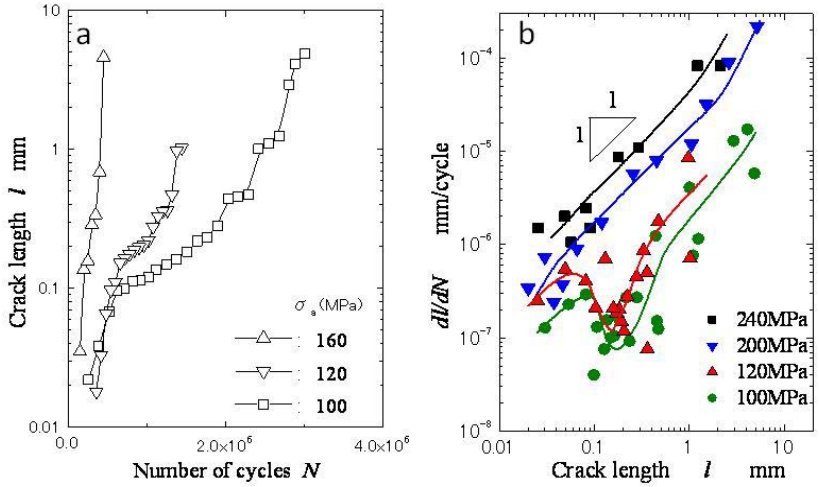


Figure 5: (a) Crack growth curve; (b) dl/dN versus l relation.

To study qualitatively the growth behavior of both UFG copper, a CGR (*dl/dN*) was calculated based on raw growth data. Fig. 5(b) shows *dl/dN* vs. *l* indicating a lower CGR. This retarded growth ceased before the crack length reached about l = 0.3 to 0.5 mm depending on the applied stress amplitude. In tertiary stage the growth curve becomes steeper and is approximated by linear relation. The linear relation nearly holds at high stress amplitudes, however the relations at low stress amplitudes exhibit a significant fluctuation, namely the CGR tends to increase initially with an increasing crack length. When the CGR reached at around $dl/dN = 3 \times 10^{-7}$ mm/c, however, a temporary drop of CGR occurs. The CGR after temporary drop gradually increases with further cycling. Once the CGR exceeds $dl/dN = 10^{-6}$ mm/c, the CGR linearly increased with an increase in crack length, holding the relation that CGR is nearly proportional to crack length.

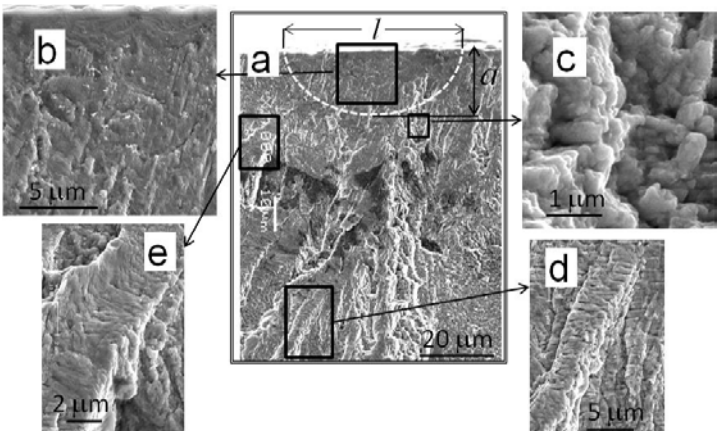


Figure 6: SEM micrographs of the fractured surface formed at σ_a = 100 MPa: (a) macroscopic view beneath the crack initiation site, (b) planar surface formed at about 5 μm beneath the surface, (c) granulated surface formed at about 22 and 45μm beneath the surface, respectively, (d) striation formed below 80 μm beneath the surface, and (e) a striated plateau surrounded with granulated surfaces.

To clarify the reason for transient retarded crack growth, SEM analysis of fracture surface were carried out. Fig. 6 shows the post-fatigued fracture surface at σ_a = 100 MPa. Fig. 6a shows a whole macroscopic view of the crack initiation site. Fig. 6b shows a magnified view of the fracture surface, a few micrometers beneath the surface (a = 5 μm, a: crack depth). A flat fracture surface is observed. With further crack propagation the morphological features of the fracture surface changed as a granulated surface is observed at 22 μm beneath the crack initiation site (Fig. 6c). Interestingly, the grain size on the granulated fracture surface is roughly equivalent to the grain size of the material (Fig. 6c). At about 80 μm below the surface, striation features that are nearly perpendicular

to the macroscopic growth direction are recognized (Fig. 6d). In addition, a striated plateau (Fig. 6e) surrounded with granulated surface is observed at equivalent crack depth where granulated surfaces are formed.

3.4 Crack growth model and evaluation of crack growth rate

Fatigue cracks usually propagate as a result of the accumulation of irreversible plastic deformation at a crack tip. In addition, the formation of a plastically deformed zone is closely related to microstructural inhomogeneity of materials. To study the crack growth mechanism a relation between the reversible plastic zone (RPZ) size at the crack tip and microstructural factors such as grain (cell) sizes, slip orientation, etc., should be investigated. The crack growth mechanism is thus discussed paying particular attention to the interrelationship between the RPZ size at the crack tip and the grain (cell) size of the material.

Here, the reversible plastic zone size, r_{rp}, under a plane strain condition was defined as:

$$r_{rp} = \frac{1}{2\sqrt{2}\pi}\left(\frac{\Delta K_{eff}}{2\sigma_{0.2c}}\right)^2 \tag{1}$$

where ΔK_{eff} is the effective stress intensity factor range and $\sigma_{0.2c}$ is the cyclic 0.2% proof stress. ΔK_{eff} was calculated from the relation $\Delta K_{eff} = U\Delta K$. U and ΔK are the crack opening ratio and the stress intensity factor range, respectively. Jono et al. [19] conducted plane-bending fatigue test (with a stress ratio, R = -1) of smooth specimens, for structural steels. They measured the opening-closing behavior of small surface-cracks by using the unloading elastic compliance method. The measurements showed that U-values of crack depths under 0.1 mm are between 0.8 to 0.6. In the present calculations, U = 0.7 was used. The solution for ΔK was taken from Fonte and Freitas for semi-elliptical surface cracks in round bars under bending [20]. Calculated values of r_{rp} and ratios of the grain size of the material, r_{rp}/d, are shown in Table 1 for the three crack depths. Essentially, different fracture surfaces were observed (Fig. 6). Consequently, planer, granular and striated fracture surfaces were formed under a range of $r_{rp}/d < 1$, $r_{rp}/d \div 1$ - 2 and $r_{rp}/d > 2$, respectively.

Table 1: Values of reversible plastic zone sizes and their ratio to the grain size.

Sample (Value of d)	σ_a MPa	l μm	a μm	r_{rp} nm	r_{rp}/d	Fracture surface*
UFG12 (d = 295 nm)	100	15	4.8	52	0.18	Fig. 6b
		100	32	348	1.18	Fig. 6c
		250	80	863	2.93	Fig. 6d

*: Corresponding micrographs in Fig. 6.

Figure 7 schematically illustrates the interrelationship between the RPZ and grains in the vicinity of the crack tip. When the RPZ size is smaller than the grain size (Fig. 7a), a crack appears to grow by a mechanism that conforms to the individual deformation mode of the localized area inside the grain and adjacent to grain boundaries (GBs). Segal [21] has shown oriented distributions of defects along the shearing plane of the pressing in a material processed by ECAP. Here, the defects refer to dislocations, grain boundaries and cellular substructure. This means that diffusion is accelerated considerably in regions along the shear direction of the final pressing, where GB sliding (GBS) can occur easily [22–24]. Thus the crack may propagate with the assistance of sliding along the maximum shear stress direction showing a planar fracture surface and a straight growth

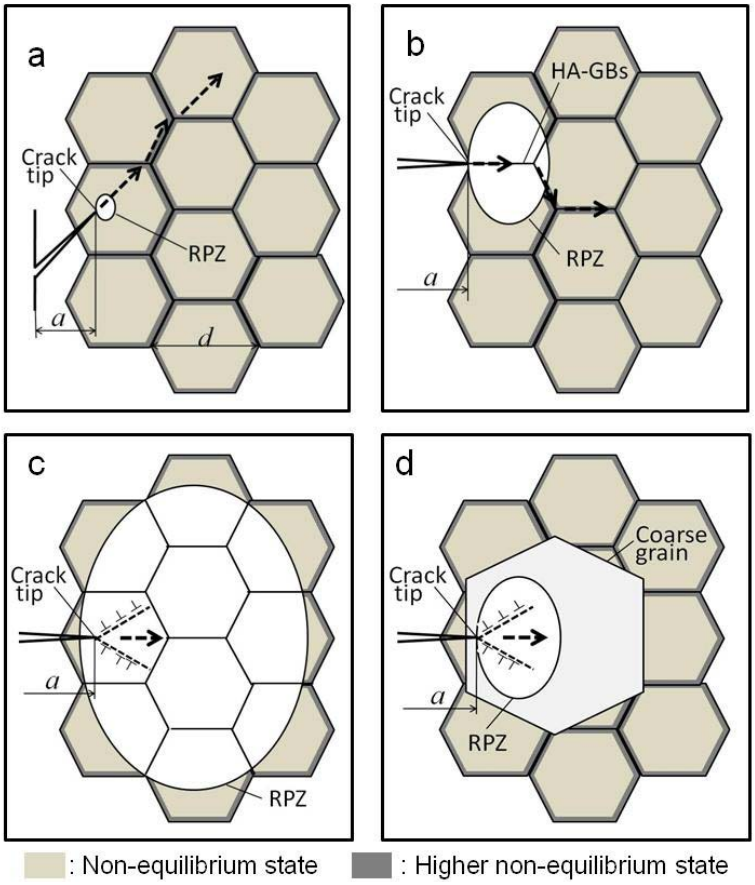


Figure 7: Schematic illustration of the crack growth mechanism in UFG12: (a) $r_{rp}/d < 1$: planar surface formation, (b) $r_{rp}/d \approx 1$: granular surface formation, (c) $r_{rp}/d >> 1$: striation formation, and (d) striated plateau formation within coarsened grains.

path along the shear direction of the final pressing. When the RPZ is 1 to 2 times the grain size (Fig. 7b) the crack should grow along GBs where an incompatibility of deformation in adjacent grains is concentrated, showing an intergranular crack growth path and a granulated fracture surface. Such a growth path may be convenient for roughness-induced crack closure which contributes to a decrease in CGR. When the RPZ size is more than 3 to 4 times the grain size size (Fig. 7c) the crack propagates because of the striations formation mechanism which is associated with crack tip retardation and blunting [25]. In addition to above stated mechanisms, for coarse grains which are evolved during cyclic stressing and are larger than a few times RPZ size, the crack grows with the formation of striated surface inside the grains as a result of irreversible plastic deformation at the crack tip, showing striated plateaus (Fig. 7e). It can thus be concluded that a temporary decrease in CGR, results from a change in the crack growth mechanism associated with the interrelationship between RPZ size and grain sizes.

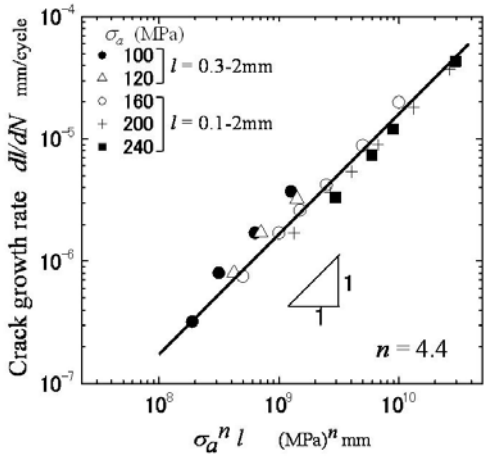


Figure 8: Growth data of small cracks: dl/dN vs. $\sigma_a^n l$ relation (n = 4.4).

It has been shown that the growth rate of small cracks cannot be treated by stress intensity factor range ΔK. In such a case, the CGR of small crack is determined uniquely by a term $\sigma_a^n l$, which is derived from an assumption that the CGR is proportional to RPZ size [11]. Fig. 8 shows the growth data of a mechanically small crack; dl/dN vs. $\sigma_a^n l$ relation. The value of n is material constant and was 4.4. All growth data plotted based on $\sigma_a^n l$ are on a straight line. The CGR of small cracks growing by striation formation mechanism is estimated by the term $\sigma_a^n l$.

4 Conclusions

The main findings of this study can be summarized as follows:

(1) The surface damage of UFG coppers was gradually formed with cycling, but significant extension of damaged areas was occurred in the latter half of fatigue life. Correspondingly, surface hardness exhibited an initial mild dropping term and subsequent severe drop in the latter half of fatigue life. The initial hardness drop strongly depends on the absorption of mobile dislocations at non-equilibrium GBs. The subsequent severe drop may be attributed to the grain coarsening.
(2) The CGR temporarily dropped at around $dl/dN = 3 \times 10^{-7}$ mm/c, and then was gradually recovered with subsequent cycling.
(3) The fracture surface showed a planar, granular and striated surface as the crack continued to grow. The ratio of the RPZ size at the crack tip to the grain size, r_{rp}/d, was calculated for crack lengths where a planar, granular and striated surface was observed. The values of r_{rp}/d for crack lengths of the planer, granular and striated fracture surfaces corresponded to a range of $r_{rp}/d < 1$, r_{rp}/d between 1 and 2, and $r_{rp}/d > 3$, respectively.
(4) To understand the change in fracture surface morphologies, a quantitative model describing the crack growth mechanism was developed based on the RPZ size and microstructural factors. The changes in the CGR and morphological features of the fracture surface were successfully explained by this model.
(5) The CGR of a mechanically small surface crack could not be estimated by the stress intensity factor range, but it was uniquely determined by a term, $\sigma_a^n l$, which is derived from an assumption that CGR is proportional to RPZ size. The n is material constant and was 4.4.

Acknowledgements

This study was supported by a Grant-in-Aid (20560080) for Scientific Research (C) from the Ministry of Education, Science, and Culture of Japan, and a grant from the Fundamental R&D Program for Core Technology of Materials funded by the Ministry of Commerce, Industry and Energy, Republic of Korea.

References

[1] Agnew, S.R. & Weertman, J.R., *Materials Science and Engineering*, **A 244**, pp. 145-153, 1998.
[2] Vinogradov, A. & Hashimoto, S., *Materials Transactions*, **42**, pp. 74-84, 2001.
[3] Höppel, H.W., Zhou, Z.M., Mughrabi, H. & Valiev, R.Z., *Philosophical Magazine*, **A 87**, pp. 1781-1794, 2002.
[4] Mughrabi, H., Höppel, H.W. & Kautz, M., *Scripta Materialia*, **51**, pp. 807-812, 2004.
[5] Vinogradov, A., Nagasaki, S., Patlan, V., Kitagawa, K. & Kawazoe, M., *NanoStructure Materials*, **11**, pp. 925-934, 1999.
[6] Goto, M., Han, S.Z., Yakushiji, T., Lim, C.Y. & Kim, S.S., *Scripta Materialia*, **54**, pp. 2101-2106, 2006.

[7] Pao, P.S., Jones, H.N., Cheng, S.F. & Feng, C.R., *International Journal of Fatigue*, **27**, pp. 1164-1169, 2005.
[8] Chung, C.S., Kim, J.K., Kim, H.K., & Kim, W.J., *Materials Science and Engineering*, **A337** (2002) pp. 39-44, 2002.
[9] Kim, H.K., Choi, M-I., Chung, C.S. & Shin, D.H., *Materials Science and Engineering*, **A340**, pp. 243-250, 2002.
[10] Hanlon, T., Tabachnikova, E.D. & Suresh, S., *International Journal of Fatigue*, **27**, pp. 1147-1158, 2005.
[11] Nisitani, H., Goto, M. & Kawagoishi, N., *Engineering Fracture Mechanics*, **41**, pp. 499-513, 1992.
[12] Kunz, L., Lukáš, P. & Svoboda, M., *Materials Science and Engineering*, **A424**, pp. 97-104, 2006.
[13] Maier, H.J., Gabor, P. & Karaman, I., *Materials Science and Engineering*, **A410-11**, pp. 457-461, 2005.
[14] Agnew, S.R., Vinogradov, A., Hashimoto, S. & Weetman, J.R., *Journal of Electronic Materials*, **28**, pp. 1038-1044, 1999.
[15] Xu, C., Wang, Q., Zheng, M., Li, J., Huang, M., Jia, Q., Zhu, J., Kunz, L. & Buksa, M., Fatigue behavior and damage characteristic of ultra-fine grain low-purity copper processed by equal-channel angular pressing (ECAP), *Materials Science and Engineering*, **A475**: 249-256, 2008.
[16] Mughrabi, H. & Höppel, H.W., Cyclic deformation and fatigue properties of very fine-grained metals and alloys, *International Journal of fatigue*, **32**, pp. 1413-1427, 2010.
[17] Molodova X, Gottstein G, Winning M, Hellmig RJ., Thermal stability of ECAP processed pure copper, *Materials Science and Engineering*, **A460-461**, pp. 204-213, 2007.
[18] Furukawa, Y., Fujii, T., Onaka, S. & Kato, M., *Materials Transactions*, **50**, pp. 70-75, 2009.
[19] Jono, M., Song, J., Yama, Y., Nisigaichi, N., Okabe, M. & Kikukawa, M., *Transactions of Japan Society of Mechanical Engineers*, Ser-A, **51**, pp. 1677-1686, 1985.
[20] Fonte, M da. & Freitas, M de., *International Journal of Fatigue*, **21**, pp. 457-463, 1999.
[21] Segal, V.M., *Materials Science and Engineering*, **A197**, pp. 157-164, 1995.
[22] Valiev, R.Z., Kozlov, E.V., Ivanov, Yu F. Lian, J., Nazarov, A.A. & Baudelet, B., *Acta Metallurgical Materials*, **42**, pp 2467-2475, 1994.
[23] Chinh, N.Q., Szommer, P., Horita, Z. & Langdon, T.G., *Advanced Materials*, **18**, pp. 34-39, 2006.
[24] Goto, M., Han, S.Z., Kim, S.S., Kawagoishi, N. & Lim, C.Y., *Scripta Materialia*, **57**, pp. 293-296, 2007.
[25] Lynch, S.P., Progression markings, striations, and crack-arrest markings on fracture surfaces, *Materials Science and Engineering*, A4**68-470**, pp.74-80, 2007.

Structural characterization of electro-thermally driven micro-actuators with immeasurable temperature-dependent material characteristics

W. Szyszkowski & D. Hill
Department of Mechanical Engineering, University of Saskatchewan, Canada

Abstract

Multi-cell cascaded Electro-Thermal Micro-Actuators (ETMA) made of nickel alloys are analyzed by the finite element (FE) method. The computer simulation runs over the electrical, thermal, and mechanical phases of the ETMA operations. The main challenges of modeling are discussed. Some of the material's parameters such as the electrical resistivity, thermal expansion coefficient and emission are strongly temperature dependent. Furthermore, any measurements of such dependences are complicated by a magnetic phase transition occurring in nickel within the operating range of temperature. All the properties are sensitive to a particular composition of the material. The surface convection is additionally shape-dependent and, mainly due to small dimensions of the actuators, cannot be determined experimentally with sufficient accuracy. For the above reasons the use of the material data estimated from the available literature usually does not render reliable simulations.

In the approach proposed the material characteristics of the ETMA considered are determined by utilizing the fact that for a given applied voltage the total current and displacement in the real actuators (performance parameters) are measured with a relatively high precision. Similar performance parameters, i.e. the total current and displacement can be obtained as output from the FE simulation in which some important material properties of the actuator's model are assumed in a parametric form (material's parameters). The FE simulation procedure is integrated with these real measurements in such a way that the

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110301

material parameters are iteratively adjusted to match the simulated and experimental performance parameters for the sampled actuators.

Keywords: micro-actuators, finite elements, electro-thermal-mechanical fields, uncertain material characteristics.

1 Introduction

Electro-thermal micro-actuators (ETMA) are constructed as monolithic compliant mechanisms. They have the displacement/force outputs typically higher than other micro-actuators, and should be easier to control [1, 2]. However, mainly due to relatively high temperatures of the operating regimes [3, 4], any analytical predictions and numerical simulations are rather challenging. Several parameters to characterize three different physical environments involved, i.e. electrical, thermal, and mechanical, may vary quite substantially over the temperature ranges the actuators usually experience.

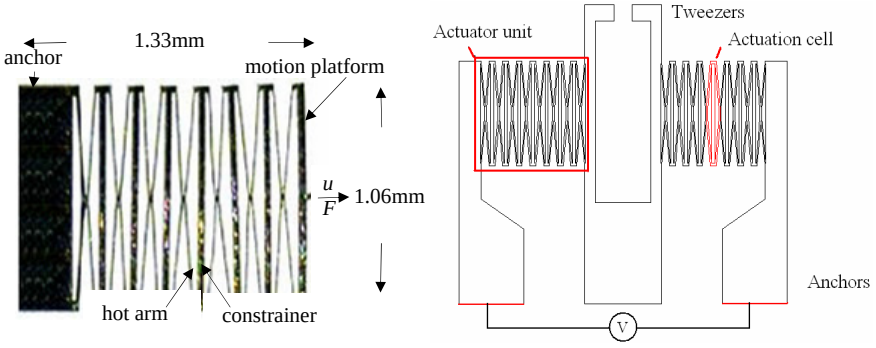
While the finite element (FE) technique is capable of handling complicated geometrical shapes, obtaining data and modeling these material parameters for a particular ETMA remains difficult and seems to be the main source of the discrepancies between the computed and measured results. Some of the parameters are also scale dependent and almost impossible to measure for such small devices, or may be affected by the changes in the material's microstructure triggered at some transition temperatures difficult to detect with a sufficient precision. Consequently, analytical predictions of the structural characteristics of such devices have not been particularly accurate so far, and that's why any new designs still have to rely heavily on costly, time consuming, and numerous experimental testing.

Most of the FE analysis of ETMA reported in the literature adopted the material parameters as temperature-independent constants at best 'averaged' over the expected operating temperature range [5, 6], while in fact their values may vary several times over that range.

The above issues are discussed here on the example of cascaded ETMA made of nickel alloy and manufactured by using laser micro-fabrication technology [7]. All the main material parameters are assumed temperature-dependant. The material's description is assessed by comparing the ETMA's measured and the FE simulated performance parameters. In order to improve accuracy the FE simulation is combined with experimental testing to modify the values of some uncertain material parameters. The procedure is iterative in which the changes in temperature distribution, indicating the coupling between the electrical and thermal fields, and the ETMA performance, which is sensitive to the current values of the material parameters, are monitored. It is demonstrated that only the measurements pertaining to the electrical and mechanical fields are needed, while rather cumbersome, unreliable, or just impossible measurements of the heat transfer and temperatures can be avoided.

2 The ETMA and the governing equations

The multi-cascaded ETMA of the overall dimensions shown in Figure 1a was manufactured from a 25μm thick nickel alloy foil by a laser-material removal process [7]. An application of two actuators to drive micro-tweezers is sketched in Figure 1b. Each actuation unit, which has a fixed electrical anchor to apply power to, consists of several *actuation cells* connected serially. The cell is composed of two actuation beams to be referred to as *hot arms*, and one constraining beam to be referred to as a *constrainer*. Electro-thermal micro-actuators utilize the effects of thermal expansion of metals upon heating due to electrical current when voltage V_a is applied. The thermal expansion is to be converted into either the vertical displacement u of the motion platform, or the force F, or both. For best effects, the temperature field should be such that hot members expand as freely as possible, while the thermal expansion in constrainers is reduced as much as possible. These requirements are not easy to meet, because both the thermal and structural effects in hot arms and constrainers are coupled.



a) A single actuator b) Micro-tweezers

Figure 1: Multi-cell cascaded ETMA.

The distributions of voltage V, current density i, and power Q in the actuator are governed by the equations:

$$\nabla \cdot (\tfrac{1}{\rho}\nabla V) = 0$$
$$i = -\rho \nabla V \qquad (1a,b,c)$$
$$Q = \tfrac{1}{\rho}\nabla V \nabla V$$

where $\rho(T)$ is a temperature dependent resistivity of the material.

Power Q is converted into Joule's heat generated per unit of the volume. The temperature raise, T, is governed by the equations representing the heat transfer via conduction in the body and through convection/radiation on the surface in the form:

$$\nabla(K\nabla T) + Q = 0 \qquad (2a,b)$$

$$K\nabla T + h(T - T_0) + eB(T^4 - T_0^4) = 0$$

where $K(T)$, $h(T)$, and $e(T)$ are the temperature-dependant conduction, convection, and emissivity coefficients respectively.

Finally, the stress/strain/displacement (denoted by σ, ε, and u respectively) field is governed by equilibrium, constitutive law, and kinematics equations in the form:

$$\sigma_{ij,j} + f_i = 0$$
$$\sigma_{ij} = D_{ijkl}(\varepsilon_{kl} - \alpha\Delta T\delta_{kl}) \quad (3abc)$$
$$\varepsilon_{ij} = \tfrac{1}{2}(u_{i,j} + u_{j,i})$$

where D_{ijkl} is defined in terms of the temperature-dependent elasticity modulus $E(T)$ and Poisson's ratio $\nu(T)$, while $\alpha(T)$ is the thermal expansion coefficient.

As can be seen from equations (1) and (2), the temperature T and the voltage V distributions are both affected by the power Q effectively coupling the electrical and thermal fields. To analyze these two phases the following material parameters are required $\rho(T)$, $K(T)$, $h(T)$, and $e(T)$. Once the temperature is determined the corresponding stress/deformations can be obtained. In this final phase only $E(T)$ and $\nu(T)$ have to be known.

The temperature in the actuator considered varies from point to point within the range of about 20-600ºC. As discussed in the next section, some of the above parameters may vary quite substantially over that range. Also, the convection coefficient, which varies the most, is generally almost impossible to obtain from experimental measurements.

3 Effects of temperature on the material characteristics

Although nickels alloys are popular materials used in microelectronics, no data were available for the alloys matching exactly the material used in this particular ETMA (90% Nickel, 6% Copper, and 4% trace materials). However, the published data [8–11] for similar compositions have generally indicated the following:

- The resistivity and the convection coefficients are affected by temperature much more than the other parameters.
- In particular the temperature dependence of the resistivity is highly sensitive to material composition.
- Within the temperature range of the ETMA operations a magnetic phase change may occur in the material, which will affect its properties.

Another important factor is that the convection coefficient is also strongly shape-dependent, which means, for example, that it should be measured on 'micro-samples' of the geometry similar to the geometry of the hot arm ($0.010 \times 0.025\,mm$ rectangle), the task that practically appears to be almost impossible.

A 'macro-sample' of dimensions $22 \times 6 \times 0.025\,mm$ kept at a uniform temperature was used to determine the resistivity and emissivity. The

temperature was measured with the help of a hot plate and thermal glasses. It was found that the resistivity varies quadratically with temperature up to about 306ºC, above this temperature the variation is linear. It was concluded that this is due to the phase change that takes place at $T_p = 306\,^{\circ}\mathrm{C}$. The following relationship was used in the FE simulation (SI units are used):

$$\rho(T) = \begin{cases} 6.77 \cdot 10^{-5} \cdot (1 + 0.00476T(1 + 0.00303T)) & if \quad T \le 306 \\ 16.4 \cdot 10^{-5} \cdot (1 + 0.001872T) & if \quad T > 306 \end{cases} \tag{4a,b}$$

In should be noticed that the magnitude of ρ (T) changes by some 420% over the range of 20–600ºC.

The emissivity was found to vary pretty much linearly with temperature. The variation was approximated by:

$$e(T) \cong 0.1 \cdot (1 + 0.0008T) \tag{5}$$

Note that the magnitude of this parameter varies only about 50% over the same range of temperature.

According to [10], variation of the thermal conduction coefficient should be less significant over that temperature range. Typically its value first slightly drops with temperature, and then starts to increasing, the change most probably associated with the magnetic phase change already mentioned.

$$K(T) = \begin{cases} 92.4 - 0.093T & if \quad T < 306 \\ 56.3 + 0.025T & if \quad T > 306 \end{cases} \tag{6}$$

It should be noted that this coefficient varies about 30% over the whole temperature range.

The relations $h(T)$ is more difficult to establish. This is because generally the convection coefficient dependents on the geometry of heat transferring body. On the other hand, any direct temperature or heat flux measurements on a 'micro-sample' with the dimensions comparable to the hot arms dimensions are very challenging (too small for thermal glasses, for example). Nevertheless the 'macro-sample' mentioned above was used to determine this coefficient as well. The results are denoted by the curve $h_t(T)$ shown in Figure 2. This curve is generally in a range for a flat surface undergoing free convection [12]. Note that for this test the value of the convection coefficient changed about 8 times over the temperature range from 20ºC to 600ºC.

What was the most important, however, was that $h_t(T)$ used in the FE simulation were giving consistently much *higher* then expected temperature of the hot arms (around 2400ºC), which in turn indicated that the convection coefficient for small dimensions members of the actuator should be much higher than those obtained for a relatively flat surface of the macro-sample. However, it was noticed that the convection coefficient data for small diameter wires

published in [12] are generally of similar character, but numerical values were higher. For example, for the wire of 100μm this coefficient, denoted by $h_w(T)$ in Figure 2, is about 5 times higher than $h_t(T)$. Also, one should realize that the convection coefficient values shown in Figure 2 vary about 10 times over the temperature range of 20-600ºC.

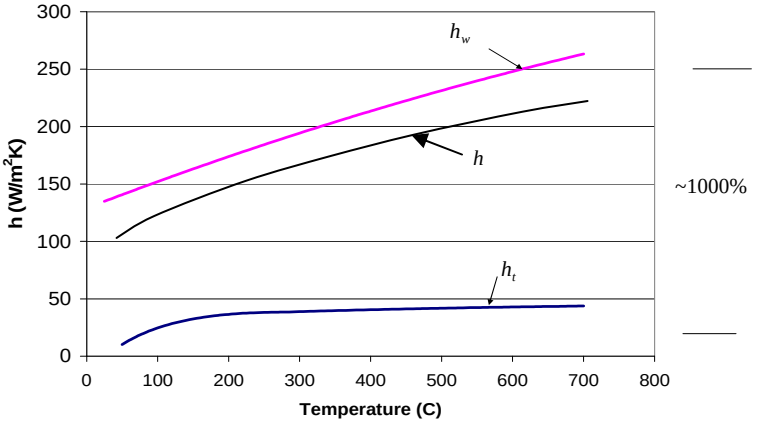


Figure 2: The convection coefficient.

When the curve $h_w(T)$ was used in the ETMA's simulation, the resulting temperature was in turn *lower* than expected. It indicated, when combined with the numerical test performed on $h_t(T)$, that for the members of the actuator the convection coefficient $h(T)$ should be limited by $h_t < h < h_w$. Therefore this coefficient is assumed in the form:

$$h(T) = \eta h_w + (1-\eta)h_t, \tag{7}$$

where the value of parameter η is to be determined by comparing the simulation and experimental results.

The mechanical properties such as Young's modulus, *E*, and Poisson's ratio *v*, as well as the thermal expansion coefficient α are less affected by the temperature and the alloy composition.

The variation of *E* with temperature up to 600ºC were adopted from [8] in the form:

$$E(T) = 206.4 \cdot 10^9 \cdot (1 - 0.000286T) \tag{8}$$

The Poisson's ratio *v*, according to [11], can be considered approximately constant and equal to 0.31.

Reference [8] also suggested the linear variation of α with temperature as:

$$\alpha(T) = 13 \cdot 10^{-6}(1 + 0.000343T) \tag{9}$$

However, the detailed data reported in [13,14] for a similar Nickel alloy shown that α increased with temperature up to the magnetic phase transition, then dropped down, and started increasing again for temperatures above

approximately 450ºC. The data as well as the linear formula (11) are plotted in Figure 3. In the FE simulation the $\alpha(T)$ relation is approximated by piecewise linear functions also indicated in Figure 3, with the parameters $\alpha_0, \alpha_1, \alpha_2, \alpha_3, T_1$, T_2, and T_3. The values of these parameters indicated in this figure were determined by matching the simulation and experimental displacement results as closely as possible, as explained in the next section.

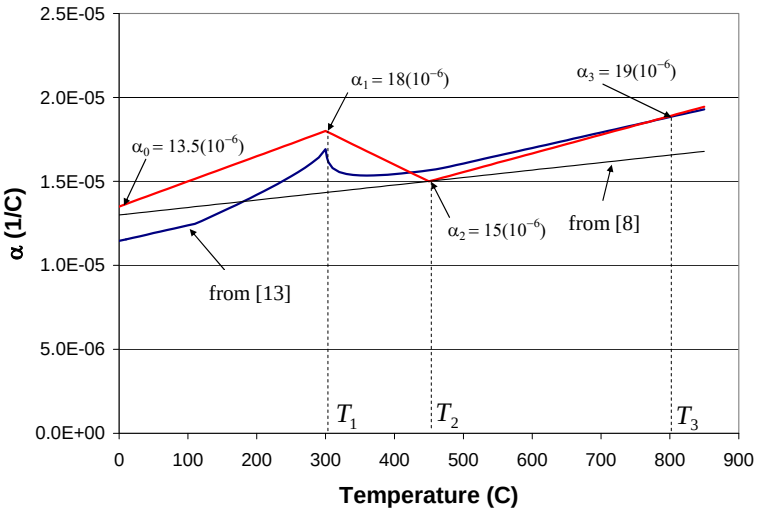


Figure 3: The thermal expansion coefficient.

Note that for the temperature range considered the values of $E(T)$ or $\alpha(T)$ parameters change only about 20%. Therefore any inaccuracies in these properties should have much less effects on the response of the simulated actuator than the inaccuracies in the convection or resistivity coefficients.

4 The FE modeling and simulation procedure

The finite element simulation of the cascaded actuation unit would have to depict accurately the response of the system across three main physics environments; electrical, thermal and mechanical. The flow chart in Figure 4 outlines the main blocks of the analysis. The complete simulation procedure was setup in the ANSYS FE program, which allows running the three blocks with a relatively smooth flow of data between them. It has been found that 2-D elements provide almost identical accuracy at significantly less calculations (in this case over 20 times faster than the model with 3-D elements). A heat transfer in the direction perpendicular to the actuator's plane is included in the ANSYS' 2-D elements SHELL157, which were used within the electrical and thermal blocks. Next, the element SHELL63 was used within the mechanical block. In the areas where

large amounts of Joule heat is produced elements are smaller and placed more densely (Figure 7). A particular meshing of the FE model was considered accurate (the model accuracy) when it was found insensitive to increased number of elements.

All the three environments include temperature dependent material properties, and therefore are coupled. However, as mentioned before, the iteration loops running only the electrical and thermal block are required to obtain the temperature solution.

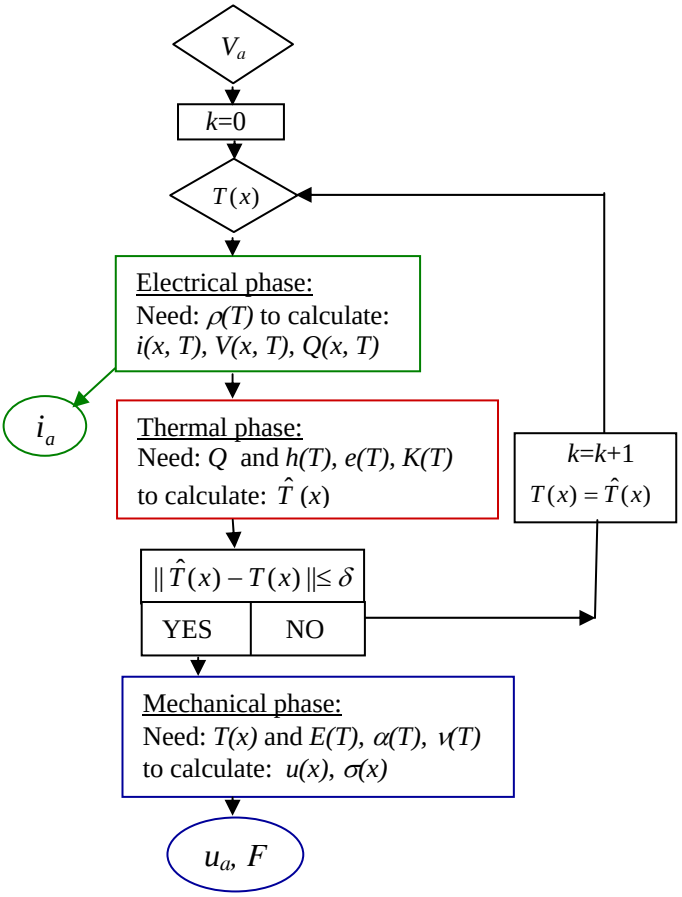


Figure 4: Flow chart of the FE analysis.

To start the simulation, an external voltage V_a is applied across a cascaded actuation unit. Resistivity $\rho(x,T(x))$ at each point x of the actuator is calculated using an initially assumed (or calculated in the previous loop) temperature distribution $T(x)$. The electrical procedures are then used to obtain current

density $i(x,T)$, voltage $V(x,T)$ and the Joule heat energy $Q(x,T)$ distributions. The latter is sent to the block solving the heat transfer problem, in which a best guess of the value for η in Eqn. (7) is used. In this block the thermal properties such as convection $h(T(x))$, emissivity $e(T(x))$, conduction $K(T(x))$ are evaluated at the current temperature distribution $T(x)$.

With this information a new temperature distribution $\hat{T}(x)$ is calculated. If this new temperature is not within some preset convergence limits δ, then $\hat{T}(x)$ is substituted for $T(x)$ and the electrical and thermal analyses are repeated. Actually, it was solved by the standard Newton-Raphson in which the convergence was set in such a way that $\left\| i(x,\hat{T})/i(x,T)-1 \right\| < 0.001$ for the electrical block and $\left\| q(x,\hat{T})/q(x,T)-1 \right\| < 0.001$ for the thermal block (where q is the heat flux). Once a convergent temperature distribution is determined, the total current for the actuator i_a is calculated and the temperature field is sent to the mechanical block. This temperature is used to assign the values of such material constants as the elastic modulus $E(T(x))$, the thermal expansion coefficient $\alpha(T(x))$ to calculate the stress and displacement fields, which includes the displacement of the motion platform u_a.

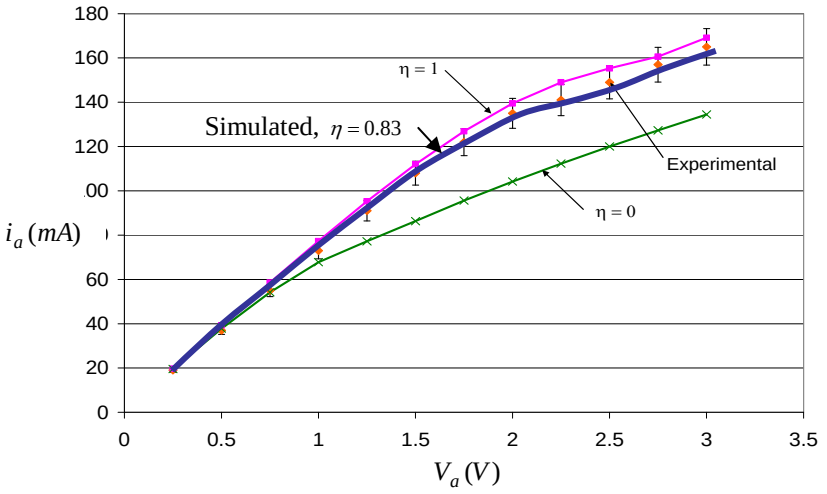


Figure 5: The current match.

Note that the magnitude of this displacement depends on the values assigned to parameters $\alpha_0, \alpha_1, \alpha_2, \alpha_3, T_1$, T_2, and T_3 discuss before. If the platform motion is constrained then force F generated by the actuation unit to constrain it is calculated.

The real actuators are usually tested for the $i_a^{\exp}(V_a)$ and $u_a^{\exp}(i_a)$ relationships (see Figures 5 and 6).

In order to achieve as close an agreement as possible with these experimental result the parameters η, α_i, and T_i characterizing the material properties in the simulation presented in Figure 4 should be properly adjusted. Note that the difference $i_a - i_a^{\exp}$ depends only on the value of parameter η. The 'best' match for the voltage from $0.25V$ to $3.0V$ was obtained for η =0.83 and is shown in Figure 5. Also included, for comparison, are the results when using $h_t(T)$ or $h_w(T)$. In turn, to minimize the difference $u_a - u_a^{\exp}$ the values of parameters α_i and T_i indicated in Figure 3 were modified. The best match is shown as a solid line in Figure 6.

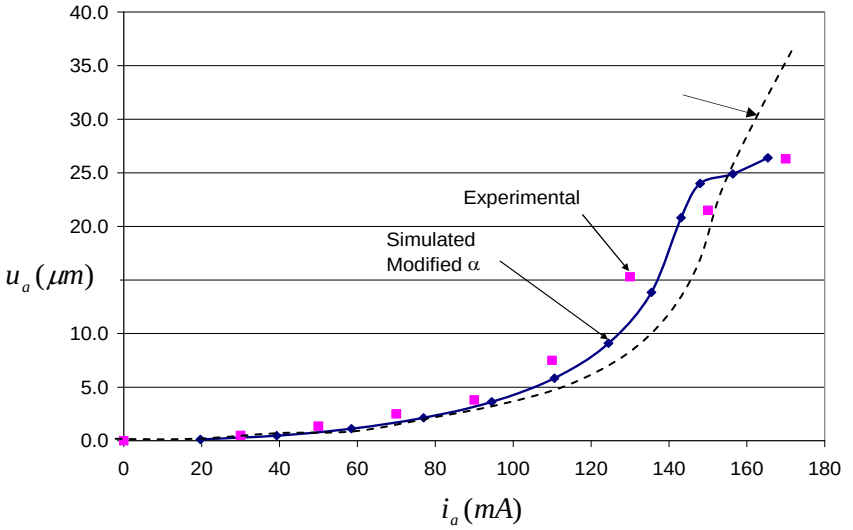


Figure 6: The displacement match.

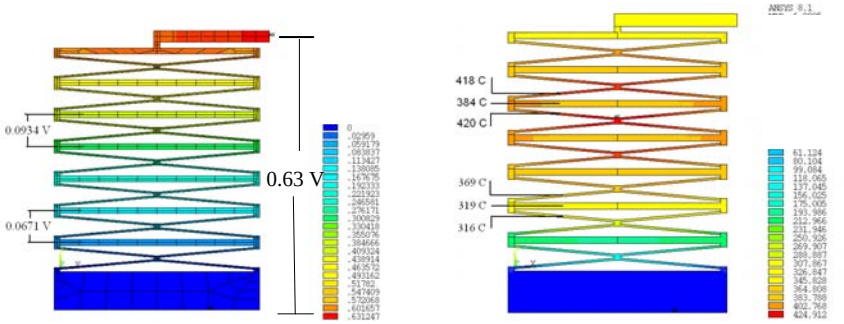
The displacement obtained by the applying the linear relation (9) for the thermal expansion coefficient is also shown, for comparison.

5 Other results of the simulation

The voltage distribution for $V_a = 0.63$ V is depicted in Figure 7a. The average voltage drop over an actuation cell is 0.09V. However, in the bottom cells a voltage drop is ~0.067V, while in the top cells a drop is ~0.093V. This unequal voltage distribution can be attributed to different temperatures of these cells (this is shown in Figure 7b). Bottom cells are cooler then the top cells due to

conduction of heat into the anchor. Since resistivity is lower in colder cells and the same current passes through each cell, the voltage drop is not uniform.

The corresponding temperature distribution is presented in Figure 7b. It should be noted that roughly half of the actuator's material undergone the transition phase occurring at $T_a = 306\ ^{o}C$.



a) The voltage distribution b) The temperature distribution

Figure 7: The FE results.

On average the constrainers are cooler by about 35ºC than the hot arms. Closer analysis reveals that this difference is responsible for about 90% of the displacement generated by the unit at the motion platform

6 Conclusions

Detailed knowledge of the temperature dependency of several electro-thermal-mechanical material properties is necessary to accurately simulate the ETMA by the FE method. A great deal of caution should be exercised when selecting data from the literature, because some of these properties are very sensitive to a particular composition of the material. Also, some changes in the material internal structure may have significant effect on these properties. For example, the resistivity of the nickel alloys used in the ETMA presented varies quadratically below the temperature of the magnetic phase change, but linearly above it. Similarly, the conduction drops with temperature below the transition point, and increases above it.

The simulation results appear to be mostly affected by the uncertainty in the convection coefficient, the values of which may vary several times over the temperature range involved. This coefficient happens to be also scale dependant and impossible to measure accurately for the members of the actuator due to its small dimensions.

The 'uncertain' material characteristics can be 'tuned up' by combining the FE simulation with the experimental measurements of the ETMA performance. It should be noticed that the electrical and thermal characteristics should be adjusted simultaneously due to coupling, while the mechanical properties could be modified independently.

The FE simulation procedure is iterative, runs across the three physics phenomena, and is computationally extensive. However, proper choices of elements, meshing patterns, and iterative schemes may substantially reduce the calculations and alleviate the convergence problems.

References

[1] T. Moulton, G.K. Ananthasuresh. Micromechanical devices with embedded electro-thermal-compliant actuation. *Sensors and Actuators* A, 90, 2001, p 38-48.
[2] D. Hill, W. Szyszkowski, and E. Bordachev, On Modeling and Computer Simulation of an Electro-thermally Driven Cascaded Nickel Micro-actuator, *Sensors and Actuators* A, 126, 2006, pp. 253-263.
[3] C.P. Hsu, W.C. Tai, W. Hsu. Design and Analysis of an Electro-Thermally Driven Long-Stretch Micro Drive with Cascaded Structure. *Proceedings of IMECE2002*, 2002.
[4] H. Du, C. Su, M. K. Lim, W. L. Jin. A micro-machined thermally-driven gripper: a numerical and experimental study. *Smart Mater. Struct*, 8, 1999, p. 616-622.
[5] P. Lerch, C.K. Slimane, B. Romanowicz, P. Renaud. Modelization and characterization of asymmetrical thermal micro-actuators. *J. Micromech. Microeng*. 6, 1996, p 134-137.
[6] N. Mankame, G.K. Ananthasuresh. Comprehensive thermal modeling and characterization of an electro-thermal-compliant microactuator. *J. Micromech. Microeng*. 11, 2001, p 452-462.
[7] E.V. Bordatchev, S.K. Nikumb, and W. Hsu, Laser micromachining of the miniature functional mechanisms, *Proceedings of SPIE, Vol. 5578* (SPIE, Bellingham, WA, 2004), paper # 5578D-77, pp. 579-588.
[8] King, J., *Materials Handbook for Hybrid MicroElectronics*. Artech House, Boston, 1988.
[9] Everhart, J., *Engineering Properties of Nickel and Nickel Alloys*. Plenum Press, New York, 1971.
[10] Temperature Dependent Elastic & Thermal Properties Database, MPDB (JAHM) Software, Inc., USA, 1999.
[11] Yao, Y D, Tsai, J H. Magnetic Phase Transition in Nickel-Rich Nickel-Copper Alloys, *Chinese Journal of Physics*, Vol. 16, No 4. p. 189 – 195, 1978.
[12] Incropera F.P. and Dewitt D.P. *Fundamentals of Heat and Mass Transfer*, Fifth Edition. John Wiley and Sons, New York, 2002.
[13] T A Faisst. Determination of the critical exponent of the linear thermal expansion coefficient of nickel by neutron diffraction. *J. Phys.: Condens. Matter 1*, 1989, p 5805-5810.
[14] T G Kollie. Measurement of the thermal-expansion coefficient of nickel from 300 to 1000 K and determination of the power-law constants near the curie temperature. *Physical Review B*, V16, N11, 1977, p 4872-4882.

Impact loading of inelastic elliptical plates

J. Lellep & A. Mürk
Institute of Mathematics, University of Tartu, Estonia

Abstract

Inelastic response of elliptical plates to impact and initial impulsive loading is studied. For determination of the deflected shape of plates the concept of mode form motions amalgamated with the conical velocity field in used. Theoretical predictions of residual deflections are developed for plates of piece wise constant thickness. The cases of circular and annular plates subjected to initial impulsive loading are studied as particular cases of an elliptical plate.

Keywords: impulsive loading, plasticity, plate, elliptical plate, mises condition.

1 Introduction

Thin plates and shells are important elements of structures. In accidental situations the plates can be subjected to impact and shock loadings. This involves the need for methods of evaluation of maximal residual deflections caused by initial impact and impulsive loading.

Exact and approximate theoretical predictions and experimental results regarding to the behaviour of inelastic structures have been presented by several authors. Reviews of these studies can be found in the books by Jones [2], Kaliszky [5], Stronge and Yu [17], also in papers by Kaliszky [4], Jones [3], Kaliszky and Logo [6], Nurick and Martin [12], Yu and Chen [21]. Shen and Jones [15], also Wen *et al*. [20] studied the dynamic plastic response of fully clamped circular plates in the cases of rate sensitive and rate insensitive materials. Liu and Stronge [11] considered simply supported circular plates subjected to dynamic pressure at the central part of the plate. Wang *et al*. [19] used the concept of the unified strength theory in the dynamic plastic analysis. Lellep and Hein [7], Lellep and Mürk [8, 9] studied stepped plates and shallow shells. In papers [10, 9] the concept of plates with stable cracks located of the re-entrant corners of steps was used for determination of optimal parameters of the plate.

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110311

Approximate methods of determination of residual displacement fields have been suggested by Jones [2], Symonds [18], Kaliszky [4] and others. Theoretical predictions obtained for axisymmetric and rectangular plates are in good agreement with experimental data [2]. It is somewhat surprising that only a little attention is paid to non-axisymmetric problems, excluding the case of rectangular plates.

In the present paper an attempt is made to evaluate residual deflections of stepped elliptical plates subjected to impulsive and impact loading.

2 Formulation of the problem and governing equations

Consider a thin walled plate of variable thickness $h = h(r,\theta)$. Here r,θ stand for polar coordinates, which are related to Cartesians as

$$x = r\cos\theta,\ y = r\sin\theta\ . \tag{1}$$

Let the boundary of the midplane of the plate be an ellipse (fig. 1)

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1, \tag{2}$$

where *a* and *b* are given constants.

Substituting eqn. (1) to eqn. (2) one obtains the equation of the plate boundary as $r = r_*$ where

$$r_* = \frac{ab}{\sqrt{a^2\sin^2\theta + b^2\cos^2\theta}}\ . \tag{3}$$

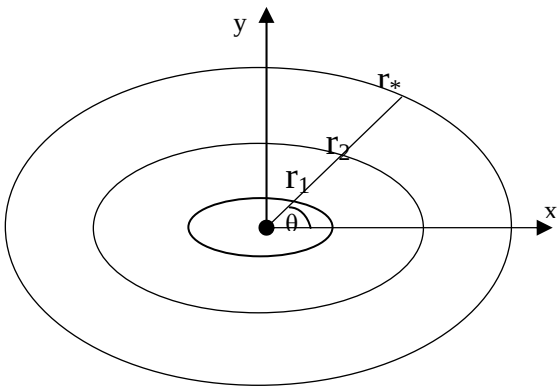


Figure 1: An elliptical plate.

The plates under consideration have piece wise constant thickness or the thickness can be approximated with a stepped distribution, e.g. $h = h_j$ for $r \in [\, r_j(\theta), r_{j+1}(\theta)\,]$ where $j = 0,...,n$ and $h_j = const$. The curves $r = r_j(\theta)$ are assumed to be given closed smooth curves.

We assume that the plate under subjected to initial impulsive loading and that the initial kinetic energy K_0 of the plate is given. The subsequent motion is due to inertia only so that initial kinetic energy will finally be absorbed in the plastic work. As regards the distribution of the initial velocity we assume that it is unidentified. However, it is expected that it can be approximated with a conical distribution which has its apex at a given point. The vertical projection of the apex on the middle plane of the plate is taken the origin of polar coordinates r and θ.

Material of the plate is assumed to be an ideal rigid-plastic one obeying von Mises yield condition.

The aim of the paper is to develop approximate theoretical predictions of residual deflections for plates of elliptical shape subjected to impact loading. For evaluation of residual deflections the method of mode form motions will be employed.

Von Mises yield condition can be presented in the form $\Phi \le 0$ where (see Chakrabarty [1])

$$\Phi = M_1^2 - M_1 M_2 + M_2^2 + 3M_{12}^2 - M_0^2 . \tag{4}$$

In eqn. (4) M_1. M_2 are the bending moments with respect to axes *Ox* and *Oy*, respectively, M_{12} is the shear moment on the *xy*-plane whereas M_0 stands for the yield moment. In the case of a solid plate $M_0 = \sigma_0 h^4 / 4$; *h* being the thickness of the plate and σ_0 – the yield stress of the material.

According to the associated flow law one has

$$\begin{aligned} \dot{\kappa}_1 &= \overline{\lambda}(2M_1 - M_2), \\ \dot{\kappa}_2 &= \overline{\lambda}(2M_2 - M_1), \\ \dot{\kappa}_{12} &= 6\overline{\lambda} M_{12}, \end{aligned} \tag{5}$$

where $\overline{\lambda}$ is a non-negative scalar multiplier and

$$\dot{\kappa}_1 = -\frac{\partial^2 \dot{W}}{\partial x^2},\ \dot{\kappa}_2 = -\frac{\partial^2 \dot{W}}{\partial y^2},\ \dot{\kappa}_{12} = -2\frac{\partial^2 \dot{W}}{\partial x \partial y} . \tag{6}$$

Making use of eqns. (6) one can present the power of the plastic dissipation per unit area of the middle surface as

$$\dot{d}_i = M_1 \dot{\kappa}_1 + M_2 \dot{\kappa}_2 + M_{12} \dot{\kappa}_{12} \tag{7}$$

The total internal energy rate is

$$\dot{D}_i = \iint_S \dot{d}_i dS ,$$

where S stands for the area of the ellipse.

For determination of residual deflections an approximate method suggested by Jones [2, 3] will be used. According to this method the rate of the work done by inertial forces is equalized to the internal energy dissipated in continuous deformation fields and on discontinuous hinge lines. This method is combined with the concept of mode form motions. Thus the foregoing analysis is based on the equality (Jones [2])

$$\dot{D}_i = \dot{A}_e , \tag{8}$$

where $\dot{A}_e$ is the power of external forces. Note that the work done by inertial forces is included in $\dot{A}_e$.

3 Determination of residual deflections

Consider a stepped elliptical plate simply supported at the boundary. The boundary may be defined with the equation

$$r = r_*(\theta),$$

where $r_*(\theta)$ is specified with eqn. (3). According to the latter we can write

$$\dot{W}(r,\theta,t) = \dot{W}_0(t) f(r,\theta), \tag{9}$$

where W is the transverse deflection and $\dot{W}_0$ stands for the deflection rate at a specific point. Here dots denote time derivatives. As the particular case of eqn. (9) one can take

$$f(r,\theta) = 1 - \frac{r}{r_*(\theta)} \tag{10}$$

for plates simply supported at their boundaries at $r = r_*(\theta)$.

The internal energy dissipation corresponding to the velocity field (9), (10) can be calculated as

$$\dot{D}_i = \sum_{j=0}^{n} \int_{r_j}^{r_{j+1}} \sum_{i=0}^{m} M_{0j} \Delta\varphi_{ij} r dr , \tag{11}$$

where the simplifications suggested in [2] are taken into account. In eqn. (11) M_{0j} stands for the yield moment for the plate with thickness h_j and $\Delta\varphi_{ij}$ is the slope discontinuity at the hinge line located at $\theta = \theta_i$ in the region $r \in [\, r_j, r_{j+1} \,]$. Evidently

$$\Delta\varphi_i = \frac{\partial \dot{W}}{\partial \theta}(r, \theta_i +, t) - \frac{\partial \dot{W}}{\partial \theta}(r, \theta_i -, t).$$

However, in the case of a continuous field of straight yield lines, called yield fan, it is judicious to calculate the internal energy dissipation according to eqn. (7). Making use of eqns. (5)-(7) and eqn. (10) one obtains

$$\dot{d}_i = \frac{2M_{0j}}{\sqrt{3}} \frac{\dot{W}_0}{r r_*} \left[\, 1 + 2\left(\frac{r_*'}{r_*}\right)^2 - \frac{r_*''}{r_*} \,\right] \tag{12}$$

for $r \in (r_j, r_{j+1})$; $j = 0,...,n$.

In eqn. (12) and henceforth primes denote the differation with respect to the polar angle θ. Note that that the relation (12) can be reached by different ways (Skrszypek and Hetnarski [16]; Sawczuk and Sokol-Supel [14]; Rżanitsyn [13]). Since the middle plane of the elliptical plate covers the area $0 \le r \le r_*$, $0 \le \theta \le 2\pi$ and $dS = r dr d\theta$ it follows from eqn. (12) that

$$\dot{D}_i = \frac{2}{\sqrt{3}} \dot{W}_0 \int_0^{2\pi} \sum_{j=0}^{n} \int_{r_j}^{r_{j+1}} \frac{M_{0j}}{r_*} \left(1 + 2\left(\frac{r_*'}{r_*}\right)^2 - \frac{r_*''}{r_*} \right) dr d\theta .$$

Thus the internal energy dissipation can be presented as

$$\dot{D}_i = \frac{2}{\sqrt{3}} \dot{W}_0 \sum_{j=0}^{n} \int_0^{2\pi} M_{0j} \frac{r_{j+1} - r_j}{r_*} \left(1 + 2\frac{r_*'^2}{r_*^2} - \frac{r_*''}{r_*} \right) d\theta . \tag{13}$$

The external work rate done by inertial forces can be calculated as

$$\dot{A}_e = \iint_S -\mu \ddot{W} \dot{W} dS , \tag{14}$$

where μ stands for the density of the material. Calculating accelerations from eqns. (9)-(10) and substituting in eqn. (14) yields

$$\dot{A}_e = -\mu \ddot{W}_0 \dot{W}_0 \int_0^{2\pi} \sum_{j=0}^{n} \int_{r_j}^{r_{j+1}} h_j \left(1 - \frac{r}{r_*}\right)^2 r dr d\theta \,, \tag{15}$$

where $\dot{W}_0$ and $\ddot{W}_0$ are functions of t only. Implementing integration in eqn. (15) gives

$$\dot{A}_e = -\mu \ddot{W}_0 \dot{W}_0 \int_0^{2\pi} \left[\sum_{j=1}^{n} B_j \left(h_{j-1} - h_j\right) + \frac{h_n r_*^2}{12} d\theta \right] \tag{16}$$

where

$$B_j = \frac{1}{r_*^2} \left(\frac{r_*^2}{2} r_j^2 - \frac{2}{3} r_* r_j^3 + \frac{1}{4} r_j^4 \right) \tag{17}$$

for $j = 0,..,n$.

Making use of eqns. (13)-(17) one can determine

$$\ddot{W}_0 = -\frac{\frac{2}{\sqrt{3}} \sum_{j=0}^{n} M_{0j} \int_0^{2\pi} \left(r_{j+1} - r_j\right) \frac{1}{r_*^3} \left(r_*^2 + 2r_*'^2 - r_* r_*''\right) d\theta}{\mu \int_0^{2\pi} \left[\sum_{j=1}^{n} B_j \left(h_{j-1} - h_j\right) + \frac{h_n}{12} r_*^2 \right] d\theta} \tag{18}$$

In the particular case of the plate of constant thickness $h = h_0$ one has

$$\ddot{W}_0 = -\frac{\frac{24}{\sqrt{3}} M_0 \int_0^{2\pi} \left(1 + \frac{2r_*'^2}{r_*^2} - \frac{r_*''}{r_*}\right) d\theta}{\mu h_0 \int_0^{2\pi} r_*^2 d\theta} \tag{19}$$

where M_0 is the yield moment for the plate with thickness h_0.

Making use of eqn. (3) one can easily calculate derivatives

$$r_*' = \frac{ab(a^2-b^2)\cos\theta\sin\theta}{\left[\ a^2+(b^2-a^2)\cos^2\theta\ \right]^{3/2}}$$

$$r_*' = \frac{ab(b^2-a^2)\cos\theta\sin\theta}{\left[\ a^2+(b^2-a^2)\cos^2\theta\ \right]^{5/2}}\Big\{(\cos^2\theta - \\ -\sin^2\theta)\left[\ a^2+(b^2-a^2)\cos^2\theta\ \right]+3\cos^2\theta c\sin^2\theta\Big\} \tag{20}$$

and integrals

$$\int_0^{2\pi}\left(1+\frac{2r_*'^2}{r_*^2}-\frac{r_*''}{r_*}\right)d\theta = 4\pi$$

$$\int_0^{2\pi} r_*^2 d\theta = 2\pi ab \tag{21}$$

Substituting eqn. (21) in eqn. (19) leads to the simple expression

$$\ddot{W}_0 = -\frac{24M_0}{\sqrt{3}ab\mu h_0} \tag{22}$$

Evidently, if $a=b=R$ then eqn. (22) presents the acceleration for a circular plate of radius R.

It can be easily seen from eqns. (18), (19) that $\ddot{W}_0 = const$. Integrating twice with respect to time under initial conditions

$$\dot{W}_0(0) = V_0,\ \dot{W}_0(0) = V_0 \tag{23}$$

one obtains

$$\dot{W}_0 = \ddot{W}_0 t + V_0, \tag{24}$$

and

$$W_0 = \frac{1}{2}\ddot{W}_0 t^2 + V_0 t\ . \tag{25}$$

The motion of the plate stops at $t = t_1$ when $\dot{W}_0(t_1) = 0$. From eqns. (24) and (25) one easily obtains that

$$t_1 = -\frac{V_0}{\ddot{W}_0},\ W_f = -\frac{V_0^2}{2\ddot{W}_0}\ . \tag{26}$$

The initial kinetic energy corresponding to the velocity field (9), (10) can be evaluated as

$$K_0 = \frac{\mu}{2}\int_0^{2\pi}\int_0^{r_*} h(r,\theta)V_0^2\left(1-\frac{r}{r_*}\right)^2 r\,dr\,d\theta\,. \tag{27}$$

where $h = h_j$ for $r \in [\, r_j, r_{j+1} \,]$; $j = 0,..,n$ and $\dot{W}_0(0) = V_0$. From eqn. (27) one can define

$$V_0^2 = \frac{2K_0}{\mu\int_0^{2\pi}\left[\sum_{j=1}^{n} B_j\left(h_{j-1}-h_j\right)+\frac{h_n}{12}r_*^2\right]d\theta}\,. \tag{28}$$

Making use of eqns. (26)-(28) one obtains the maximal residual deflection

$$W_1 = \frac{\sqrt{3}K_0}{2\sum_{j=0}^{n} M_{0j}\int_0^{2\pi}\frac{1}{r_*^3}\left(r_{j+1}-r_j\right)\left(r_*^2+2r_*'^2-r_*r_*''\right)d\theta}\,. \tag{29}$$

4 Discussion

The accuracy of the approximate approach suggested in the present paper is evaluated in the particular case of a circular plate in fig. 2. Note that the present method is accommodated for the case of a Tresca material. In fig. 2 maximal permanent deflections of a circular plate simply supported at the edge versus the impulse are presented. The highest curve 1 in fig. 2 corresponds to the exact solution for a circular plate made of a Tresca material and subjected to impulsive loading [2], whereas the curve labeled with triangles is obtained in the present study. Intermediate curves in fig. 2 are obtained for different values of the parameter η in the case of a rectangular pressure pulse. In this case following notation is used

$$I = \frac{p_0\tau}{\mu V_0},\ \eta = \frac{p_0}{p_c},\ \lambda = \frac{\eta V_0 R^2}{M_0 h}\,.$$

Here p_0 is the load intensity, p_c stands for the static load carrying capacity and τ is the time instant when the loading is removed.

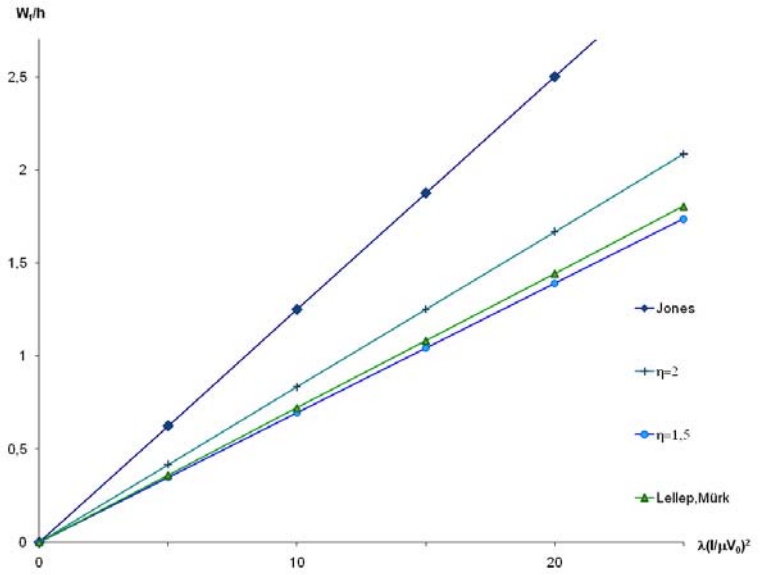


Figure 2: Maximal permanent deflections of a circular plate.

It can be seen from fig. 2 that in the case of a circular plate the method used in the present study in the case of impulsive loading leads to the results which are comparable to those corresponding to a rectangular impulse of medium value.

Maximal residual deflections versus α_1 are presented in fig. 3. It is assumed herein that the step is located at the ellipse with the semiaxes $a_1 = \alpha_1 a$ and $b_1 = \alpha_1 b$, e.g. $r_1 = \alpha_1 r_*$. Here and henceforth

$$W_1 = \frac{\sqrt{3}}{2} W_0(t_1) \frac{M_0}{K_0}.$$

Different curves in figs. 3, 4 correspond to different values of $\gamma = h_1 / h_0$. To lowest curve (straight line) in fig. 3 corresponds to elliptical plates of constant thickness. It can be seen from fig. 3 that the maximal permanent deflections monotonically decrease when the step radius r_1 (or the semiaxis a_1) increases.

In fig. 4 similar results are presented for elliptical plates with cutouts. The lines of cutouts are assumed to be ellipses with semiaxis $a_0 = \alpha_0 a$ and $b_0 = \alpha_0 b$, provided *a*, *b* are the parameters of external boundary of the plate and $\alpha_0 = 0.2$. In the case of plates with cutouts W_1 presents the residual deflection at the internal boundary of the plate.

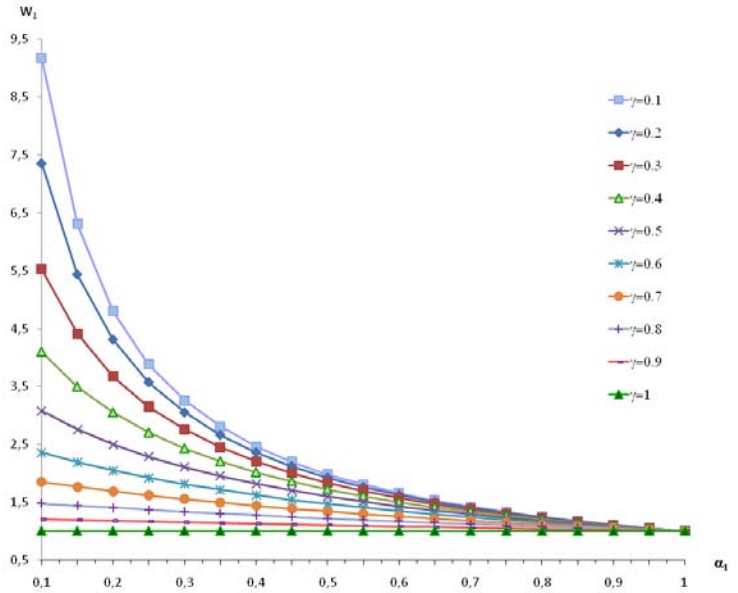


Figure 3: Maximal permanent deflections of an elliptical plate.

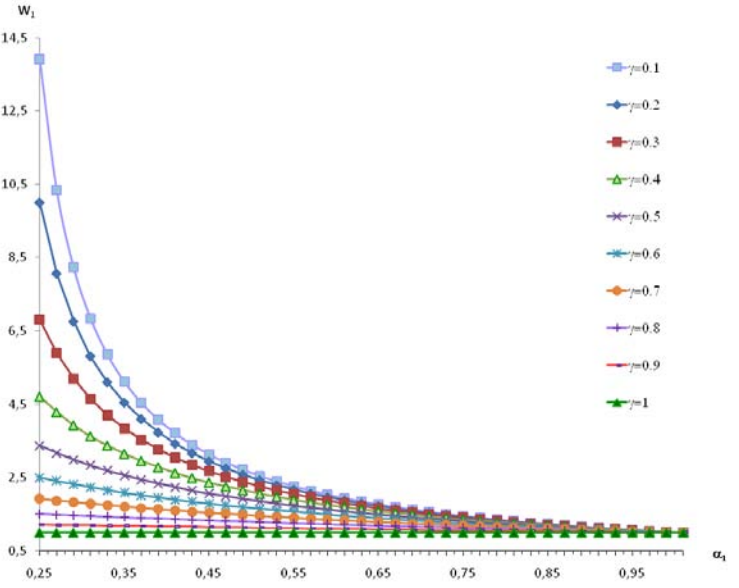


Figure 4: Maximal deflections of a plate with cutout.

5 Concluding remarks

Dynamic plastic response of elliptical plates to impulsive and impact loading was considered. An approximate method for evaluation of residual deflections of plates with elliptical boundaries has been developed. The method can be easily extended for stepped plates of arbitrary shape with arbitrary number of steps. Calculations carried out in the case of elliptical plates showed that maximal residual deflections can be remarkably shortened under given weight when re-distributing the material in the plate.

Acknowledgement

The support of the Estonian Science Foundation (grant ETF7461) is acknowledged.

References

[1] Chakrabarty, J., *Applied Plasticity*, Springer: Berlin, 1989.
[2] Jones, N., *Structural Impact*, CUP: Cambridge, 1989.
[3] Jones, N., Recent progress in dynamic plastic behaviour of structures. *The Shock and Vibration Digest*, **17**, pp. 33-47, 1985.
[4] Kaliszky, S., Dynamic plastic response of structures. Plasticity Today: *Modelling, Methods and Applications*, eds. A. Sawczuk, G. Bianchi, Elsevier: Science Publishers, London, 1984.
[5] Kaliszky, S., Plasticity. *Theory and Engineering Applications*, Elsevier: Amsterdam, 1989.
[6] Kaliszky, S. and Logo, J., Layout optimization of rigid-plastic structures under high intensity, short-time dynamic pressure. *Mechanics Based Design of Structures and Machines*, **31**, pp. 131-150, 2003.
[7] Lellep, J. & Hein, H., Optimization of clamped plastic shallow shells subjected to initial impulsive loading. *Engineering Optimization*, **34**, pp. 545-556, 2002.
[8] Lellep, J. & Mürk, A. Inelastic behaviour of stepped square plates. *Theories of Plates and Shells. Critical Review and New Applications*, eds. R. Kienzler, H. Altenbach & I. Ott, Euromech Colloquium 444. Springer: Berlin, pp. 133-140, 2004.
[9] Lellep, J. & Mürk, A., Inelastic response of axisymmetric plates with cracks. *International Journal of Mechanics and Material Design*, **3(3)**, pp. 237-251, 2006.
[10] Lellep, J. & Mürk, A., Optimization of inelastic annular plates with cracks, *Structural and Multidisciplinary Optimization*, **35(1)**, pp. 1-10, 2008.
[11] Liu, D. & Stronge, W., Deformation of a simply supported plate by central pressure pulse. *International Journal of Solids and Structures*, **33(2)**, pp. 283-299, 1996.

[12] Nurick, G.N. & Martin, J.B., Deformation of thin plates subjected to impulsive loading- a review. *International Journal of Impact Engineering*, **8**:159-170, 1989.
[13] Ržanitsyn A.R., Limit Analysis of Plates and Shells, Nauka: Moscow, (in Russian) 1983.
[14] Sawczuk, A. & Sokol-Supel, J., *Limit Analysis of Plates*, PWN: Warszawa, 1993.
[15] Shen, W.Q. & Jones, N., Dynamic response and failure of fully clamped circular plates under impulsive loading. *International Journal of Impact Engineering*, **13**, pp. 259-278, 1993.
[16] Skrszypek, J.J. & Hetnarski, R.B., *Plasticity and Creep. Theory, Examples and Problems*, CRC Press: Boca Raton, 1993.
[17] Stronge, W. & Yu, T.X., *Dynamic Models for Structural Plasticity*, Springer: New York, 1993.
[18] Symonds, PS., Elastic, finite deflection and rate effects in a mode approximation technique for plastic deformation of pulse loaded structures. *Journal of Mechanics and Engineering Sciences*, **22**, pp. 189-197, 1980.
[19] Wang, Y., Yu, M., Xiao, Y & Li, L., Dynamic plastic response of a circular plate based on unified strength theory. *International Journal of Impact Engineering*, **31**, pp. 25-40, 2005.
[20] Wen, HM, Yu, TX and Reddy, TY. A note on circular clamped plates under impulsive loading. *Mechanics of Structures and Machines*, **23**, pp. 331-342, 1995.
[21] Yu, TX and Chen, FL. Failure modes and criteria of plastic structures under dynamic loading: a review. *Metals and Materials*, **4**, pp. 219-226, 1998.

A fast incremental-iterative procedure for ultimate strength analysis of composite cross-sections of arbitrary shape

C. G. Chiorean
Technical University of Cluj-Napoca, Faculty of Civil Engineering, Romania

Abstract

A new computer method for bi-axial ultimate strength analysis of composite steel-concrete cross-sections of arbitrary shape subjected to axial force and biaxial bending moments is developed. An incremental-iterative procedure based on arc-length approach is proposed in order to determine, in a unitary formulation, both interaction diagrams and moment capacity contours, overcoming the difficulties and inaccuracies of the previously published methods. This procedure adopts a tangent stiffness strategy for the solution of the non-linear equilibrium equations thus resulting in a high rate and unconditionally convergence. An object oriented computer program, to obtain the ultimate strength of composite cross-sections under combined biaxial bending and axial load was developed. Examples run and comparisons made have proved the effectiveness and time saving of the proposed method of analysis.
Keywords: composite cross-sections, ultimate strength, arc-length method, bi-axial bending.

1 Introduction

In recent years, some methods have been presented for the ultimate strength analysis of various concrete and composite steel-concrete sections such as rectangular, L and T -shape, polygonal and circular under biaxial moments and axial loads [1–5]. Among several existing techniques, two are the most common; the first consists of a direct generation of points of the failure surface by varying the position and inclination of the neutral axis and imposing a strain distribution corresponding to a failure condition. This technique generates the failure surface

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110321

through 3D curves (Fig. 1) that are difficult to plot, having limited applicability in the practical applications. The second approach is based upon the solution of the non-linear equilibrium equations according to the classical Newton's scheme to obtain either the strain equilibrium plane or the location and inclination of the neutral axis. In general, these methods generates plane interaction curves and gives fast solutions but are sensitive to the origin of the loading axes and some problems in convergence may arise, particularly when the initial or starting values of variables are not selected properly and under large axial forces [2]. These methods can be further distinguished by how they plot the interaction curves. These curves may be computed, indirectly, by explicit computation of the moment-curvature response and the failure of the cross-section corresponds to the top of the moment-curvature diagram when any of the materials reaches its predefined maximum allowable strain [3, 5]. This method can be used under any loading mode, but is rather time-consuming and the basic equations of equilibrium are not always satisfied [3]. On the other hand, to determine directly the interaction diagrams or load contours, the exact location of the neutral axis may be determined by solving three coupled nonlinear system of equations by an iterative approach [1, 2, 4].

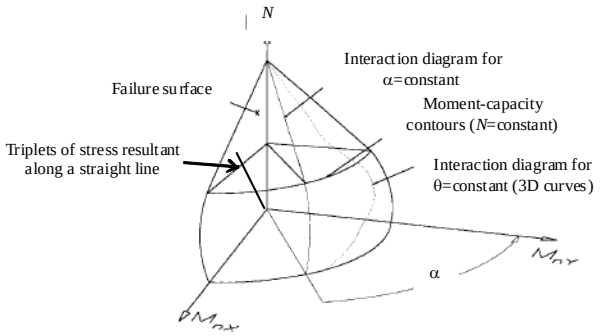


Figure 1: Failure surface and interaction diagrams.

These algorithms are not straightforward to implement, starting or initial values are not always simply to choose, and problems of convergence may arise especially when strain softening of the concrete in compression is taken into account, and they may become unstable near the state of the pure compression. There exist three different methods to generate plane interaction curves for cross-sections under biaxial bending: (1) interaction curves for a given bending moment ratio [1], (2) load contours for a given axial load [2] and (3) generate triplets of stress-resultants on the failure surface by extending an arbitrary oriented straight line [4]. The plane interaction curves generated by these methods are depicted in Fig. 1.

The main objective of the present paper is to present a new formulation by which the biaxial interaction diagrams and moment capacity contours of an arbitrary composite cross-section can be determined. Essentially, the inelastic

cross-section analysis method employed herein uses the accuracy of the fibber element analysis through the use of path integral approach for numerical integration of the cross-sectional nonlinear characteristics, and addresses its efficiency and modelling shortcomings both to failure surface generation procedure, overcoming the difficulties and inaccuracies of the previously proposed methods, and to postprocessing procedure of the axial force and bending moments obtained at a cross-section level, in order to check directly that they fulfil the ultimate limit state condition.

2 Mathematical formulation

2.1 Assumptions and problem definition

Consider the arbitrary cross-section shape subjected to the action of the external bending moments about each global axes and axial force as shown in Figure 2. It is assumed that plane section remains plane after deformation. This implies perfect bonding between the steel and concrete components of a composite cross section. Thus resultant strain distribution corresponding to the curvatures about global axes $\mathbf{\Phi}=\{\phi_x, \phi_y\}$ and the axial compressive strain ε_0 can be expressed at a generic point (x, y) in a linear form as:

$$\varepsilon = \varepsilon_0 + \phi_x y + \phi_y x \tag{1}$$

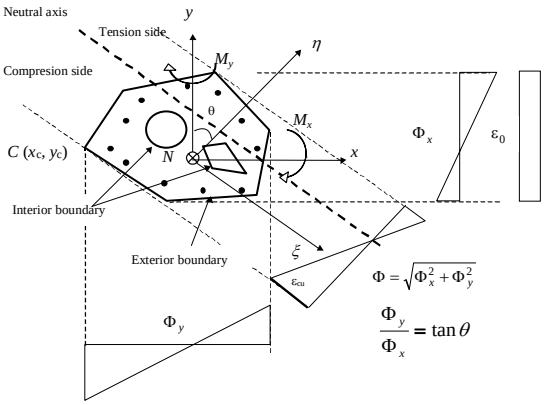


Figure 2: Model of arbitrary composite cross-section.

The constitutive relations for concrete under compression are represented by a combination of a second-degree parabola and a straight line as depicted in Fig. 3(*a*). The parameter γ represents the degree of confinement in the concrete, and allows for the modelling of creep and confinement in the concrete by simply varying the crushing strain ε_{c0} and γ respectively. The tensile strength of concrete is neglected. A multi-linear elasto-plastic stress-strain relationship, both in tension and in compression, is assumed for the structural steel and the steel

reinforcing bars (Fig. 3(*b*)). At ultimate strength capacity the equilibrium is satisfied when the external forces are equal to the internal ones and in the most compressed or tensioned point the ultimate strain is attained. These conditions can be represented mathematically in terms of the following nonlinear system of equations as:

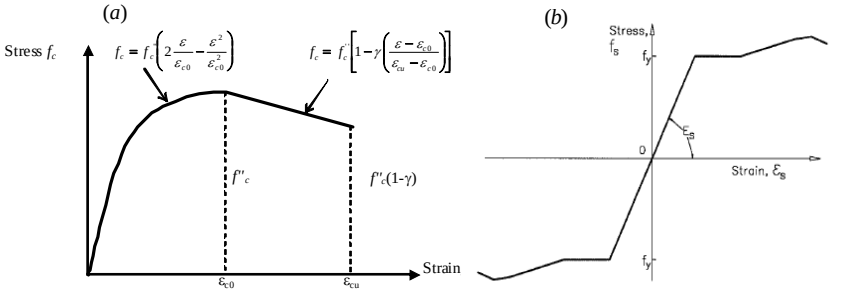


Figure 3: Stress-strain relationships: (*a*) for concrete in compression; (*b*) for steel.

$$\begin{cases} \int_A \sigma(\varepsilon_0,\phi_x,\phi_y)dA - N = 0; \int_A \sigma(\varepsilon_0,\phi_x,\phi_y)ydA - M_x = 0; \int_A \sigma(\varepsilon_0,\phi_x,\phi_y)xdA - M_y = 0 \\ \varepsilon_0 + \phi_x y_c(\phi_x,\phi_y) + \phi_y x_c(\phi_x,\phi_y) - \varepsilon_u = 0 \end{cases} \quad (2)$$

and in which N, M_x, M_y, $\varepsilon_0, \phi_x, \phi_y$ represent the unknown. In the Eqs. (2) the first three relations represent the basic equations of equilibrium for the axial load N and the biaxial bending moments M_x, M_y respectively, given in terms of the stress resultants. The last equation represents the ultimate strength capacity condition; that is, in the most compressed or most tensioned point the ultimate strain is attained, and in which $x_c(\phi_x,\phi_y)$ and $y_c(\phi_x,\phi_y)$ represent the coordinates of the point in which this condition is imposed. The coordinates of the "constrained" point can be always determined for each inclination of the neutral axis defined by the parameters ϕ_x and ϕ_y, and ε_u represents the ultimate strain either in most compressed concrete point or in most tensioned reinforcement steel fibre. Under the above assumptions, the problem of the ultimate strength analysis of composite cross-sections can be formulated as: *With strain distribution corresponding to a failure condition, find the ultimate resistances N, M_x, M_y so as to fulfil the basic equations of equilibrium and one of the following linear constraints:*

$$(a)\begin{cases} L_1(N,M_x,M_y) \equiv M_x - M_{x0} = 0 \\ L_2(N,M_x,M_y) \equiv M_y - M_{y0} = 0 \end{cases}; \quad (b)\begin{cases} L_1(N,M_x,M_y) \equiv N - N_0 = 0 \\ L_2(N,M_x,M_y) \equiv M_x - M_{x0} = 0 \end{cases} \quad (3)$$

where N_0,M_{x0},M_{y0} represents the given axial force and bending moments, respectively.

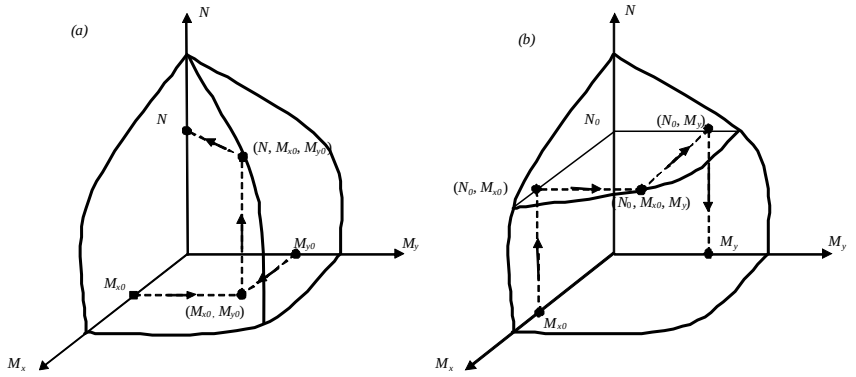


Figure 4: General solution procedures. (*a*) Interaction diagrams for given bending moments; (*b*) Moment-capacity contours for given axial force and bending moment M_x.

The general solution procedure is organized around solving the nonlinear system (2) for one of two linear constraints defined by the Eqs. (3*a*,*b*) Corresponding to each linear constraints we can define a point on the failure surface as: (I) when the constraints (*a*) are injected in the nonlinear system (2), a point on the failure surface is defined computing the axial resistance N associated to a failure criterion and for a fixed value of bending moments (M_x, M_y); (II) when constraints (*b*) are used, the point is associated to a fixed axial load (N) and a given bending moment M_x about x axis. All these situations are graphically illustrated in Figure 4.

2.2 Method of solution

An incremental-iterative procedure based on arc-length constraint equation is proposed in order to determine the biaxial strength of an arbitrary composite steel-concrete cross section accordingly to the already described situations. The failure diagrams corresponds either to maximum strains attained at the outer compressed point of the concrete section (i.e. ε_u equal to the compressive strain at failure) or to maximum strains attained in most tensioned reinforcement steel fibre (i.e. ε_u equal to the tensile steel strain at failure). Consider an irregular composite section as shown in Figure 2. The global x, y -axes of the cross section could have their origin either in elastic or plastic centroid of the cross-section. For each inclination of the neutral axis defined by the parameters ϕ_x and ϕ_y the farthest point on the compression side (or the most tensioned steel bar position) is determined (i.e. the point with co-ordinates x_c, y_c). We assume that at this point the failure condition is met, and consequently the axial compressive strain ε_0 can be expressed as:

$$\varepsilon_0 = \varepsilon_u - \left(\phi_x y_c + \phi_y x_c\right) \tag{4}$$

Thus, resultant strain distribution corresponding to the curvatures ϕ_x and ϕ_y can be expressed in linear form as:

$$\varepsilon(\phi_x, \phi_y) = \varepsilon_u + \phi_x(y - y_c) + \phi_y(x - x_c) \tag{5}$$

In this way, substituting the strain distribution given by the Eq. (5) in the basic equations of equilibrium, the unknown ε_0 together with the failure constraint equation can be eliminated from the nonlinear system (2). Thus the basic equations of equilibrium together with the linear constraints Eqs. (3(*a*))–(3(*b*)) forms a determined nonlinear system of equations (i.e. 5 equations and 5 unknowns):

$$\begin{cases} \int_A \sigma(\varepsilon(\phi_x, \phi_y))dA - N = 0 \\ \int_A \sigma(\varepsilon(\phi_x, \phi_y))ydA - M_x = 0 \\ \int_A \sigma(\varepsilon(\phi_x, \phi_y))xdA - M_y = 0 \end{cases} + \begin{cases} L_1(N, M_x, M_y) = 0 \\ L_2(N, M_x, M_y) = 0 \end{cases} \tag{6}$$

and the solutions can be obtained iteratively following an approach outlined in the next sections.

2.2.1 Interaction diagrams for given bending moments

In this case introducing the constraints (3*a*) in the system (6) the problem of the ultimate strength analysis of cross-section can be expressed mathematically as:

$$\int_A \sigma(\varepsilon(\phi_x, \phi_y))dA - N = 0;\ \int_A \sigma(\varepsilon(\phi_x, \phi_y))ydA - \lambda M_{x0} = 0;\ \int_A \sigma(\varepsilon(\phi_x, \phi_y))xdA - \lambda M_{y0} = 0 \tag{7}$$

in which axial load N and curvatures ϕ_x and ϕ_y represents the unknowns and λ represents the load parameter defining the intensity of the bending moments. If we regard the curvatures as independent variables in axial force equation, the curvatures and the load amplifier factor λ are given by solving the following nonlinear system of equations:

$$\begin{cases} \int_A \sigma(\varepsilon(\phi_x, \phi_y))ydA - \lambda M_{x0} = 0 \\ \int_A \sigma(\varepsilon(\phi_x, \phi_y))xdA - \lambda M_{y0} = 0 \end{cases} \tag{8}$$

This can be rewritten in terms of non-linear system of equations in the following general form:

$$\mathbf{F}(\lambda, \mathbf{\Phi}) = \mathbf{f}^{\text{int}} - \lambda \mathbf{f}^{ext} = \mathbf{0} \tag{9}$$

To traverse a solution path a proper parametrization is needed. A common setting of a continuation process is to augment the equilibrium equations (9) with a constraint [6]. In this case the curvature-moment constraint can be defined by equation g in the following form:

$$\mathbf{H}(\lambda, \mathbf{\Phi}) = \begin{cases} \mathbf{f}^{int} - \lambda \mathbf{f}^{ext} = \mathbf{0} \\ g(\lambda, \mathbf{\Phi}) = 0 \end{cases} \tag{10}$$

In this procedure, commonly called arc-length method, these equations are solved in a series of steps or increments, usually starting from the unloaded state $(\lambda = 0)$, and the solution to (10) is referred to as equilibrium path. Instead of solving Eqs. (10) directly, an indirect solution scheme for the constraint equation may be introduced. According to the indirect arc-length technique [6], the iterative changes of curvature vector $\delta\mathbf{\Phi}$ for the new unknown load level $\Delta\lambda_{k+1} = \Delta\lambda_k + \delta\lambda$, is written as:

$$\delta\mathbf{\Phi} = -\mathbf{K}_T{}^{-1}\mathbf{F} + \delta\lambda\mathbf{K}_T{}^{-1}\mathbf{f}^{ext} = \delta\mathbf{F} + \delta\lambda\delta\mathbf{\Phi}_T \tag{11}$$

where $\mathbf{F}$ represents the out-of-balance force vector (Eq. 9) and $\mathbf{K}_T$ represents the tangent stiffness matrix of the cross-section:

$$\mathbf{K}_T = \left(\frac{\partial\mathbf{F}}{\partial\mathbf{\Phi}}\right) = \begin{bmatrix} \dfrac{\partial M_x{}^{int}}{\partial\phi_x} & \dfrac{\partial M_x{}^{int}}{\partial\phi_y} \\ \dfrac{\partial M_y{}^{int}}{\partial\phi_x} & \dfrac{\partial M_y{}^{int}}{\partial\phi_x} \end{bmatrix} \tag{12}$$

in which the partial derivatives are with respect to the strains and stresses evaluated at current iteration *k*. Assuming the strain distribution given by the Eq.(5), the coefficients of the stiffness matrix can be symbolically evaluated as:

$$\begin{aligned}
k_{11} &= \frac{\partial M_x^{int}}{\partial\phi_x} = \frac{\partial}{\partial\phi_x}\int_A \sigma\left(\varepsilon\left(\phi_x,\phi_y\right)\right)y dA = \int_A \frac{\partial\sigma}{\partial\varepsilon}\frac{\partial\varepsilon}{\partial\phi_x}y dA = \int_A E_T y\left(y-y_c\right)dA \\
k_{12} &= \frac{\partial M_x^{int}}{\partial\phi_y} = \frac{\partial}{\partial\phi_y}\int_A \sigma\left(\varepsilon\left(\phi_x,\phi_y\right)\right)y dA = \int_A \frac{\partial\sigma}{\partial\varepsilon}\frac{\partial\varepsilon}{\partial\phi_y}y dA = \int_A E_T y\left(x-x_c\right)dA \\
k_{21} &= \frac{\partial M_y^{int}}{\partial\phi_x} = \frac{\partial}{\partial\phi_x}\int_A \sigma\left(\varepsilon\left(\phi_x,\phi_y\right)\right)x dA = \int_A \frac{\partial\sigma}{\partial\varepsilon}\frac{\partial\varepsilon}{\partial\phi_x}x dA = \int_A E_T x\left(y-y_c\right)dA \\
k_{22} &= \frac{\partial M_y^{int}}{\partial\phi_y} = \frac{\partial}{\partial\phi_y}\int_A \sigma\left(\varepsilon\left(\phi_x,\phi_y\right)\right)x dA = \int_A \frac{\partial\sigma}{\partial\varepsilon}\frac{\partial\varepsilon}{\partial\phi_y}x dA = \int_A E_T x\left(x-x_c\right)dA
\end{aligned} \tag{13}$$

where the coefficients k_{ij} are expressed in terms of the tangent modulus of elasticity E_t. Thus the incremental curvatures for the next iteration can be written as:

$$\Delta\mathbf{\Phi}_{k+1} = \Delta\mathbf{\Phi}_k + \delta\mathbf{\Phi} \tag{14}$$

This procedure is iterated until convergence upon a suitable norm is attained. Assuming that a point ($^{\tau}\mathbf{\Phi}$, $^{\tau}\lambda$) of the equilibrium path has been reached, the next point ($^{\tau+\Delta\tau}\mathbf{\Phi}$, $^{\tau+\Delta\tau}\lambda$) of the equilibrium path is then computed updating the loading factor and curvatures as:

$$\begin{aligned}
{}^{\tau+\Delta\tau}\lambda &= {}^{\tau}\lambda + \Delta\lambda_{k+1} \\
{}^{\tau+\Delta\tau}\mathbf{\Phi} &= {}^{\tau}\mathbf{\Phi} + \Delta\mathbf{\Phi}_{k+1}
\end{aligned} \tag{15}$$

In this way with curvatures and loading factor known, the axial force resistance N is computed based on the resultant strain distribution corresponding to the curvatures ϕ_x and ϕ_y through Equation (7), and the ultimate bending moments, M_x and M_y, are obtained scaling the reference external moments M_{x0} and M_{y0} through current loading factor λ given by Equation (15). Graphical

representation of the present method according to Equation (10) is depicted in Figure 5. It is important to note that, the stiffness matrix of cross-section K_T given by the Equations (13), could become negative definite or singular during the iterative process therefore the above procedure based on arc-length constraint equation is essential to overcome these difficulties.

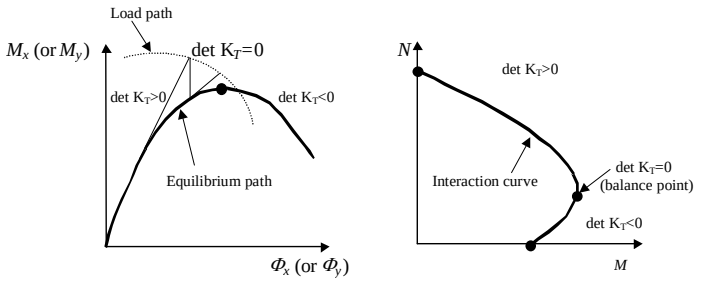


Figure 5: Geometrical representation of the present method.

2.2.2 Moment capacity contour for given axial force *N* and bending moment M_x

In this case, injecting the linear constraints (3*b*) in the nonlinear system (6), and arranging the system accordingly with the decoupled unknowns, we obtain:

$$\begin{cases} \int_A \sigma\left(\varepsilon\left(\phi_x,\phi_y\right)\right)dA - N_0 = 0 \\ \int_A \sigma\left(\varepsilon\left(\phi_x,\phi_y\right)\right)ydA - \lambda M_{x0} = 0 \end{cases} + \int_A \sigma\left(\varepsilon\left(\phi_x,\phi_y\right)\right)xdA - M_y = 0 \qquad (16)$$

in which bending moment M_y and curvatures ϕ_x and ϕ_y represents the unknowns. Following a similar approach as presented above, the curvatures are obtained solving the first two equations and then with this strain distribution the bending moment resistance about *y* axis is computed with the last equation of the system.

2.3 Evaluation of tangent stiffness and stress resultant

Based on Green's theorem, the integration of the stress resultant and stiffness coefficients over the cross-section will be transformed into line integrals along the perimeter of the cross-section. For this purpose, is necessary to transform the variables first, so that the stress field is uniform in a particular direction, given by the current position of the neutral axis [7]. This is achieved by rotating the reference axes *x*, *y* to ξ, η oriented parallel to and perpendicular to the neutral axis, respectively as shown in Fig. 2. Thus, the stress field is uniform in direction parallel with the neutral axis and strains and the internal forces carried on by the compressive concrete and structural steel can be obtained by the following expressions:

$$N_{\text{int}} = \iint \sigma(x,y)dxdy = \iint \sigma(\eta)d\xi d\eta$$
$$M_{x,\text{int}} = \iint \sigma(x,y)ydxdy = \iint \sigma(\eta)(-\xi\sin\theta + \eta\cos\theta)d\xi d\eta = M_{\xi,\text{int}}\cos\theta - M_{\eta,\text{int}}\sin\theta$$
$$M_{y,\text{int}} = \iint \sigma(x,y)xdxdy = \iint \sigma(\eta)(\xi\cos + \eta\sin\theta)d\xi d\eta = M_{\xi,\text{int}}\sin\theta - M_{\eta,\text{int}}\cos\theta \qquad (17)$$

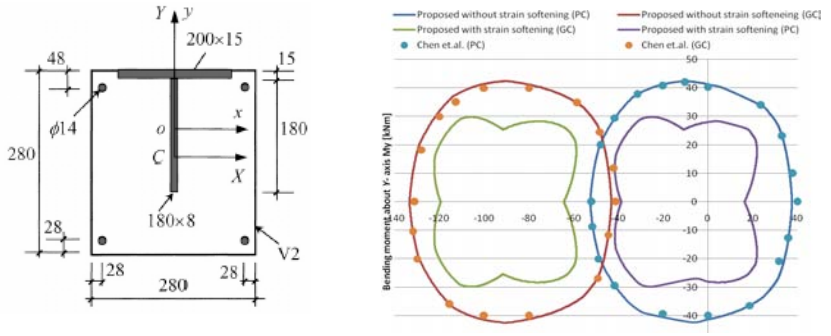
where N_{int}, $M_{\xi,int}$ and $M_{\eta,int}$ are the internal axial force and bending moments about the ξ and η axis respectively and can be computed based on the Green's path integral approach. The tangent stiffness matrix coefficients are computed in the same way. In order to perform the integral of a determined side of the contour, polygonal or circular, of the integration area, the interpolatory Gauss-Lobatto method is used. Though this rule has lower order of accuracy than customary Gauss-Legendre rule, it has integration points at each ends of interval, and hence performs better in detecting yielding. However, because the stress filed is defined by a step function and there is no continuity in the derivative, the polynomial interpolation can produce important integration errors. In this case, an adaptive quadrature strategy can be applied. In this context of the adaptivity quadratures, the Lobatto integration scheme has another advantage over the Legendre integration scheme, observing that the point corresponding to the left end in one interval is the same as the point corresponding to right end in the next. Consequently, the cost of evaluating a Lobatto rule is reduced by about one integrand evaluation comparing with Legendre rule. The steel bars are assumed discrete points with area A_{sj}, co-ordinates x_{sj}, y_{sj} and stress f_{sj}.

3 Computational example

Based on the analysis algorithm just described, a computer program **ASEP** has been developed to study the biaxial strength behaviour of arbitrary concrete-steel cross sections. In order to demonstrate the validity, accuracy, unconditionally convergence and time saving of the analytic procedure developed here, the interaction diagrams and moment capacity contours of a rectangular cross-section with asymmetrically placed structural steel (Fig. 6(*a*)) are determined and compared with the numerical procedure developed in [2].

Characteristic strength for concrete in compression is f_c=31.79 Mpa and the stress-strain curve which consists of a parabolic and linear- part was used in the calculation, with the crushing strain ε_0=0.002 and ultimate strain ε_{cu}=0.0035. The Young modulus for all steel sections was 200GPa while the maximum strain was ε_{su}=±1%. The yield strength of steel reinforcing bars is f_y=420 MPa, whereas for the structural steel the following values has been considered $f_{y,flange}$ =255 MPa $f_{y,web}$ =239 MPa. In order to demonstrate the unconditionally convergence of the algorithms developed in the current paper the cross-section has been analysed, drawing the interaction diagrams and moment capacity contours for axial loads near the axial load capacity, considering both geometric and plastic centroid of the cross-section. Convergence problems have been experienced by the Chen *et al.* [2] in this portion of the moment capacity contour when the geometrical centroid of the cross-section has been taken as the reference loading axes.

Figure 6(*b*) presents the moment capacity contours, obtained by the present algorithm and those obtained by the Chen *et al*. [2] for axial load *N*=3000kN, considering as reference loading axes, geometric (GC) and plastic centroid (PC) respectively.



(*a*) (*b*)

Figure 6: (*a*) Rectangular section with asymmetrically placed structural steel. (*b*) Moment capacity contour with axial load *N*=3000kN.

No convergence problems have been experienced by the proposed approach, even when the geometric centroid has been chosen as reference axes or the strain-softening of the concrete in compression has been taken into account (γ=0.15), and a maximum of just two iterations has been required during the incremental-iterative process. As it can be seen the results obtained in the current paper and those reported in [2] agree closely in both cases. The effect of the strain-softening is not taken into account in the method proposed in [2]. However it is important to note that, although claimed, the method proposed by Chen does not generate genuinely plane moment-capacity curves. The method proposed in [2] fails in some circumstances to draw the moment capacity contour, under a fixed axial load, and in order to overcome some divergences, axial load value is slightly adjusted. Figure 7(*a)* shows the complete interaction diagrams under uniaxial bending moment about *x* axis with and without the effect of the strain softening of the concrete. As it can be seen near the axial load capacity multiple solutions exists in the *N-M* space when the strain softening is modelled. This can be clearly observed on the moment-ultimate curvature diagrams (Fig. 7(*b*)), when multiple "snap-through" phenomenons occur in these situations. When the strain-softening is ignored (γ=0), the moment-ultimate curvature curve is flat in that region (Fig.7.*b*) that explains the uniqueness of the solution in the *N-M* space in this case (Fig.7(*a*)). Figure 8 shows the comparative interaction diagrams for α=15° and 30° respectively (M_y=tan(α)M_x). The bending moments are computed about axes *x-y*, which pass through the geometric centroid of the cross-section.

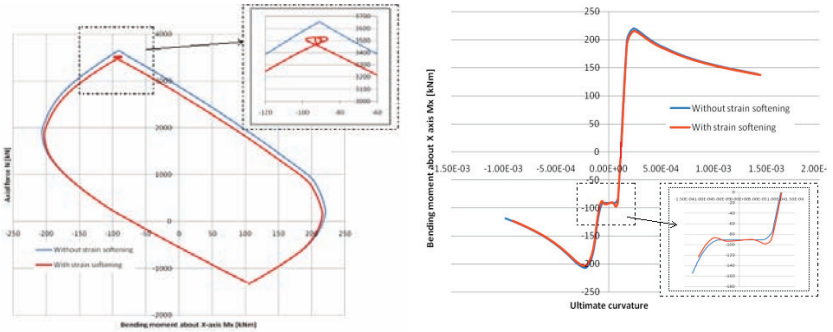


Figure 7: (*a*) Interaction diagrams under uniaxial bending moment about *x* axis; (*b*) Bending moment-ultimate curvature variation.

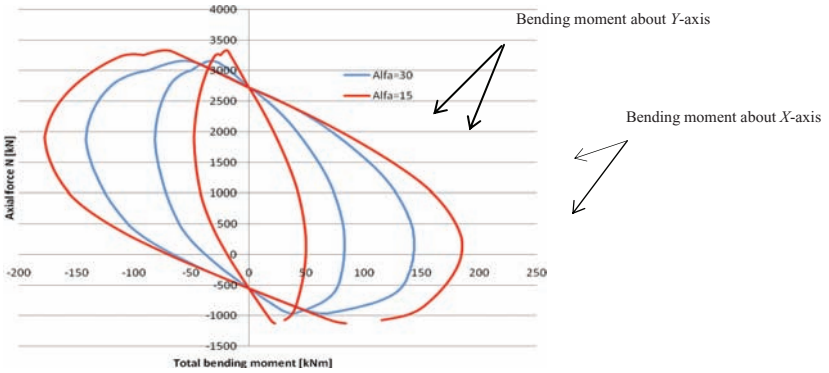


Figure 8: Biaxial interaction diagrams.

4 Conclusions

A new computer method based on incremental-iterative arc-length technique has been presented for the ultimate strength analysis of composite steel-concrete cross-sections subjected to axial force and biaxial bending. Comparing the algorithm presented in the current paper with the existing methods it can be concluded that the proposed approach is general and complete, can determine both interaction diagrams and moment capacity contours, and, of great importance, it is fast, the diagrams are directly calculated by solving, at a step, just two coupled nonlinear equations, and assures convergence for any load case, even near the state of pure compression or tension and is not sensitive to the initial/starting values, how the origin of the reference loading axes is chosen or to the strain softening effect for the concrete in compression. Furthermore, the proposed method as is formulated can be applied to provide directly the ultimate resistances of the cross-section, supposing that one or two components of the section forces are known, without the need of knowing in advance the whole

interaction diagram or moment capacity contour. The method has been verified by comparing the predicted results with the established results available from the literature. It can be concluded that the proposed numerical method proves to be reliable and accurate for practical applications in the design of composite steel-concrete beam-columns and can be implemented in the advanced analysis techniques of 3D composite frame structures.

Acknowledgement

The writer gratefully acknowledges the support from Romanian Research Foundation (CNCSIS- Grant PNII-IDEI No. 193/2008) for this study.

References

[1] Rodrigues, J.A. & Aristizabal-Ochoa, J.D. Biaxial interaction diagrams for short RC columns of any cross section, Journal of Structural Engineering, ASCE, 125(6): 672-683, 1999.
[2] Chen S.F., Teng J.G., Chan S.L. Design of biaxially loaded short composite columns of arbitrary section, Journal of Structural Engineering, ASCE, 127(6), 678-685, 2001.
[3] Sfakianakis, M.F., Biaxial bending with axial force of reinforced, composite and repaired concrete cross sections of arbitrary shape by fibber model and computer graphics, Advances in Engineering Software, 33, 227-242, 2002.
[4] Rosati, L., Marmo, F., Serpieri, R., Enhanced solution strategies for the ultimate strength analysis of composite steel-concrete sections subject to axial force and biaxial bending, Computer methods in applied mechanics and engineering, 197(9-12), 1033-1055, 2008.
[5] Charalampakis, A.E., Koumousis, V.K., Ultimate strength analysis of composite sections under biaxial bending and axial load, Advances in Engineering Structures, 39(11), 923-936, 2008.
[6] Crisfield, M.A. Non linear finite element analysis of solids and structures, Wiley, Chichester, 1991.
[7] Chiorean, C.G., A fast incremental-iterative procedure for inelastic analysis of RC cross-sections of arbitrary shape, Acta Technica Napocensis, 47, 85-98, 2004.

Section 5
Composite materials

This page intentionally left blank

Compression property of waste polyurethane rubber/unsaturated polyester composite cubes

S. Enoki[1], T. Tsujitan[2] & J. Yamashita[2]
[1]*Department of Mechanical Engineering, Nara National College of Technology, Japan*
[2]*Nara National College of Technology, Japan*

Abstract

This paper describes a method to recycle waste polyurethane rubbers. Waste rubbers after manufacturing rubber products and most of the used rubber products have affected bad effects on the environment. In the case of larger rubber products, the waste rubbers and the used products have been recycled. However the waste rubbers of smaller rubber products and the used products are not recycled and are incinerated. An aseismatic mat is one of the smaller products and it is made of polyurethane rubber. The mats are manufactured a lot in Japan, so a lot of waste rubbers occur and are incinerated. In order to recycle the waste polyurethane rubbers, we created a composite cube as a structural material by using the waste rubbers. The composite cube had 15mm on a side. Unsaturated polyester was used as a matrix, and then the waste polyurethane rubbers were put in the unsaturated polyester. The waste polyurethane rubbers were minced and the minced rubbers were made into a ball. Some composite cubes on different quantities of the rubbers were prepared. Static compression tests were carried out with the composite cubes and with a cube made of only unsaturated polyester. Comparing with the load-displacement curves, it has been shown that a maximum load of the 0.1g-rubber-containing composite cube is larger than that of the rubber-free cube. However it has been shown that a maximum load of the 0.5g-rubber-containing composite cube is not larger than that of the rubber-free cube. So it is possible that polyurethane rubber reinforce unsaturated polyester, if the rubber is minced and is made into a ball and is used in the appropriate contained amount.
Keywords: waste rubber, recycle, composite, compression property.

WIT Transactions on Modelling and Simulation, Vol 51,
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110331

1 Introduction

Waste rubbers after manufacturing rubber products and most of the used rubber products have affected bad effects on the environment. In the case of larger rubber products which are usually manufactured in big companies, the waste rubbers and the used rubber products have been recycled [1, 2]. For example, waste tire rubbers are used for a pavement material. In the case of smaller rubber products which are usually manufactured in small companies, the wastes and the used smaller products are not recycled and are incinerated.

An aseismatic mat is one of the smaller products and it is made of polyurethane rubber. The mat was developed for keeping up home electric appliances and furniture in quake-prone Japan. A lot of aseismatic mats are manufactured in Japan, so a lot of waste polyurethane rubbers occur and are incinerated. It is difficult to recycle them, because the polyurethane rubber is thermosetting and is not able to re-thermoform. Additionally, all of the waste rubbers are too small and different in shape and in size to be recycled.

In order to recycle the waste polyurethane rubbers, we created a composite cube as a structural material by using the waste rubbers. Unsaturated polyester was used as a matrix, and then the waste polyurethane rubbers were put in the unsaturated polyester. If compression properties of the polyurethane rubber/unsaturated polyester composite cubes were better than that of a rubber-free cube, the composite cubes will be able to recycle the waste polyurethane rubbers.

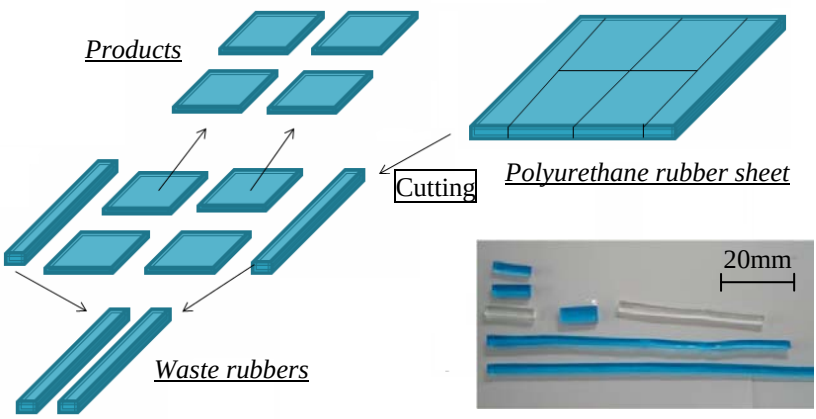
In this study, therefore, compression tests were carried out to examine the compression properties of the composite cubes and the rubber-free cube.

2 Composite cubes

The composite cube had 15mm on a side and unsaturated polyester was used as a matrix, and then the waste polyurethane rubbers were put in the unsaturated polyester. In this section, characteristics of the polyurethane rubber and how to put the waste rubbers in the matrix were described. Then a procedure for preparing the composite cubes was proposed.

2.1 Waste polyurethane rubbers

Aseismatic mats are manufactured in Japan and a lot of waste polyurethane rubbers occur as shown in Figure 1(a). The polyurethane rubber is thermosetting and is not able to re-thermoform. And all of the waste rubbers are different in shape and in size as shown in Figure 1(b). To solve the problems, we focused on characteristics of the polyurethane rubber. The characteristics are rubber elasticity and high-adhesive property. After the waste polyurethane rubbers were minced, the minced rubber can be made into a ball because of the high-adhesive property. It was thought that behaviour of the balled-up rubber was as same as the rubber elasticity of a solid polyurethane rubber when they were put in the unsaturated polyester.



(a) Production flow of aseismatic mats.

(b) Waste polyurethane rubbers. (different in size and in shape)

Figure 1: Generating waste polyurethane rubbers and photo of the wastes.

A 0.1g balled-up rubber, a 0.3g balled-up rubber and a 0.5g balled-up rubber were prepared as shown in Figure 2 because of the minced rubbers was able to be adjusted. The different quantities of balled-up rubbers were put in unsaturated polyester and composite cubes contained different quantities of rubber were prepared.

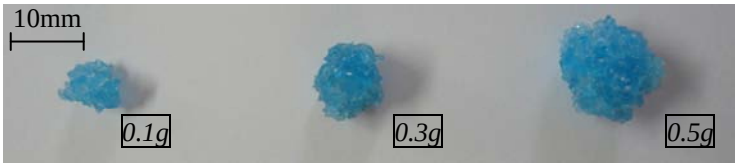


Figure 2: Photo of balled-up minced waste polyurethane rubbers.

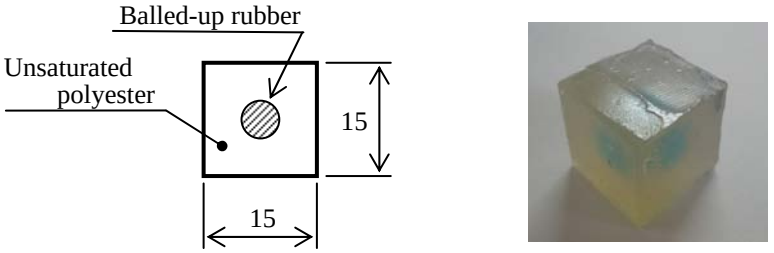
2.2 Preparing composite cubes

Prepared cubes are shown in Table 1. A mold was used for preparing the cubes which size was 15mm on a side. Unsaturated polyester mixed with a hardener was casted into the mold. When rubber-containing composite cubes were prepared, the balled-up minced waste polyurethane rubbers shown in Section 2.1 were put in the unsaturated polyester in the middle of the casting. All of the balled-up rubbers put in the centre of the each cube as shown in Figure 3.

For comparison, rubber-free cubes which were made of only unsaturated polyester were prepared. Moreover, the other rubber-containing composite cubes, using balled-up rubbers that debased adhesive, were prepared. The low-adhesive rubbers were made by leaving balled-up rubbers in the atmosphere for about 2 months. It is predicted that the high-adhesive rubber/unsaturated

Table 1: Prepared rubber-containing composite cubes and rubber-free cube.

	Waste Polyurethane Rubber		Matrix
	Quantities	Adhesive	
Rubber-free Cube	-	-	Unsaturated Polyester
Composite Cubes contained balled-up rubber	Minced 0.1g	High	
	Minced 0.3g		
	Minced 0.5g		
	Minced 0.1g	Low	
	Minced 0.3g		
	Minced 0.5g		



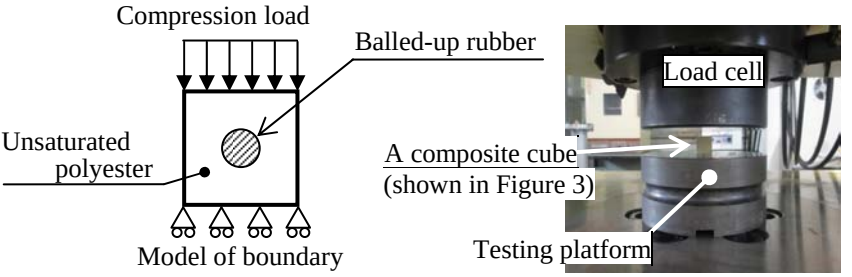
(a) A section of composite cube. (b) A composite cube.

Figure 3: Schematic view and photo of polyurethane rubber/unsaturated polyester composite cube.

polyester composite cubes have higher compression properties than the low-adhesive rubber/unsaturated polyester composite cubes.

3 Compression test

Static compression tests were carried out with the cubes shown in Section 2. Hydraulic type universal testing machine (SHIMADZU CORPORATION Model: UH-500kNI) was used for the static compression tests. The test speed is set on 2mm/min. and the stroke is set on 2mm. Mechanical model of the compression test and a photo of the testing situation are shown in Figure 4. Lubricant agent was applied on upper and lower planes of the cubes. It reduces friction on the planes faced on upper testing head and testing platform.



(a) Mechanical model (b) Testing situation.

Figure 4: Mechanical model of balled-up rubber/unsaturated polyester composite cube and a photo of compression test for the cube.

4 Results and discussion

Compression property was evaluated with load-displacement curves and maximum compression stress values.

4.1 Load-displacement curve

Comparing with load-displacement curves shown in Figure 5, a maximum load of the composite cube contained 0.1g balled-up rubber is larger than that of the rubber-free cube. However it is shown that maximum loads of the composite cubes contained 0.3g and 0.5g balled-up rubbers are not larger than that of the rubber-free cube. Moreover the maximum loads of the composite cubes decrease with increasing contained amount of the rubbers.

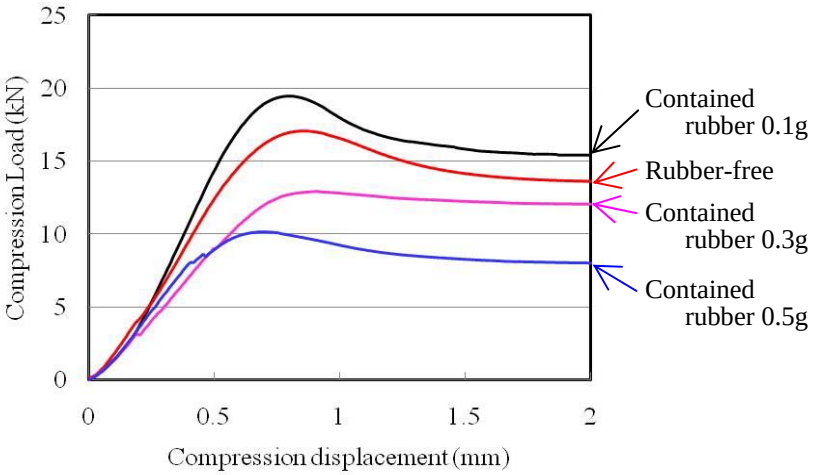
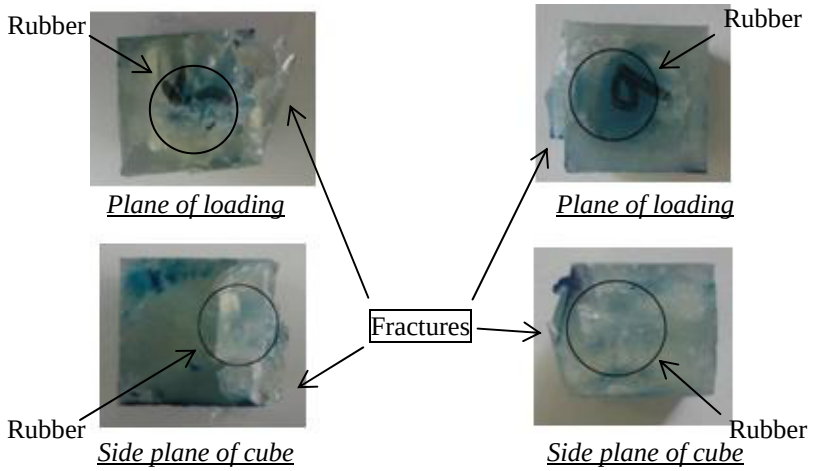


Figure 5: Comparing with load-displacement curves of cubes with different contained amount of rubbers.



(a) Contained 0.3g balled-up rubber. (b) Contained 0.5g balled-up rubbers.

Figure 6: Typical of fracture morphology after compression tests.

Composite cubes contained 0.3g and 0.5g balled-up rubbers after compression tests are shown in Figure 6. The side planes were broken in both composite cubes. This fracture morphology is thought to be typical.

An anticipate fracture mechanism of balled-up polyurethane rubber/unsaturated polyester composite cubes is thought as shown in Figure 7. The balled-up rubbers are thought to deform widely to side planes of the cubes and spread the unsaturated polyester from inside after compressive loading. It is because the rubber Poisson ratio 0.45 is larger than the unsaturated polyester Poisson ratio 0.34. The situation is similar to occurring hydrostatic pressure. Therefore it is thought that the composite cube can support higher loading, comparing with a rubber-free cube. This case agrees with the 0.1g-rubber-containing composite cubes. However, in the case that quantities of the contained rubber are a lot, thickness of a side plane of the cubes is small. Therefore it is easy to be occurred plastic deformation and to be broken at the side plane. This case agrees with the composite cubes contained 0.3g and 0.5g balled-up rubbers.

4.2 Maximum compression stress

The anticipate fracture mechanism, shown in Figure 7, is thought to be based on high-adhesive property of the polyurethane rubber. The minced rubbers can be made into a ball by using the high-adhesive property. To consider the influence of adhesive property on the compression property of the composite cubes, therefore, maximum compression stress values were calculated. Nominal stress was used in this study. Figure 8 shows the maximum compression stress values compared rubber-free cubes with composite cubes contained high-adhesive minced rubbers and low-adhesive ones.

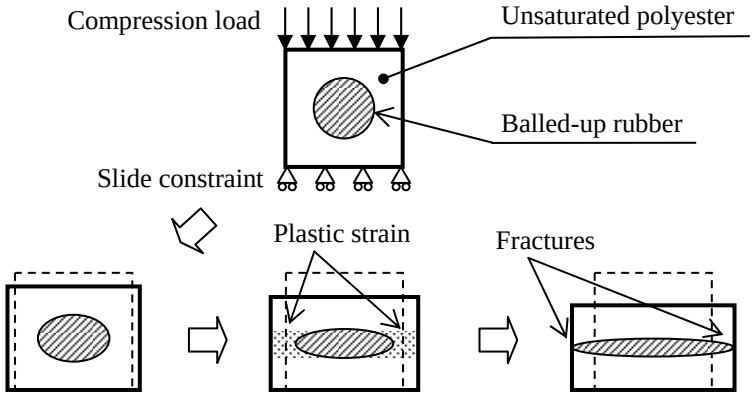


Figure 7: Anticipated fracture mechanism of balled-up rubber/unsaturated polyester composite cube.

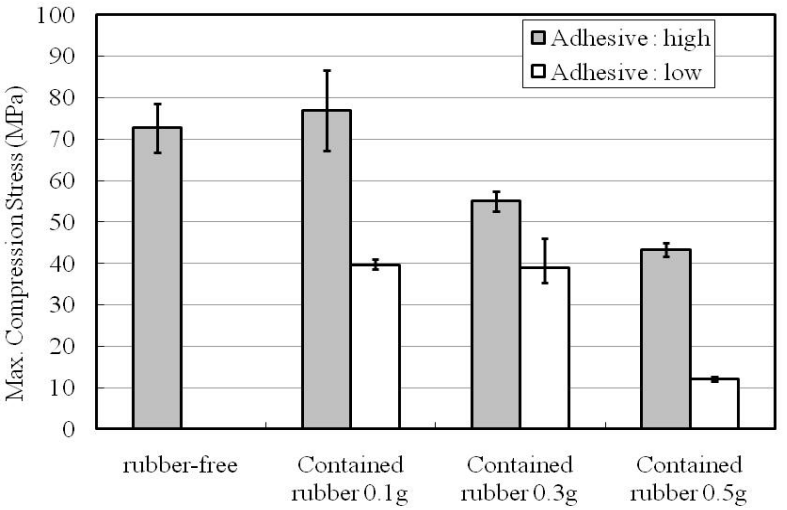


Figure 8: Comparing with maximum compression stresses of cubes having different quantities of contained rubbers and different adhesive properties of contained rubbers.

In the case of containing the high-adhesive minced rubbers, maximum compression stress values show a same trend of the maximum loads shown in Section 4.1. However, maximum compression stress values of the composite cubes contained low-adhesive minced rubbers are smaller than that of the composite cubes contained high-adhesive ones independently of contained quantities.

The balled-up minced polyurethane rubber is similar to an aggregate of fine polyurethane rubber fragments as shown in Figure 9. In the case of high-

adhesive rubber, there are few voids in the aggregate and the aggregate has rubber elasticity in the unsaturated polyester. In the case of low-adhesive rubber, there are a lot of voids in the aggregate and it is thought that the balled-up rubber dose not has rubber elasticity.

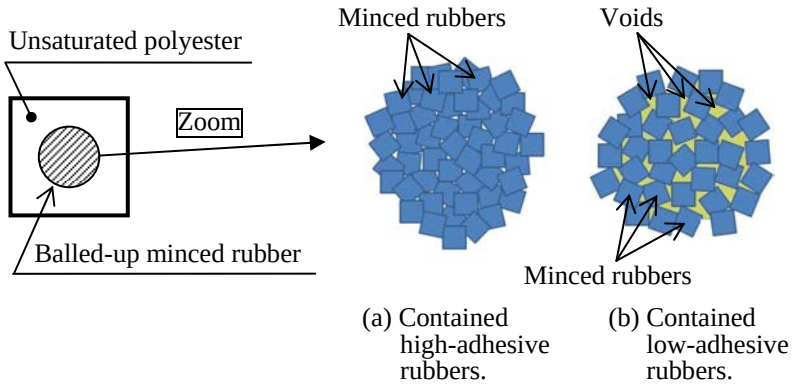


Figure 9: Model of minced rubber in unsaturated polyester.

5 Conclusion

It is possible that polyurethane rubber reinforce unsaturated polyester if the rubber is minced and is made into a ball and is used in the appropriate quantity. The quantity is 0.1g in the case of composite cubes which size is 15mm on a side. It is thought that waste polyurethane rubbers can recycle as a structural material.

Acknowledgements

We thank SHINAGAWA Co., Ltd. (Rubber Manufacturing) and RODAN21 Co., Ltd. (Manufacturing Coordination) for the provision of waste polyurethane rubbers.

References

[1] Masahito FUJII, Kenji HIMENO, Kenichi KOUGO & Masato MURAYAMA, Physical Properties of Asphalt Rubber Mixtures, 6th ICPT, Sapporo, Japan, July, 2008.
[2] Kenzo FUKUMORI & Mitsumasa MATSUSHITA, Material Recycling Technology of Crosslinked Rubber Waste, R&D Review of Toyota CRDL Vol. 38 No.1, p.39-47, 2003.

Development of interaction diagrams for RC sections confined with CFRP composites

P. Christou, A. Michael & Z. Neofytou
Department of Civil Engineering, Frederick University, Cyprus

Abstract

The interaction diagram is a surface which defines the maximum capacity of compression members that are subjected to axial force and bending moments. As a result, these diagrams provide the engineers with an additional tool for the design of such members. When the compression members are confined with FRP their capacity increases, however in many cases the increase in capacity is normally neglected which sometimes can lead to very conservative designs. This work includes the development of interaction diagrams for circular compression members confined with CFRP using the fiber model. The longitudinal reinforcement is considered to be symmetric whereas the confinement can vary. The method presented herein defines the location of the neutral axis and based on that calculates the axial force and bending moment. A comparison of the unconfined to the confined section shows a considerable difference in the interaction diagram plot in the compression controlled region.
Keywords: interaction diagrams, confinement, section equilibrium, RC section strength.

1 Introduction

The analysis of concrete columns using an analytical solution is not trivial. As a result the analysis of columns is basically reduced to the development of the interaction diagram and the plot of the load condition in order to define failure or not for the section. Normally the confinement for compression reinforced concrete sections is provided either by ties or spirals. However, other methods and materials are used in the later years which can provide increased confinement and thus satisfy the requirement for increased ductility. The column wrapping with CFRP composites is a popular alternative for improving the

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110341

ductility and thus the seismic resistance of columns. Fiber fabrics and prefabricated FRP composite jackets or tubes cover the entire area of the concrete element and therefore cannot be embedded in concrete. Another technique is the use of a CFRP composite grid (Michael *et al.* [1]). The carbon grid has approximately 69% open surface area allowing the grid to be embedded in the concrete. Light grids are easily formed into a round shape and can provide more effective confinement than wraps that are forced to follow the column cross-section, which might be square or rectangular.

The work presented herein refers to a numerical procedure for the development of interaction diagrams that are confined with CFRP and provide a comparison with similar sections without confinement.

1.1 Approximate analyses for columns

A search in the literature reveals a number of numerical approximations for the development of such diagrams. In the three dimensional case these methods rely on using single axis bending response in the two principal directions for the approximation of the biaxial bending. Some of these methods are:

- The Bressler load contour method [2]
- The Bressler reciprocal load method [2]
- The PCA load contour method [2]
- The Weber design charts [3]

1.2 Confined concrete

Confinement can improve both the compressive strength and ductility of concrete. Steel has typically been used to confine the concrete in reinforced concrete columns. Steel can be internal reinforcement, usually used as a spiral or ties, or it can be external such a steel jacket that is bonded to the outside face of the column. When fiber reinforced polymer (FRP) materials became widely available in the civil sector they started replacing steel as external confinement reinforcement. One of the primary applications of FRP materials is retrofit of concrete elements, primarily columns, to improve their ductility. This is done mainly in seismic regions where concrete structures experience large deformations. Column wrapping improves the strength and ductility of the concrete and improves its performance under earthquake loads. Xiao and Wu [4, 5] wrapped concrete cylinders using various materials and methods. Some were wrapped with carbon fiber fabrics, while others were wrapped with glass fiber fabrics. They also wrapped cylinders using a machine that tightly wound carbon fibers around the cylinders. The results showed a significant improvement both in strength and ductility. Lam and Teng [6], Li *et al.* [7] and Harries and Kharel [8] wrapped cylinders with carbon fabrics with similar results as Xiao and Wu [4]. Li and Hadi [9] and Campione and Miraglia [10] tested round concrete columns wrapped with either glass or carbon fiber fabric sheets in a polymer matrix. In doing so they improved the ductility of the columns. Campione and Miraglia [10] also wrapped, in the same manner, square columns and square columns with round corners with the same results. It was found that the round

shape is the most effective shape for confinement while the square with sharp corners the least effective of the three. Teng *et al*. [11] wrapped bridge columns in the field using FRP wraps. Laboratory specimens were also tested with the columns exhibiting a ductile behavior. Shahawy *et al*. [12] tested standard concrete cylinders wrapped with carbon fiber fabrics in an epoxy matrix. The results varied depending on the number of carbon layers applied. For an unconfined concrete strength of 41.4 MPa the confined strength of cylinders was increased to 70 MPa for the 1-layer wrap and 110 MPa for the 4-layer wrap. The ultimate strain for the 1-layer wrap was 0.007 and for the 4-layer wrap 0.016. Prefabricated FRP tubes can be filled with concrete and serve at the same time as formwork, flexural reinforcement and confinement reinforcement. Davol *et al*. [13] tested prefabricated round shells filled with concrete in flexure with satisfactory results. The concrete filled FRP shells exhibited a ductile behavior. Michael *et al*. [1] used a light CFRP composite grid to confine concrete. Through a series of cylinder tests they found that the grid provides light confinement to concrete. The crushing strain of confined concrete was twice as high compared to the unconfined concrete tested. Michael *et al*. [1] used the CFRP composite grid in a series of flexural members and had improvements in the member ductility of more than 30% with minimal confinement reinforcement.

2 Interaction diagrams

The interaction diagram (fig. 1) is a graphical representation of the ultimate capacity of a column subjected to axial load (Pn) and uniaxial bending (Mn). The

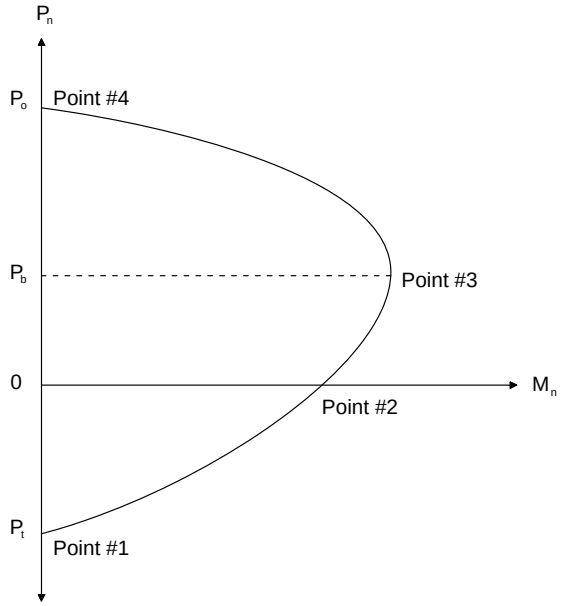


Figure 1: Interaction diagram.

interaction diagram depends on the concrete cross sectional area, the material properties (stress and strain) and also the amount and distribution of reinforcement. Therefore, each concrete section with a specific reinforcement distribution is characterized by a unique interaction diagram representing failure as the crushing of concrete at the maximum compressive strain. After the determination of design loads (P, M) three possible load conditions plotted as points can be defined once the interaction diagram for a section is obtained:

- The load condition coincides with the interaction diagram curve: represents the limit state.
- The load condition falls inside the interaction diagram curve: causes no failure in the section.
- The load condition falls outside the interaction diagram curve: causes failure in the section.

The interaction diagrams can also be extended to three dimensional surfaces to account for biaxial bending. The principle regarding the load conditions remains the same.

3 Requirements for the development of interaction diagrams

The following conditions, assumptions and limitations, the definition of the stress vs strain relations of the material as well as the definition of the plastic centroid of the section are required for the development of interaction diagrams.

3.1 Conditions

The following conditions must be satisfied in the development of the interaction diagram.

- Force equilibrium
- Strain Compatibility
- Stress vs Strain relationships

3.2 Assumptions and limitations

The following assumptions and limitations are applied.

- Plane sections remain plane
- The strain in the reinforcement is the same as that of the adjacent concrete interface
- The tensile strength of concrete is neglected
- Failure occurs in the concrete at maximum compressive strain

3.3 Stress vs strain properties

In this section the stress strain relationships for the materials are presented. The sections that are examined refer to reinforced column sections without any confinement and then compared to similar sections that are confined with CFRP.

3.3.1 Concrete

The stress – strain relation in the concrete that is used in this work is represented by the parabola defined by Hognestad as this is defined in the literature [14]. The tensile part of the graph is neglected. In order to define the curve it is required to have the concrete strength (f'_c), the strain at peak stress, ε_o, and the concrete modulus of elasticity (E_c).

3.3.2 Steel

The stress – strain relation is assumed to be elastic-plastic and it is the same in tension and compression [14]. In order to define this curve it is required to define the steel yield stress (f_y) and the modulus of elasticity of steel (E_s).

3.3.3 Experimental confinement model

Most models for concrete confined with CFRP reinforcement are based on the fact that in most cases even one layer of carbon fabric or a carbon jacket will provide enough reinforcement to have highly confined concrete. Therefore, the confinement effectiveness is high leading to a failure of the CFRP jacket or encasement at the peak axial stress. When the CFRP grid is used as confinement reinforcement the confining pressure and confinement effectiveness is low and therefore models developed using data from relatively high confined concrete may not be adequate. To model the behavior of CFRP grid confined concrete existing models were used. Existing models are based on a constant thickness of the FRP material that covers all of the surface area of the confined concrete core. Michael *et al.* [1] used the modified Hognestad stress–strain curve to model the behavior of CFRP grid confined concrete as shown in fig. 2 [1]. In fig 2 ε_c is the

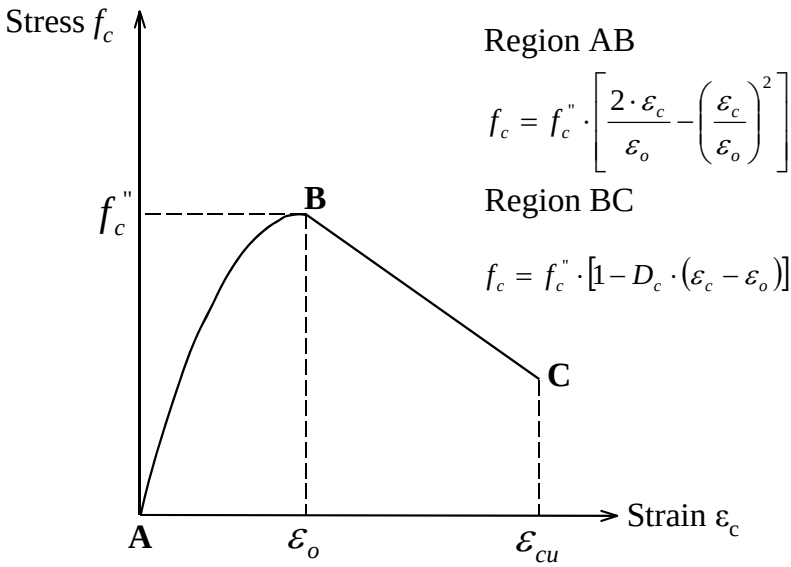


Figure 2: Modified Hognestad parabola used for the modeling of confined concrete [1].

concrete strain, ε_o is the strain at peak stress of unconfined concrete and ε_{cu} is the ultimate strain. The modified Hognestad parabola consists of two regions. In region AB ($\varepsilon_c < \varepsilon_o$) the Hognestad parabola is used and in region BC ($\varepsilon_o < \varepsilon_c < \varepsilon_{cu}$) a linearly descending curve. The equation for region BC is based on the deterioration constant (D_c) that controls the slope of the line. The equations for the two regions were modified to model the behavior of CFRP grid concrete. The material properties of the CFRP grid strands were used in the process of constructing the stress-strain curve of the CFRP grid confined concrete. The average strength of the control cylinders tested in deflection control model was taken as the strength of unconfined concrete (f'_c). The ultimate concrete strain ε_{cu} was assumed to be 0.00725 mm/mm. The deterioration constant was taken equal to 130 to match post peak experimental data. All three curves are depicted in fig. 3. The modified Hognestad matches well with the experimental curve.

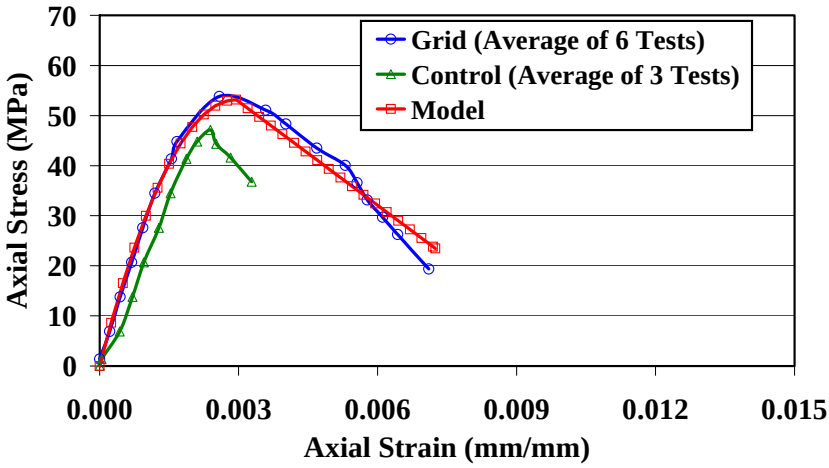


Figure 3: Concrete modelling as obtained by experimental data [1].

3.4 Plastic centroid

Reference must be made as to the point about which the moments are calculated. We refer to this point as the plastic centroid and it is the point of action of the axial load when there is a uniform maximum compressive strain. This assures that the moment capacity of the section is zero at maximum axial load capacity.

4 Development of the interaction diagram

The development of the interaction diagram is based on the calculation of a series of points representing failure in the concrete subjected to maximum compressive strain and a specified strain in the extreme steel fiber. A number of points, axial force vs bending moment (Pn, Mn) are calculated in order to define

the failure curve of the section. The calculation of individual points ensures equilibrium of the section and it includes:

- The definition of the neutral axis location
- The calculation of the plastic centroid
- The definition of the strain plane over the entire section
- The calculation of strains using compatibility and the corresponding stresses based on the stress vs strain relation
- The integration of stresses over the section to calculate the axial force and the bending moment

4.1 Neutral axis location

The neutral axis location is calculated using the values of the maximum compressive strain in the concrete, ε_{cu}, and a variable value for the strain in the extreme reinforcing steel fiber, ε_{st}. Each combination of strains (ε_{cu}, ε_{st}) will define a strain distribution over the section at failure and thus a point on the interaction diagram (axial load vs bending moment). The calculation of the neutral axis in a circular section can take advantage of the symmetry of the section. One point on the section (P_1) is assigned the maximum compressive strength and it is considered the extreme concrete compression fiber. The extreme steel fiber is located at the steel bar which is located at the maximum distance from the extreme compression fiber. Having the location of the two extreme fibers and the values of the corresponding strains the neutral axis can be defined as shown in fig. 4.

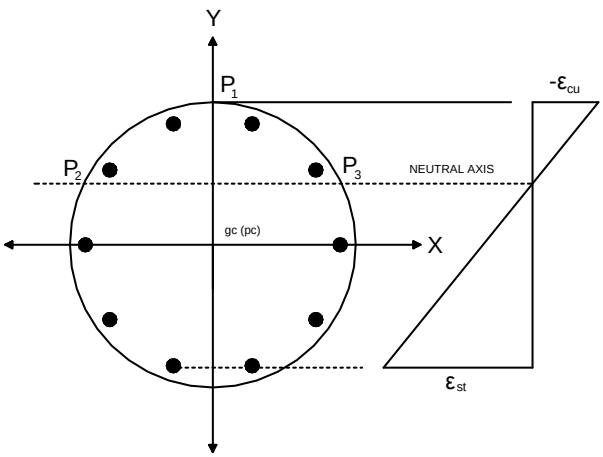


Figure 4: Rectangular cross sectional data.

4.2 Plastic centroid location

For symmetrical sections both in concrete geometry as well as the reinforcement distribution the plastic centroid (pc) coincides with the geometric centroid (gc).

When this is not the case then the location of the plastic centroid has to be calculated accounting for the concrete geometry as well as the area and location of each reinforcing bar.

4.3 Strain plane and calculation of strains

When the neutral axis is defined the distribution of the strain over the whole section can be defined with the calculation of a strain plane. The strain plane is defined using two points on the line of the neutral axis (P_2, P_3) and the point at the extreme compression fiber (P_1) as shown on fig. 4.

Using the strain plane equation a strain value for any point on the section can be defined. Based on the strain and the stress vs strain relation of the material the stress at each point can be obtained using eqn (1).

$$dF_i = \sigma_i dA_i \tag{1}$$

4.4 Integration of stresses to calculate axial load and bending moment

The axial load (P_n) and bending moment (M_n) can be calculated by the integration of stresses over the section. This can be done using eqn (2) and eqn (3):

$$P = \iint_A dF_i = \iint_A \sigma_i dA \tag{2}$$

$$M = \iint_A dF_i \; x = \iint_A \sigma_i \; x \, dA \tag{3}$$

4.5 Numerical procedure for the generation of the interaction diagram

The direct integration for the calculation of the axial load as well as the moment is not trivial. For this reason a numerical procedure (and accompanying software) was developed and used in this work for the generation of the interaction diagrams. The range of the axial loads spans from the maximum compressive axial load to the axial load of pure tension. In order to numerically generate the diagram four points are identified on it (fig. 1).

- Point #1: Fracture failure point (pure tension)
- Point #2: Zero axial load point
- Point #3: Balanced point
- Point #4: Maximum compressive axial load point (pure compression)

These points are calculated independently and they define three (3) sub regions on the diagram. For each Point the important element to be known is the value of the net tensile strain at the extreme tension reinforcement fiber. The strains at Point #1, Point #3 and Point #4 are known directly from material properties. The other one has to be calculated. The strain at Point#2 represents the point with zero axial load. However, the strain in the extreme reinforcement bar is not known. As a result an iteration convergence procedure (secant method)

is used to calculate the strain in the extreme steel fiber when the axial load equals to zero. Once the strains for the boundary points of the sub regions are defined, the diagram can be generated by assigning different values of strains for the extreme steel fiber in each sub region and thus calculating intermediate points within the sub regions on the interaction diagram. Fig. 5 shows the flowchart of the numerical procedure.

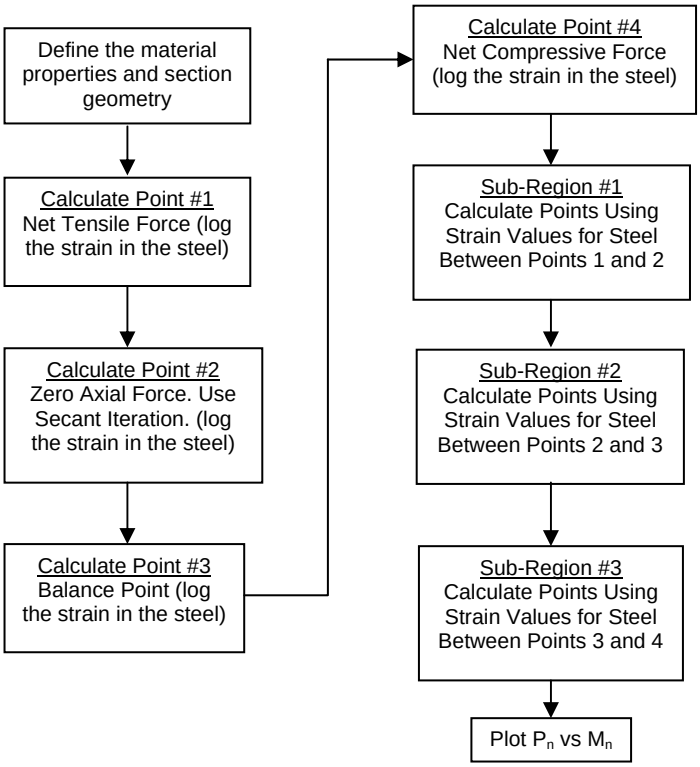


Figure 5: Flowchart describing the numerical procedure.

5 Example

The presented procedure has been used for the development of the interaction diagram of different sections. Fig. 6 shows the interaction diagram of the same section with three different levels of concrete strength. The inner line shows the unconfined section whereas the intermediate line shows the section with confinement strength as described in section 3.3.3 (Experimental results). The outer line shows the section using the same model for confinement as that of section 3.3.3 but with a different value of the maximum compressive strength. Particularly the data is shown below:

5.1 Section data

Section Radius	25 cm
Reinforcement	10 bars (dia=16 mm) distributed uniformly
ε_{cu} (Unconfined)	0.003 mm/mm
ε_{cu} (Confined)	0.00725 mm/mm
f'$_c$ (Unconfined)	40 MPa
f'$_c$ (Confined – 1)	54 MPa
f'$_c$ (Confined - 2):	60 MPa
D_c (Deterioration Constant)	130

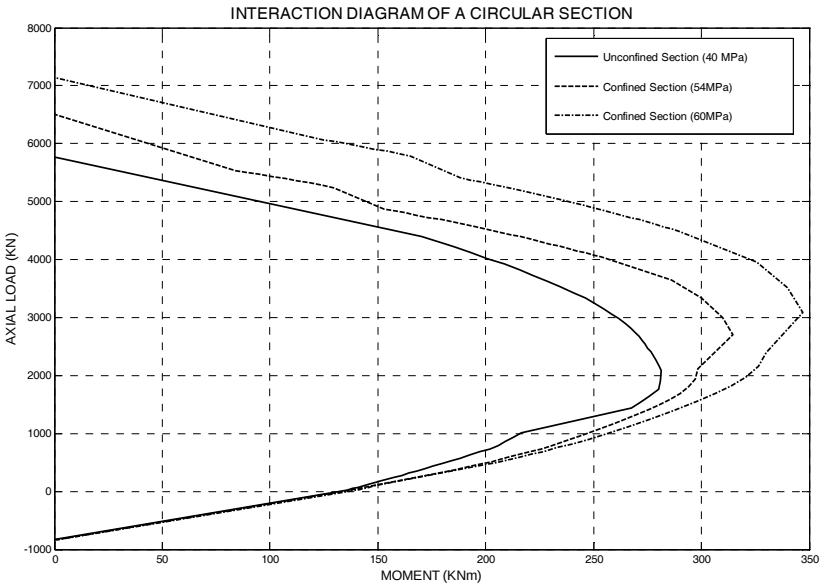


Figure 6: Section comparison with variable concrete strength.

5.2 Discussion

Looking at the plots on fig. 6 it is obvious that there is a trend defined as the value of the maximum compressive strength is increased. Specifically we see that the plots look virtually the same at the tension controlled regions and they diverge in the compression controlled regions as the maximum compressive strength increases. The maximum compressive strength obviously increases as the level of confinement increases. It is interesting to point out that the value of the maximum compressive strain, ε_{cu}, does not have an effect on the interactive diagram. Therefore the increase of ductility which is gained due to confinement does not play a significant role in the maximum capacity of the section. The decisive factor that affects the section capacity is the maximum compressive strength.

6 Conclusions

The following conclusions have been drawn at the end of this work:

- Confinement Increases the maximum compressive strength of the section
- Increase in the confinement reinforcement increases the capacity of the section in the compression controlled region
- Confinement affects significantly the capacity of the section when the section is in the compression controlled region (pure compression to balance point)
- The effect of confinement is small in the region between pure bending and the balance point
- Confinement has no effect in the region between pure tension and pure bending since concrete is primarily in tension. Therefore the presence of reinforcement in the hoop direction offers no improvement in concrete strength

References

[1] Michael, A. P., H. R. Hamilton III, Ansley, M. H, Concrete Confinement Using Carbon Fiber Reinforced Polymer Grid, *7th International Symposium on Fiber Reinforced Polymer (FRP) Reinforcement for Concrete Structures* (ACI 2005 Fall Convention), American Concrete Institute, Kansas City, MO, Vol. 2, pp. 991-1010, 2005.

[2] Bresler, B., Design Criteria for Reinforced Concrete Columns under Axial Load and Biaxial Bending, ACI Journal, Proceedings, Vol. 57, 1960.

[3] Parme, A. L., Nieves, J.M., Gouwens, A., Capacity of Reinforced Rectangular Columns Subjected to Biaxial Bending, ACI Journal, Proceedings, Vol. 63, No 9, 1966.

[4] Xiao, Y. and Wu, H., Compressive Behavior of Concrete Confined by Carbon Fiber Composite Jackets, *Journal of Materials in Civil Engineering*, Vol. 12, No 2, pp. 139-146, 2000.

[5] Xiao, Y. and Wu, H., A Constitutive Model for Concrete Confinement with Carbon Fiber Reinforced Plastics, *Journal of Reinforced Plastics and composites*, Vol. 22, No 13, pp. 1187-1201, 2003.

[6] Lam, L., and Teng, J. G., Ultimate Condition of Fiber Reinforced Polymer-Confined Concrete, *Journal of Composites for Construction*, Vol. 8, No 6, pp 539-548, 2004.

[7] Li, Y., Lin, C. and Sung, Y., Compressive Behavior of Concrete Confined by Various Types of FRP Composite Jackets, *Mechanics of Materials*, Vol. 35, No 3-6, pp. 603-619, 2002.

[8] Harries, K. A., and Kharel, G., Experimental Investigation of the Behavior of Variably Confined Concrete, *Cement and Concrete Research*, Vol. 33, No 6, pp. 873-880, 2002.

[9] Li, J., Hadi, M.N.S., Behaviour of Externally Confined High-Strength Concrete Columns Under Eccentric Loading, *Composite Structures*, Vol. 62, No 2, pp 145-153, 2003.
[10] Campione, G., Miraglia, N., Strength and Strain Capacities of Concrete Compression Members Reinforced with FRP, *Cement and Concrete Composites*, Vol. 25, No 1, pp 31-41, 2003.
[11] Teng, M., Sotelino, E. D., and Chen, W., Performance Evaluation of Reinforced Concrete Bridge Columns Wrapped with Fiber Reinforced Polymers, *Journal of Composites for Construction*, Vol. 7, No 2, pp 83-92, 2002.
[12] Shahawy, M, Mirmiran, A. and Beitelman, T., Tests and Modeling of Carbon-Wrapped Concrete Columns, *Composites Part B: Engineering*, Vol. 31, No 6-7, pp 471-480, 2000.
[13] Davol, A., Burgueno, R. and Seible, F., Flexural Behavior of Circular Concrete Filled FRP Shells, *Journal of Structural Engineering*, Vol. 127, No 7, pp 810-817, 2001.
[14] Park, R. and Paulay, T., Ultimate Deformation and Ductility of Members with Flexure (Chapter 6), *Reinforced Concrete Structures*, John Wiley & Sons, New York, pp 195-269, 1975.

Constitutive formulation of strain-rate sensitive deformation in orthotropic composites

K. Micallef[1], A. Soleiman-Fallah[1], P. T. Curtis[2], D. J. Pope[2] & L. A. Louca[1]
[1]*Department of Civil Engineering, Imperial College, London, UK*
[2]*Physical Sciences Department, DSTL Porton Down, UK*

Abstract

The aim of the current study is to develop a constitutive model which captures the full orthotropic behaviour of a laminated composite by employing 9 material parameters and also taking into account strain-rate sensitivity to loading.

The formulation is an extension of the work by Ogihara and Reifsnider (DOI: 10.1023/A:1016069220255), whose model considers 4 parameters, and with the inclusion of strain-rate effect considerations using the method employed by Thiruppukuzhi and Sun (DOI: 10.1016/S0266-3538(00)00133-0).

A plastic potential function which can describe plasticity in all directions, including fibre plasticity, is chosen and using an associated flow rule, the plastic strain-rate components are derived. The plastic compliance matrix is assembled, using a rate-dependent visco-plastic modulus. The elastic compliance matrix is combined with its plastic counterpart to give a rate-form constitutive law.

It is found that the proposed model accounts for strain-rate dependence and by correct choice of model parameters, the model can also be used for various composite architectures, including woven and uni-directional architectures.

Keywords: Composites, orthotropic, constitutive modelling, strain-rate effects.

1 Introduction

The formulation of a comprehensive constitutive model is imperative for the proper understanding of a material's behaviour under different loading conditions. This includes fibre-reinforced laminated polymeric composites. For efficient use of composite materials under extreme loading, it is necessary to

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110356

consider their response beyond the linear-elastic regime and for different loading rates.

Composites are, in general, non-isotropic materials and tend to exhibit highly non-linear behaviour under various kinds of dynamic loadings. The severe loading nature of dynamic loads also implies that strain-rate effects have a repercussion on the stress-strain response of the material. In addition, it is common for a material to be loaded beyond its elastic limit and deform plastically, giving rise to plastic strains which lead to progressive damage of the material.

Thus, it is necessary to develop a constitutive relationship which can predict non-linear, large strain, strain-rate dependent behaviour to failure of anisotropic uni-directional and/or woven composites which are subjected to dynamic loads.

When a material, including a composite, is subjected to severe loads, it is subjected to significantly high stresses and strains. Thus, when one analyses a composite structure using a linear elastic technique and using a conventional failure criterion to describe failure at a ply level, one would find that failure is attained immediately and throughout the structure. Designing a structure not to fail and to remain elastic under high loads will make the structures thicker and thus heavier and more expensive.

Thus, it is required to have a non-linear response which can describe the material performance beyond the first ply failure and which can model plastic damage sustained by the composite when loaded at high strain-rates.

1.1 Constitutive models for composite materials

The constitutive model of a material can be defined as a relation used to characterise its physical properties and such a model will be used to describe the material response of the material when subjected to loading.

The non-linear response of composites has long been recognised to be an important consideration in constitutive modelling. It is common in non-linear constitutive theory to develop the concept of effective stress, $\bar{\sigma}$, and effective plastic strain, $\bar{\varepsilon}^p$. This is similar to classical plasticity theory in metals, where the non-linearity is defined on the basis of the effective quantities only.

In 1973, Hahn and Tsai [1] used a complementary elastic energy density function to derive a stress-strain relation which is linear in the longitudinal and transverse directions but non-linear in the in-plane shear. The non-linearity is taken into account by the inclusion of a bi-quadratic term.

A non-linear plastic theory was first applied to composites by Sun and Chen [2], who suggested that a plastic potential function could be applied to uni-directional composites. Using a simple one parameter plastic flow rule, the potential function adopted was essentially a derivative of Hill's [3] generalised plastic potential for a 3D orthotropic material. Sun and Chen, however, assumed a plane stress scenario, imposing that there is no plastic strain in the fibre direction and limiting plasticity only to the matrix.

This assumption is unacceptable in the case of woven composites, since experimental data [4] has shown that non-linearity did occur when material was loaded in the 0° and 90° directions.

In fact, Ogihara and Reifsnider [4] expanded the work of Sun and Chen by using a more general plastic potential with four unknown parameters. These parameters were determined by a number of simple tension experiments for different specimen angles. The effective stress and effective plastic strain were found for each test angle and the parameter combinations resulted in all effective stress-effective plastic strain curves to converge into a single master curve.

The concept of using a master curve has been shown to be valid for various materials by Sun, both in his original 1989 work [2] and also in subsequent works [5–8]. Non-linearity is expressed by a function representing the master curve. Sun and Chen [2] had proposed a power law relating effective stress to effective strain:

$$\bar{\varepsilon}^p = A\,\bar{\sigma}^n \tag{1}$$

Weeks and Sun [6] found that the exponent n in the above relation was appropriate at all tested strain-rates and thus rate independent. Thus, the rate sensitivity is described solely by the parameter A.

For composites to be used under dynamic loading, the effects of strain-rate need to be included in the constitutive model. Thiruppukuzhi and Sun [5, 7] proposed a developed power law describing the strain-rate dependence of A:

$$A = \chi\,(\dot{\bar{\varepsilon}}^p)^m \tag{2}$$

Thus:

$$\bar{\varepsilon}^p = \chi\,(\dot{\bar{\varepsilon}}^p)^m\,\bar{\sigma}^n \tag{3}$$

More recently, Hufner and Accorsi [9] have extended the four parameter plastic potential function of Ogihara and Reifsnider to include strain-rate effects using the strain-dependent power law of Thiruppukuzhi and Sun described above.

It should be noted that the non-linear behaviour is characterised on a macroscopic level. Although it is desirable to predict non-linear response of woven composites on a micromechanical scale, i.e., based on the properties of the fibres and the matrix, many workers have shown that it is difficult to achieve this, even for uni-directional composites [4].

1.2 Choice and method of formulation

The present work will extend the formulation of Ogihara and Reifsnider from a model with 4 parameters to consider all possible parameters (i.e., 9) and taking into account the strain-rate effects as in the work of Thiruppukuzhi and Sun.

2 Derivation of a rate-dependent constitutive model

In the present study, a plastic potential function which has the most general form with 9 unknown parameters, that could be applied to both anisotropic uni-directional and woven composites is proposed, namely:

$$2f(\sigma_{ij}) = a_{11}{\sigma_{11}}^2 + a_{22}{\sigma_{22}}^2 + a_{33}{\sigma_{33}}^2 + 2a_{44}{\sigma_{23}}^2 + 2a_{55}{\sigma_{13}}^2 + 2a_{66}{\sigma_{12}}^2 + 2a_{12}\sigma_{11}\sigma_{22} + 2a_{13}\sigma_{11}\sigma_{33} + 2a_{23}\sigma_{22}\sigma_{33} \tag{4}$$

Using appropriate values for the various parameters, the function could be used to describe a range of material systems. For example, choosing $a_{11} = a_{12} = 0$ and $a_{22} = 1$ reduces the above potential function to the one parameter potential of Sun and Chen [2] for uni-directional composites. Similarly, by setting $a_{11} = a_{12} = a_{22} = 0$ gives the function used by Thiruppukuzhi and Sun [7] for woven composites.

The generalised anisotropic constitutive equations, in rate form, are expressed as follows:

$$[\dot{\boldsymbol{\varepsilon}}] = [\boldsymbol{C}][\dot{\boldsymbol{\sigma}}] \tag{5}$$

In expanded form, this is written as:

$$\begin{Bmatrix} \dot{\varepsilon}_{11} \\ \dot{\varepsilon}_{22} \\ \dot{\varepsilon}_{33} \\ \dot{\varepsilon}_{23} \\ \dot{\varepsilon}_{13} \\ \dot{\varepsilon}_{12} \end{Bmatrix} = \begin{bmatrix} C_{11} & C_{12} & C_{13} & C_{14} & C_{15} & C_{16} \\ C_{21} & C_{22} & C_{23} & C_{24} & C_{25} & C_{26} \\ C_{31} & C_{32} & C_{33} & C_{34} & C_{35} & C_{36} \\ C_{41} & C_{42} & C_{43} & C_{44} & C_{45} & C_{46} \\ C_{51} & C_{52} & C_{53} & C_{54} & C_{55} & C_{56} \\ C_{61} & C_{62} & C_{63} & C_{64} & C_{65} & C_{66} \end{bmatrix} \begin{Bmatrix} \dot{\sigma}_{11} \\ \dot{\sigma}_{22} \\ \dot{\sigma}_{33} \\ \dot{\sigma}_{23} \\ \dot{\sigma}_{13} \\ \dot{\sigma}_{12} \end{Bmatrix} \tag{6}$$

The strain-rate tensor is decomposed into two components, namely the elastic and the plastic strain-rate components:

$$\varepsilon_{ij} = \varepsilon_{ij}^{e} + \varepsilon_{ij}^{p} \tag{7}$$

Thus, the compliance matrix is expressed as the sum of the elastic and plastic components:

$$[\dot{\boldsymbol{\varepsilon}}] = [\boldsymbol{C}][\dot{\boldsymbol{\sigma}}] = [\boldsymbol{C^e} + \boldsymbol{C^p}][\dot{\boldsymbol{\sigma}}] \tag{8}$$

Each part of the compliance matrix will be derived in turn in the following sections.

2.1 Elastic compliance matrix

The elastic part of the compliance matrix is the standard orthotropic compliance relation obtained from standard textbooks, such as Jones [10]:

$$[\boldsymbol{C^e}] = \begin{bmatrix} \frac{1}{E_1} & -\frac{\nu_{21}}{E_2} & -\frac{\nu_{31}}{E_3} & 0 & 0 & 0 \\ -\frac{\nu_{12}}{E_1} & \frac{1}{E_2} & -\frac{\nu_{32}}{E_3} & 0 & 0 & 0 \\ -\frac{\nu_{13}}{E_1} & -\frac{\nu_{23}}{E_2} & \frac{1}{E_3} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{G_{23}} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{G_{31}} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{G_{12}} \end{bmatrix} \tag{9}$$

This matrix is symmetric, i.e. $C_{ij}^e = C_{ji}^e$ and has effectively only nine unequal constants since:

$$\frac{\nu_{ij}}{E_i} = \frac{\nu_{ji}}{E_j} \tag{10}$$

The elastic constants in this matrix are assumed to be rate-dependent in the work of Hufner and Accorsi [9] but Thiruppukuzhi and Sun [7] state that there is no strain-rate dependence in the elastic part of loading. This is assumed to be a valid simplification for the purpose of this study.

2.2 Plastic compliance matrix

The plastic part of the compliance matrix is assembled by first computing the plastic strain-rate components from the potential function using an associated flow rule:

$$\dot{\varepsilon}_{ij}^p = \dot{\lambda} \frac{\partial f\left(\sigma_{ij}\right)}{\partial \sigma_{ij}} \tag{11}$$

where $\dot{\lambda}$ is the proportionality factor rate.

Applying this flow rule to the chosen plastic potential, the plastic strain-rate components are:

$$\dot{\varepsilon}_{11}^{p} = \dot{\lambda}\,(a_{11}\sigma_{11} + a_{12}\sigma_{22} + a_{13}\sigma_{33}) \quad (12a)$$
$$\dot{\varepsilon}_{22}^{p} = \dot{\lambda}\,(a_{22}\sigma_{22} + a_{23}\sigma_{33} + a_{12}\sigma_{11}) \quad (12b)$$
$$\dot{\varepsilon}_{33}^{p} = \dot{\lambda}\,(a_{33}\sigma_{33} + a_{23}\sigma_{22} + a_{13}\sigma_{11}) \quad (12c)$$
$$\dot{\varepsilon}_{23}^{p} = \dot{\lambda}\,(2a_{44}\sigma_{23}) \quad (12d)$$
$$\dot{\varepsilon}_{13}^{p} = \dot{\lambda}\,(2a_{55}\sigma_{13}) \quad (12e)$$
$$\dot{\varepsilon}_{12}^{p} = \dot{\lambda}\,(2a_{66}\sigma_{12}) \quad (12f)$$

The proportionality factor rate, $\dot{\lambda}$, is derived using the equivalence of the rate of plastic work, $\dot{W}^{p}$, namely:

$$\dot{W}^{p} = \bar{\sigma}\dot{\bar{\varepsilon}}^{p} = \sigma_{ij}\dot{\varepsilon}_{ij}^{p} \quad (13)$$

Rearranging the equation for the flow rule:

$$\dot{\lambda} = \frac{\bar{\sigma}\,\dot{\bar{\varepsilon}}^{p}}{2f} \quad (14)$$

From classical plasticity theory, $\bar{\sigma} = \sqrt{3f}$ and thus:

$$f = \frac{\bar{\sigma}^{2}}{3} \quad (15)$$

This implies that $\dot{\lambda}$ can be expressed as a function of $\dot{\bar{\varepsilon}}^{p}$ and $\bar{\sigma}$ only:

$$\dot{\lambda} = \frac{\bar{\sigma}\,\dot{\bar{\varepsilon}}^{p}}{2f} = \frac{3\,\bar{\sigma}\,\dot{\bar{\varepsilon}}^{p}}{2\bar{\sigma}^{2}} = \frac{3\,\dot{\bar{\varepsilon}}^{p}}{2\,\bar{\sigma}} \quad (16)$$

A rate-dependent visco-plastic modulus, H_p, is defined as:

$$H_p = \frac{\dot{\bar{\sigma}}}{\dot{\bar{\varepsilon}}^{p}} = \frac{d\bar{\sigma}}{d\bar{\varepsilon}^{p}} \quad (17)$$

Thus, the proportionality factor rate is expressed as:

$$\dot{\lambda} = \frac{3\,\dot{\bar{\sigma}}}{2\,H_p\,\bar{\sigma}} \quad (18)$$

Using the definition of $\bar{\varepsilon}^{p}$ proposed by Thiruppukuzhi and Sun [5, 7] given in eqn (3) earlier and differentiating it with respect to $\bar{\sigma}$,

$$\frac{d\bar{\varepsilon}^{p}}{d\bar{\sigma}} = n\,\chi\,(\dot{\bar{\varepsilon}}^{p})^{m}\,\bar{\sigma}^{n-1} \quad (19)$$

This implies that:

$$H_p = \frac{d\bar{\sigma}}{d\bar{\varepsilon}^p} = \frac{1}{n\,\chi\,(\bar{\varepsilon}^p)^m\,\bar{\sigma}^{n-1}} \tag{20}$$

The effective stress-rate is given as:

$$\begin{aligned}\dot{\bar{\sigma}} = \frac{3}{2\,\bar{\sigma}}[&(a_{11}\sigma_{11} + a_{12}\sigma_{22} + a_{13}\sigma_{33})\dot{\sigma}_{11}\\ &+ (a_{22}\sigma_{22} + a_{23}\sigma_{33} + a_{12}\sigma_{11})\dot{\sigma}_{22}\\ &+ (a_{33}\sigma_{33} + a_{23}\sigma_{22} + a_{13}\sigma_{11})\dot{\sigma}_{33}\\ &+ (2a_{44}\sigma_{23})\dot{\sigma}_{12} + (2a_{55}\sigma_{13})\dot{\sigma}_{13}\\ &+ (2a_{66}\sigma_{12})\dot{\sigma}_{23}]\end{aligned} \tag{21}$$

Finally, after substitution, the proportionality factor rate is given by:

$$\begin{aligned}\dot{\lambda} = \frac{3\,\dot{\bar{\sigma}}}{2\,H_p\,\bar{\sigma}} = \frac{3}{2\,H_p\,\bar{\sigma}}\Big\{\frac{3}{2\,\bar{\sigma}}[&(a_{11}\sigma_{11} + a_{12}\sigma_{22}\\ &+ a_{13}\sigma_{33})\dot{\sigma}_{11}\\ &+ (a_{22}\sigma_{22} + a_{23}\sigma_{33} + a_{12}\sigma_{11})\dot{\sigma}_{22}\\ &+ (a_{33}\sigma_{33} + a_{23}\sigma_{22} + a_{13}\sigma_{11})\dot{\sigma}_{33}\\ &+ (2a_{44}\sigma_{23})\dot{\sigma}_{12} + (2a_{55}\sigma_{13})\dot{\sigma}_{13}\\ &+ (2a_{66}\sigma_{12})\dot{\sigma}_{23}]\Big\}\end{aligned} \tag{22}$$

Thus:

$$\begin{aligned}\dot{\lambda} = \frac{9}{4\,H_p\,\bar{\sigma}^2}[&(a_{11}\sigma_{11} + a_{12}\sigma_{22} + a_{13}\sigma_{33})\dot{\sigma}_{11}\\ &+ (a_{22}\sigma_{22} + a_{23}\sigma_{33} + a_{12}\sigma_{11})\dot{\sigma}_{22}\\ &+ (a_{33}\sigma_{33} + a_{23}\sigma_{22} + a_{13}\sigma_{11})\dot{\sigma}_{33}\\ &+ (2a_{44}\sigma_{23})\dot{\sigma}_{12} + (2a_{55}\sigma_{13})\dot{\sigma}_{13}\\ &+ (2a_{66}\sigma_{12})\dot{\sigma}_{23}]\end{aligned} \tag{23}$$

By substituting this expression for $\dot{\lambda}$ into the plastic strain-rate components derived earlier in eqn (12), it is possible to find the individual terms of the plastic compliance matrix.

The derivation requires extensive algebraic manipulation and, for the sake of brevity, only one full derivation of the first row of the plastic compliance matrix ($\dot{\varepsilon}^p_{1i}, i = 1-6$) is presented as an example.

$$\begin{aligned}\dot{\varepsilon}_{11}^{p} &= \dot{\lambda}\,(a_{11}\sigma_{11} + a_{12}\sigma_{22} + a_{13}\sigma_{33}) \\ &= \left\{\frac{9}{4\,H_p\,\bar{\sigma}^2}[(a_{11}\sigma_{11} + a_{12}\sigma_{22} \right. \\ &+ a_{13}\sigma_{33})\dot{\sigma}_{11} \\ &+ (a_{22}\sigma_{22} + a_{23}\sigma_{33} + a_{12}\sigma_{11})\dot{\sigma}_{22} \\ &+ (a_{33}\sigma_{33} + a_{23}\sigma_{22} + a_{13}\sigma_{11})\dot{\sigma}_{33} \\ &+ (2a_{44}\sigma_{23})\dot{\sigma}_{12} + (2a_{55}\sigma_{13})\dot{\sigma}_{13} \\ &\left. + (2a_{66}\sigma_{12})\dot{\sigma}_{23}]\right\}(a_{11}\sigma_{11} + a_{12}\sigma_{22} \\ &+ a_{13}\sigma_{33})\end{aligned} \tag{24}$$

Upon expansion and simplification,

$$\begin{aligned}\dot{\varepsilon}_{11}^{p} = \frac{9}{4\,H_p\,\bar{\sigma}^2}[&(a_{11}^2\sigma_{11}^2 + 2a_{11}a_{12}\sigma_{11}\sigma_{22} + 2a_{11}a_{13}\sigma_{11}\sigma_{33} \\ &+ 2a_{12}a_{13}\sigma_{22}\sigma_{33} + a_{12}^2\sigma_{22}^2 + a_{13}^2\sigma_{33}^2)\dot{\sigma}_{11} \\ &+ (a_{11}a_{12}\sigma_{11}^2 + a_{22}a_{12}\sigma_{22}^2 + a_{13}a_{23}\sigma_{33}^2 \\ &+ a_{11}a_{22}\sigma_{11}\sigma_{22} \\ &+ a_{11}a_{23}\sigma_{11}\sigma_{33} + a_{12}a_{13}\sigma_{11}\sigma_{33} \\ &+ a_{12}a_{23}\sigma_{22}\sigma_{33} + a_{13}a_{22}\sigma_{22}\sigma_{33} \\ &+ a_{12}^2\sigma_{22}\sigma_{33})\dot{\sigma}_{22} \\ &+ (a_{11}a_{13}\sigma_{11}^2 + a_{23}a_{12}\sigma_{22}^2 + a_{13}a_{33}\sigma_{33}^2 \\ &+ a_{11}a_{23}\sigma_{11}\sigma_{22} \\ &+ a_{12}a_{13}\sigma_{11}\sigma_{22} + a_{11}a_{13}\sigma_{11}\sigma_{33} \\ &+ a_{12}a_{33}\sigma_{22}\sigma_{33} + a_{13}a_{23}\sigma_{22}\sigma_{33} \\ &+ a_{13}^2\sigma_{11}\sigma_{33})\dot{\sigma}_{33} \\ &+ 2(a_{11}a_{44}\sigma_{11}\sigma_{23} + a_{12}a_{44}\sigma_{22}\sigma_{23} \\ &+ a_{13}a_{44}\sigma_{33}\sigma_{23})\dot{\sigma}_{23} \\ &+ 2(a_{11}a_{55}\sigma_{11}\sigma_{13} + a_{12}a_{55}\sigma_{22}\sigma_{13} \\ &+ a_{13}a_{55}\sigma_{33}\sigma_{13})\dot{\sigma}_{13} + 2(a_{11}a_{66}\sigma_{11}\sigma_{12} \\ &+ a_{12}a_{66}\sigma_{22}\sigma_{12} + a_{13}a_{66}\sigma_{33}\sigma_{12})\dot{\sigma}_{12}]\end{aligned} \tag{25}$$

Extracting the co-efficients of each $\dot{\sigma}_{ij}$ gives the entries for the plastic compliance matrix. For example, the first entry of the first row of the compliance matrix would be:

$$\begin{aligned}C_{11}^{p} = \frac{9}{4\,H_p\,\bar{\sigma}^2}(&a_{11}^2\sigma_{11}^2 + 2a_{11}a_{12}\sigma_{11}\sigma_{22} + 2a_{11}a_{13}\sigma_{11}\sigma_{33} \\ &+ 2a_{12}a_{13}\sigma_{22}\sigma_{33} + a_{12}^2\sigma_{22}^2 + a_{13}^2\sigma_{33}^2)\end{aligned} \tag{26}$$

A similar procedure yields the rest of the terms in the plastic compliance matrix, which are presented in Appendix 1. It is noted that the plastic compliance matrix is also symmetric, i.e. $C_{ij}^{p} = C_{ji}^{p}$ and thus the elasto-plastic compliance matrix is overall symmetric.

3 Conclusions and future work

The framework for a constitutive model which captures the full orthotropic behaviour of a laminated composite has been developed. This model employs 9 material parameters and also takes into account strain-rate sensitivity to loading. Thus, the rate form constitutive relationship is defined in terms of a compliance matrix which also includes a visco-plastic modulus.

The devised model will be implemented in a user-defined material subroutine (VUSDFLD and/or VUMAT) in the finite element code ABAQUS/Explicit, using material properties obtained from small-specimen basic tensile, compressive and shear tests at different strain-rates on specimens of various composite layups under dynamic loads.

The proposed constitutive model will also be validated using larger 300x300mm panels of the same composites tested earlier. A series of parametric studies will be performed using the numerical model to assess the overall response sensitivity of the proposed material model to the parameters defining it uniquely, such as modulii, Poisson ratios or yield locus. The study will include both small specimens as well as large panels.

Acknowledgements

This work is part of a research project jointly funded by the Defence Science and Technology Laboratory (DSTL) and the Engineering and Physical Sciences Research Council (EPSRC).

Appendix

The complete list of the elasto-plastic compliance matrix terms are given below:

$$C_{11} = \frac{1}{E_1} + \frac{9}{4\,H_p\,\bar{\sigma}^2}(a_{11}^2\sigma_{11}^2 + 2a_{11}a_{12}\sigma_{11}\sigma_{22} + 2a_{11}a_{13}\sigma_{11}\sigma_{33} + 2a_{12}a_{13}\sigma_{22}\sigma_{33} + a_{12}^2\sigma_{22}^2 + a_{13}^2\sigma_{33}^2)$$

$$C_{12} = C_{21} = -\frac{\nu_{21}}{E_2} + \frac{9}{4\,H_p\,\bar{\sigma}^2}(a_{11}a_{12}\sigma_{11}^2 + a_{22}a_{12}\sigma_{22}^2 + a_{13}a_{23}\sigma_{33}^2 + a_{11}a_{22}\sigma_{11}\sigma_{22} + a_{11}a_{23}\sigma_{11}\sigma_{33} + a_{12}a_{13}\sigma_{11}\sigma_{33} + a_{12}a_{23}\sigma_{22}\sigma_{33} + a_{13}a_{22}\sigma_{22}\sigma_{33} + a_{12}^2\sigma_{22}\sigma_{33})$$

$$C_{13} = C_{31} = -\frac{\nu_{31}}{E_3} + \frac{9}{4\,H_p\,\bar{\sigma}^2}(a_{11}a_{13}\sigma_{11}^2 + a_{23}a_{12}\sigma_{22}^2 + a_{13}a_{33}\sigma_{33}^2 + a_{11}a_{23}\sigma_{11}\sigma_{22} + a_{12}a_{13}\sigma_{11}\sigma_{22} + a_{11}a_{13}\sigma_{11}\sigma_{33} + a_{12}a_{33}\sigma_{22}\sigma_{33} + a_{13}a_{23}\sigma_{22}\sigma_{33} + a_{13}^2\sigma_{11}\sigma_{33})$$

$$C_{14} = C_{41} = \frac{9\,a_{44}}{2\,H_p\,\bar{\sigma}^2}(a_{11}\sigma_{11}\sigma_{23} + a_{12}\sigma_{22}\sigma_{23} + a_{13}\sigma_{33}\sigma_{23})$$

$$C_{15} = C_{51} = \frac{9\,a_{55}}{2\,H_p\,\bar{\sigma}^2}(a_{11}\sigma_{11}\sigma_{13} + a_{12}\sigma_{22}\sigma_{13} + a_{13}\sigma_{33}\sigma_{13})$$

$$C_{16} = C_{61} = \frac{9\,a_{66}}{2\,H_p\,\bar{\sigma}^2}(a_{11}\sigma_{11}\sigma_{12} + a_{12}\sigma_{22}\sigma_{12} + a_{13}\sigma_{33}\sigma_{12})$$

$$C_{22} = \frac{1}{E_2} + \frac{9}{4\,H_p\,\bar{\sigma}^2}(a_{22}^2\sigma_{22}^2 + 2a_{22}a_{23}\sigma_{22}\sigma_{33} + 2a_{12}a_{22}\sigma_{11}\sigma_{22} + 2a_{12}a_{23}\sigma_{11}\sigma_{33} + a_{12}^2\sigma_{11}^2 + a_{23}^2\sigma_{33}^2)$$

$$C_{23} = C_{32} = -\frac{\nu_{32}}{E_3} + \frac{9}{4\,H_p\,\bar{\sigma}^2}(a_{33}^2\sigma_{33}^2 + a_{22}a_{33}\sigma_{22}\sigma_{33} + a_{13}a_{22}\sigma_{11}\sigma_{22} + a_{13}a_{23}\sigma_{11}\sigma_{33} + a_{12}a_{33}\sigma_{11}\sigma_{33} + a_{12}a_{23}\sigma_{11}\sigma_{22} + a_{12}a_{13}\sigma_{11}^2 + a_{23}^2\sigma_{22}\sigma_{33} + a_{22}a_{23}\sigma_{22}^2)$$

$$C_{24} = C_{42} = \frac{9\,a_{44}}{2\,H_p\,\bar{\sigma}^2}(a_{12}\sigma_{11}\sigma_{23} + a_{22}\sigma_{22}\sigma_{23} + a_{23}\sigma_{33}\sigma_{23})$$

$$C_{25} = C_{52} = \frac{9\,a_{55}}{2\,H_p\,\bar{\sigma}^2}(a_{12}\sigma_{11}\sigma_{13} + a_{22}\sigma_{22}\sigma_{13} + a_{23}\sigma_{33}\sigma_{13})$$

$$C_{26} = C_{62} = \frac{9\,a_{66}}{2\,H_p\,\bar{\sigma}^2}(a_{12}\sigma_{11}\sigma_{12} + a_{22}\sigma_{22}\sigma_{12} + a_{23}\sigma_{33}\sigma_{12})$$

$$C_{33} = \frac{1}{E_3} + \frac{9}{4\,H_p\,\bar{\sigma}^2}(a_{33}^2\sigma_{33}^2 + 2a_{23}a_{33}\sigma_{22}\sigma_{33} + 2a_{13}a_{33}\sigma_{11}\sigma_{33} + 2a_{13}a_{23}\sigma_{11}\sigma_{22} + a_{13}^2\sigma_{11}^2 + a_{23}^2\sigma_{22}^2)$$

$$C_{34} = C_{43} = \frac{9\,a_{44}}{2\,H_p\,\bar{\sigma}^2}(a_{13}\sigma_{11}\sigma_{23} + a_{23}\sigma_{22}\sigma_{23} + a_{33}\sigma_{33}\sigma_{23})$$

$$C_{35} = C_{53} = \frac{9\, a_{55}}{2\, H_p\, \bar{\sigma}^2}(a_{13}\sigma_{11}\sigma_{13} + a_{23}\sigma_{22}\sigma_{13} + a_{33}\sigma_{33}\sigma_{13})$$

$$C_{36} = C_{63} = \frac{9\, a_{66}}{2\, H_p\, \bar{\sigma}^2}(a_{13}\sigma_{11}\sigma_{12} + a_{23}\sigma_{22}\sigma_{12} + a_{33}\sigma_{33}\sigma_{12})$$

$$C_{44} = \frac{1}{G_{23}} + \frac{9\, a_{44}^2\sigma_{23}^2}{H_p\, \bar{\sigma}^2}$$

$$C_{45} = C_{54} = \frac{9\, a_{44}a_{55}\sigma_{23}\sigma_{13}}{H_p\, \bar{\sigma}^2}$$

$$C_{46} = C_{64} = \frac{9\, a_{44}a_{66}\sigma_{23}\sigma_{12}}{H_p\, \bar{\sigma}^2}$$

$$C_{55} = \frac{1}{G_{31}} + \frac{9\, a_{55}^2\sigma_{13}^2}{H_p\, \bar{\sigma}^2}$$

$$C_{56} = C_{65} = \frac{9\, a_{55}a_{66}\sigma_{13}\sigma_{12}}{H_p\, \bar{\sigma}^2}$$

$$C_{66} = \frac{1}{G_{12}} + \frac{9\, a_{66}^2\, \sigma_{12}^2}{H_p\, \bar{\sigma}^2}$$

References

[1] Hahn, H.T. and S.W. Tsai, *Nonlinear elastic behaviour of unidirectional composite laminae.* Journal of Composite Materials, 1973. 7(1): p. 102-118.

[2] Sun, C.T. and J.L. Chen, *A simple flow rule for characterizing nonlinear behaviour of fibre composites.* Composites, 1989. 23(10): p. 1009-1020.

[3] Hill, R., *A theory of the yielding and plastic flow of anisotropic metals.* Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences, 1948. 193(1033): p. 281-297.

[4] Ogihara, S. and K. L. Reifsnider, *Characterization of nonlinear behavior in woven composite laminates.* Applied composite materials, 2002. 9(4): p. 249.

[5] Thiruppukuzhi, S.V. and C.T. Sun, *Testing and modeling high strain rate behavior of polymeric composites.* Composites Part B: Engineering, 1998. 29(5): p. 535-546.

[6] Weeks, C.A. and C.T. Sun, *Modeling non-linear rate-dependent behavior in fiber-reinforced composites.* Composites Science and Technology, 1998. 58(3-4): p. 603-611.

[7] Thiruppukuzhi, S.V. and C.T. Sun, *Models for the strain-rate-dependent behavior of polymer composites.* Composites Science and Technology, 2001. 61(1): p. 1-12.
[8] Tsai, J. and C.T. Sun, *Constitutive model for high strain rate response of polymeric composites.* Composites Science and Technology, 2002. 62(10-11): p. 1289-1297.
[9] Hufner, D.R. and M.L. Accorsi, *A progressive failure theory for woven polymer-based composites subjected to dynamic loading.* Composite Structures, 2009. 89(2): p. 177-185.
[10] Jones, R., *Mechanics of composite materials*. 1999: Taylor & Francis Inc.

Section 6
Detection and signal processing
(Special session organised by Prof. A. Kawalec)

This page intentionally left blank

The synthesis of a radar signal having nonlinear frequency modulation function

C. Lesnik, A. Kawalec & M. Szugajew
Institute of Radioelectronics, Military University of Technology, Poland

Abstract

A method of signal synthesis is presented. After theoretical introduction an algorithm of frequency modulated (FM) signal synthesis is presented. Simulation results made in Matlab are presented in the last chapter.
Keywords: NLFM, signal synthesis, autocorrelation function.

1 Introduction

This paper presents the problem of the synthesis of signals modulated in frequency with an autocorrelation function that implements an optimal approximation to a given autocorrelation function.

The output signal of the matched filter is proportional to the autocorrelation function of the expected signal. Because of that one would want to use a signal whose autocorrelation function $R(\tau)$ would be "similar" to certain "perfect" $R_{opt}(\tau)$ in the sense of a criterion that would provide a desirable property or properties. In this case it is the square criterion of similarity

$$f = \int_{-\infty}^{\infty} \left[\left| R_{opt}(\tau) \right| - \left| R(\tau) \right| \right]^2 d\tau . \tag{1}$$

In addition we are assuming, that the energy spectrum of the signal $s(t)$,

$$G(\omega) = 0 \quad \text{for} \quad |\omega| > \Omega \tag{2}$$

is non zero in a finite range of frequencies.

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110361

Realized autocorrelation function does not determine in explicit way signal $s(t)$, it can only determine the signals amplitude spectrum:

$$R(\tau)=\frac{1}{2\pi}\int_{-\infty}^{\infty}|G(\omega)|^2 e^{j\omega\tau}d\omega, \tag{3}$$

$$|G(\omega)|^2=\int_{-\infty}^{\infty}R(\tau)e^{-j\omega\tau}d\tau, \tag{4}$$

where $|G(\omega)|$ - amplitude spectrum.

From the above one can see that any signal with a given amplitude spectrum and a random phase spectrum can have the desired autocorrelation function. In the next paragraph the signal synthesis algorithm will be presented.

2 Signal synthesis algorithm

The signals synthesis task is to determine a signal x, that is an element of a set of frequency modulated signals *X*, on the basis of the set of signals *Y* with desired a property or properties of the autocorrelation function. In other words one should synthesize a signal $x_{opt}(t)$, which is closest to the signal $y_{opt}(t)$ in the meaning of the square criterion. The synthesis of the optimum signal $x_{opt}(t)$ is equivalent to the determination of the shortest distance $d(x,y)$, in the sense of the square criterion, between the *X* and *Y* sets

$$d_{\min}=\min_{x\in X} d(x,y). \tag{5}$$

The square criterion specifies a rule according to which to each pair of functions x and y distance $d(x,y)$ is assigned to each other. The distance $d(x,y)$ is often called a function space metric and should not be interpreted as the geometric distance. Signals belonging to this space are governed in terms of energy as mentioned earlier.

There are several ways to solve the problem of the signal synthesis i.e., finding a signal x with lowest value of distance $d(x,y)$. The problem was divided into two parts. The first part (first three blocks in fig. 1) was carried out only once during the initial stage of the algorithm and its purpose was to find the form of equation (iv) from fig. 1. The second part of the signal synthesis problem was solved, by building two separate algorithms (last two blocks in fig. 1) executed consecutively. In the first algorithm one chooses a facultative signal $x\in X$ and determines its best approximation on *Y* set of signals, [1, 2] (main program from fig. 1).

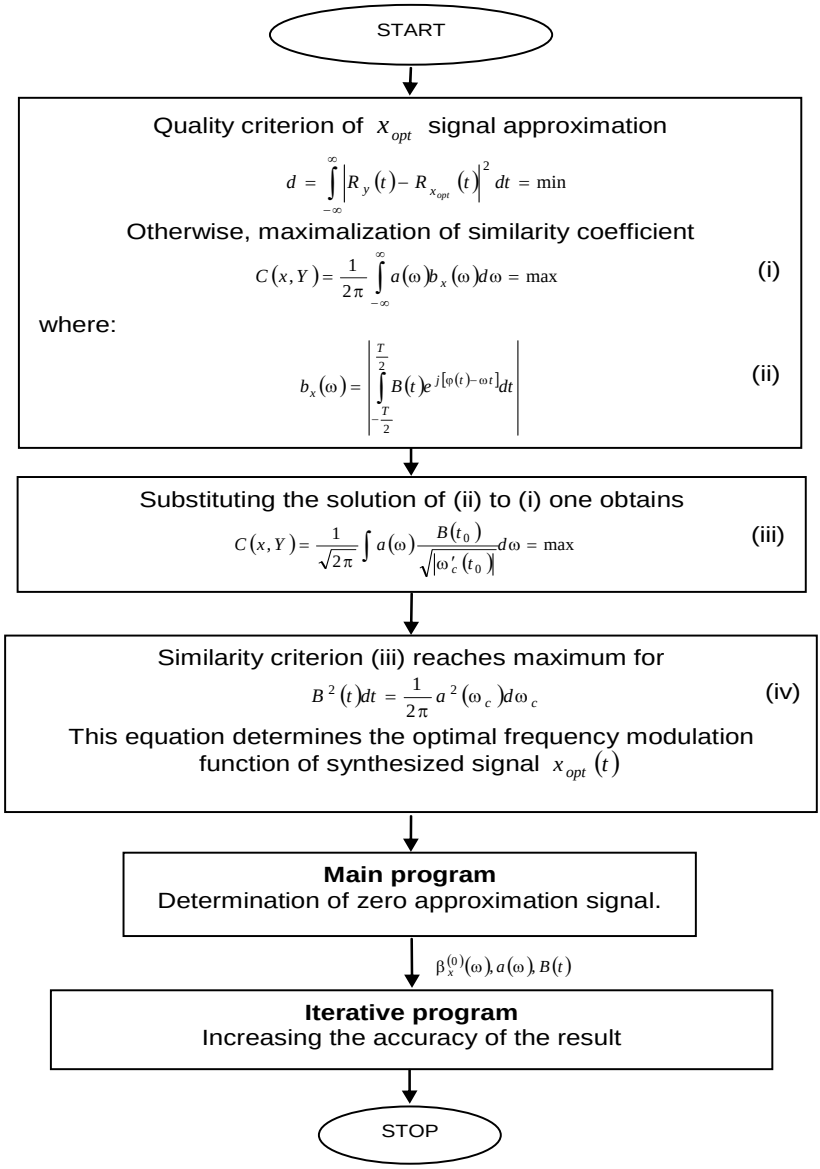


Figure 1: Signal synthesis algorithm.

The quality of approximation is characterized by a distance expressed by the following

$$d(x,Y) = \|x - y\|, \tag{6}$$

where $\|\cdot\|$ denotes the norm of the signal.

This distance corresponds to a specific signal y_0

$$y_0 = a(\omega)e^{-j\beta_{0x}}. \tag{7}$$

If the signal does not have the desired properties, the signal x is changed (moving inside X set), and the distance $d(x,y)$ is determined again, this task is carried out by iterative program from fig. 1. Successive distance values should form a decreasing sequence

$$d_1 > d_2 > d_3 > \ldots .$$

This operation is repeated until the minimum of eqn (6) is achieved. Both algorithms were implemented in Matlab using numerical methods.

The task of the signal synthesis is to synthesize the frequency modulation function $\omega_c(t)$ that implements the best approximation to the desired signal $y(t)$. Suppose that the elements of the set of possible signals *X* are given as

$$x(t) = B(t)\cdot e^{j\varphi(t)} \text{ for } |t| \le \frac{T}{2}, \tag{8}$$

where: T – duration of the pulse, $B(t)$– signal envelope, $\varphi(t)$– phase modulation function.

The spectrum of the signal $x(t)$ is given by

$$\tilde{X}(\omega) = b_x(\omega)\cdot e^{-j\beta_x(\omega)}, \tag{9}$$

where: $b_x(\omega)$ – amplitude spectrum, $\beta_x(\omega)$– phase spectrum.

In the process of the synthesis, it is assumed that the envelope of the signal ($B(t)$) is set, and the function $\varphi(t)$ is arbitrary. The frequency modulation function is expressed as follows

$$\omega_c(t) = \frac{d\varphi(t)}{dt} \tag{10}$$

Elements of *Y* have the following form

$$y(t) = A(t)e^{j\Phi(t)}, \tag{11}$$

and their spectra

$$\tilde{y}(\omega) = a(\omega)e^{j\alpha(\omega)}, \tag{12}$$

where:

$A(t)$ - envelope of the signal $y(t)$,
$\Phi(t)$ - phase function of the signal $y(t)$,
$\alpha(\omega)$ - phase spectrum of s the signal $\tilde{y}(\omega)$,
$a(\omega)$ - amplitude spectrum of the signal $\tilde{y}(\omega)$.

Signals $x(t) \in X$ differ only in the form of the phase modulation function $\varphi(t)$, but have a given envelope $a(\omega)$. With given functions $a(\omega)$ and $B(t)$ one synthesizes the phase modulation function $\varphi(t)$ (or $\beta_x(\omega)$), which minimizes the distance between the X and Y sets [3]. This means that one must determine the similarity coefficient:

$$C(x,Y) = \frac{1}{2\pi}\int_{-\infty}^{\infty} a(\omega)b_x(\omega)d\omega = \max , \tag{13}$$

$$b_x(\omega) = |\tilde{x}(\omega)| = \left|\int B(t)e^{j[\varphi(t)-\omega t]}dt\right| . \tag{14}$$

The optimal frequency modulation function, which maximizes the similarity coefficient (13), shall be determined on the basis of the differential equation

$$B^2(t)dt = \frac{1}{2\pi}a^2(\omega_c)d\omega_c . \tag{15}$$

After determining the frequency modulation function $\omega_c(t)$ from the above equation, in the next step the phase modulation function is determined as

$$\varphi(t) = \int \omega_c(t)dt . \tag{16}$$

In order to determine the phase spectrum of signal $x(t)$ one must resolve the integral

$$\tilde{X}(\omega) = \int_{-T/2}^{T/2} B(\omega)e^{j[\varphi(t)-\omega t]} , \tag{17}$$

Using the method of stationary phase [2, 4] one obtains a solution of the form

$$\beta_x(\omega) = -\left[\varphi(t) - \omega t_0 \pm \frac{\pi}{4}\right] , \tag{18}$$

where t_0 - point of stationary phase.

In the next step, from the above equation a phase spectrum $\beta_x(\omega)$ should be determined, which is then mapped to a specified amplitude spectrum $a(\omega)$, in order to obtain the signal $y(t)$ closest to the signal $x(t)$

$$\tilde{y}_0(\omega) = a(\omega)e^{j\beta_x(\omega)} . \tag{19}$$

The determined phase spectrum $\beta_x(\omega)$ is both the phase spectrum of the signal $x_{opt}(t)$ and the signal $y_0(t)$. The signal $y_0(t)$ is obtained by taking the

inverse Fourier transform from the signal $\tilde{y}_0(\omega)$. At this stage, namely after determination of the signal $x_{opt}(t)$, the main algorithm ends. The results of the main program are affected by errors. These errors are coming from the stationary phase method and numerical methods used in the Matlab program that implements the algorithm. In order to reduce them an iterative method was build and the signal $\tilde{y}_0(\omega)$ was used as the input signal for this method.

In the next paragraph the simulation results of both programs, the main one and the iterative one, are presented.

3 Simulation results

With given functions $a(\omega)$ and $B(t)$ the phase function $\varphi(t)$ (or $\beta_x(\omega)$) of the optimal signal $x_{opt}(t)$ is synthesized. This function minimizes the distance between the sets X and Y. On the basis of the shape of $x_{opt}(t)$ signal, having the optimal phase function $\varphi(t)$, the resulting autocorrelation function $R_{opt}(t)$ is determined.

In order to verify the correctness of the program's performance, the synthesis of a signal with a nonlinear frequency modulation (NLFM) was made. The signal has a bell shaped amplitude spectrum, given by

$$a(\omega) = \frac{\sqrt{\frac{2}{\Omega}}}{\sqrt{1+\frac{\omega^2}{\Omega^2}}}, -\infty < \omega < +\infty . \tag{20}$$

The envelope of synthesized signal was rectangular

$$B(t) = \begin{cases} \frac{1}{\sqrt{T}} & \text{for } -\frac{T}{2} \le \omega \le \frac{T}{2} \\ 0 & \text{for } |t| > \frac{T}{2} \end{cases}.$$

Signal parameters that were used during the simulation:

- Ω = 200,
- T = 4.

The goal of this simulation was to confirm the correctness and usefulness of the iterative procedure for the synthesis of the signal having a nonlinear frequency modulation function. On the basis of the given amplitude spectrum a signal with non-linear frequency modulation function was obtained (fig. 2).

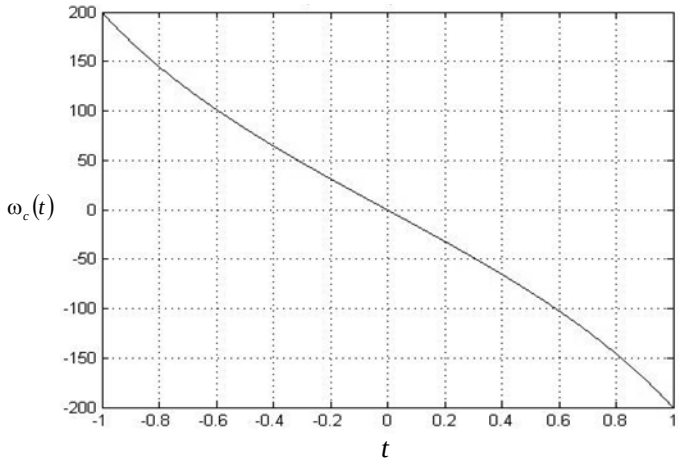


Figure 2: Frequency modulation function.

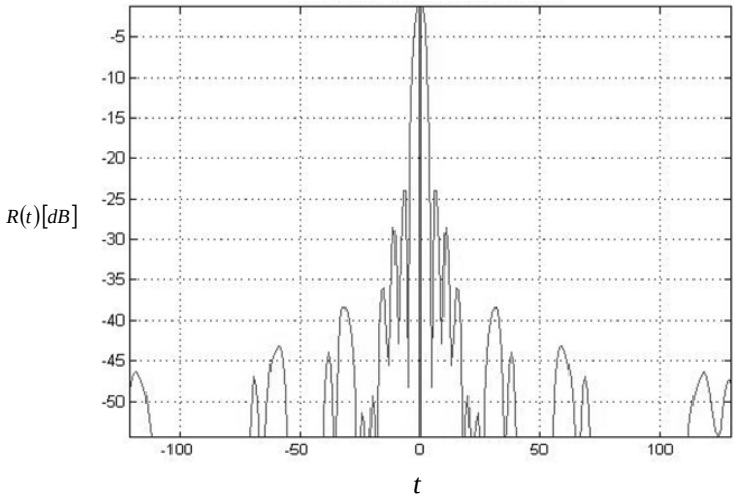


Figure 3: Autocorrelation function after one iteration ($\Omega = 200$).

In fig. 3 the obtained spectrum and autocorrelation function is presented.

As can be seen in figures 3 and 4, the result of the synthesis is not satisfactory. Although the level of the first side lobe of the autocorrelation functions is -24 dB, the amplitude spectrum does not have the bell shape. The amplitude spectrum is almost rectangular. Because of that the result is passed to

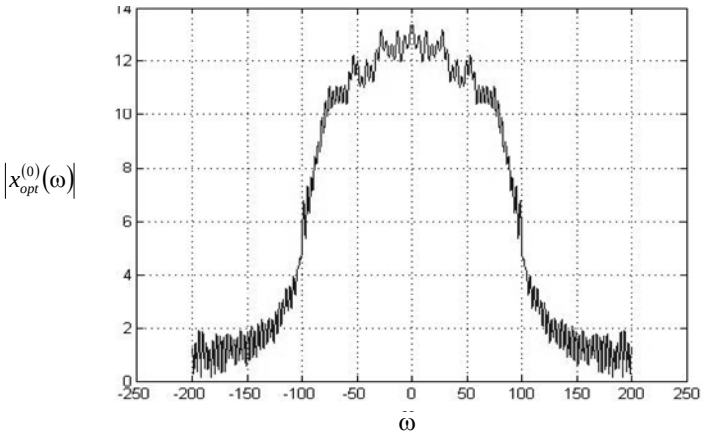


Figure 4: Amplitude spectrum after one iteration ($\Omega = 200$).

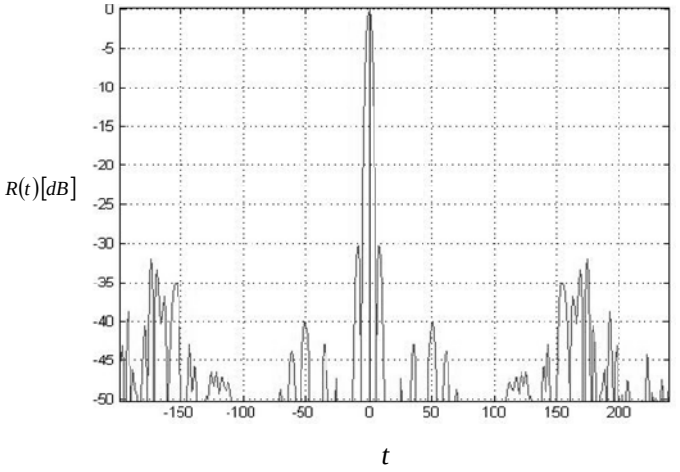


Figure 5: Autocorrelation function after thirty iterations ($\Omega = 200$).

the second (iterative) program. After the execution of thirty iterations significant improvement of the properties of the autocorrelation function can be seen (fig. 5 and fig. 6).

In fig. 5 and fig. 6 one can see the effectiveness of the iterative method. After thirty iterations the level of the first side lobe decreased by 7 dB (to –31 dB). In fig. 6 one can see how the shape of the power spectral density (PSD) function PSD of thirtieth iteration rise much slower than after the first iteration and in the same time both PSD-s overlap in $\omega = -50 \ldots 50$ region.

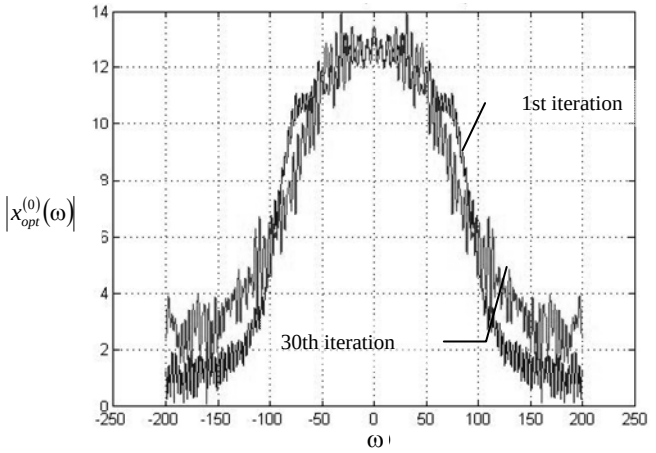


Figure 6: Power spectral density after first iteration and after thirty iterations ($\Omega = 200$).

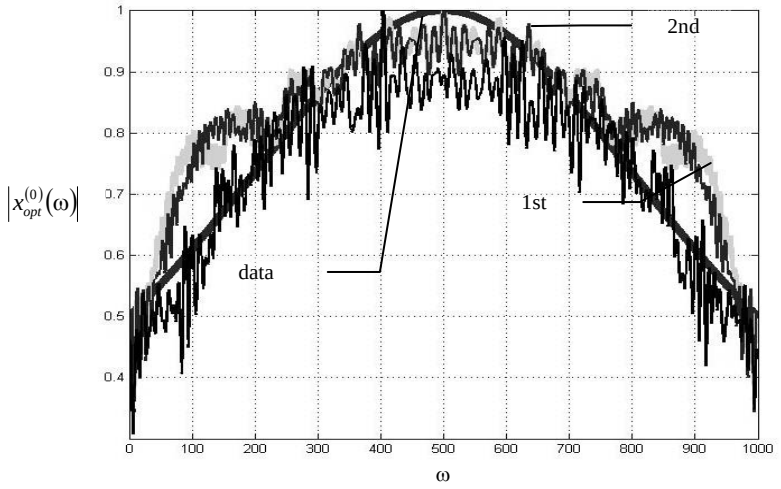


Figure 7: Comparison of PSD-s after the first, second and thirtieth iterations.

Fig. 7 shows how the iterative method changes the PSD function in a chosen number of iterations. The curve of the thirtieth iteration is filtered to improve the clarity of the figure.

The final autocorrelation function, after 150 iterations, and the PSD are shown in fig. 8 and fig. 9 respectively.

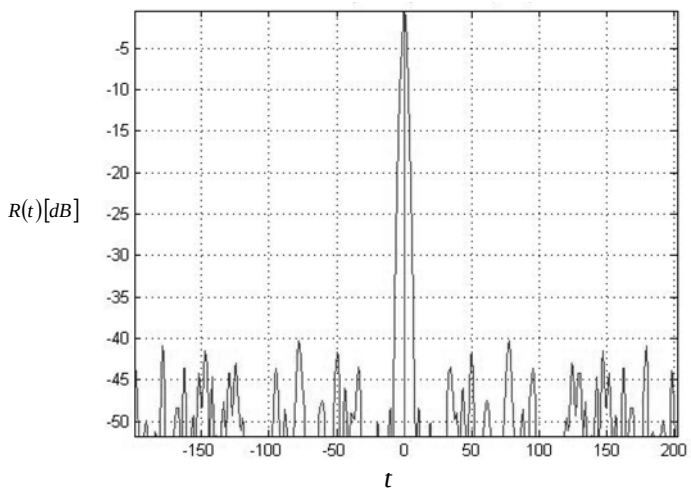


Figure 8: Autocorrelation function after 150 iterations ($\Omega = 200$).

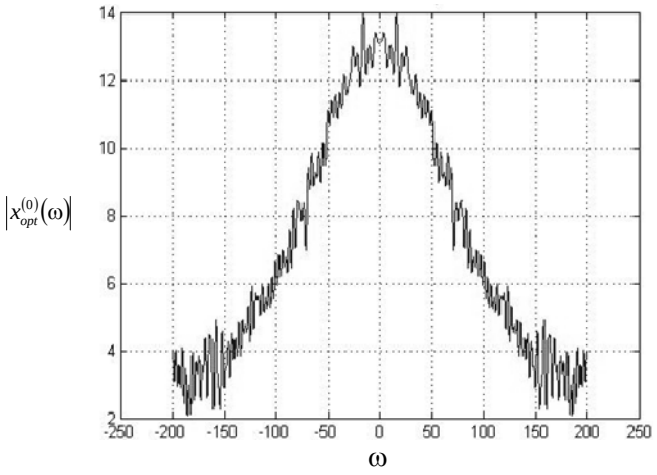


Figure 9: Power spectral density after 150 iterations ($\Omega = 200$).

In fig. 8 the resulting autocorrelation function was plotted. The side lobes are on equal level of -40 dB. The first side lobe is barely noticeable and its level is –48 dB. That is an improvement of 24 dB in comparison to the first iteration. Also the PSD plotted in fig. 8 is almost the ideal bell shaped. This proves the effectiveness of the iterative method. The noise that can be seen in fig. 9 comes from used numerical methods.

4 Conclusion

This paper discusses the key aspects of the signal synthesis needed for the selection of signals with a desired autocorrelation function, for example in radar technology. The results of previous theoretical studies and numerical results confirm the usefulness of the method discussed in the article from the standpoint of the signal optimisation. The iterative method reduces the errors introduced by the method of the stationary phase and the errors that are coming from used numerical methods. The presented method of the signal synthesis is very useful for cases where the desired autocorrelation function and the subsequent results of the calculations cannot be represented in a strict analytical form. Another key advantage of the presented algorithm is that the used numerical methods allow us to find the optimal solution, having at the entrance, only a discrete set of desired signals PSD points without any prior knowledge about the signal itself.

Acknowledgement

This work was supported by the Polish Ministry of Science and Higher Education from sources for science in the years 2009-2011 under project OR00007509.

References

[1] Sołowicz, J., Institute of Radioelectronics Technical Report, Warsaw, 2008, (in Polish).
[2] Vakman, D.E. & Sedleckij, R.M., Voprosy sinteza radiolokacionnych signalov, Sovietskoye Radio: Mosva, 1973 (in Russian).
[3] Kawalec, A., Lesnik, C., Solowicz, J., Sedek, E. & Luszczyk, M., Wybrane problemy kompresji i syntezy sygnałów radarowych. *Elektronika*, **3**, pp. 76-83, 2009 (in Polish).
[4] Cook, C.E. & Bernfeld, M., *Radar Signals: An Introduction to Theory and Application.* Artech House: Boston and London, 1993.

This page intentionally left blank

Synthetic aperture radar raw signal simulator for both pulsed and FM-CW modes

P. Serafin, C. Lesnik & A. Kawalec
Institute of Radioelectronics, Military University of Technology, Poland

Abstract

Simulated raw radar signals prove to be very useful at the first stages of testing radar signal processing algorithms and procedures. This is particularly true for synthetic aperture radar (SAR), where the costs of real signals acquisition are very high due to the costs of building the system as well as the costs of mission (air or space-borne). This paper describes a multifunctional SAR raw signal simulator that has been used for the verification of SAR image synthesis algorithms. Generated signals can be imitated for pulsed as well as FM-CW radars both in SLAR and squinted cases, it is also possible to choose between video and intermediate frequency signals. The simulator allows us to generate echo signals from stationary and moving targets. The user is able to differentiate the statistic properties of received echo signals for each target, thus allowing us to generate different types of reflecting surfaces. If present, a real raw SAR signal can be merged with a simulated one to produce more complicated scenarios. The paper presents results of the simulation of raw signals and their image after SAR processing.
Keywords: synthetic aperture radar, SAR signal simulator, FM-CW SAR.

1 Introduction

Airborne and space borne radar imaging allows us to collect the pictures of a terrain fragment independently of the time of the day, the weather or visibility conditions over the imaged scene. Unfortunately, due to relatively long waves used by the radar (compared to photography), it needs to employ large antennae in order to achieve the desired image resolution. This can make the system impossible to build onboard an airplane or a satellite. The solution to this problem is the technique of the synthetic aperture radar (SAR) that is able to

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110371

generate “virtual” antenna by scanning the observed area along a known route. After covering a desired distance gathered signals are processed as though they came from a real antenna giving a radar image with the resolution comparable with the wavelength of used signals.

Due to its advantages, SAR systems became very useful and widely applied reconnaissance tools and thanks to the rapid development of the computational techniques, their capabilities have grown steadily over past years.

The development of SAR signal processing algorithms requires the verification of their results with test signals. This can be difficult in the early stages of the project, especially during system feasibility verification, when no physical realisation of the radar that would provide data exists. Acquiring the signals from third party institutions if possible is rather expensive due to the expenses put into building the systems as well as conducting the mission.

Therefore one of the basic sources of radar signals used in SAR image synthesis algorithms testing is the computer simulations. Simulated signals are incomparably cheaper than real ones and their additional advantage is the ability to configure the observed scene and the radar mission arbitrarily, which allows us to test the algorithms in a wide range of parameters.

The main drawback of simulations, however, is the simplifications made in order to keep computational costs as low as possible, which sometimes render the results unreliable. The simulation design must be then optimised for maximum reality and minimum complexity.

This paper presents a simulated SAR raw signal generator developed for the verification SAR image synthesis algorithms in Military University of Technology (MUT) in Warsaw, Poland.

2 Principle of SAR

A SAR system is typically a radar installed onboard an airplane or a spaceship whose antenna system is illuminating the Earth’s surface perpendicularly or at some squint angle to the carrier’s route.

A typical SAR configuration for the airborne case is Side-looking Airborne Radar (SLAR) whose geometry is presented in fig. 1.

The carrier is moving along a set, well known and, preferably, straight, route. During the operation, as its carrier moves the radar emits the electromagnetic energy and receives echo signals reflected from the terrain objects. Received signals are then pre-processed and stored in system memory.

This signal pre-processing includes filtration and down-conversion to intermediate frequency or can be extended to baseband conversion and subsequently range compression, depending on the type of the algorithm used.

After the radar has covered the distance equal the antenna length L needed to achieve the desired azimuth resolution ρ_A, signals stored in memory are processed as if they were received by a real phased array of a length L.

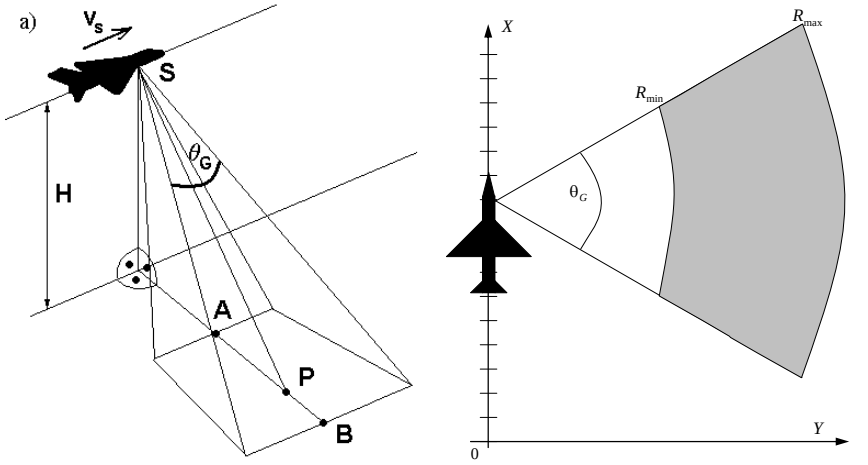


Figure 1: The geometry of a SAR system.

Let us assume the transmitted sounding signal to be of form

$$s_T(t) = A(t)e^{-j(\omega_0 t)}, \tag{1}$$

where: $A(t)$ is a modulation signal, ω_0 is the centre angular frequency.

Then the echo signal reflected from a single point target at the range R will take form [1]

$$s(t) = s_T(t-\tau)e^{-j\frac{4\pi}{\lambda}R}, \tag{2}$$

where: λ is the wavelength of sounding signal, τ is a two way electromagnetic wave propagation time between the radar antenna and the target given by

$$\tau = \frac{2R}{c}, \tag{3}$$

where c is the wave propagation velocity.

If one takes into account that the radar is moving and denote the time t associated with wave propagation as "fast-time" and the time T associated with the radar movement, as "slow-time" one will obtain the complete equation describing received signal

$$s(T,t) = s_T(t-\tau(T))e^{-j\frac{4\pi}{\lambda}R(T)}. \tag{4}$$

In the above equation the instantaneous (in the slow-time) range $R(T)$ is defined as follows

$$R(T)=\sqrt{\left(x_R(T)-x_T\right)^2+\left(y_R(T)-y_T\right)^2+\left(z_R(T)-z_T\right)^2}\,, \tag{5}$$

where: $\left(x_R(T), y_R(T), z_R(T)\right)$ is the instantaneous radar position and $\left(x_T, y_T, z_T\right)$ is the position of the target.

Assuming the coordination system from fig. 1 one can define the coordinates as:

$$x_R(T)=x_0+v_R T\,, \tag{6}$$

$$y_R(T)=const=0\,, \tag{7}$$

$$z_R(T)=const=H_R\,, \tag{8}$$

where: x_0 is the initial radar position along the X axis, v_R is the radar speed and H_R is the radar height above the ground. It is also assumed that the target lies on the ground meaning

$$z_T=0\,. \tag{9}$$

Therefore one can reduce eqn (5) to

$$R(T)=\sqrt{\left(x_0+v_R T-x_T\right)^2+y_T^2+H_R^2} \tag{10}$$

and then substitute the result to eqn (4) which takes then the following form

$$s(T,t)=s_T\left(t-\frac{2R(T)}{c}\right)e^{-j\frac{4\pi}{\lambda}\sqrt{\left(x_0+v_R T-x_T\right)^2+y_T^2+H_R^2}}\,. \tag{11}$$

If the signal is sampled with sampling frequency f_s and define the range resolution cell dimension as

$$dR=\frac{c}{2f_s}\,, \tag{12}$$

one can obtain

$$s(m,l)=s_T\left(\left\lfloor l-\frac{R(T)}{dR}\right\rfloor\right)e^{-j\frac{4\pi}{\lambda}\sqrt{\left(x_0+ld-x_T\right)^2+y_T^2+H_R^2}}\,, \tag{13}$$

where: m is a number of the sounding impulse emitted by the radar with the Pulse Repetition Frequency (PRF), l is a number of range cell and d is distance between two consecutive sounding positions of the radar

$$d = \frac{v_R}{PRF}. \tag{14}$$

In order to obtain a high resolution radar image one should compress the signal in the range (fast-time) and azimuth (slow-time) domains.

The range compression is performed by a filter matched to the form of the sounding signal s_T and can be done either before the azimuth compression or after it. Some of the SAR image synthesis algorithms combine those two compressions in one block. They will be, however, considered here as separate processing steps.

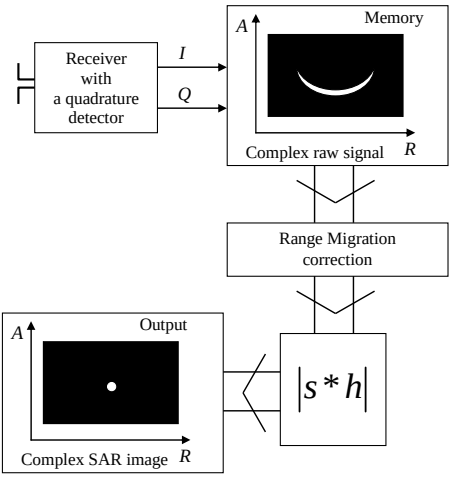


Figure 2: Structure of the TDC SAR processor.

It will be assumed that the obtained raw SAR signal is already range-compressed.

The azimuth compression is in fact matched filtering [1, 2], where filter impulse response is so called “azimuth chirp” being actually the complex conjugate to the exponential form in eqn (13).

Therefore the operation of azimuth compression of the SAR image would be a convolution of the signal with the azimuth chirp. However simultaneously with the changes of signal phase due to the changes of range the signal position in system memory changes as well. This phenomenon, called Range Migration (RM), if not compensated for seriously decreases maximum achievable synthetic aperture length and consequently the maximum image resolution. Range Migration correction procedure consists of computing the actual number of range

cell where the signal for a given image pixel for each element of the synthetic aperture.

Therefore the structure of the SAR processor (shown in fig. 2) consists the above mentioned convolution block following the RM correction step.

The described algorithm is called Time Domain Correlation (TDC) and although it is computationally very expensive, it is also the most accurate due to the lack of any simplifying assumptions. This makes this algorithm a useful reference tool for verification of other SAR algorithms.

3 The simulation strategy

The simulated raw SAR signal generator developed in MUT was designed to evaluate the accuracy and efficiency of SAR algorithms implemented in an FPGA structure.

The generator can provide the single- or multi-channel raw SAR signals reflected from user-defined objects placed in the observed space. Objects are flat and lay on the Earth surface. They can, however, move in the X and Y directions.

Each object consists of a number of elementary scatterers distributed within its boundaries.

User can define the shape (rectangle, ellipse or line) of the object, its size, the number of scatterers within the object and the v_x and v_y velocity components.

There are three types of objects in the program:

1) with randomly distributed scatterers that have defined parameters such as minimal and maximal values of the Radar Cross Section (RCS) over the object, variance of RCS and phase fluctuations of all scatterers from pulse to pulse; additionally each scatterer can have omnidirectional scattering characteristic in the azimuth and elevation domain or it can have an eight shaped one in azimuth domain; in the latter case user can also define the range of angles for azimuthal position of the scatterers; the probability density function (PDF) for RCS fluctuations is assumed to be Gaussian, whereas the phase distribution is uniform and those fluctuations are uncorrelated,
2) with uniformly distributed elementary reflecting surfaces; the statistical properties of such object are described by PDF and autocorrelation function (AF) of those surfaces,
3) jamming sources emitting white Gaussian noise with an omnidirectional characteristic.

The user has also to define all the radar parameters such as number of receiving channels $q_{\max}$, distance between the antenna elements d_a, centre frequency f_0 and wavelength λ, pulse repetition frequency PRF, LFM deviation Δf, pulse duration T_i, sampling frequency f_s, intermediate frequency f_{IM} and the emitted peak power. Also the carrier parameters such as

height, velocity, observation distance or initial coordinates need to be established.

Radar can have a number of receiving channels whose antennae are uniformly distributed along the carrier movement direction.

The sounding signal in the simulated SAR system can be either pulsed or continuous wave with linear frequency modulation (LFM-CW), the raw signal after simulation will differ in each case. For the pulsed radar it is either video or IM signal with or without range compression, and for the LFM-CW case it is the Discrete Fourier Transform (DFT) of dechirped (down-converted) received signal. Both types of signal are stored in a text or a binary file as complex samples arranged in range lines (signals received during one sounding).

The user can also define the standard deviation of the receiver thermal noise, which will have added to the signal at the end of the simulation.

4 The simulation algorithm

The echo signal reflected from each element is computed accordingly to the following algorithm

1) For each m-th position of radar (for each sounding) the distance $R_{m,n}$ between the radar antenna and the n-th scatterer is computed accordingly to the following equation

$$R_{m,n} = \sqrt{\left(x_{Rm} - x_{Tm,n}\right)^2 + \left(y_{Rm} - y_{Tm,n}\right)^2 + \left(H_R - h_{Tm,n}\right)^2}\,, \tag{15}$$

where: x_{Rm} and y_{Rm} are the radar position coordinates in the m-th sounding, and $x_{Tm,n}$ and $y_{Tm,n}$ are the coordinates of the n-th scatterer in the m-th sounding, H_R is the height of the carrier above the ground.

If the position of no. 0 antenna element would be assumed as the position of radar x_{Rm}, then the position of the q-th antenna element can be defined as

$$x_{aqm} = x_{Rm} - qd_a\,. \tag{16}$$

As it was mentioned all the objects are flat and are placed at the height equal

$$h_{Tm} = h_T = 0\,. \tag{17}$$

Considering the above and the radar movement the statement for the distance between n-th scatterer and q-th antenna element in the m-th sounding can be rewritten as follows

$$R_{m,n,q} = \sqrt{\left(x_{R0} - qd_a + md - x_{Tm,n}\right)^2 + \left(y_{Rm} - y_{Tm,n}\right)^2 + H_R^2}\,. \tag{18}$$

As in accepted model the radar Y coordinate y_{Rm} is constant and equal to zero, and the objects can be moving with the velocity having components v_{Tx} and v_{Ty} the eqn (18) should take the following form

$$R_{m,n,q} = \sqrt{\left(x_{R0} - qd_a + md - \left(x_{Tn0} + \frac{mv_{Tnx}}{PRF} \right) \right)^2 + \left(\frac{mv_{Tny}}{PRF} - y_{Tn0} \right)^2 + H_R^2} \,, \quad (19)$$

where: x_{Tn0} and y_{Tn0} are the coordinates of the n-th scatterer.

2) The received echo signals reflected from the simulated objects are created either as impulses with linear frequency modulation (LFM) or the LFM-CW signals after dechirping. The pulsed signals can have zero central frequency (video signals) or else their spectrum can be centred around a non-zero intermediate frequency f_{IM}. The video signals are assumed to be after quadrature down-conversion and therefore are stored as complex samples whereas signals with nonzero f_{IM} are real with imaginary part equal to zero.

3) After having computed the distance $R_{m,n,q}$, voltage samples in receiving channel are determined

$$U_{m,n,q}(lt_s) = A_{m,n,q} s_T(lt_s - t_{m,n,q}) e^{-j\frac{4\pi R_{m,n,q}}{\lambda}} \,, \quad (20)$$

where: $s_T(t)$ is the function describing the transmitted sounding signal, $t_{m,n,q}$ is the time of the two way propagation of the electromagnetic wave from the q-th antenna element to the n-th scatterer and back to antenna in the m-th sounding and is equal

$$t_{m,n,q} = \frac{2R_{m,n,q}}{c} \,, \quad (21)$$

$A_{m,n,q}$ is the amplitude of the echo signal dependant on scatterer reflecting characteristic $\sigma_{m,n,q}$ (RCS), antenna characteristic G_A, range and the transmitted power P_N and is equal to [2]

$$A_{m,n,q} = \frac{G_A^2 \lambda}{R_{m,n,q}^2} \sqrt{\frac{P_N \sigma_{m,n,q}}{(4\pi)^3}} \,. \quad (22)$$

The form of the transmitted sounding signal $s_T(t)$ is dependant on radar parameters set by the user. In particular if the value of f_{IM} is non-zero this signal is generated as a series of real valued samples having the form of

$$s_T(l) = \cos\left(2\pi\left(f_{IM} - \frac{\Delta f}{2} + \frac{\Delta f}{2T_i} lt_s\right) lt_s\right). \tag{23}$$

In the case of the baseband (i.e. video) signals it takes the form of a series of complex valued samples

$$s_T(l) = \cos\left(2\pi\left(\frac{\Delta f}{2T_i} lt_s - \frac{\Delta f}{2}\right) lt_s\right) + j\sin\left(2\pi\left(\frac{\Delta f}{2T_i} lt_s - \frac{\Delta f}{2}\right) lt_s\right). \tag{24}$$

If the scatterer under consideration is not fluctuating and it has an omnidirectional reflecting characteristic the $\sigma_{m,n,q}$ parameter is constant. However for scatterers with eight-shaped reflecting characteristic a $\theta_{m,n,q}$ angle between radar-target line and the azimuthal position of the scatterers characteristic is established and then on its basis the $\sigma_{m,n}(\theta_{m,n})$ is computed.

The signal can be range compressed by matched filtering in a FIR filter whose coefficients are the samples of the sounding pulse replica. The filtering itself if done in the form of so called fast convolution employing the Fast Fourier Transform (FFT) on both signal and the impulse replica, then multiplication of the spectra, and finally the Inverse Fast Fourier Transform (IFFT).

In the case of LFM-CW radar the received signal is computed as the result of downconversion with the transmitted signal and for a single scatterer it takes the form of real valued series of samples [3]

$$U_{m,n,q}(lt_s) = A_{m,n,q}\cos\left(2\pi(f_0\tau + \mu)\tau lt_s - \frac{\mu(lt_s)^2}{2}\right). \tag{25}$$

After computing the values of the signal samples and their position in the system memory the signals are added to the respective cells (their values are summed with those already existing it the memory).

If simulated radar is working as LFM-CW the FFT in the range domain is performed in order to obtain the range compression.

The program has the functionality of synthesising the SAR image using described earlier the TDC algorithm, which can be used as the tool for verification of the generated raw signals. It is also possible to merge new situation with earlier generated signals or even with real radar signals if only all of the needed parameters are known.

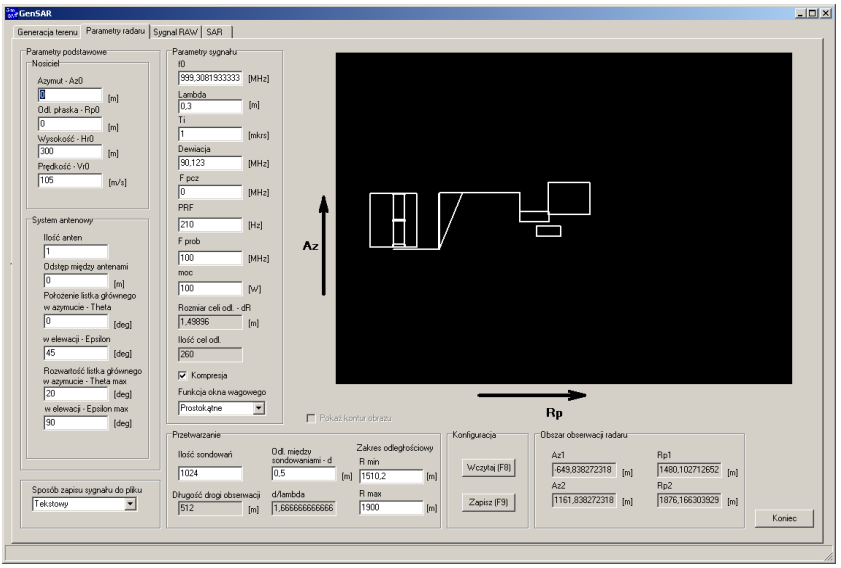


Figure 3: Program window with radar parameters for the simulation tab.

Fig. 3 presents the program window with radars parameters tab showing also the distribution of the simulated objects. In fig. 4 the tab with generated raw SAR signal is presented and in fig. 5 the tab with SAR image computed with the TDC SAR image synthesis algorithm.

Figure 4: Program window with SAR raw signal generated.

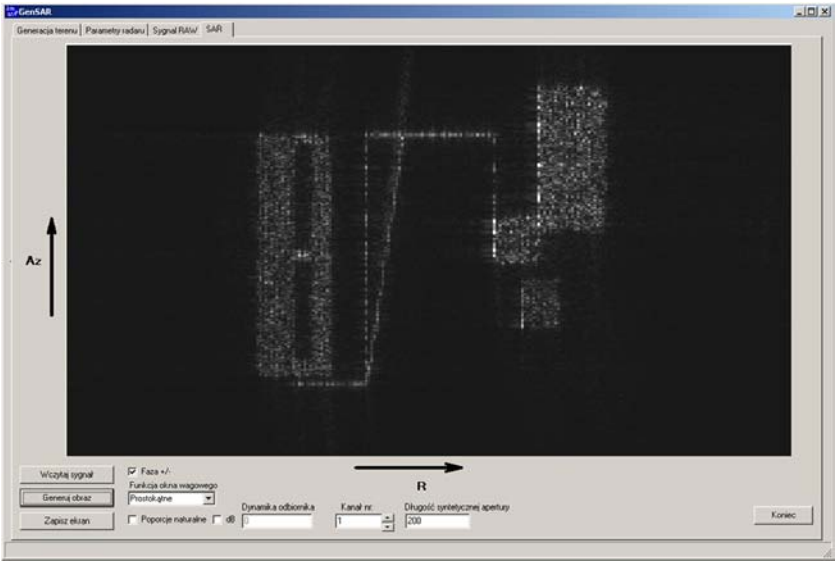


Figure 5: Program window with SAR image synthesised.

5 Conclusions

A Synthetic Aperture Radar raw signal generator developed in Military University of Technology in Warsaw, Poland was descried in this paper. Algorithm of the generation of echo signals from both stationary and moving targets was presented as were the simulation results.

Acknowledgement

This work was supported by the Polish Ministry of Science and Higher Education from sources for science in the years 2009-2011 under project OR00007509.

References

[1] Cumming, I.G. & Wong, F.H., *Digital Processing of Synthetic Aperture Radar Data. Algorithms and Implementation*, Artech House: London and Boston, 2005.
[2] Skolnik, *M., Radar Handbook, Second Edition,* McGraw-Hill Book Company: New York, 1990.
[3] Stove, A.G., Linear FMCW radar techniques. *IEE Proceedings F Radar and Signal Processing*, **139(5)**, pp. 343-350, 1992.

This page intentionally left blank

A real time SAR processor implementation with FPGA

C. Lesnik, A. Kawalec & P. Serafin
Institute of Radioelectronics, Military University of Technology, Poland

Abstract

Great numerical complexity is a characteristic of synthetic aperture radar (SAR) image synthesis algorithms that poses a particularly serious problem for real-time application. Advances in the operating speed and density of the field programmable gate arrays (FPGA) have allowed many high-end signal processing applications to be solved in commercially available hardware. A real-time SAR image processor was designed and implemented with the commercial off the shelf (COTS) hardware. The hardware was based on the Xilinx Virtex 5 FPGA devices. Under the assumption of squinted SAR geometry and range migration effect present the SAR image synthesis algorithm was developed and implemented. The results of the processor tests conducted with simulated and real raw SAR signals are presented in the paper.
Keywords: SAR, radar, real-time processing, FPGA, COTS.

1 Introduction

Airborne radar systems constitute the essential part of radio-electronic terrain imaging and recognition systems. Their primary advantage is the insensitivity to the time of the day or atmospheric conditions. However the images obtained by radar characterises much lower than in the photographic case resolution.

In the radar imaging two resolutions can be distinguished: the range resolution called also the fast-time resolution and the azimuth or the slow-time resolution.

If the task of achieving the high range resolution is relatively easy - it is realized by using sounding signals with internal frequency or phase modulation or manipulation, the high azimuth resolution is much harder to achieve. It

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110381

requires specific and refined method of image synthesis called Synthetic Aperture Radar (SAR).

This technique proved to be very useful in both civil and military applications. One of the many applications is the employment of a SAR sensor onboard Unmanned Aerial Vehicle (UAV) for an immediate terrain monitoring for industrial, military, antiterrorist and crisis management purposes. The specific nature of those areas of the practical use of a radar system requires the image synthesis to be performed in the real-time.

The process of the radar image synthesis can be reduced to the two-dimensional matched filtration problem both in range and azimuth domains. There are a variety of different algorithms implementing this function such as Range-Doppler (RDA), Chirp Scaling (CSA), Omega-K, SPECAN or Time Domain Correlation (TDC) Algorithms. The last one is also called "the naive" algorithm. This one was described for example by Cumming and Wong [1] and Wang [2]. Those algorithms differ in accuracy and computational complexity. In this work the authors made an attempt at a real-time hardware implementation of the time domain matched filtration algorithm, which is often used as a reference algorithm due to its accuracy, and yet the one that has the greatest requirements for computational power among the SAR algorithms.

2 The algorithm description

For the considerations as well as for the practical implementation an assumption of strip-map mode of the SAR sensor was made. The geometry of the system during the observation of a single point object is presented in fig. 1.

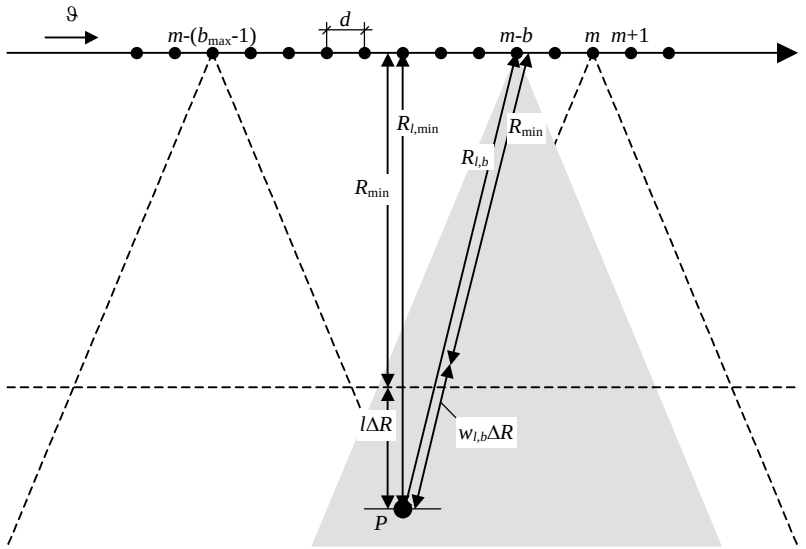


Figure 1: Side-looking SAR in stripmap mode with single point target.

The radars carrier (a plane or a UAV) is moving with a constant velocity ϑ along a straight line at a constant height H. The distance d between the two consecutive radar positions in which the radar emits the sounding impulses with the pulse repetition frequency PRF is given by

$$d = \frac{\vartheta}{PRF}. \tag{1}$$

The raw SAR signal $s(t)$ received in consecutive observations is sampled and denoted as a series of samples $s(l)$, where l - is a number of a sample. The size of the range cell ΔR, resulting from the sampling operation is equal to

$$\Delta R = \frac{ct_s}{2} = \frac{c}{2f_s}, \tag{2}$$

where: c - speed of light, t_s - sampling time interval and f_s - sampling frequency.

A the first stage of raw SAR signal processing the matched filtration in range is performed, as a result of which we obtain for each m-th sounding and for each l-th range cell discrete signal $s_R(m,l)$. As a reference signal for the matched filtration in range the sounding impulse replica is used.

The second and the main stage of SAR signal processing is matched filtration in azimuth, implemented in time domain. This operation is also called SAR image synthesis. It is performed in a synthesis window having azimuthal size equal to $b_{\max}$, expressed in the number of sounding periods. The size of the synthesis window is also the length of the synthetic aperture. For each m-th sounding period the length of synthesis window is defined in the limits from $m-(b_{\max}-1)$ to m the sounding.

The nature of the pulsed radar assumes some minimal range of reception $R_{\min}$.

For an exemplary point object P (fig. 1), laying in the l-th range cell the minimal distance radar - object $R_{l,\min}$ can be defined as

$$R_{l,\min} = R_{\min} + l\Delta R. \tag{3}$$

The distance between the radar and the point object in an arbitrary b-th sounding period inside the synthesis window is equal to

$$R_{l,b} = \sqrt{R_{l,\min}^2 + \left[\left(\frac{b_{\max}-1}{2} - b\right)d\right]^2}. \tag{4}$$

Due to the changes the radar-object distance taking place during the flight the echo signal form the point object P lying at the distance $R_{l,\min}$ from the radar's route, in an arbitrary b-th sounding period will fall into the range cell having number

$$w_{l,b} = \left\lfloor \frac{R_{l,b}}{\Delta R} \right\rfloor, \tag{5}$$

where symbol $\lfloor . \rfloor$ denotes the maximal integer number not greater than the argument.

The effect of the change of the signal position in radar memory as a function of the number of sounding period inside of the synthesis window (number of the synthetic aperture element) is referred to as Range Migration (RM).

The main effect of the change of this distance, however, is an additional, dependant on the number of the synthetic aperture element, phase shift $\Delta\phi_{l,b}$, equal to

$$\Delta\phi_{l,b} = 2k\left(R_{l,b} - R_{l,\min}\right), \tag{6}$$

where: $k = 2\pi/\lambda$, λ - wavelength of the sounding signal.

Finally, on the basis of eqns. (5) and (6), for each m-th sounding period, by integration in azimuth domain, a signal for a single m-th range line of the SAR image $s_{RA}(m,l)$ is determined. Each l-th range cell in this line take the complex value equal to

$$s_{RA}(m,l) = \sum_{b=0}^{b_{\max}-1} s_R\left(m-b, w_{l,b}\right)\exp\left(-j\Delta\phi_{l,b}\right). \tag{7}$$

Eqn. (7) constitutes the basis of the hardware implementation of the SAR image synthesis algorithm.

3 The application description

The considerations presented above indicate very large requirements for computational power of the hardware platform realising the above algorithm. This task is achievable by the FPGA structures. Modern FPGA structures are not only reconfigurable logic cells matrices, but also integrated in one chip hundreds of computing blocks such as CMAC (complex multiplier accumulator), and integrated one or more DSP (Digital Signal Processors), e.g. PowerPC. Moreover the resultant capacity of the integrated RAM blocks still growing. Those components aid the computations and help to perform the DSP functions. If we take into account the capability of the FPGA structures of performing parallel signal processing, the real-time hardware implementation of SAR algorithms becomes quite feasible.

Due to the experimental nature of the project, the decision was made on the purchase of the COTS (Commercial Off-The-Shelf) FPGA modules and the dedicated software enabling development of a specialised signal processing system with the use of PC.

Hardware platform was build on the basis of Sundance DSP/FPGA functional modules utilising Xilinx FPGA Virtex-5 SX95T chips and Texas Instruments C6455 DSP. The following module types were used:

1. SMT362 – 1 piece; DSP/FPGA module containing 2 C6455 digital signal processors and Virtex-4 FX FPGA,
2. SMT351T – 2 pieces.; a module containing Virtex-5 SX95T FPGA and (external to FPGA chip) 2 GB of DDR2SDRAM memory.

These modules are designed to mount on a SMT310Q motherboard, made as a long PCI card.

The SMT362 module was used for organisational tasks in the SAR application. Mainly it was used to load the application code to the FPGA modules and as an interface between the code executed in the FPGA modules and the base PC (host).

The SMT351T were used as the main hardware resources for algorithm implementation - processors. One of them was designed for the range compression and the other one for the azimuth compression.

Fully mounted hardware platform is presented in fig. 2. This one was described in [3–5].

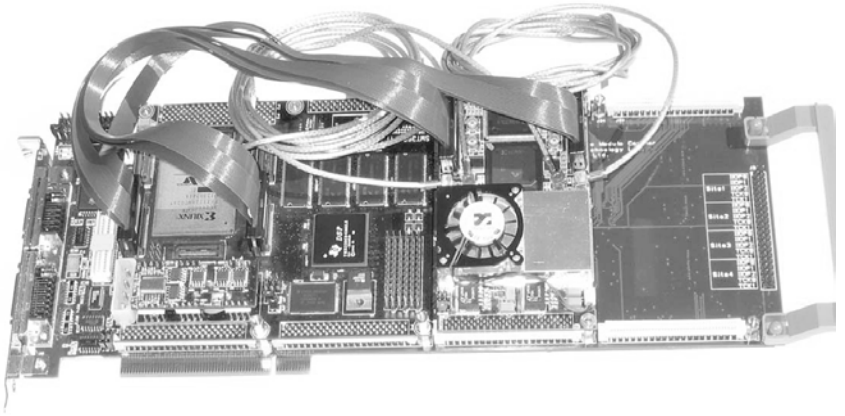


Figure 2: Hardware platform for real-time SAR image synthesis algorithm implementation.

The platform used for the project requires four packets of the utility software:

1. the Sundance hardware dedicated packet, containing mainly the drivers,
2. the FPGA software development and diagnostic packet,

3. the Code Composer Studio packet designed for the development and diagnostic of the DSP C6000 series processors software,
4. the 3L Diamond packet, aiding the development of applications combining the C written software for the DSP and VHDL written software for the FPGA.

4 Research results

The correctness of the hardware implementation of the SAR signal processing module has been verified with a simulated raw SAR echo signal of an isolated single point object. Signal was created in simulated raw SAR signal generator described by Serafin [6]. As the sounding signal a Linear Frequency Modulated (LFM) pulse was used. Such form of the testing signal simplified the application diagnostic and debugging process.

Figure 3a presents the raw SAR signal echo of a single point object. The RM effect is clearly visible. Moreover the effects of the real antenna sidelobe reception are visible at the edges of the image.

In fig. 3b, the matched filtration effect of the above raw SAR signal in the range domain is presented. The matched filtration operation was carried without the weighting function, therefore relatively high range sidelobes are visible.

In fig. 3c, the result of SAR image synthesis is presented. As it would be expected, a single point in the image is visible, but also the range and the azimuth sidelobes are present.

Another testing signal used for the testing contained echo signals of some complex objects having forms of flat geometrical figures consisting of a large number of elementary point objects. The raw form of this signal is presented in fig. 4a. The testing signal after the range compression is presented in fig. 4b. Figure 4c presents the SAR image synthesised from the simulated raw signal.

In order to test the accuracy and the quality of the hardware implementation of the SAR echo signal processing algorithm, besides the simulated signals, real measure data from SAR/GMTI sensor AER-II was used. This data has been obtained by the courtesy of the Director of the Fraunhofer Institute for High Frequency Physics and Radar Techniques FHR, Germany.

Figure 5a presents the real raw SAR signal of an exemplary fragment of a land infrastructure. Figure 5b presents the same signal after the range compression.

In fig. 5c, the real SAR image of an exemplary terrain fragment picturing the intersection of a highway and a local road. The next picture (fig. 5d) presents the same intersection, but in the process of image synthesis the squint angle of the main beam of the radar antenna was taken into account. This allowed for a better resolution of fine details in the image.

The results obtained during the algorithm verification are consistent with our expectations i.e. the developed hardware implementation generates the images of a quality comparable to the ones obtained with the application written in C language and ran on a PC in floating point format. The main difference between the two applications lies in the time of computations.

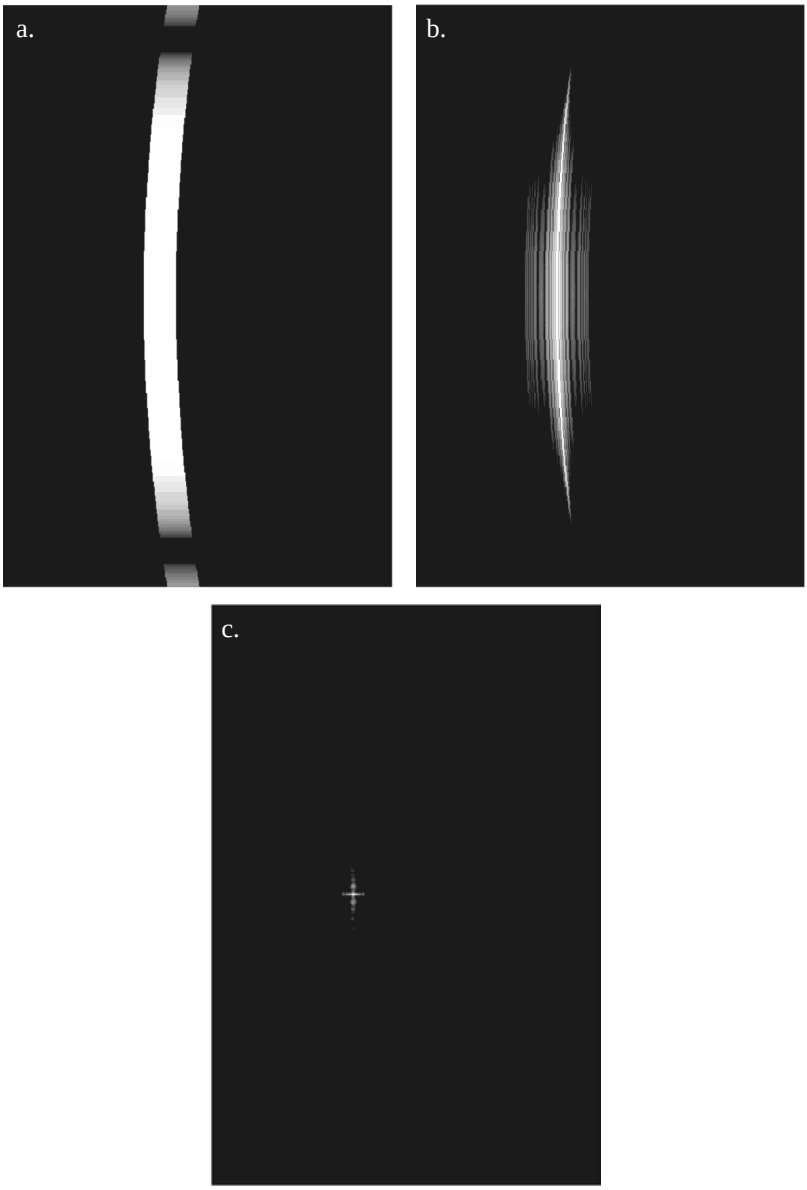


Figure 3: The raw SAR signal echo of a single point object (a), signal after range compression (b) and SAR image of a simulated single point object (c).

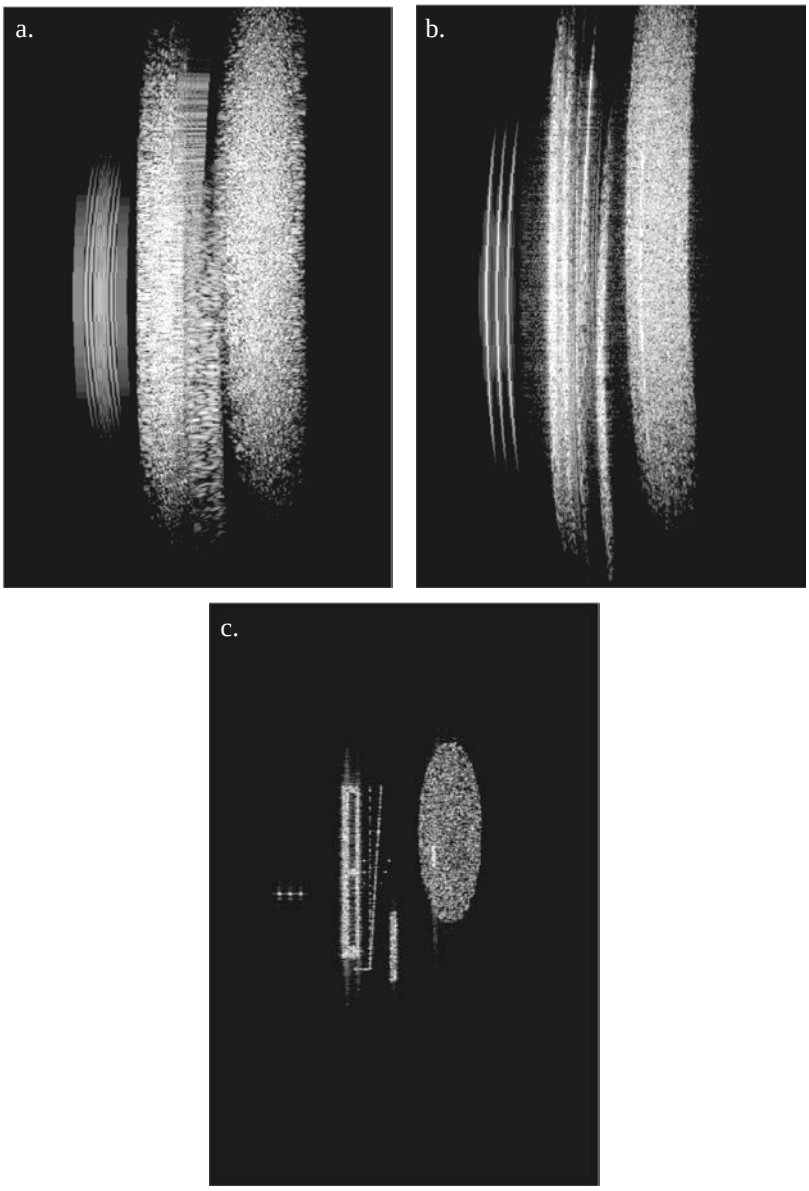


Figure 4: Simulated raw SAR echo signal of a set of complex objects (a), signal after range compression (b) and SAR image of a set of simulated complex objects (c).

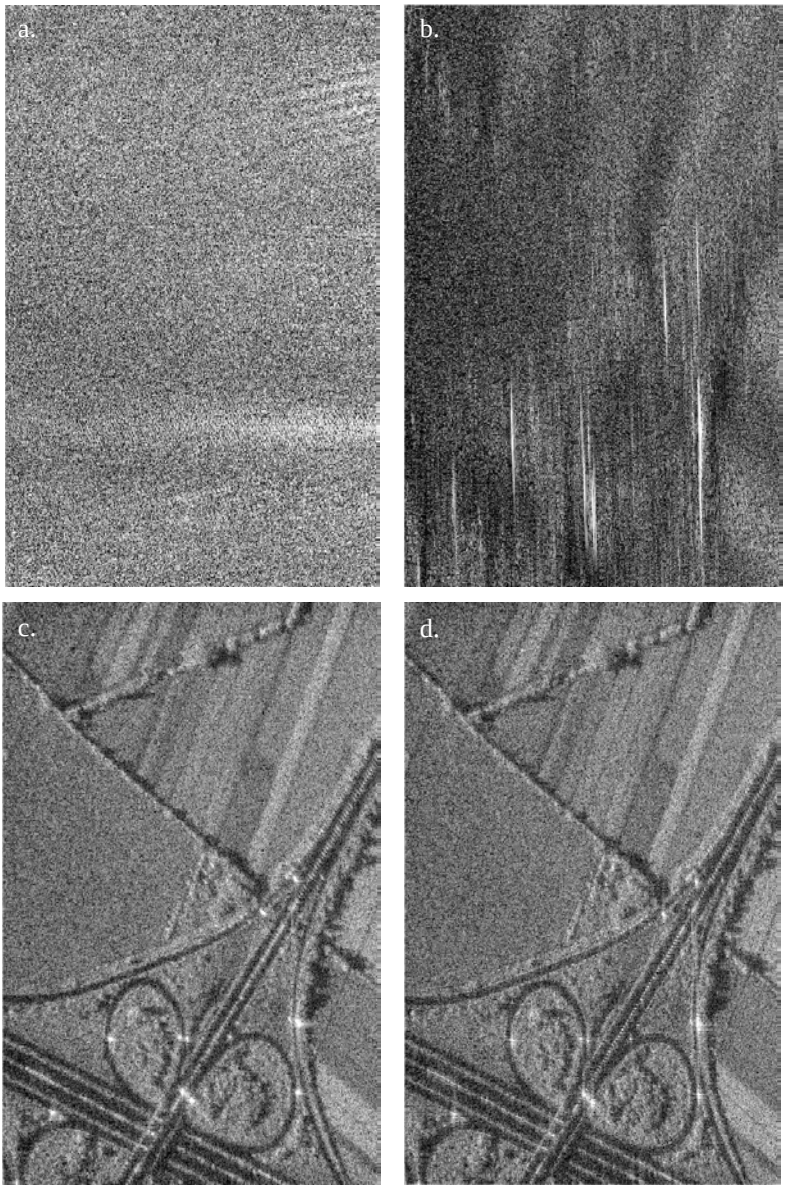


Figure 5: Real raw SAR echo signal (a), real raw SAR echo signal after range compression (b), real SAR image without taking the squint angle into account (c) and real SAR image with a proper value of the squint angle applied during the processing (d) (courtesy of the Director of the Fraunhofer FHR).

5 Conclusions

The hardware implemented in FPGA modules SAR image synthesis algorithm proved to be able to generate SAR image of a width of about 0.5 km in the real time, with the PRF of an order of a few hundreds of Hertz. Those values are acceptable for the SAR sensors mounted onboard UAV's. We should, however, mention that the implemented algorithm is very computationally demanding, and its current implementation is the first one without any optimisations.

Despite very high complexity of the implemented algorithm, the report from the resultant code generation indicates on a relatively low degree of the logical FPGA resources occupied by the application (about 30%), with the exception of the block memory (RAMB) whose usage exceeds 90%.

Presented research results of the hardware implementation of the SAR image synthesis module confirmed the feasibility of single FPGA implementation of the algorithm. This leads to creation of compact and low energy consuming applications working in the real time, especially attractive for the UAV applications.

Acknowledgement

This work was supported by the Polish Ministry of Science and Higher Education from sources for science in the years 2009-2011 under project OR00006909.

References

[1] Cumming, I.G. & Wong, F.H., *Digital Processing of Synthetic Aperture Radar Data. Algorithms and Implementation*, Artech House: London and Boston, 2005.
[2] Wang, Bu-Chin, *Digital Signal Processing Techniques and Applications in Radar Image Processing*, John Wiley & Sons: Hoboken, USA, 2008.
[3] SMT310Q User Manual, version 2.1, Sundance Multiprocessor Technology Ltd. 2005.
[4] SMT351T User Guide, version 1.0.6, Sundance Multiprocessor Technology Ltd. 2008.
[5] User Manual for SMT362, version 2.2, Sundance Multiprocessor Technology Ltd. 2009.
[6] Serafin, P., Institute of Radioelectronics Technical Report, Warsaw, 2009, (in Polish).

Finite element method application for simulation of ground penetrating radar response

M. Pasternak, R. Kędzierawski & J. Pietrasiński
Military University of Technology, Warsaw, Poland

Abstract

The software for the simulation of electromagnetic waves propagation through the soil structure with buried objects has been described in the paper. The calculations are based on the finite element method (FEM) and allow prediction of the ground penetrating radar (GPR) echo structure from different buried objects. Such a virtual radar seems to be useful for test of chosen objects detection possibility in different soil structure configurations as well as for test of different kinds of radar signals. The comparison of simulation results and real measurements data has been also presented and discussed in the paper.
Keywords: finite element method (FEM), ground penetrating radar (GPR).

1 Introduction

The most popular method for simulation of electromagnetic wave propagation inside the soil structure is finite difference time domain (FDTD) method [1, 2]. The main reason for the FDTD method popularity is its relative simplicity and good enough accuracy. An alternative approach is possible using finite element method (FEM). The solutions in this case are easier for interpretation and it gives more precise information about material discontinuities. The last feature is very important as far as ground penetrating radar working conditions are concerned. Additionally, the construction of the simulation boundaries in the FEM approach is easier and more precise [3]. Contrary to the FDTD in the case of FEM method the main field quantity characterizes all elements of volume instead of nodes. The material properties ε, μ and σ are assumed to be constant inside the single simulation cells.

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110391

2 Simulation description

To obtain the discrete centered differential equations spatial shift between magnetic and electric field components is assumed. The magnetic field components operate among simulation cells around [*i*, *j*, *k*] node and occupy 1/8 volume of elements [*i*, *j*, *k*], [*i*+1, *j*, *k*], [*i*, *j*+1, *k*],[*i*, *j*, *k*+1], [*i*+1, *j*+1, *k*], [*i*+1, *j*, *k*+1], [*i*, *j*+1, *k*+1] and [*i*+1, *j*+1, *k*+1].

For time derivative centering the Yee method was used. The electric field is defined for $t = n\Delta t$ and magnetic one for $t = (n+1/2)\Delta t$.

For simulation the integral forms of Maxwell equations was used:

$$\oint \mathbf{E} \cdot \mathbf{dl} = -\iint dA\,\mu \frac{\partial}{\partial t}(\mathbf{H} \cdot \mathbf{n}) \tag{1}$$

$$\oint \mathbf{H} \cdot \mathbf{dl} = \iint dA \left[\varepsilon \frac{\partial}{\partial t}(\mathbf{E} \cdot \mathbf{n}) + \mathbf{J}_0 + \sigma\, \mathbf{E} \cdot \mathbf{n} \right] \tag{2}$$

The contour integral of magnetic field on the left hand side of equation (2) is calculated with positive rotation along the path presented in the Fig 1.

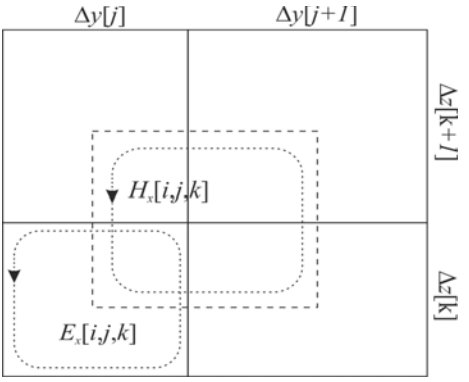


Figure 1: Magnetic and electric fields calculation path example [3].

The magnetic field component overlapping neighbour elements is averaged in *x* direction. Summing up of all components the following integral formula around [*i*, *j*, *k*] node can be obtained:

$$\begin{aligned}
IntH_x^{n+1/2}[i,j,k] = \{ & H_y^{n+1/2}[i-1,j-1,k-1] + H_y^{n+1/2}[i,j-1,k-1] + \\
& + H_y^{n+1/2}[i-1,j,k-1] + H_y^{n+1/2}[i,j,k-1] - H_y^{n+1/2}[i,-1j,k] + \\
& - H_y^{n+1/2}[i,j,k] - H_y^{n+1/2}[i-1,j-1,k] - H_y^{n+1/2}[i,j-1,k] \} \Delta y[j]/4 + \\
& + \{ H_z^{n+1/2}[i-1,j,k-1] + H_z^{n+1/2}[i,j,k-1] + H_z^{n+1/2}[i-1,j,k] + \\
& + H_z^{n+1/2}[i,j,k] - H_z^{n+1/2}[i-1,j-1,k] - H_z^{n+1/2}[i,j-1,k] + \\
& - H_z^{n+1/2}[i-1,j-1,k-1] - H_z^{n+1/2}[i,j-1,k-1] \} \Delta z[k]/4.
\end{aligned} \tag{3}$$

The field component of the surface integral on the right-hand side of (1) is equal to:

$$\left\{\varepsilon[i,j,k]\frac{E_x^{n+1}[i,j,k]-E_x^n[i,j,k]}{\Delta t}+J_{0x}^{n+\frac{1}{2}}+\right.$$
$$\left.+\sigma[i,j,k]\frac{E_x^{n+1}[i,j,k]+E_x^n[i,j,k]}{2}\right\}\Delta y[j]\Delta z[k] \tag{4}$$

In the form (4) the terms containing electric field are centered in time $t=(n+1/2)\Delta t$. From (3) and (4) the electric field component can be obtained in the following form:

$$E_x^{n+1}[i,j,k]=\frac{1}{\frac{2\varepsilon[i,j,k]}{\Delta t}+\sigma[i,j,k]}\left\{\left(\frac{2\varepsilon[i,j,k]}{\Delta t}-\sigma[i,j,k]\right)E_x^n+\right.$$
$$\left.-2J_{0x}^{n+1/2}+\frac{2IntH_x^{n+1/2}}{\Delta y[j]\Delta z[k]}\right\} \tag{5}$$

The electric field can be calculated from (2). It has the form as follows:

$$\begin{aligned}IntE_x^n[i,j,k]=&\left\{E_y^n[i,j,k]+E_y^n[i+1,j,k]+E_y^n[i,j+1,k]+\right.\\&+E_y^n[i+1,j+1,k]-E_y^n[i,j,k+1]-E_y^n[i+1,j,k+1]+\\&\left.-E_y^n[i,j+1,k+1]-E_y^n[i+1,j+1,k+1]\right\}\Delta y[j]/4+\\&+\left\{E_z^n[i,j+1,k]+E_z^n[i+1,j+1,k]+E_z^n[i,j+1,k+1]+\right.\\&+E_z^n[i+1,j+1,k+1]+E_z^n[i,j,k]-E_z^n[i+1,j,k]+\\&\left.-E_z^n[i,j,k+1]-E_z^n[i+1,j,k+1]\right\}\Delta z[k]/4.\end{aligned} \tag{6}$$

The surface integral from right-hand side of (2) contains the magnetic permittivity averaged from 8 neighbour cells. Using time centered derivative it has got the following form:

$$\begin{aligned}&\frac{1}{8}[(\mu[i,j,k]+\mu[i+1,j,k])\Delta y[j]\Delta z[k]+\\&(\mu[i,j+1,k]+\mu[i+1,j+1,k])\Delta y[j+1]\Delta z[k]+\\&(\mu[i,j,k+1]+\mu[i+1,j,k+1])\Delta y[j]\Delta z[k+1]+\\&(\mu[i,j+1,k+1]+\mu[i+1,j+1,k+1])\Delta y[j+1]\Delta z[k+1]]\times\\&\frac{H_x^{n+\frac{1}{2}}[i,j,k]-H_x^{n-\frac{1}{2}}[i,j,k]}{\Delta t}\end{aligned} \tag{7}$$

Based on (6) and (7) the magnetic field components obtain the form as follows:

$$H_x^{n+\frac{1}{2}}[i,j,k] = H_x^{n-\frac{1}{2}}[i,j,k] - \frac{\Delta t}{MuA_x} IntE_x^n[i,j,k], \tag{8}$$

where

$$\begin{aligned} MuA_x = \frac{1}{8}\Big[&\big(\mu[i,j,k]+\mu[i+1,j,k]\big)\Delta y[j]\Delta z[k]+ \\ &\big(\mu[i,j+1,k]+\mu[i+1,j+1,k]\big)\Delta y[j+1]\Delta z[k]+ \\ &\big(\mu[i,j,k+1]+\mu[i+1,j,k+1]\big)\Delta y[j]\Delta z[k+1]+ \\ &\big(\mu[i,j+1,k+1]+\mu[i+1,j+1,k+1]\big)\Delta y[j+1]\Delta z[k+1]\Big]. \end{aligned} \tag{9}$$

The remaining components of ***E*** and ***H*** were calculated in the same way.

In the first step of simulation the electric field components are actualized by directly replacing old values by new ones, because they are independent of neighbour values of the field. Next the magnetic field is actualized.

The time step of the method has to be shorter than minimum field propagation time through the smallest simulation element.

If v is the electromagnetic wave velocity inside the element the well-known from the FDTD method Courant criterion has the following form:

$$\Delta t \le \min\left(\frac{\Delta x[i]}{v[i]}, \frac{\Delta y[j]}{v[j]}, \frac{\Delta z[k]}{v[k]}\right). \tag{10}$$

For appropriate time step, centre frequency f_0 and excitation pulse bandwidth f_w of the discretized excitation wave has the form:

$$u(n) = \exp\left(-\left((n-t_0)/T\right)^2\right)\sin\left(2\pi f_0 n\Delta t\right) \tag{11}$$

where $T = 1/(f_w\Delta t)$, $t_0 = 5T$.

In the FEM method the boundary conditions formulation is relatively simple. The simulated region is usually surrounded by one layer with suitable material parameters. Out of the layer the nodes having fixed magnetic field values (usually equal to zero) are placed.

The ideal absorber has to behave like an open circuit with current value equal to the current induced by the incident field. It can be obtained by appropriate selection of boundary layer specific conductivity.

When flat wave incidents perpendicularly to the boundary layer with specific conductivity σ and thickness Δ and the layer has the same permittivities like neighbour material, $\Delta << \lambda$ and electric component E_x is constant in all layer

thickness, then electric field induces current density $J_x = \sigma \Delta E_x$ in the layer. For $\boldsymbol{H} = 0$ outside of the simulated space the magnetic field component H_y inside is equal to J_x. Using well-known formula:

$$\frac{E_x}{H_y} = \sqrt{\frac{\mu}{\varepsilon}} \tag{12}$$

one can obtain the specific layer conductivity necessary to terminate the wave.

$$\sigma = \frac{1}{\Delta\sqrt{\mu/\varepsilon}} \tag{13}$$

If the incident angle is different from 90º then effective thickness of the layer is equal to $\Delta/\sin\theta$.

3 Model verification

The model shortly described above was implemented to the software operating in Windows 7 x64 system and prepared using a six-core Pentium processor.

The software allows us to define up to ten soil layers with different thickness, permittivities and conductivities, as well as a random number of cuboidal objects. It is also possible to define the basic properties of antennas as well as radar signals. Thanks to the features mentioned above, the reconstruction of real conditions inside the simulated medium is possible with a fidelity good enough.

Significant increase of the layers or objects number makes the simulation duration time very long. The calculations concerning for example air and two soil layers with two buried objects takes about 14 hours in 0.7 m^3virtual sandbox.

Numerical calculations were verified in laboratory conditions using dry river sand and two buried objects: metal plate and wooden box as a surrogate of antipersonnel mine PMD – 7 (Fig. 2).

Figure 2: Test objects: 1 – metal plate, 2 – wooden box as an antipersonnel mine PMD-7 surrogate.

The measurements were carried out using vector network analyser and Vivaldi antennas inside anechoic chamber (Fig. 3). The stepped frequency continuous wave (SFCW) signal was generated in the range from about 1 to 3

GHz and it consists of 512 steps with equal duration time and 4 MHz single step. The results obtained in the frequency domain was recalculated into the time domain using inverse discrete Fourier transform with appropriate window function and zero –padding technique.

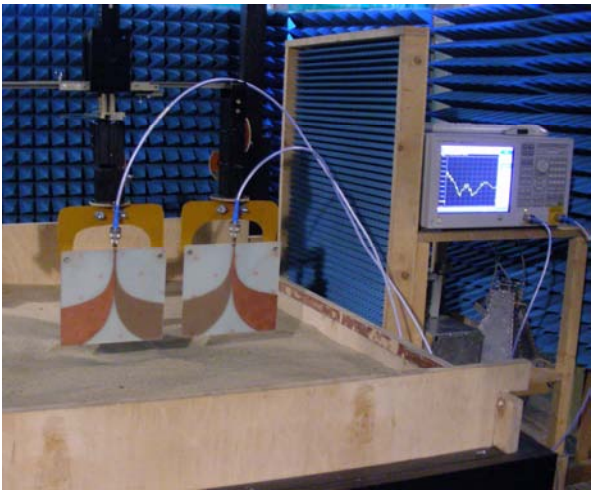


Figure 3: The measurement set.

The comparison of measurement and calculation of exemplary results for the objects shown in Fig. 2 is illustrated in Fig. 4.

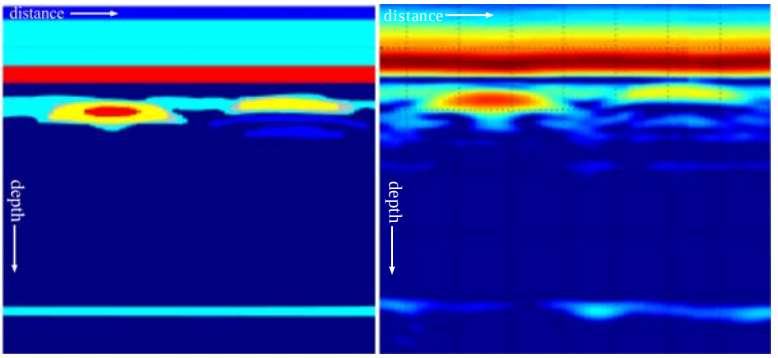


Figure 4: Comparison of calculation (on the left) and measurements (on the right).

The virtual and real environment conditions were similar as much as possible. The results both theoretical and measured data have very good similarity.

Conclusions

The FEM simulations of the GPR imaging seem to be more precise than in the FDTD case. Main reason of the difference is caused by numerical boundary conditions construction that is much easier for FEM method. The FDTD method models tested earlier by authors required much bigger simulated volume for absorbing boundary conditions implementation (PML) to obtain similar effects as FEM with one layer only.

The simulation results give very good similarity with real data from the laboratory environment. Moreover, it is possible to simulate GPR imaging in more realistic environment because the software is flexible enough. The simulation results analysis is very important from GPR image recognition problem point of view.

Acknowledgement

This work was supported by the Polish Ministry of Science and Higher Education from sources for science in the years 2009-2011 under project OR00006909.

References

[1] Taflove, A. & Hagness, S. C., *Computational electrodynamics the finite-difference time domain method*, Artech House, 2000.
[2] Pasternak, M. & Silko, D., Software for simulation of electromagnetic waves propagation through the soil with buried objects, Proc. of the 11th International Radar Symposium, pp. 524-527, 2010.
[3] Humphries, S. Jr., *Field Solutions on Computers*, CRC Press, 1997.

This page intentionally left blank

Non-contact breath sensor based on a Doppler detector

Z. R. Szczepaniak & M. Łuszczyk
Przemysłowy Instytut Telekomunikacji S.A. Warsaw, Poland

Abstract

Respiratory activity is an important parameter for observation of human body activity. A normal adult has a respiratory rate of 12–15 breaths/min in normal conditions. Breath sensor based on non-contact Doppler detector allows for the measurement and detection and the absence of the breathing action, which may cause death in a few minutes. The monitoring and detection of some respiratory abnormality presents interest in certain situations (i.e. patients from intensive therapy, of newborn children and many others). The paper covers a new technical solution of the low-cost breath sensor, which consists of the microwave generator together with a resonant patch antenna. There is only one antenna for transmitting and receiving. The technical solution of the oscillator is based on a single FET transistor. The microwave oscillator may be tuned by change of the antenna dimensions. The solution presented here is designed for 2.4GHz (ISM band). The respiratory activity is mainly detected due to the motion of the body. The wave reflected by the moving body surface is mixed with the oscillating frequency by FET transistor junction. Filtering the low frequency signals gives a component that represents the Doppler frequency due to body surface motion. Next, it is processed with high a resolution analogue-digital converter and digitally filtered, and time-frequency analyzed as well. The processing enables the detection of the respiration rate with accuracy 1 beat/min or 0.016Hz. Signal processing in digital domain includes removal of DC offset and out-of-band noise. Experimental results confirm the possibility of using a microwave Doppler detector to measure and analyze respiratory activity signals. This method permits to measure the breathing rate period.
Keywords: respiratory sensor, microwave Doppler detector.

WIT Transactions on Modelling and Simulation, Vol 51,
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110401

1 Introduction

Recently, interest in non-contact breath sensors has increased rapidly. Advances in microwave technology make it possible to design very small and simple devices for detecting vital signs. New applications of non-contact sensing of human beings have also been conceived in medical applications and as tools for detecting human beings hidden behind walls for instance, or as rescue tools for finding survivors trapped under rubble [1].

The research activities are currently focussed on the use of two different techniques for vital signs detection by means of microwave sensors: ultra-wideband (UWB) radars and continuous wave (CW) radars.

UWB radars transmit short pulses with a pulse duration of the order of nanoseconds [6]. This type of radar, as well as CW radars, is able to detect the movement of a target by measuring the low-Doppler variation that affects the received backscattered signal. UWB radars also provide a range resolution that permits to eliminate the interfering pulses due to reflections of other targets in the background. However, that characteristic requires a fast switching time discriminator that opens the receiver when the wanted reflected pulse arrives. If the distance is changed, the delay of the time window of the discriminator must be adjusted.

CW radars are simpler systems than the UWB radars and the received signal processing is independent on the target distance [3]. But in order to measure the displacements due to breathing, other movements of the subject under observation, different from that of respiratory, should be avoided. Several CW radar transducer configurations have been developed to deal with sensitivity and detection capability. Sensors working in various frequency ranges (i.e. S-band, C-band and K-band) have been tested to adapt the transmitted wavelength to the chest movements. Quadrature direct-conversion systems or double sideband indirect-conversion systems have been carried out to resolve the null point problem, which causes an accuracy decrease related to the distance between the sensor and the chest. However, all the above-mentioned systems involve a typical transducer with transmitting and receiving parts.

2 CW radar sensor features

Doppler-radar motion-sensing systems typically transmit a continuous-wave (CW) signal (sometimes frequency modulated), which is reflected off a target and then demodulated in the receiver [4]. According to Doppler theory, a target with a time-varying position but a net zero velocity, will reflect the signal with its phase modulated proportionally to the time-varying target position. A stationary person's chest has a periodic movement with a zero velocity, and therefore, a CW radar with the chest as the target will receive a signal similar to the transmitted signal with its phase modulated by the time-varying chest position. Demodulating the phase will then give a signal directly proportional to the chest position, which contains information about movement due to heartbeat and respiration, from which heart and respiration rates and signatures can be

determined [1]. Based on this principle, a noncontact heart and respiration monitor can be envisioned. By leveraging existing wireless technology, it would be possible to manufacture an inexpensive and portable device to perform such monitoring. Assuming that there is no amplitude noise, a CW radar typically transmits a single-tone signal:

$$s_n(t) = A_n \cos[2\pi ft + \phi(t)]$$

where: f – signal frequency, ϕ(t) – phase noise of the oscillator, t –time.

The transmitted signal reflects from body surface and returns to the receiving antenna. After amplification it is led to a phase detector. The received signal is phase-shifted and this phase shift is a function of emitted wavelength and distance between radar and monitored object. Moreover, when the surface is being radiated (e.g. chest surface) changes its position periodically $x(t)$ (with mean value equal to zero), the resulting phase shift is periodically modulated.

$$S_o(t) = A_o \cos\left[2\pi ft - \frac{4\pi L}{\lambda} - \frac{4\pi x(t)}{\lambda} + \phi\left(t - \frac{2L}{c}\right)\right]$$

where: c -speed of EM wave in vacuum, λ - wavelength in the air

The output signal of phase detector is a function of phase difference between transmitted and received signals.

$$S_D(t) = \cos\left[\theta_0 + \frac{4\pi L}{\lambda} + \frac{4\pi x(t)}{\lambda} + \Delta\phi(t)\right]$$

where: $4\pi L/\lambda$ - phase shift depending on target distance,

θ_0 – constant phase shift resulting from signal delays in the transmit/receive circuitry (for given structure of transmitter /receiver), $\Delta\phi(t)$ – component resulting from conversion of phase noise to baseband.

The single-channel Doppler radar exhibits minima of sensitivity for specific target distances expressed as:

$$\theta_0 + \frac{4\pi L}{\lambda} = n\pi$$

3 Microwave sensor description

The sensor presented here has very simple structure. The main assumption is one channel operation without splitting of the received signal into the in-phase and

quadrature channels. Typically, the sensor setup consists of microwave signal source, splitters, amplifiers, mixers and two antennas. On the other hand, there are solutions using signal generating circuitry based on IMPATT diode and rectangular waveguide resonator. Transmitting and receiving of the signals is accomplished by means of one horn antenna. The IMPATT diode forms an active circuit for oscillator and short-circuited section of rectangular waveguide plays a role of microwave resonator, and therefore a passive circuit for oscillator. The advantage of this solution is that the mixing of transmitted and received signals takes place in nonlinear junction of semiconductor diode without any additional mixer circuit. The disadvantage is that this simple one-channel circuit suffers from the nulling effect.

The solution presented here relies on the loop-type microwave oscillator. Here, the active circuit is formed with the use of microwave HEMT transistor (2, see fig. 1). The active circuit is a two-port network. The passive circuit is formed by two-port resonant circuit which simultaneously plays a role of patch antenna. Therefore, this solution is more compact and simpler than presented earlier [5]. The use of radiating element as a resonant circuit in the oscillator gives additional substantial advantage. When the distance from antenna to monitored object is decreased the object enters near field area and antenna resonant circuit starts to be tuned by presence of the object material in the field region. The oscillator should be properly designed; it means taking into account the presence of object at possible closest position to the antenna. As an effect, the "target-tuned" oscillator is obtained.

For different target distance from the antenna the oscillator conditions of generation are still fulfilled. This means that for every target position oscillator generates output signal with different position-dependant frequency. For small periodic variation of position the frequency modulation will occur.

Finally, the component at intermediate frequency will appear in the active nonlinear element bias circuit (1, see fig. 1). In that way, for close range operation, there is no nulling effect due to distance between object and antenna.

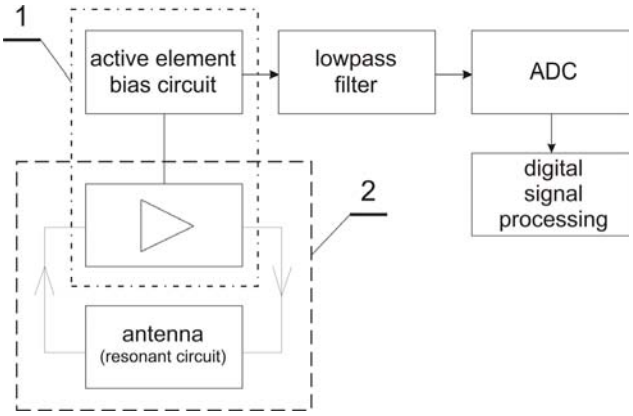


Figure 1: Microwave Doppler sensor structure.

4 Signal processing in Doppler sensor

In Doppler respiration measurement signal processing is realized in analogue and digital domain. The initial part of the signal processing in discussed Doppler sensor is baseband signal conditioning. Before a signal is digitized it must be lowpass filtered to avoid out-of-band interference alias signals into baseband signal. The bit resolution of the analog-to-digital converter determines kind of preprocessing steps. 16-bit resolution converter used in measurement set-up is enough to resolve respiration signal. In Doppler sensor the dc offset is sufficiently large that the level of amplification must be set carefully not to saturate the converter. When the signal is small and noisy the dc offset must be removed before or simultaneously with amplification.

Anti-aliasing filter is required to limit a bandwidth less than half the sampled rate to avoid aliasing effect. This problem is resolved with active analog lowpass filter with cutoff frequency below Nyquist frequency (sampling frequency is typically performed at five times the signal bandwidth. In measurement setup frequency sampling is 100Hz. The filter is realized as Sallen-Key lowpass filter with cutoff frequency equals 25Hz (fig. 2). The cutoff frequency is sufficient to limit any sampling frequency above 50Hz [1].

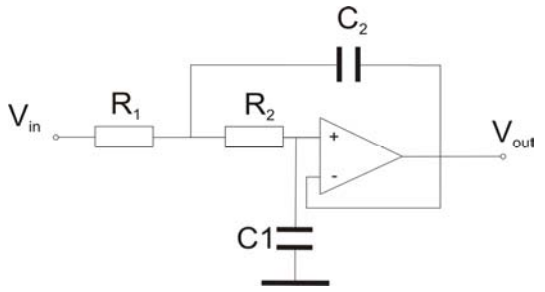


Figure 2: Anti-aliasing Sallen-key lowpass filter circuit.

The dc blocking circuit and amplification circuit is shown in figure 2. Typical signal amplitude is between 0.5 and 1mV. To amplify the signal to full scale of ADC (0-5V or ±5 typically) the amplification of about 80dB is required. Due to dc offset, gain below is 40dB suitable. The dc-block circuit efficiency depends on Doppler sensor bias stability.

In the Doppler sensor the respiration and heartbeat signal are superimposed on each other. The chest move amplitude due to breath is much bigger than this of the heart beating. The respiration signal is typically about 100 times greater than that of the signal due to the heartbeat. The respiration signal can be detected and processed without filtering but detection of heartbeat signal requires filtering.

Digital signal processing (DSP) includes digitization process and digital filtering. Digitization in discussed system is 100Hz and has 16-bit resolution. Signal conditioning steps include removal of dc offset and filtering which removes out-of band noise and separates breath signal and other signal (i.e. heartbeat signal or backscatter signal or clutter signal).

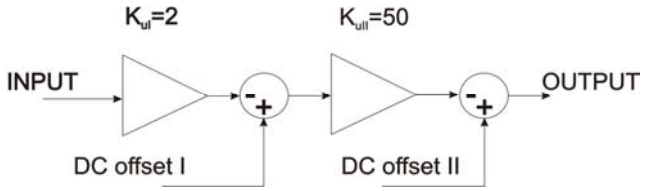


Figure 3: DC offset reduction circuit.

The DSP is based on fact that required respiration signal and other signals are separated in frequency domain. Resting respiration rate is between 0.15 and 0.4Hz what corresponds to 9 and 24 breaths per minute [2]. Heart rate is between 0.83 and 1.5Hz what corresponds to 50 and 90 beats per minute. This means that the breath signal can be isolated by a lowpass filter. A sample of respiration and heart signals measured with a microwave Doppler sensor are shown in fig. 4 and fig. 5.

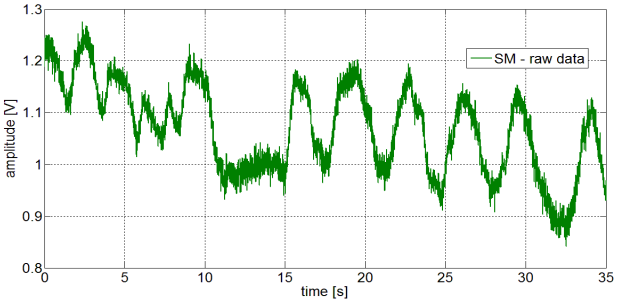


Figure 4: Respiration raw signal measured at 20cm distance (chest).

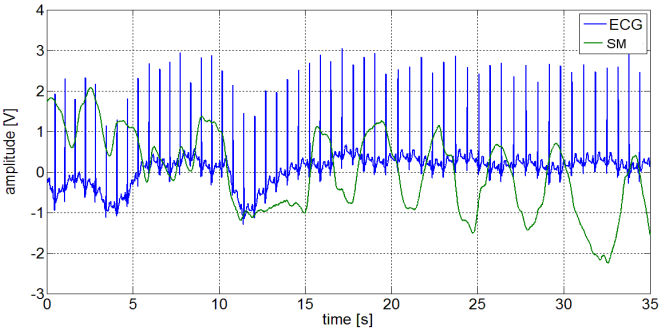


Figure 5: Respiration signal after the lowpass filter with 1Hz cut-off frequency (SM) and reference signal ECG (variation of isoline).

Respiration and heart signals are separated with digital filters. Filter isolating heartbeat must attenuate the respiration signal at last 50dB. Signal processing with finite impulse response (FIR) is presented. The FIR filters use current and past input samples. The number of coefficients ant its values determine the

filter's properties in frequency and time domain (i.e. cut-off frequency, the steepness of transition between the passband and stopband, group delay and how long the filter has an output signal given a step at the input). The phase shift of an FIR filter is linear within passband and group delay is constant and its value depends on filter order. To separate heartbeat and respiration signal FIR 250-order Kaiser filter with parameter $\beta=2$ with 3 second group delay is used.

The Fourier Transform is used to determine the frequency characteristics of respiration signal (fig. 6). When the frequency characteristics varies in time (non-stationary signals) what is typical for physiological signals short-time Fourier transform should be used for analysis.

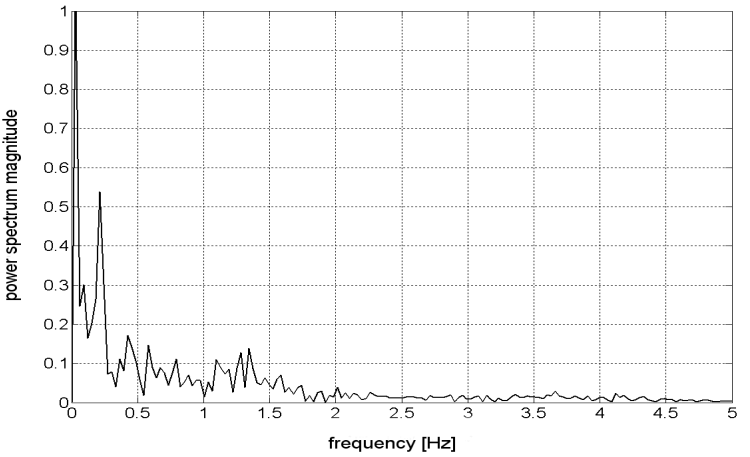


Figure 6: Respiration signal in frequency domain.

For heart monitoring the heartbeat signal must be isolated by filtering with highpass filter. The Doppler sensor signal filtered with 400-order Kaiser ($\beta=2.5$) filter is shown in fig. 7.

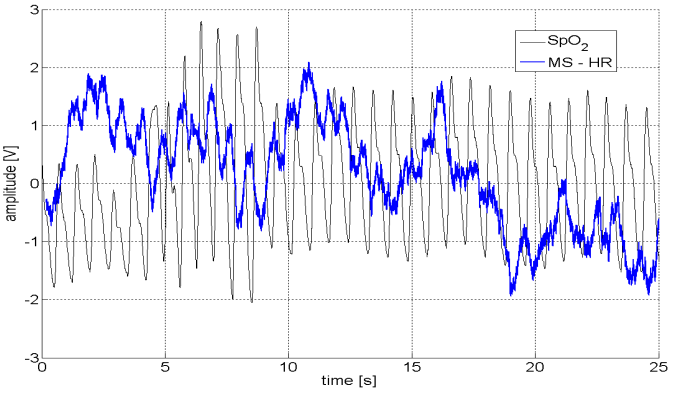


Figure 7: Heartbeat signal measured at 20cm (chest MS-HR) with reference from pulse oximeter (SpO_2)

5 Conclusions

A low cost microwave sensor for human breath and heart rate activity is presented. The circuit concept is based on a loop-type microwave oscillator with resonant circuit acting as antenna. The oscillator is designed to be frequency-tuned by the change of target distance from the antenna. This feature allows avoidance of the nulling effect. The sensor structure is very compact. The concept has been verified by the measurement of the model sensor with the use ECG and pulse oximeter as the reference.

Proper detection of breath action has been obtained form the distance of 20 cm from the chest surface.

This kind of sensors may be used in distributed systems for monitoring the patient activity in the field of health-care or to verify the condition of machine operator (pilot, drive, etc.).

References

[1] Droitcour A., Lubecke V., Lin J.C., Boric-Lubecke O.: A microwave radio for Doppler radar sensing of vital signs. 2001 IEEE MTT-S Int. Microwave Symp. Dig., 2001, vol. 1, pp. 175-178.

[2] Droitcour A. D., Boric-Lubecke O., Lubecke V. M., Lin J., Kovacs G.: Range correlation and I/Q performance benefits in single-chip silicon Doppler radars for noncontact cardiopulmonary monitoring. IEEE Transactions on Microwave Theory and Techniques, vol. 52, no. 3, pp. 838-848, 2004

[3] Lin J.C.: Microwave sensing of physiological movement and volume change: a review. Bioelectromagnetics, vol. 13, no. 6, pp. 557-565, 1992

[4] Ichapurapu R.: A 2.4GHz Non-Contact Biosensor System for Continuous Vital-Signs Monitoring, Proc. WAMICON 2009

[5] Baltag O.: Microwaves Doppler Transducer for Noninvasive Monitoring of the Cardiorespiratory Activity, IEEE Transactions On Magnetics, vol. 44, No. 11, November 2008

[6] Immoreev I.J.: Practical Application of Ultra-Wideband Radars. International Conference "Ultra Wideband and Ultra Short Impulse Signals" (UWBUSIS'06), 18-22 September, 2006, Sevastopol, Ukraine.

N_2O detection with the CEAS method

J. Wojtas[1], B. Rutecka[1], T. Stacewicz[2], J. Mikolajczyk[1]
& Z. Bielecki[1]
[1]*Institute of Optoelectronics, Military University of Technology, Poland*
[2]*Department of Physics, Warsaw University, Poland*

Abstract

Nitrous oxide (N_2O) plays a significant role in many different fields therefore its monitoring is an important task. In this paper the opportunity of application of cavity enhanced absorption spectroscopy for N_2O detection is presented. This method is a modification of cavity ring down spectroscopy. The laser radiation tuned to absorption line of N_2O is injected into an optical cavity under a very small angle in respect to its axis. In the case of lack of the absorption, the result is determined by the mirrors reflectivity coefficient, diffraction losses and length of the cavity. When the absorber is present in the cavity, the result additionally depends on absorption and scattering of light in cavity. The method provides the determination of a very weak absorption coefficient as well as the concentration of the absorbing gas.

Our N_2O sensor consisted of a pulsed radiation source, optical cavity, detection module and a digital oscilloscope. As the light source anoptical parametric generator was applied. It enabled the wavelength tuning in a broad spectral range with resolution of 1 nm. The optical cavity was composed of two high-reflective spherical mirrors. Optical signal registration was done with detection module equipped with HgCdTe photodetector.

The spectral range of 4.52–4.53 μm is the best for N_2O detection. Operation at these wavelengths provides opportunity avoiding of interferences with other atmosphere gases, like CO_2 and H_2O. Assuming 2% uncertainty of measurements and the effective value of the absorption cross section of about $6 \cdot 10^{-19}$ cm^2 the detection limit of 10 ppb was achieved.

Keywords: CEAS, N_2O detection, optoelectronic sensor.

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110411

1 Introduction

Nitrous oxide (N_2O) is a colorless gas with a slightly sweet odor. N_2O is an important greenhouse gas and the major natural source of NO. In consequence, it initiates the catalytic NO_x ozone destruction cycles in the stratosphere. The gas is used as an anesthetic, especially in dentistry and minor surgery. It produces mild hysteria and laughter preceding the anesthetic effect. Thus it is also known as 'laughing gas'. Excessive exposure may cause headache, nausea, fatigue and irritability. Simultaneously, N_2O is a strong oxidizer above 300°C and it self-explodes at higher temperatures. Nitrous oxide is also a characteristic compound emitted by majority of explosive materials. Therefore, measuring and monitoring of N_2O concentration is very important [1, 2].

N_2O can be analyzed by gas chromatography (GC) on a molecular sieve column using a thermal conductivity detector (TCD). For this method detection limit of 4 ppm is achieved [3]. In the case of the electron capture detector (ECD) application, detection limit is about 50 ppb [4]. Nitrous oxide may be identified by GC/MS basing on its mass spectra as well. For solid phase microextraction GC/MS the sensitivity of 72 ppb for N_2O is reported [5]. Other nitrous oxide detection system with detection limit of 50 ppb is described by Hellebrand [6]. It consists of Fourier transform infrared spectrometer (FT-IR), heated measuring cell with an optical path length of 20 m and HgCdTe-detector (MCT).

During the last several years some spectroscopic methods for gas detection were alsosignificantly developed. In the paper the opportunity of application of cavity enhanced absorption spectroscopy (CEAS) for N_2O detection is presented. This method is a modification of cavity ring down spectroscopy (CRDS).

2 Cavity enhanced measurements analyses

Spectroscopic measurements based on light absorption are the most popular methods of matter investigation. Absorption spectroscopy is a simple, non-invasive, in situ technique for obtaining information about different species. However, the sensitivity of traditional spectroscopic methods is approximately limited to values of 10^{-4}–10^{-5} cm^{-1}. It is mainly caused by detection limits of small differences of light intensities between input and output of the measuring system. Others reasons are the instabilities in light source intensity and the fluctuation of photodetector detectivity, which provide noises and interferences of absorption signals [7]. To improve the sensitivity, a longer absorption path length should be used. Such solution is used in multipass spectroscopy, e.g. in White or Herriott cells. Also the absorption spectroscopy with a modulation technique could be applied. O'Keefe and Deacon proposed novel measuring technique called cavity ring-down spectroscopy [8]. It allows to increase the sensitivity of about 3–5 orders of magnitude or even higher. CRDS system requires high–finesse stable optical cavity.

In this technique, the laser pulses are injected into an optical cavity (resonator) consisting of two spherical high-reflective mirrors. The optical radiation is multiplied by reflected inside the resonator. After each reflection, a

small part of laser radiation leaves the optical cavity due to residual transmission of mirrors. The transmitted light is registered with a photodetector. The signal from the photodetector can be measured e.g. with digital oscilloscope. The amplitude of single–mode radiation trapped within the cavity decays exponentially over time with a time constant τ. The time constant is often referred to as the decay time or ring–down time. The decay of light intensity $I(t)$ can be described as

$$I(t) = I_0 \cdot e^{-\frac{t}{\tau}}. \tag{1}$$

The value of decay time constant depends on mirror reflectivity R, resonator length L, and extinction factor α (absorption and scattering of light in cavity) [9]

$$\tau = \frac{L}{c \cdot [(1-R)+\alpha L]}. \tag{2}$$

In this way determination of the absorption coefficient is possible by the decay time measurement. Firstly, it is measured for the case without absorption (α=0), when the decay time τ_0 depends only on mirrors reflectivity and the cavity length. Secondly for cavity filled the analyzed gas. By comparison of these two times, the absorber concentration C can be achieved from

$$C = \frac{\propto}{\sigma} = \frac{1}{\sigma \cdot c}\left(\frac{1}{\tau} - \frac{1}{\tau_0}\right), \tag{3}$$

where σ denotes the absorption cross section.

The sensitivity of this spectroscopic method is determined with the lowest detectable concentration of analyzed gas molecules C_L, which causes measurable change of the output signal. The concentration limit of absorbing molecules can be described with the formula

$$C_L = \frac{1}{c \cdot \sigma \cdot \tau_0} \cdot \delta_\tau = \frac{(1-R)}{\sigma \cdot L} \cdot \delta_\tau, \tag{4}$$

where δ_τ is the relative precision of the decay time measurement. The relationship between uncertainty δ_τ and τ_0 can be described as

$$\delta_\tau = \frac{\tau_0 - \tau_L}{\tau_0} \cdot 100\%, \tag{5}$$

where τ_L denotes a decay time of the optical cavity for minimal absorber concentration. The analysis presented in Figs. 1 and 2 show that sensitivity of the N_2O experimental setup better than 10 ppb can be obtained.

Effective storage of light in the resonator is ensured only when laser frequency is well-matched to a cavity mode. Then the best sensitivity can be achieved. However the major disadvantage of this method is a strong dependence of cavity modes frequency on mechanical instabilities. They damage cavity Q-factor and provide fluctuations of the output signal [10].

Such disadvantage is eliminated in cavity enhanced absorption spectroscopy. This modification of CRDS technique was described in 1998 by Engel et al. [11].

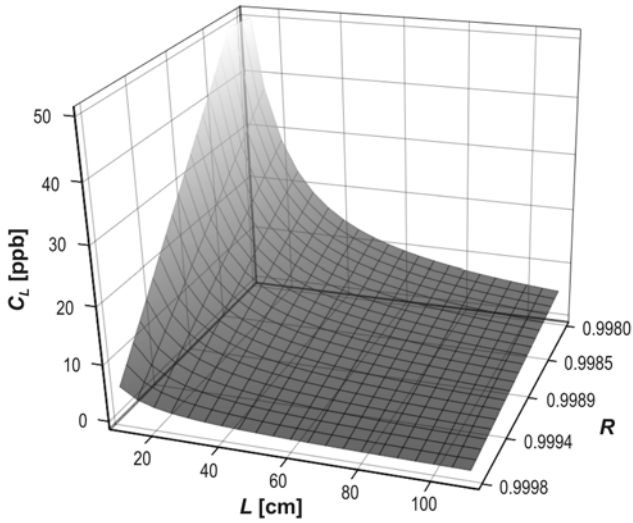


Figure 1: Dependence of the concentration limit C_L on the cavity length L and the reflectivity of mirrors R (calculations were performed for $\delta_\tau = 2.5\%$ and $\sigma = 4 \cdot 10^{-18}$ cm^2).

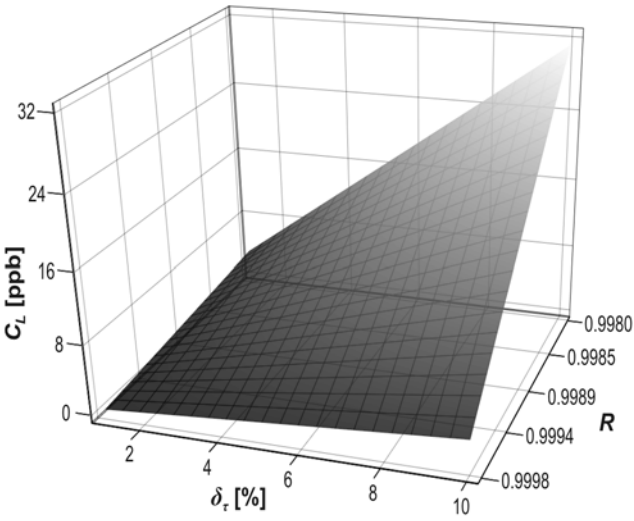


Figure 2: Dependence of the concentration limit C_L on uncertainty δ_τ and the reflectivity of mirrors R (calculations were performed for $L = 60$cm and $\sigma = 4 \cdot 10^{-18}$ cm^2).

CEAS is based on off-axis arrangement of the cavity and laser beam. The beam is injected under a very small angle in respect to the cavity axis. Usually the beam is repeatedly reflected by the mirrors, however, the reflection points are spatially separated. As a result a dense structure of weak modes is obtained or the modes do not occur due to overlapping. The system is much less sensitive to mechanical instabilities. CEAS sensors attain the detection limit of about 10^{-9} cm^{-1} [12, 13]. Another advantage is that due to off-axis illumination of the front mirror, the source interference by the optical feedback from the cavity is eliminated.

3 Laboratory N_2O sensing setup

Detection of nitrous oxide could be done at both the electronic and vibronic transitions. The electronic transitions correspond to ultraviolet range, below 200 nm, where neither the suitable laser sources nor the high reflectivity mirrors are available. Therefore N_2O detection was performed at vibronic transitions related to infrared range. The spectral region of 4.52–4.53 μm was chosen. In this wavelength range N_2O is characterized by relatively large peak cross section ($4 \cdot 10^{-18}$ cm^2) versus weak interference by the atmospheric compounds like CO_2 or H_2O. Comparison of these spectra based on *US Standard Atmosphere 1976* is presented in Fig. 3. According to this standard usual CO_2 concentration is 330 ppm, N_2O–773 ppm, while N_2O–0.32 ppm (at temperature 300 K and normal pressure).

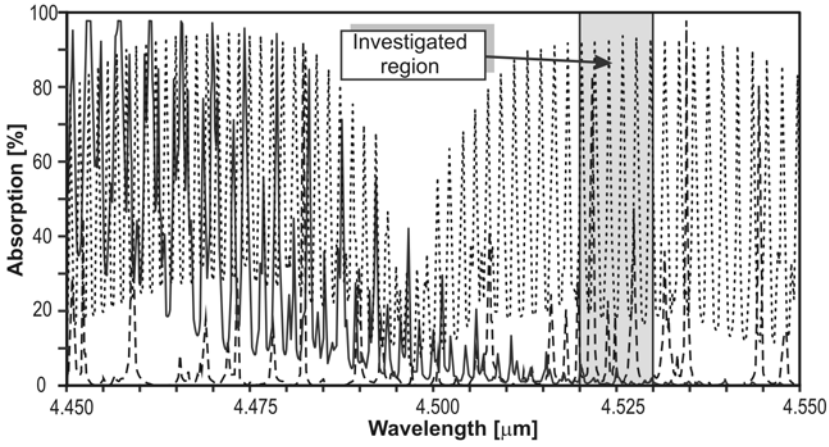


Figure 3: Absorption spectrum for N_2O (dotted line), H_2O (continuous line) and CO_2 (dashed line) (from HITRAN 2008).

Nitrous oxide sensing setup consisted of: source of infrared pulses, optical cavity, detection module and an oscilloscope. Nowadays, QCL lasers just elaborated by Alpes Lasers are the best light sources for this application. FWHM duration time of their pulses reaches hundreds of microseconds pulses while the repetition rate might be of some kHz. The emission wavelength can be easily

tuned to the maxima of N_2O cross section, e.g. 4.5239 μm or 4.5297 μm. However, in our preliminary experiments optical parametric generator (OPG) model PG711-DFG produced by Ekspla (Lithuania) was used. It enabled wavelength tuning within a broad spectral range with 1 nm resolution. The OPG was pumped with Nd:YAG pulsed laser. The output energy at the wavelength of 4.5 μm was about of 8.9 μJ. Because of high beam divergence two-lens collimator was applied. The scheme of lab setup for preliminary investigation of CEAS sensor is presented in Fig. 4.

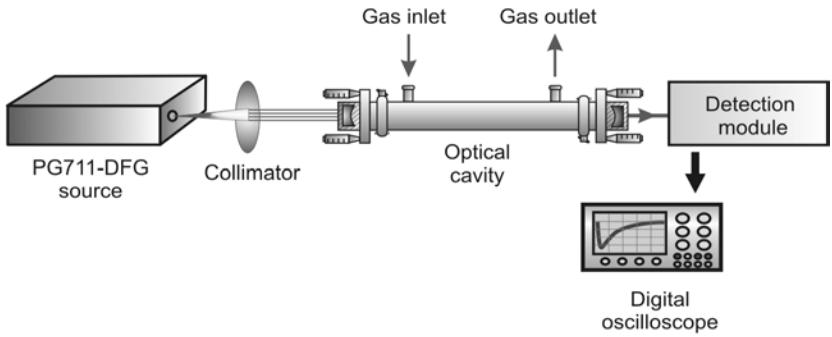


Figure 4: The scheme of the laboratory CEAS system with PG711-DFG radiation source.

The optical cavity was built of two spherical mirrors, which reflectivity reaches about 0.9998 at the wavelength of interest (Los Gatos Research, Inc., USA).

The radiation leaving the cavity was registered with the low noise detection module PVI-2TE (VIGO System S.A., Poland). In the module, a photodetector (photodiode), cooling system and a preamplifier were integrated in a common housing [14]. Such construction provides opportunity of room temperature operating. Both mechanical and spectral parameters of the module were optimized to the N_2O sensor application. Photodetector maximum responsivity corresponds to the absorption bands of nitrous oxide (Fig. 5).

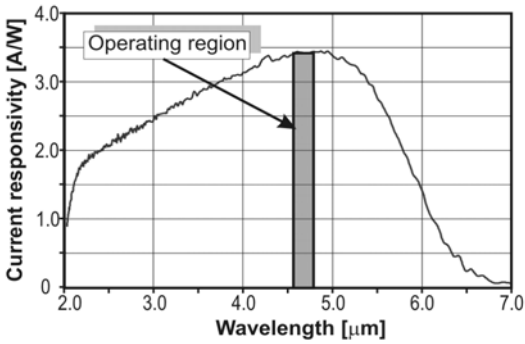


Figure 5: Spectral characteristics of the HgCdTe photodiode current sensitivity.

The current signal from the photodetector was converted by a low-noise, broadband transimpedance preamplifier. The preamplifier ensured the best possible S/N ratio and matched the signal to the input circuits of the digital oscilloscope (DPO 4104, Tektronix). The experimental setup is presented in Fig. 6.

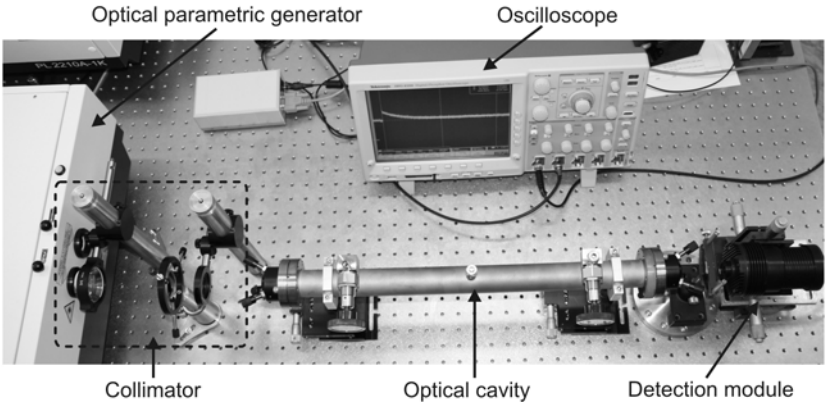


Figure 6: Experimental setup for N_2O detection.

Increase in mirrors reflectivity improves the detection limit. However it also reduces the intensity of the optical signal from the cavity. As a consequence it increases the requirements about detection module parameters. The most important factor describing optical signal detection capability is the signal to noise ratio (S/N). Fig. 7 presents the noise equivalent scheme of the first stage of a detection module using a transimpedance preamplifier.

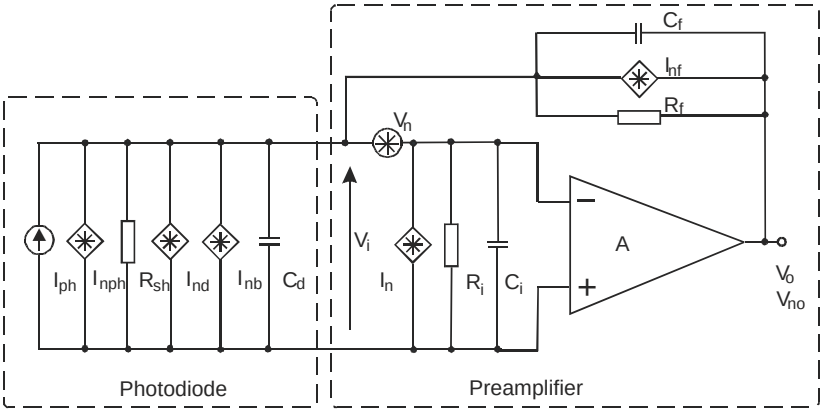


Figure 7: Scheme of the first stage of a detection module [15].

The signal current generator I_{ph} represents the detected signal. Noises in a photodiode are represented by three noise generators: I_{nph} – the shot noise associated with photocurrent, I_{nd} - the shot noise of a dark current, while I_{nb} - the shot noise from a background current. In the scheme, the value of the load resistance of photodetector depends on the feedback resistance R_f and the preamplifier gain A. The resistor R_f affects both the level of preamplifier output signal and its noise. The noise current generator I_{nf} is the thermal noise current and excess noise of the feedback resistance. Since the thermal noise of I_{nf} is inversely related to the square root of the resistance, R_f should be of great value.

In the first stage of detection module the noise is a function of frequency. At very low frequency, the photodiode is the dominant noise source. However preamplifier noise becomes increasingly important at higher frequency. The R_{sh} is the shunt resistance of a photodiode. Preamplifier noise is represented by the voltage source V_n and the current source I_n. From the arrangement in Fig. 7 it can be shown that the equivalent input noise is the square root of the sum of squares of noise components from: the photocurrent I_{ph}, the dark current of a photodiode I_{nd}, the background current I_{nb}, thermal noise of the resistor R_f, the noise sources from a preamplifier: I_n *and* V_n. Thus, the signal-to-noise ratio can be expressed by

$$\frac{S}{N}=\frac{I_{ph}}{\left[I_{nph}^2+I_{nd}^2+I_{nb}^2+I_n^2+\frac{4kT\Delta f}{R_f}+\left(\frac{V_n}{R_f}\right)^2\right]^{1/2}}. \tag{6}$$

In the N_2O detection system, the shot noise from a background current I_{nb} can be negligible, then, equation (6) becomes

$$\frac{S}{N}=\frac{I_{ph}}{\left[I_{nph}^2+I_{nd}^2+I_n^2+\frac{4kT\Delta f}{R_f}+\left(\frac{V_n}{R_f}\right)^2\right]^{1/2}}. \tag{7}$$

In the case of high frequency, the last term of the denominator should contain additional combination of impedances across the input of the preamplifier, i.e. $(\omega C_d)^{-1}$ and $(\omega C_i)^{-1}$. Assuming that V_{nd} is the voltage noise of the serial resistance of the photodiode and the preamplifier capacitance C_i is conveniently grouped with photodiode capacitance C_d, the noise equivalent signal current, $I_{n\,total}$, is given by

$$I_{ntotal}^2=\left[I_{nph}^2+I_{nd}^2+I_n^2+\frac{4kT\Delta f}{R_f}\right]+\omega^2\left[V_{nd}^2C_d^2+V_n^2\left(C_d+C_i\right)^2\right]. \tag{8}$$

There are two terms: a ‘white’ noise term in the first square brackets, and a second term which gives frequency proportional increase in a noise current. Although a capacitor does not add noise, the photodetector noise voltage (V_{nd}), and preamplifier noise voltage (V_n) is increased by the C_d and the C_d+C_i

respectively. Analyzing equation (8), we see that for matching a preamplifier to a photodetector, it is important to minimize the sum of $I_n + V_n^2 \omega^2 \left(C_d + C_i\right)^2$.

Thus, in order to ensure registration of low-level signals and accurate measurements of decay time, special preamplifier was developed. Moreover, further increase in *S/N* can be improved by the use of one of the advanced methods of signal detection, i.e. coherent averaging. This technique can be implemented in software for digital processing circuit. Thanks to this increase in *S/N* is directly proportional to a root of a number of the averaging samples n_{smpl},

$$\left(\frac{S}{N}\right)' = S\left(\frac{N}{\sqrt{n_{smpl}}}\right)^{-1}. \tag{9}$$

4 Experimental results

The main task of experiments was to check the opportunities of N_2O detection by constructed CEAS system. The research was performed for wavelengths range from 4.519 μm to 4.536 μm. Measurement procedure began from setting the appropriate wavelength of interest. Then τ_0 (3) was determined for the cavity filled with pure N_2. Subsequently the flow of N_2O-N_2 mixture was set to the cavity and the corresponding decay time τ was registered. The comparison of both signals is presented in Fig. 8.

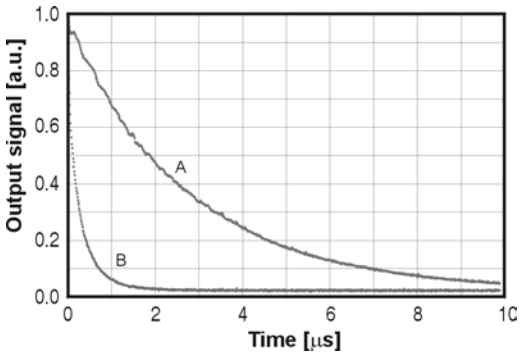


Figure 8: The exemplary output signals from the cavity filled with reference gasN_2 (τ_0) (line A), and from the cavity filled with 10 ppm N_2O(τ) (line B).

The values of decay times were used to calculate the concentration of N_2O accordingly to the relationship (3). However the line width of optical parametric generator was about 0.001 μm. Thus it overlapped several absorption peaks of N_2O (Fig. 9). Therefore, in order to determine the absorber concentration the effective absorption cross section was taken into account. In the vicinity of the investigated lines the mean value of this parameter reached about $6 \cdot 10^{-19}$ cm^2.

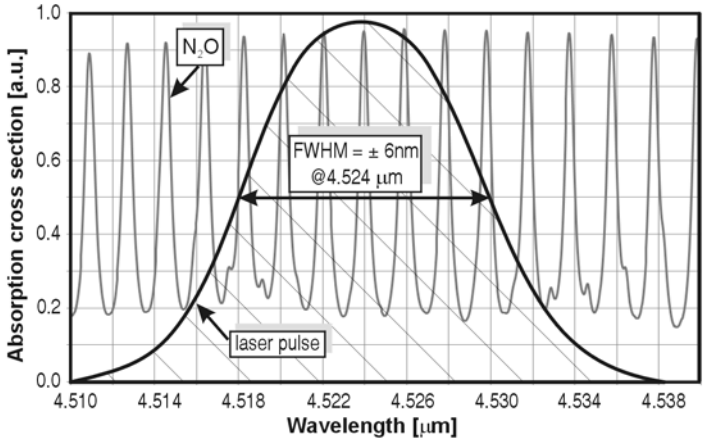


Figure 9: The spectral characteristics of N_2O absorption cross section and the laser radiation.

During the experiments the gas samples containing about 10 ppm N_2O were investigated. Concetration measurements were carried out at the different wavelengths. A good agreement between various results was achieved (Fig. 10). The differences are caused by uncertainity of gas sample preparations (~10% precision). Furthemore, 2% uncertainty of the decay time determination was obtained. In accordance with equation (4) the value of concentration limit of detection C_L of 10 ppb was achieved.

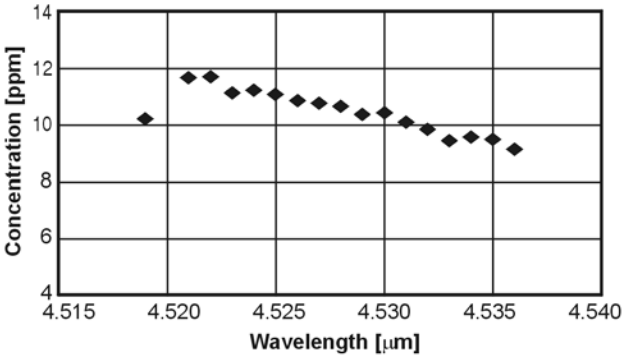


Figure 10: Measured N_2O concentration at the different wavelength.

5 Summary and conclusion

The paper presents the preliminary research of a nitrous oxide detection system based on cavity enhanced absorption spectroscopy. The best sensitivity was obtained in the wavelength range of 4.52 μm–4.53 μm. During experiments with

a broadband parametric oscillator the sensitivity of 10 ppb was achieved. One can expect that with a specialized narrowband light source (like QCL laser), with a wavelength well matched to the peak of the absorption spectrum the sensitivity might be increased about one order of magnitude.

Our N_2O sensor will be applied for detection of vapours from explosive materials. Some successful research with nitrogen dioxide sensor to detect nitroglycerine and nitrocellulose has been already performed. CEAS technique with blue violet laser diodes was applied [9, 16, 17]. We expect that an additional channel, i.e. the nitrous oxide sensor, would enable detection of other types of explosives. In Fig. 11 block diagram of two-channel CEAS sensor is presented.

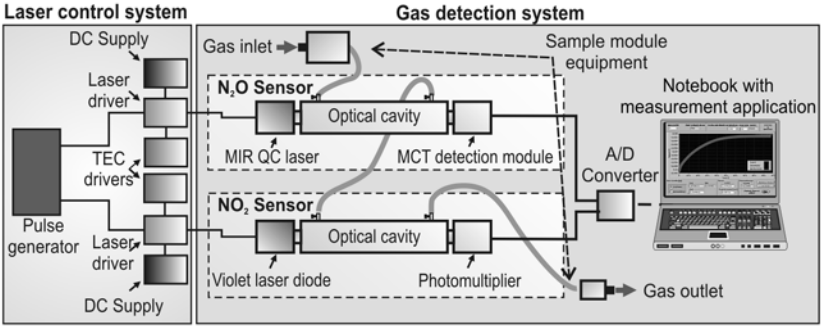


Figure 11: Block diagram of NO_2 and N_2O optoelectronic sensor.

Acknowledgement

This work is supported from the Ministry of Science and High Education of Poland (Project No. OR00002807).

References

[1] Seinfeld J. H. & Pandis S. N. *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*, 2nd Edition, John Wiley & Sons, Inc., New Jersey 2006.

[2] Pohanish R. P., *Sittig's Handbook of Toxic and Hazardous Chemicals and Carcinogens (5th Edition)*, William Andrew Publishing, 2008.

[3] Mosier A. R. & Klemedtsson L., *Methods of Soil Analysis, Part 2—Microbiological and Biochemical Properties*, Soil Science Society of America Inc., Madison, Wisconsin 1994.

[4] Shimadzu Scientific Instruments, http://www.mandel.ca/application_notes/SSI_GC_Green_Gasses_Lo.pdf

[5] Drescher S. R. & Brown S. D., Solid phase microextraction-gas chromatographic-mass spectrometric determination of nitrous oxide evolution to measure denitrification in estuarine soils and sediments, *Chroma. A*1133(1–2), pp. 300–304, 2006.

[6] Hellebrand H. J., Emission of Nitrous Oxide and other Trace Gases during Composting of Grass and Green Waste. *J. Agric. Engng Res.* **69**, 365–375, 1998.
[7] Zare R. N., Paldus B. A., Herb C. C. & Spence T., Cavity –Loked Ring Down Spectroscopy, US Patent No 6,084,682, 2000.
[8] O'Keefe A. & Deacon D. A. G., Cavity ring-down optical spectrometer for absorption measurements using pulsed laser sources. *Rev. Sci. Instrum.*, **59**, pp. 2544–2551, 1988.
[9] J. Wojtas, Z. Bielecki. Signal processing system in the cavity enhanced spectroscopy. *Opto-Electron. Rev.*, **16(4)**, pp. 44–51, 2008.
[10] Berden G., Peeters R. & Meijer G., Cavity ring-down spectroscopy: Experimental schemes and applications. *Int. Rev. Phys. Chem.*, **19(4)**, pp. 565–607, 2000.
[11] Engel R., Berden G., Peeters R. & Meijer G., Cavity enhanced absorption and cavity enhanced magnetic rotation spectroscopy.*Rev. Sci. Instrum*., **69**, pp. 3763–3769, 1998.
[12] Kasyutich V.L., Bale C.S.E., Canosa-Mas C.E., Pfrang C., Vaughan S. & Wayne R.P., Cavity-enhanced absorption: detection of nitrogen dioxide and iodine monoxide using a violet laser diode. *Appl. Phys. B*, **76(6)**, pp. 691–698, 2003.
[13] Courtillot I., Morville J., Motto-Ros & Romanini D., Sub-ppb NO_2 detection by optical feedback cavity-enhanced absorption spectroscopy with a blue diode laser. *Appl. Phys. B*, **85**, pp. 407–412, 2006.
[14] Piotrowski A., Madejczyk P., Gawron W., Klos K., Romanis M., Grudzien M., Rogalski A. & Piotrowski J., MOCVD growth of Hg1xCdxTe heterostructures for uncooled infrared photodetectors. *Opto-Electron. Rev.*, **12**, 453–458, 2004.
[15] A. Rogalski, Z. Bielecki, Detection of optical radiation (chapter A1.4), *Handbook of optoelectronics*, Taylor & Francis, New York, London pp. 73–117, 2006.
[16] Wojtas J., Czyzewski A., Stacewicz T. & Bielecki Z., Sensitive detection of NO_2 with Cavity Enhanced Spectroscopy, *Optica Applicata*, **36(4)**, pp. 461–467, 2006.
[17] Nowakowski M., Wojtas J., Bielecki Z., & Mikolajczyk J., Cavity enhanced absorption spectroscopy sensor, *Acta Phys. Pol. A*, **116**, 363–367, 2009.

Identity management in VHF radio systems

Z. Piotrowski & P. Gajewski
Telecommunications Institute, Faculty of Electronics,
Military University of Technology, Poland

Abstract

For up-to-date commercial short-distance communication systems, e.g. RF systems of VHF bandwidth or WiFi wireless networks, the development of efficient and reliable algorithms and procedures for management of the radio network subscriber identity is becoming still more important and urgent need. The existing systems for that purpose are limited to checking of a simple authorization code against the tables for supervising the RF information. Identity of a radio network subscriber is verified on the basis of the rule that hidden messages and hidden responses are generated during the communication sessions. The concept of the centralized system for management of identities is based on the rule that a unique set of PIN numbers is assigned to each subscriber of the RF communication system. Each PIN number is repeatedly changed during a communication session in order to avoid the attack when a part of the signal including the PIN code is copied from the communication signal for further redistribution.
Keywords: VHF radio, hidden communication, hidden authorisation.

1 Introduction

Electronic identity understood as an unambiguous assignment of a digital ID to an individual person fulfils many important functions in contemporary telecommunications. Mechanisms of identification, authentication and authorization ensure the function of incontrovertibility of sent information and the use of abbreviation function mechanisms makes it possible also to verify the integrity of the received information. Safe data transmission through a telecommunication network requires the sent data to be provided with the following information: who sent it, who the addressee is; often, also confirmation is required of: who received the data and whether the received data

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110421

was modified during transmission. These are the most basic functions implemented as a standard in telecommunication systems; moreover, it is often demanded that the information be classified, i.e. inaccessible for third parties that are not authorized to receive it. Confidentiality functions are realized by means of encoding in accordance with the accepted cryptology standards.

However, there exist telecommunication systems of special purpose in which the above-described mechanisms are expanded by an additional function: the function of hidden transmission. A hidden transmission means hiding the fact of sending additional data through a telecommunication network by means of standard links, protocols and open transmission. Thus, the special-purpose telecommunication system may be one which uses a standard telecommunication network of a set architecture in order to carry out a hidden transmission.

Subscriber's identification involves recognizing the person by the system in accordance with the name declared by the subscriber, while authentication is the process of checking whether the declared subscriber name is coherent with the unique data assigned to the user and that stored by the system. Proper authentication takes place when the declared user name (so-called login) is coherent with the password assigned to it. The password depends on the adopted authentication scheme and may be the knowledge of the subscriber (password, personal data), the fact of possessing something (key, token, card), a characteristic biometric element (fingerprint, iris image, voice, anatomic face structure) or even a skill characteristic of a given subscriber (behavioral biometrics: handwriting, dynamics and manner of pressing keys). After a positive authentication process the subscriber is assigned rights to specific resources, i.e. so-called authorization.

2 Known authentication techniques and threats

According to Eric Diehl, head of an international group of experts from the Technicolor laboratory [1] which deals with the safety of multimedia material distribution, authentication is "*a hardware or software process whose aim is to determine an impossible-to-forge identification manner of two conversation or transaction actors. From this process it stands that the actors confirm and authenticate one another's identity using techniques of passwords, exchange of confidential information or digital signature.*"

Authentication may be multi-stage (expanded) or hierarchical which takes into account authorization levels. A classic example may be the scenario of authenticating an internet banking customer in which losing a one-time password for a transaction realization does not lead to the loss of previously acquired resources, e.g. preview of bank account, history of operations. Here, multi-stage authentication involves using two or more identification attributes, e.g. a password and a token, or a password and a biometrics feature (fingerprint reader). The problem of losing electronic identity (identity fraud) is one of the most frequent causes for infringing safety in existing telecommunication systems. What main threats may be identified to the loss of electronic identity? Javelin's report [2] introduces several definitions of identity loss which provide

an insight into the existing threats. They include *data breach*, i.e. an unauthorized revealing of information that compromises the safety, privacy and integrity of user identification data; *identity fraud*, i.e. unauthorized use of part of personal data for financial gain (it may be done without identity theft – familiarization with data, randomly generated credit card numbers); *identity theft*, i.e. unauthorized access to personal data; *Man-in-the-Middle (MTM)* defined as unauthorized access to personal information in which the perpetrator is able to decipher, modify and embed information between two parties during a session without their knowledge; *Synthetic Identity Fraud* denoting a fictional identity created in order to deceive an organization, e.g. identity generated using real social insurance numbers and various first and last names. Of course, there are many other definitions and notions related to electronic identity loss. Electronic identity theft is the most common internet offence and losses suffered as a result of an improper distribution and protection of electronic identity are huge.

According to the aforementioned Javelin's report and the Spendonlife service [3], in 2008 the United States of America recorded a loss of $31 billion on account of various forms of identity loss, whereas total losses in the world amount to $221 billion. In USA, 26% of identity theft cases involved an unauthorized takeover of credit card numbers and purchases of material goods made by third parties based on those numbers; 18% was theft of public utility services (gas, electricity) involving assigning a given service to a person residing outside their area of permanent residence; 17% was banking fraud (change of account assets, theft of access codes for ATM systems); 12% involved theft of social security numbers, e.g. in order to obtain employment; 5% included loan fraud (applying for a loan on someone else's behalf, e.g. assigning a social security number to other personal data); 9% was fraud related to taxes, driver's license, etc.; 13% involved other types of identity theft. In 2008, 10 million people in the US alone have become victims of identity theft, i.e. 22% more than in 2007. In the same year (2007) 1.6 million households in USA experienced theft which was unrelated to credit card losses, but instead - to breaches of bank accounts or debit card accounts. Moreover, 38 to 48% of people notice the theft of their electronic identity within 3 months, while 9 to 18% does not notice this fact for 4 or more years.

The seriousness and scale of electronic identity loss was noted also in recent years by decision-making authorities in the European Union, including the European Commission. One of the many initiatives aimed at solving the problems with assigning and distributing electronic identity is the pilot program STORK [4] (***S****ecure iden****T****ity cr****O****ss bo****R****ders lin****K****ed*) which pertains to a cross-border recognition of existing national electronic identity systems (eID), thus allowing for obtaining access to public services in member states. Several dozen million people in EU use national eID identity cards when accessing services related to social insurance, filling in tax returns and other services. Thus, the project is related to electronic identity management (eIDM) through a system which is a federation of already existing systems. This is a fundamental difference in comparison to other projects realized by EU regarding identity,

e.g. FIDS, PRIME, PRIMELIFE, PICOS, SWIFT, SWEB where attention was paid more to attempts at standardizing marked off, centralized systems. The European Commission argues that eIDM is a foundation stone in the implementation of a full scope of eGovernment services for citizens, as well as for businesses in the entire European Union [5]. It should also be mentioned that a number of interesting initiatives were created thus far which, according to their creators, are aimed at facilitating the determination of a person's identity in order to provide them with access to specific resources. One such initiative is Identity 2.0 [6]; its creator, Dick Hardt aims at co-creating and supporting an OpenID architecture [7] (architecture of dispersed authentication and distribution of identity in the Internet). The idea of the proposed mechanism is simple: instead of remembering countless numbers of login-password pairs for different internet services, it is enough that the user creates an account on an OpenID server of their choice. Wanting to log in on any internet service, the user will be redirected to an OpenID server by providing his/her OpenID ID in the form of a URL address. The OpenID server authorizes the logging-in operation. An organization supporting this technology was created under the name OpenID Europe [8].

According to Elbirt [9], the recommended measures of electronic identity protection include personal data protection (social insurance numbers, last name (including family name), date of birth, previous residence address, driver's license number). Consultations are also recommended in institutions dedicated to protecting personal data in order to familiarize oneself with forms of illegal obtaining of personal data by third parties and with recommendations pertaining to revealing one's personal data. Another protection measure is to retain information confidentiality (separating personal data from business data, restricting and not revealing data to other people or businesses, and in the case of data loss – immediate alarming and reporting the fact of data loss). In order to minimize the risk of an unauthorized leak of personal data, tracking carried out transactions is recommended, e.g. payments with credit cards by periodic printouts of statements. Moreover, it is advisable to properly store personal information by encrypting files and a periodic change of passwords. Furthermore, it is advisable to destroy old documents, printouts from bank accounts using shredders. In the case of electronic documents the disposal should occur by means of special computer programs which irretrievably delete a document from the computer's memory.

3 Hidden authentication

The need to introduce effective authentication procedures is especially noticeable in military heterogenic systems which offer speech transmission services. A complex solution to the problem of authentication is called for especially by short-term prototyping and implementation of new systems in accordance with the COTS (*commercial, off-the shelf*) rule. The use of the network's multi-service character in the concept of *Next Generation Networks* (NGN) may lead to a new approach to designing innovative and effective solutions in subscriber authentication.

The aim of this paper is to show an alternative for current solutions on verifying electronic identity and its distribution in telecommunication networks, using the technique of information hiding. The paper presents exemplary scenarios of authenticating in a VHF radio link with the use of a hidden transmission of additional information. Results are presented of tests and experiments on two algorithms of watermarking speech signals which enable the embedding of additional information (the personal ID of the user) in the call (original) signal, as well as the extraction of this additional information from the received signal. The personal identification number (PIN) is represented by a digital watermark with a determined *data payload* which is inaudible in the presence of an original call signal. One of the audio watermarking methods is called *Drift Correction Method* (DCM) [10] which has unique parameters in terms of the obtained data payload and robustness against degrading factors and which is offered together with the described new, one-stage psychoacoustic correction system as a basis of a new authentication standard in radio systems and internet telecommunications. A patent was obtained in the Patent Office of the Republic of Poland [11] for the method of one-stage psychoacoustic correction. Because of the confirmed generated high data payload of the watermark, the drift correction method opens up a path for carrying out not only hidden, but also confidential acoustic transmissions. Watermark *data payloads* higher than 128 bits enable the use of known methods for encrypting binary strings with a determined cryptographic power before they are embedded into the original signal in the form of a watermark. Thus, the use of a hidden and confidential transmission significantly increases the safety of the carried out transmission for especially important transmissions, e.g. radio transmissions.

4 Experimental results

In order to realize the procedure of subscriber authentication in telecommunication links, a decision has been made to use one of the known authentication models which involve transmitting a binary signature to subscribers who exchange correspondence between themselves. However, it must be noted that the authentication model has been supplemented by a hidden signature sent through a watermark added to the call signal, which constitutes its significant modification. This chapter presents the results of experiments related to subscriber authentication with the use of an information hiding technique for a radio VHF link. The link uses the drift correction modulation as a method of embedding a watermark in the original signal. Moreover, results were discussed from the carried out experiments.

A supplementation to the carried out experiment is a test of the hardware encoder and decoder of the watermark – a Personal Trusted Terminal (PTT) with an algorithm based on the phase angle scanner method that uses a detector of spectral line amplitude instead of a detector of phase angle mistuning values. A description of the method using the detector of spectral line amplitude, as well as a description of the hardware encoder and decoder of the watermark may be found in [12] and [13–17].

Aim: To check the robustness of the watermark embedded in an acoustic signal by a hardware encoder against a transmission in a VHF link.

Description of experiment: a signal with an embedded watermark was decoded; a hardware encoder and decoder were used which are based on the method of forming and detecting a spectral line amplitude [12]. The watermarked signal was sent through an ultrashort wave link (VHF) with the use of Radmor (TRC 9200) military radio stations. Transmission parameters: sending and receiving frequency 30MHz (simplex work), F3E modulation in an HLG working mode of the radio station at a determined frequency. A speech signal was transmitted in *real time* for a watermark data payload P=15 [b] and with a duration of the transmitted signal t=30 [s]. As a result of the experiment's realization, a chart was obtained of the effectiveness of decoding the correspondent's signature. The diagram of the VHF link was presented in Figure 1.

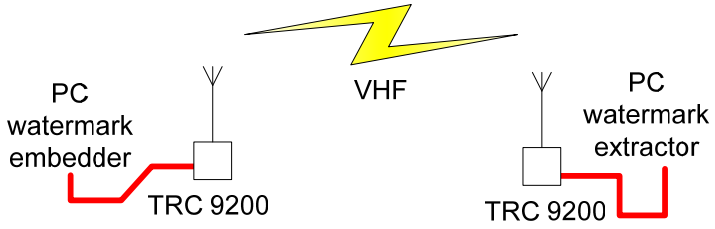


Figure 1: Scheme of the realized VHF radio link with the use of watermark encoder and decoder.

In experiment the software encoder of the watermark embedded the watermark signal into the speech signal in real time. In the application, the WM binary signature denoted the PIN number of the radio correspondent. At the receiving side of the radio link the watermark decoder compared the decoded PIN with the PIN declared in the on the receiving side. In the case of a coherence of both PIN numbers: the decoded with the declared one, the device displayed information about authenticating the radio correspondent. Tests were carried out which involved transmitting and decoding the PIN number in a radio link for 10 radio correspondents with assigned PIN values. Each correspondent read out a fragment of a radio telegram for 30 seconds with a pressed switch (radio station work mode on transmitting). Each radio correspondent was authenticated three times in sessions with numbers 1-5. Only in the case of two radio correspondents was an erroneously decoded PIN obtained in one in five sessions. In the case of the remaining eight radio correspondents, a positive authentication was obtained in each of the five sessions. Thus, the authentication effectiveness in this experiment equalled 96%. An effectiveness of 100% will be required from a professional, commercial authentication system.

4.1 Personal trusted terminal features

The aim of the device is to transmit a Personal Identification Number (PIN) through a telecommunication link in the form of a binary signature assigned to a given subscriber. The binary signature (PIN) is represented by a watermark signal which is sent together with the speech signal and which is inaudible in its presence. At the receiving side of the link, the watermark is decoded and the obtained binary signature is compared with the one that is stored in the database of the handset. In the case of a coherence of the received signature with the declared one, the caller is authorized during the realized connection. Then the LCD of the handset displays a suitable message. If only one of the interlocutors has a handset with the function of hidden authorization, conducting a conversation is possible like in a standard phone; however, the user of a handset with an authorization function receives a warning message regarding the lack of authorization due to no watermarked signal. The system is intended to work with military battlefield VHF radio stations. The system has advanced mechanisms of reducing resulting frequency mistuning and psychoacoustic correction. Using a standard handset without an authorization function the user is never sure whether he is speaking to the dedicated caller, even despite his recognizable features: the tone of the voice, intonation, timbre, base frequency, formant frequency distribution, etc. This happens because of the possibility of applying an artificial voice synthesis technique or changing the voice features of a third person during conversation. This is even more possible because the quality of the radio link is low and we do not always know precisely the voice of our interlocutor; thus, a change of the caller's identity is facilitated.

The basic task of the worked out handset is especially to ensure a safe exchange of messages between the authenticated subscribers in a radio link without openly using cryptography mechanisms. Third persons impersonating the identity of a given interlocutor are not informed about the process and result of the authentication; thus, the conversation with the unauthorized interlocutor may be continued without revealing important messages or it may be finished.

5 Conclusions

The presented authentication VHF system does not interfere with the telecommunication infrastructure of currently functioning radio systems. The lack of a special handset with a hidden authorization function does not, in any way, prevent the realization of a connection provided the correspondent has a regular handset. An attempt to discover a realized connection with a watermark sent in the background of the call signal is made more difficult because of the inaudibility of the watermark and the necessity to analyze the speech signal in terms of a hidden message in all call tracks realized at a given moment, e.g. during telecommunication rush hour. The digital watermark is perceptually transparent and inaudible on the host signal's presence and is robust against intentional and unintentional attacks. The developed system allows for transmission together with the speech signal a watermark signal with dedicated

data payload. Watermark signal represents PIN assigned to the specific subscriber using encryption based on a single-use key so that the PIN number is changing during each call session.

Acknowledgement

This paper has been co-financed from science funds granted within the years 2010-2012 as a research project of the Ministry of Science and Higher Education of the Republic of Poland No. 0181/R/T00/2010/12

References

[1] http://eric-diehl.com
[2] Javelin Strategy & Research, "2009 Identity Fraud Survey Report", Feb. 2009, www.javelinstrategy.com
[3] http://www.spendonlife.com/guide/2009-identity-theft-statistics
[4] http://www.eid-stork.eu/
[5] http://ec.europa.eu/information_society/activities/egovernment/policy
[6] http://identity20.com
[7] http://openid.net
[8] www.openideurope.eu
[9] A.J. Elbirt, *Who Are You? How to Protect Against Identity Theft,* IEEE Technology and Society Magazine, vol. 24, Issue 2, 2005, p. 5-8
[10] Z. Piotrowski, Drift Correction Modulation scheme for digital audio watermarking. Proceedings 2010 Second International conference on Multimedia Information Networking and Security MINES 2010, Nanjing, China, 4-6 November 2010, ISBN: 978-0-7695-4258-4, IEEE Computer Society, Conference Publishing Services (CPS), pp. 392-397
[11] Z. Piotrowski, Gajewski P., European Patent Application no. 09151967.8 (EP 2 085 964 A2), "*Method and apparatus for subscriber authorization and audio message integrity verification*". European Patent Office
[12] Gajewski P., Łopatka J., Piotrowski Z., *A New method of frequency offset correction using coherent averaging*, Journal of Telecommunications And Information Technology, 1/2005, National Institute of Telecommunications, ISSN 1509-4553, Warsaw 2005
[13] Piotrowski Z., Effectiveness of the frequency offset computation procedure, Elektronika, nr 3/2010, s.76-79, Wydawnictwo Sigma-NOT, 2010
[14] Piotrowski Z., Zagoździński L., Gajewski P., Nowosielski L.: *Handset with hidden authorization function*, European DSP Education & Research Symposium EDERS 2008, Proceedings, pp.201-205, Published by Texas Instruments, ISBN: 978-0-9552047-3-9
[15] Piotrowski Z., Nowosielski L., Zagoździński L., Gajewski P.: *Electromagnetic Compatibility of the Military Handset with Hidden Authorization Function Based on MIL-STD-461D Results*, Progress In Electromagnetics Research Symposium PIERS 2008 Cambridge

Proceedings, pp.103-107, Published by The Electromagnetics Academy, ISBN: 978-1-934142-06-6, ISSN: 1559-9450

[16] Z. Piotrowski, The National Network-Centric System and its components in the age of information warfare, Safety and Security Engineering III, WIT Press 2009, pp. 301-309, ISBN: 978-1-84564-193-1

[17] Z. Piotrowski, K. Sawicki, M. Bednarczyk, and P. Gajewski, *New Hidden and Secure Data Transmission Method Proposal for Military IEEE 802.11 Network* Conference Proceedings, The 6th International Conference in Intelligent Information Hiding and Multimedia Signal Processing, IIH-MSP 2010, Darmstadt Germany

This page intentionally left blank

Analytical model of a SAW gas sensor

M. Urbańczyk
Faculty of Electrical Engineering,
Silesian University of Technology, Poland

Abstract

SAW gas sensors are attractive because of their remarkable sensitivity due to changes of the boundary conditions (mechanical and electrical in the acoustoelectric effect) propagating of the Rayleigh wave, introduced by the interaction of a thin chemically active sensor film with gas molecules. This unusual sensitivity results from the fact that most of the acoustic wave energy is concentrated near the waveguide surface within approximately one or two wavelengths. In the paper a new theoretical model of analysing a SAW gas sensor is presented. The effect of SAW velocity changes depends on the profile concentration of diffused gas molecules in the porous sensor film. Basing on these analytical results, the sensor structure can be optimized. Some numerical results are shown.
Keywords: gas sensor, SAW, acoustoelectric effect, Knudsen's diffusion in porous film, numerical modelling.

1 Introduction

A very interesting feature of SAW sensors is the fact that a layered sensor structure on a piezoelectric substrate provides new possibilities of detecting gas making use of the acoustoelectric coupling between the Rayleigh waves and the free charge carriers in the semiconductor sensor layer. Using a configuration with a dual delay line and an adequately chosen active layer, a sensor with a high sensitivity and good temperature stability can be designed [1, 2]. SAW gas sensors are attractive because of their remarkable sensitivity due to changes of the boundary conditions of propagating the surface wave. This unusual sensitivity results from the simple fact that most of the acoustic wave energy is concentrated near the crystal surface within approximately one or two wavelengths.

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110431

Any change of the physical properties of the thin active layer placed on a piezoelectric substrate can affect the SAW propagation. However, from the practical point of view, only the following two affects are potentially significant, namely a change in the mass of the layer and a change of its electrical conductivity, which cause a significant change in the velocity and attenuation of SAW. These two effects occur simultaneously. In semiconductor layers the electric effect is much greater (several times, depending on the gas concentration) [2].

In order to optimize the structure of the sensor it is important to get an analytical model of the SAW sensor. This paper provides theoretical considerations concerning the acoustoelectric effect in the piezoelectric acoustic delay line and semiconducting sensor layer configuration. This will be the starting point to construct a multilayer analytical model. The sensor layer with absorbed gas molecules may be treated as a stack of partial layers with a different electrical conductivity due to the concentration profile of gas molecules in the sensor layer. These partial layers are positioned at different distances from the piezoelectric surface.

The paper summarizes the acoustoelectric theory, i.e. Ingebrigtsen's formula [3], the impedance transformation law, gas concentration profiles, and predicts the influence of a thin semiconductor sensor layer on the SAW velocity. In the paper a new theory of analyzing a SAW gas sensor is proposed. Optimizing this structure, we obtain a high sensitivity of specific gas detection. In order to reach a high sensitivity we must determine the thickness and porosity of the layer for the given gas molecules, the temperature of operation, and other parameters.

2 Acoustoelectric effect in the piezoelectric – semiconductor layer

In the case of small disturbances, both mass and the electrical "load" may be considered separately. The total effect of a relative change of the wave vector $\Delta k/k_0$ and the velocity of propagation $\Delta v/v_0$ is the sum of both these components [4]:

$$\frac{\Delta k}{k_0} \approx \left(\frac{\Delta k}{k_0}\right)_m + \left(\frac{\Delta k}{k_0}\right)_\sigma \tag{1}$$

By means of the perturbation analysis the contribution of each of these effects can be determined. We assume that the mass of the sensor layer may be neglected. Further on only the electric effect will be taken into consideration.

2.1 Electrical surface perturbations

If the electrical boundary conditions are perturbed only at the upper surface $y=0$, the perturbation formula is [5]:

$$(\Delta k)_\sigma \approx \frac{\omega\left[\varphi'(0)D_y^*(0) - \varphi^*(0)D_y'(0)\right]}{4P} \tag{2}$$

where the index σ refers to the change of the wave number due to electrical surface perturbations.

Unperturbed electrical boundary conditions are usually neither short-circuit [$\varphi(0) = 0$] nor open–circuit [$D_y(0)=0$] ones, and it is not clear whether the potential or the electrical displacement should be approximated by its unperturbed value. The best agreement between the perturbation theory and exact numerical calculations is obtained by using the so-called weak-coupling approximation, in which the stress field T is assumed to be unchanged by the perturbation [3, 5]. It is most convenient to express the electrical boundary conditions in terms of the electrical surface impedance per unit area:

$$Z_E(0) = \left(\frac{\varphi}{i\omega D_y} \right)_{y=0} \tag{3}$$

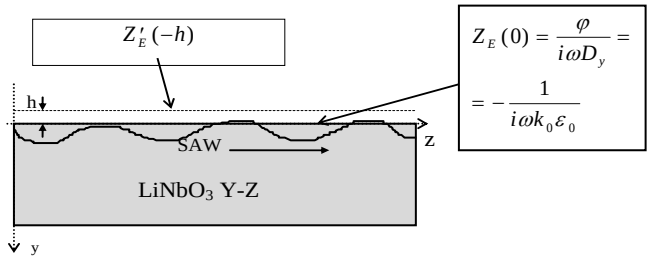


Figure 1: Electrical boundary conditions.

Let us assume that unperturbed electrical boundary conditions correspond to a free substrate surface, i.e. the region above the substrate (y<0 in fig. 1) is a vacuum and extends to y→∞. The space-charge potential satisfying the Laplace equation is reduced in this region to the form:

$$\nabla^2 \Phi = 0 \tag{4}$$

The unperturbed potential function is therefore:

$$\Phi = \Phi_R(y) e^{-ikz} = e^{ky} e^{-ikz}, \text{y<0} \tag{5}$$

and the normal component of electrical displacement is:

$$D_y = -k\varepsilon_0 e^{ky} e^{-ikz}, \text{y<0} \tag{6}$$

Consequently, the unperturbed surface impedance is:

$$Z_E(0) = -\frac{1}{i\omega k_0 \varepsilon_0} \tag{7}$$

The perturbed normalized surface impedance is:

$$z'_E(0) = \frac{Z'_E(0)}{|Z_E(0)|} = -ik_0\varepsilon_0 \left(\frac{\varphi'}{D'_y} \right)_{y=0} \tag{8}$$

where ‘ indicates perturbed quantities. The potential $\varphi'(0)$of the perturbed fields and electrical displacement $D_y'(0)$ are now related to the unperturbed fields:

$$\begin{aligned} \varphi'(0) &= \varphi(0) + A \\ D'_y(0) &= D_y(0) + k_0 \varepsilon_p^T A \end{aligned} \tag{9}$$

where $\varepsilon_p^T = \sqrt{\varepsilon_{yy}^T \varepsilon_{zz}^T - \left(\varepsilon_{yz}^T\right)^2}$ and $k_0' = k_0$ has been approximated.

Now the perturbed potential φ' and electrical displacement D'_y on the surface may be expressed by the formula:

$$\varphi'(0) = -iz'_E(0)\frac{\left(\varepsilon_0 + \varepsilon_p^T\right)}{\varepsilon_0 - i\varepsilon_p^T z'_E(0)}\varphi(0) \tag{10}$$

$$D'_y(0) = -\frac{k_0\varepsilon_0\left(\varepsilon_0 + \varepsilon_p^T\right)}{\varepsilon_0 - i\varepsilon_p^T z'_E(0)}\varphi(0) \tag{11}$$

Finally, we obtain:

$$\frac{\Delta k}{k_0} = -\left(\frac{\Delta v}{v_0}\right)_{sc}\frac{1+iz'_E(0)}{1-i\dfrac{\varepsilon_p^T}{\varepsilon_0}z'_E(0)} = \frac{K^2}{2}\frac{1+iz'_E(0)}{1-i\dfrac{\varepsilon_p^T}{\varepsilon_0}z'_E(0)} \tag{12}$$

$K^2 = 2\left(\dfrac{\Delta v}{v_0}\right)_{sc}$ - the electromechanical coupling factor, and the index *sc* indicate perturbation due to an electrical short circuit on the boundary. Eq. (12) is the Ingebrigtsen formula for electrical surface perturbations.

2.2 Thin semiconducting sensor layer

In the case of the semiconducting layer the surface impedance may be determined considering the motion of the charge carriers in the layer. Let us assume that the semiconductor layer on the piezoelectric substrate is thinner than a Debye length. In the one-dimensional case the equation of the current density in the direction *z* (fig.1) is expressed as [6]:

$$I_z = \rho_{01}\mu E_z - D\frac{d\rho_1}{dz} \tag{13}$$

where ρ_{01} and ρ_1 are the intrinsic and whole charge density per unit length, respectively. The quantities μ and D are the carrier mobility and diffusion quantity, respectively, E_z – electric intensity field in the direction *z*. For the time dependence of the E_z expressed as *exp i(ωt – kz)* the continuity equation [8], concerning current density in the sensor layer is:

$$\frac{dI_z}{dz} + i\omega\rho_l = 0 \tag{14}$$

The surface density of charge carriers $\rho_s = \rho_l/a$ (*a* – width of the layer) induced in the layer by the electric field E_z will be:

$$\rho_s = \frac{\left(\dfrac{\omega_c}{\omega}\right)\varepsilon\, kh}{1 - i\dfrac{\omega}{\omega_D}}E_z \tag{15}$$

where: $\omega_c = \dfrac{\mu|\rho_0|}{\varepsilon}$ relaxation frequency, $\omega_D = \dfrac{v_0^2}{D} = \dfrac{k_B T}{q}\mu$ diffusion frequency, k_B-Boltzman constant, *T* – temperature, *q* - electric charge.

From the boundary conditions of the space-charge field in a single semiconductor layer in the plane $y=0$ results the continuity of the potential $\varphi(0^-) = \varphi(0^+)$ and the condition of dielectric displacement $D_y(0^-) - D_y(0^+) = \rho_s$, where $D_y(0^-) = \varepsilon_0 E_y(0^-)$, and assuming the electrostatic condition we can write that $E_z^- = ik_0\varphi(y,z)$.By applying Gauss' law and the potential continuity in the plane $y=0$, the impedance $z_E'(0)$ can be expressed in the following form:

$$z_E'(0) = \left(-i + \frac{\frac{\sigma_s}{\varepsilon_0 v_0}}{1 - \mathrm{i}\frac{\omega}{\omega_\mathrm{D}}} \right)^{-1} \tag{16}$$

where $\sigma_s = \sigma h$ is the surface conductivity of the layer.

In semiconducting sensor layers the diffusion effect is usually very small (D~0) and the surface conductivity is independent of the frequency of the wave propagation. In this case the surface impedance $z_E'(0)$ is only a function of the surface conductivity σ_s:

$$z_E'(0) = \left(-i + \frac{\sigma_s}{\varepsilon_0 v_0} \right)^{-1} \tag{17}$$

3 Profile of the gas concentration in the sensor layer

In gas sensors the effect of SAW velocity changes vs. the surface conductivity depends on the profile of concentration of diffused gas molecules in the porous film. The mechanism of gas diffusion through a porous material depends on the size of the pores and the type of surface diffusion. In porous sensing material usually Knudsen's diffusion is assumed. This type of diffusion depends on the pore radius and prevails in pores ranging from 1 to 100 nm in radius [7]. The Knudsen diffusion constant, D_K, depends on the molecular weight of the diffusing gas, M, the pore radius, r, temperature, T, and the universal gas constant, R, as follows:

$$D_K = \frac{4r}{3}\sqrt{\frac{2RT}{\pi M}} \tag{18}$$

If the pore radius r>2 nm, a free molecular flow exists. That means that the gas molecules collide more frequently with the boundaries than with other gas molecules. The original studies of a free-molecule flow were limited to small holes in very thin plates. Gas molecules are consumed rapidly or slowly due to the surface reaction in the sensing layer [7].

Let us consider the gas diffusion in a porous thin semiconducting film, as shown in fig. 2.Two assumptions, i.e. Knudsen diffusion and first-order surface reaction, allow to formulate the well-known diffusion equation [7, 9]:

$$\frac{\partial C_A}{\partial t} = D_K \frac{\partial^2 C_A}{\partial x^2} - kC_A \tag{19}$$

where C_A – is the concentration of target gas, t – time, D_K – Knudsen diffusion constant, y distance from the bottom layer, counted from the piezoelectric substrate, k is the rate reaction constant.

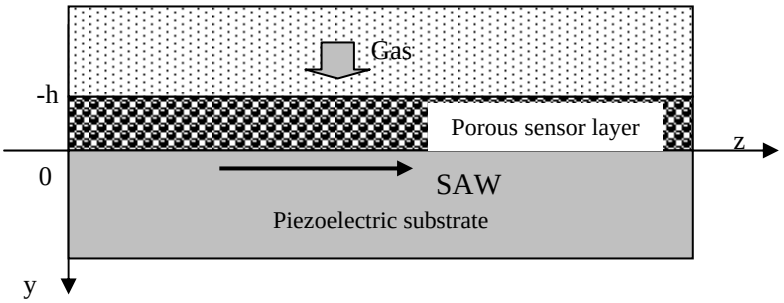


Figure 2: Model of a gas sensor with a sensitive layer on the piezoelectric substrate.

At steady–state conditions the general solution of the diffusion equation is:

$$C_A = C_1 \exp\left(y\sqrt{\frac{k}{D_K}} \right) + C_2 \exp\left(-y\sqrt{\frac{k}{D_K}} \right) \tag{20}$$

whereC_1 and C_2 are integral constants.

At boundary conditions on the surface (y =-h)$C_A = C_{As}$ and $\partial C/\partial y = 0$the profile of the gas molecules concentration in the sensor layer is:

$$C_A(y) = C_{As} \frac{\cosh(|y|\sqrt{k/D_k})}{\cosh(|-h|\sqrt{k/D_k})} \tag{21}$$

C_{As} is the target gas concentration outside the film on the surface y = -h. The concentration profile depends on the thickness of the sensor layer and the constants k and D_K. Fig. 3 presents an example of the molecule gas concentration profile. Similar as in resistance sensors let us now assume that the electrical conductance $\sigma(y)$of the sensor layer is linear to the gas concentration $C_A(y)$ [7]:

$$\sigma(y) = \sigma_0(1 \pm a \cdot C_A(y)) \tag{22}$$

where σ_0 is the initial layer conductance in the air, a is the sensitivity coefficient and the sign ± denotes the combination layer type conductivity and oxidation or reduction properties of the target gas. Experimental data for H_2 have been presented by Sakai e*t al.* [7], when C_{As} was fixed at 800 ppm at a temperature of 350°C, and assumed to be a = 1 ppm^{-1}. For these data the sensitivity data for H_2 fit fairly well to the correlation line for $\sqrt{k/D_K} = 0{,}01$ nm^{-1}.

The electrical conductance of the whole film is obtained by integrating σ(y) over the whole range of y (y = -h; 0).That treatment has been proposed by Sakai, Williams and Hilger [7, 10]. A semiconductor layer in the SAW sensor cannot be treated in the same way.Because the profile of molecule gas concentration in a semiconducting sensor layer changes with the distance from the piezoelectric substrate (fig. 3), the acoustoelectric interaction differs in every sheet of the layer. In order to analyze such a sensor layer we assume that the film is a uniform stack of infinitesimally thin sheets (fig. 4) with a variable concentration of gas molecules and a different electric conductance. Each sub-layer is in

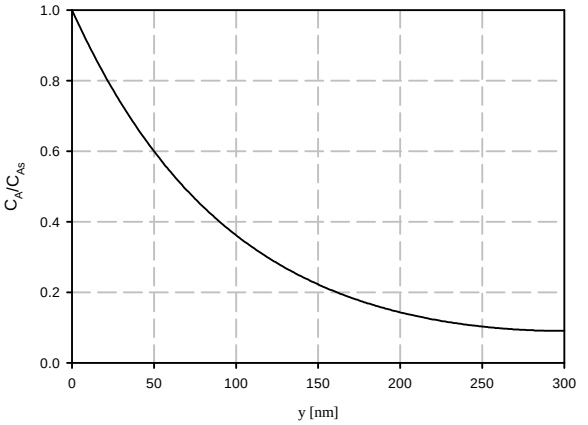


Figure 3: Gas concentration profile in a layer 300 nm thick and for $\sqrt{k/D_K} = 0,01.$

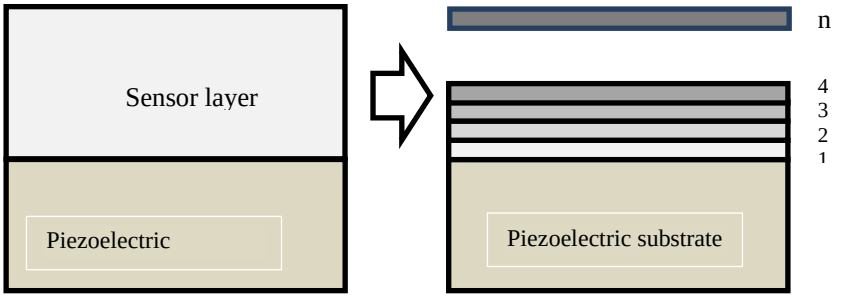


Figure 4: Model of a sensor layer divided into n-sub-layers.

another distance from the piezoelectric wave-guide. The schematic diagram in fig. 5 presents the way of analyzing the SAW gas sensor.

The resultant admittance is calculated basing on the impedance transformation law of each sub-layer to the surface of the wave guide. The surface conductance of the sub-layers is calculated applying the profile of concentration of the gas and Eq. (22). This resultant impedance applies the Ingebrigtsen formula to calculate changes in the propagation velocity of the surface wave. The transformation impedance law is derived from the general expressions for φ' and D_y' in the region *0>y>-h* [5, 11].

The normalized impedance at any arbitrary plane *y* is [5]:

$$z_E^{'}(y) = \frac{-ik_0\varepsilon_0(e^{k_0 y} + \frac{B}{A}e^{-k_0 y})e^{-k_0 y}}{-k_0\varepsilon_0(e^{k_0 y} - \frac{B}{A}e^{-k_0 y})e^{-k_0 y}} \tag{23}$$

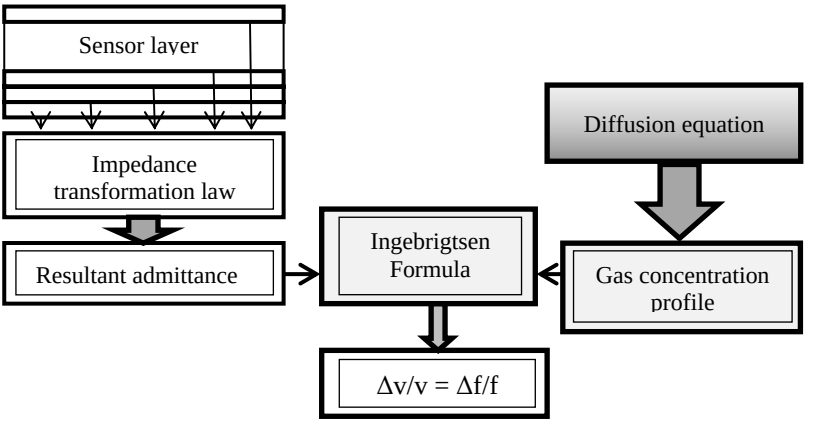


Figure 5: Analytical diagram of the SAW gas sensor.

The constant *B/A* is evaluated by setting *y*=-*h* in (23), and this gives the impedance transformation law for a single layer in the distance -*h*:

$$z_E^{'}(0) = \frac{i \cdot \tanh(k_0 h) + z_E^{'}(-h)}{1 - i \cdot z_E^{'}(-h) \cdot \tanh(k_0 h)} \tag{24}$$

It can be shown, that the Ingebrigtsen formula for *n*-sub-layers takes the following form [12]:

$$\frac{\Delta v}{v_0} = -\mathrm{Re}\left\{\frac{\Delta k}{k_0}\right\} = -\frac{K^2}{2} \frac{\sigma^2_{T_2}(1 \pm aC_A)^2 \left[1 + \sum_{i=1}^{n-1} f(y_i, \sigma(y_i))\right]^2}{\sigma^2_{T_2}(1 \pm aC_A)^2 \left[1 + \sum_{i=1}^{n-1} f(y_i, \sigma_{T_2})\right]^2 + \left[1 + \sum_{i=1}^{n-1} g(y_i, \sigma(y_i))\right]^2 (v_0 C_S)^2} \tag{25}$$

where *n* – number of sub-layers and $C_S = \varepsilon_0 + \varepsilon_p^T$, $\sigma_{T_2} = \sigma_{T_1} \exp\left(\frac{E_g}{2k} \cdot \frac{T_2 - T_1}{T_1 T_2}\right)$

$$T_1 = 300\text{K},\ \sigma_{T_1} = \sigma_0$$

$$f(y_i, \sigma_{CS}) = \sum_{i=1}^{n-1} \frac{1 - [\tanh(ky)]^2}{[1 + \tanh(ky)]^2 + \left[\tanh(ky) \cdot \frac{\sigma_{CS}}{\varepsilon_0 v_0}\right]^2}$$

$$g(y_i, \sigma_{CS}) = \sum_{i=1}^{n-1} \frac{[1 + \tanh(ky)]^2 + \tanh(ky) \cdot \left(\frac{\sigma_{CS}}{\varepsilon_0 v_0}\right)^2}{[1 + \tanh(ky)]^2 + \left[\tanh(ky) \cdot \frac{\sigma_{CS}}{\varepsilon_0 v_0}\right]^2} \tag{26}$$

4 Numerical results of modelling a SAW gas sensor

As an example of the adaptability of this theory the results of a numerical analysis have been presented. A thin layer of a semiconductor is placed on a Y-Z $LiNbO_3$ waveguide. Results show the changes in the velocity of propagation depending on the concentration of the surrounding gas (H_2, CO_2, NO_2, NH_3), the thickness of the semiconductor layer (WO_3**)**, the pore radius and the sensor temperature. In the numerical analysis the following values of the involved constants have been assumed: $\sigma_s = v_0 C_S = 1.6 \times 10^{-8} [\Omega^{-1}]$, sensitivity coefficient a = 1 [ppm^{-1}], E_g=2,7 [eV] for WO_3 as for solid material [13].

Experimental results confirm the analytical model of the SAW sensor (Fig.8).

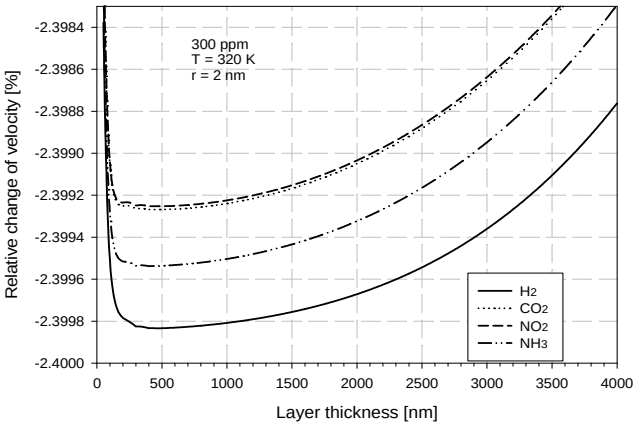


Figure 6: Changes of the SAW velocity propagation vs. the thickness of the WO_3 sensor layer at 300 ppm H_2, CO_2, NO_2 or NH_3. Pore radius r = 2 nm.

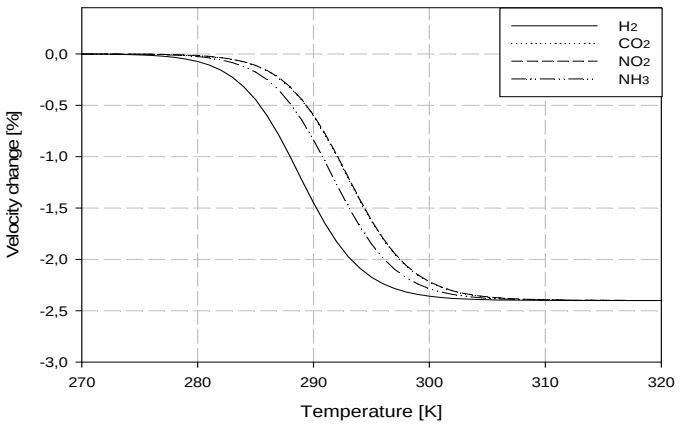
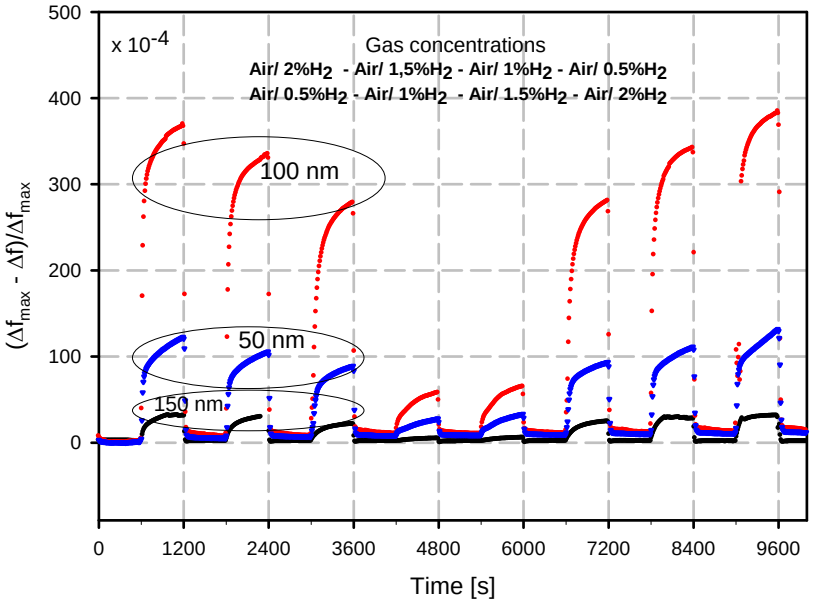
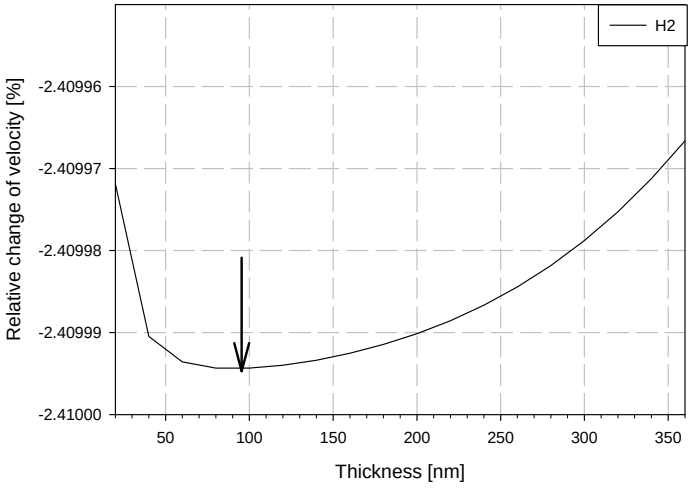


Figure 7: Changes of the SAW velocity propagation vs. the temperature of the WO_3 sensor layer for H_2, CO_2, NO_2, NH_3 gases.



(a)



(b)

Figure 8: a. Experimental data concerning a WO_3 layer with different thickness (50, 100, 150 nm). b. Numerical data, 0.5 % H_2 gas.

According to numerical results the optimal thickness of the sensor layer is about 96 nm. Experimental results reach their maximum in a 100 nm thick layer.

5 Conclusions

In this paper a new analytical model of a SAW gas sensor and numerical results have been shown for the gases H_2, CO_2, NO_2, NH_3 and sensing layers WO_3.The profile of the concentration of gas in the sensor layer has been applied in order to model the acoustoelectric effect in the SAW gas sensor. A porous semiconductor layer has been divided into sub-layers. The influence of the impedance above the piezoelectric substrate on the relative change of the SAW velocity has been calculated. In the sensor layer the Knudsen diffusion is assumed.

Analytical results are compatible with experimental results. The new analytical model may be used to optimize the structure of SAW gas sensors.

Acknowledgement

This work is financed by a grant from the Ministry of Science and Higher Education No. N N505 374237.

References

[1] W. Jakubik, M. Urbańczyk, E. Maciak, T. Pustelny, Bilayer structures of NiO_x and PdIn Surface Acoustic Wave and Electrical gas sensor systems, *Acta Phys. Pol.* A, vol. 116, no 3, pp. 315-320, 2009.

[2] W. Jakubik, Investigation of thin film structure of WO_3 and WO_3 with Pd for hydrogen detection in a surface acoustic wave sensor system, *Thin Solid Films***515**, pp. 8345-8350, 2007.

[3] G. Kino, T. Reeder, A Normal Mode Theory for the Rayleigh Wave Amplifier, *IEEE Transaction on Electron Devices*, **Ed-18**, 10, October 1971.

[4] W. Jakubik, M. Urbańczyk, The electrical and mass effect in gas sensors of the SAW type, *J. Tech. Phys*., **38(3)**, pp. 589-596, 1997.

[5] B.A. Auld, *Acoustic Fields and Waves*, vol. 2, J Willey and Sons, NY1973.

[6] D. L. White, Amplification of ultrasonic waves in piezoelectric semiconductors, *J. Appl. Phys*., Aug. 1962.

[7] Go Sakai, Naoki Matsunaga, Engo Shimanoe, Noboru Yamazone, Theory of gas-diffusion controlled sensitivity for thin film semiconductor gas sensor. *Sensors and Actuators*, **B 80**, pp. 125-131, 2001.

[8] K.M. Lakin, H.J. Shaw, Surface Wave Delay Line Amplifiers, *IEEE Trans.* **MTT-17**, pp. 912-920, 1969.

[9] J. W. Gardner, A non – linear diffusion – reaction model of electrical conduction in semiconductor gas sensors, *Sensors and Actuators*, **B1**, pp. 166-170, 1990.

[10] D. E. Williams, A. Hilger, Solid state gas sensors, Bristol, 1987.

[11] T. Hejczyk, M. Urbańczyk, W. Jakubik, Semiconductor sensor layer in SAW gas sensors configuration, *Acta Phys. Pol. A*, **118(6)**, pp. 1153-1157, 2010.
[12] T. Hejczyk, M. Urbańczyk, W. Jakubik, Analytical model of semiconductor sensor layers in SAW gas sensors, *Acta Phys. Pol. A*, **118 (6)**, pp. 1148-1152, 2010.
[13] J. Portier, H.S. Hilal, Thermodynamic correlations and band gap calculations in metal oxides, *Progress in Solid State Chemistry*, **32**, pp. 207-217, 2004.

Section 7
Advances in measurements and data acquisition

This page intentionally left blank

Advances in low cost marine technologies

M. Marcelli[1], V. Piermattei[1] & G. Zappalà[1,2]
[1]*DECOS Experimental Oceanology and Marine Ecology Laboratory, Tuscia University, Viterbo, Italy*
[2]*Istituto per l'Ambiente Marino Costiero (IAMC), National Research Council, Messina, Italy*

Abstract

In order to realize prevention and ecological risk analysis systems, the world environmental policy (UNEP, IMO, etc) is implementing complex decision systems based on economically sustainable activities including forecasting models, satellite images and sustainable observatory networks.

The oceanographic measurement networks play a key role both for satellite data calibration and mathematical models validation and feeding as well as to support the early warning systems for environmental pollution control and prevention.

The high costs of offshore mooring systems and traditional oceanographic cruises have suggested the use of VOS (Voluntary Observing Ships) to obtain affordable data.

Moreover, marine coastal areas can be monitored using small measure platforms integrating "on demand" the various measuring systems (meteorological stations, water samplers, automatic chemical analyzers, *in situ* and pumped oceanographic probes, etc).

For this purpose a big effort has been dedicated to the design, development and realization of new oceanographic devices.

This paper shows the advances in new technological devices: the TFLAP (Temperature Fluorescence LAunchable Probe) and the Automatic Multiple Launcher for expendable probes, to be used by VOS in open seas, and a coastal buoy to be moored near Civitavecchia, as a starting point of integrated coastal monitoring networks.

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110441

The described devices proved to be a cost effective solution to the need of modularity, flexibility, real time observing systems.
Keywords: observatory network, ocean color, VOS program, ocean monitoring, TFLAP, multiple launcher, expendable probes.

1 Introduction

The focus of international policy for the seas and oceans is steadily increasing due to the growth of consciousness, in the collective mentality, of their importance to man and his activities.

In addition to the use of the sea and its resources, that are traditionally exploited by humanity, and in addition to the essential role that the oceans and phytoplankton play in the climate balance, we constantly add new and important roles, such as exploitation of renewable energy and marine conservation of biodiversity, conservation meant as a pool of organic molecules of pharmaceutical compounds for possible future interest for humanity.

The oceanography sets itself as a synthesis science that can uniquely address the complex physical, geochemical and biological processes occurring in the sea: "Oceanography personifies interdisciplinary science of the 'blue' planet, Earth," as reported by Dickey and Bidigare [1].

The operational oceanography has been engaged in the development of new acquisition, transmission and assimilation systems in order to have the widest possible coverage of real time information, reflecting the guidelines of the World Meteorological Organization (WMO) and of the Intergovernmental Oceanographic Commission (IOC).

Moreover, physical and biological processes of marine ecosystems have a high spatial and temporal variability, whose study is possible only through high resolution and synoptic observations that require the simultaneous use of different platforms. "Until satellite appears, in the early 1970s, oceanography had endured a long period of "undersampling", so that, the most profund effect of the satellite oceanography was that for the first time ocean processes were adequately sampled" as stated by Munk [2].

However, the satellites, the data assimilation of mathematical models and the compliance with the standards for coastal studies, require in situ data. More than for the physical variables, the biological ones have to be observed *in situ* as reported by Marcelli *et al*. [3].

The attention to the state of marine environments is growing worldwide and the assessment of their resources needs always innovative methodologies, in order to develop policies and environmental governance for the sustainable management of the marine ecosystems.

Environment monitoring systems and networks were, in the years, designed and presented by various authors, among which, Carof *et al.* [4], Grisard [5], Eriksen [6], Griffiths *et al*. [7], Irish *et al*. [8], Paul [9], Seim *et al*. [10], Zappalà *et al.* [11], Nittis *et al.* [12], Crisafi *et al.* [13], Zappalà *et al.* [14, 15].

The methodologies for an operational monitoring of the oceans integrate satellite observations, drifting devices, Voluntary Observing Ships (VOS) and moorings:

- Satellite observation has the highest spatial and temporal sinopticity, but is limited by a series of factors: the lack of the third dimension, the low resolution in the coastal areas, the necessity of *in situ* data for sensors calibration.
- Drifter systems can measure physical, chemical and biological parameters, but their capability is limited by costs and power supply.
- The use of VOS in Ships Of Opportunity Program (SOOP) allows us to reduce operative costs of sea research, without affecting quality and increasing the spatial and temporal sinopticity of information.
- Mooring systems allow to collect a huge amount of data in near real-time for long-term monitoring.

At now, the main issue is to develop integrated measurement methods aimed at measuring physical and biological variables, integrating various networking, while simultaneously to develop new automated and low cost tools, primarily for the measurement of biological and chemical variables.

2 Materials and methods

Scientific activities were performed on two main lines: VOS equipments and moorings.

2.1 VOS systems

A Ships Of Opportunity Program (SOOP) was established in the Mediterranean Sea in September 1999, on behalf of the EC funded project Mediterranean Forecasting System – Pilot Project (MFS-PP), as described by Pinardi *et al.* [16].

The program is now evolving into a full Volunteer Observing Ships (VOS) system providing sea-temperature profiles and meteorological data.

The VOS, usually cargoes or ferries, allow us to repeat the same tracks, collecting data with a cyclic pattern.

The VOS program is based on the use of expendable probes, originally conceived for the Navy submarine warfare and later adopted by scientific community for oceanographic purposes.

The first expendable probe was the XBT (Expendable Bathy Thermograph), which was followed by XCTD (Expendable Conductivity, Temperature, and Depth profiling system) and ESV (Expendable Sound Velocimeter).

All of these probes are based on a consolidated technology: the probes can be released using manual or automatic launchers.

The depth reached by the XBT probes is estimated using a formula of the kind:

$$Z(t)=At-Bt^2$$

where Z(t) is the depth, t the time and A and B probe related coefficients.

The traditional commercial expendable probes are limited to the measurement of physical parameters, while also the biological ones are very important.

In this contest the TFLAP (Temperature Fluorescence LAunchable Probe), as described by Marcelli *et al*. [17, 18], was developed to fill this gap, ensuring both the quality of the measure and cost effectiveness.

TFLAP is an expendable probe which was developed for VOS purpose, launchable, like the others, from a moving ship. It measures pressure, temperature and fluorescence of chlorophyll *a* along the water column.

Unlike the traditional XBT probe, incorporating only the temperature sensor and having no electronics on board, the TFLAP contains a modular electronic system to manage the different measure channels. Commercial low cost components were selected and tested for the realization of the first prototypes.

The probe (Fig. 1) is built around a central tube made in anticorodal (an aluminium alloy) in which a measurement cell is obtained, hosting the sensors immersed in the water flow. An outer tube, coaxial with the first, is used to obtain a pressure case for the system electronics.

A zinc ogive is used to make the probe heavy enough to sink at a predefined rate, while a plastic tail (hosting the reel of electrical wire connecting the probe to the launching system) stabilizes its movement.

While the probe sinks, the wire dereels from the tail (lower) coil; in the meantime, the wire dereeling from the upper coil compensates the ship movement.

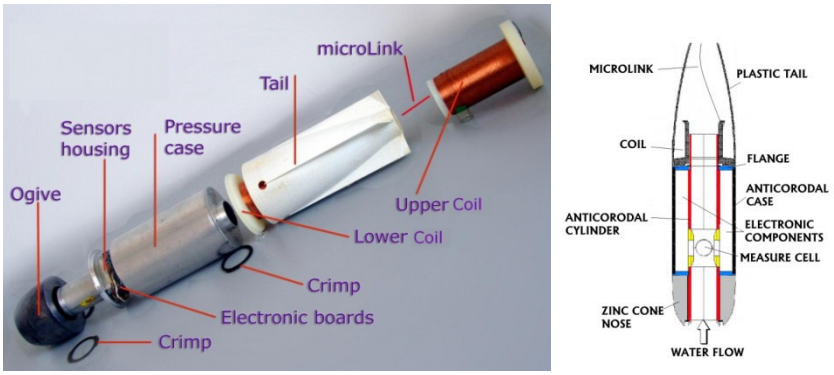


Figure 1: A partially assembled TFLAP and its section.

The measurement components include: a pressure transducer, a glass bulb temperature resistor for temperature measure; blue LEDs for fluorescence excitation, interferential filters and a photodiode for the fluorescence measure.

To better exploit the capabilities of TFLAP an enhanced version of the Automatic Multiple Launcher for expendable probes was set up.

The Multiple Launcher was designed in the Framework of MFS-TEP to enhance and automate the use of expendable probes from VOS. It has been described by Zappalà *et al*. [19], Zappalà and Manzella [20] and by Zappalà [21].

Originally designed to work with standard XBT probes, the system was enhanced to manage also TFLAPs, with a new serial interface and new communication and data display routines.
An example of Temperature and Fluorescence profiles measured using the TFLAP is shown in Fig. 2.

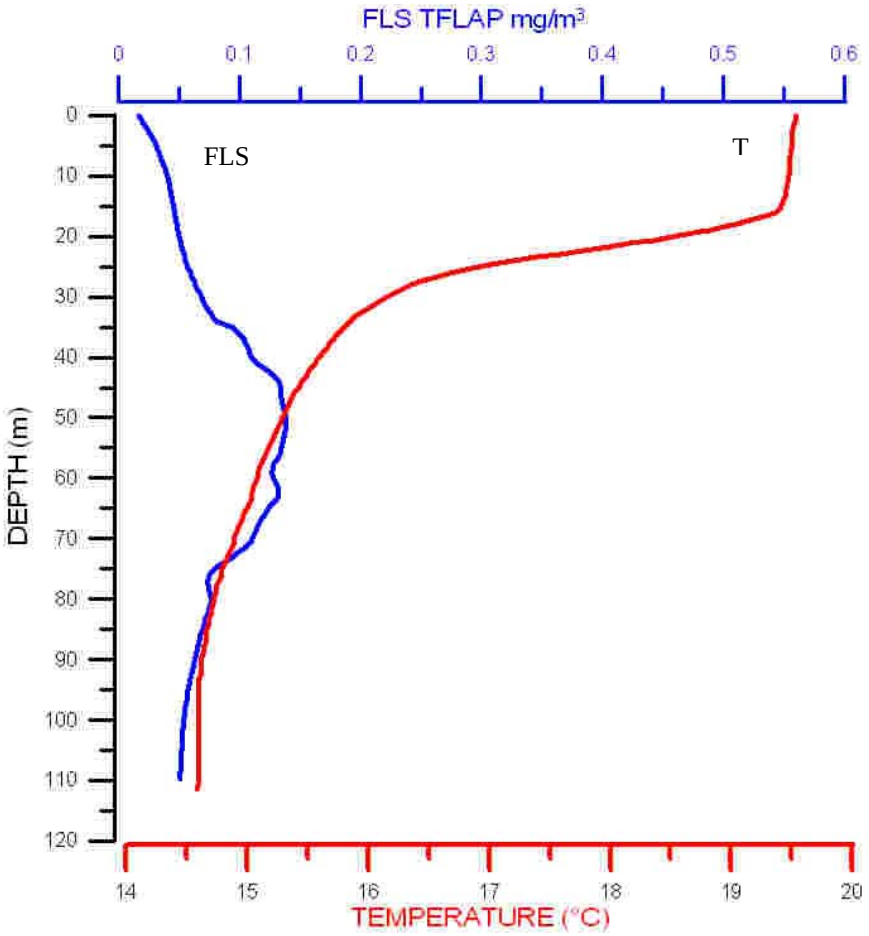


Figure 2: A plot of temperature and fluorescence profiles measured by the TFLAP.

2.2 Moorings

The buoy (or, more properly, the platform) was originally moored in Siracusa (Sicily) coastal waters as a part of a monitoring network funded by the Italian Ministry for University and Research, described by Zappalà *et al*. [22, 23].

Basically, it consisted in a 2 m x 4 m Orsogrill "floor", brought about 50 cm over the sea by ten horizontal axis cylindrical floaters distributed on the long sides; a semi-submersed chamber ("vessel") contained the measuring instruments, the batteries, the data acquisition and transmission and the pumping systems; an overhanging arc hosted meteorological sensors, the GSM antenna, the blinker and the radar reflector.

The power was given by four 12 V 50 Ah accumulators recharged using four 50 W solar panels, used also to close and protect the semi-submersed chamber.

At present, the platform is undergoing a series of major modification and a complete instrumentation refitting to be used for a new monitoring program which is going to start in Civitavecchia waters.

The platform before the modification is shown in Fig. 3 (left); the modified version is presented in Fig. 3 (right).

Figure 3: The platform before (left) and after (right) the modifications; it is evident the substitution of solar panels and the adding of the wind generator; being them totally submersed it is difficult to see the added floaters (two on each long side).

2.2.1 The power supply

To allow the use of more complex (and power consuming) instrumentation, taking into account possible problems with solar panels (long periods of bad weather, surface covering by oil film, damages caused by wave transported materials, etc.), the following improvements were made:

- a 200 W wind generator was added on the top the arc
- the 50 W panels were substituted with 80 W ones
- the power reserve was doubled adding four more accumulators.

Shunt or series voltage regulators, together with protecting diodes allow to properly recharge batteries, ensuring supply also in case of failure of one or more accumulators or generators.

The increase in weight has been balanced adding four auxiliary floaters that also, widening the submersed surface on short sides, enhance the buoy stability.

2.2.2 The data acquisition and transmission systems and the control software

The data acquisition system uses PC/104 boards that implement a PC-like architecture; the technological developments obtained in the years made available new, more powerful electronic boards to substitute the Intel 386 used in the first buoy version and the Pentium family board adopted for the first version of the "automatic multiple launcher for expendable probes".
In the current standard version up to 12 serial ports are available to connect measuring instruments, switched on and off by solid state and electromechanical relays.

To monitor power system status (battery level, solar panels and wind generator voltage) a 16 single ended channels, 12 bit resolution Analog to Digital Converter (ADC) board is installed; it is also possible to add a 16 bit resolution ADC board to connect sensors with voltage or current output.

The original GSM modem has been substituted by a GSM-GPRS one, able (thanks to an embedded TCP-IP stack) to directly connect to Internet, to transfer data as e-mail messages, so allowing to de-localize the base station and to better disseminate acquired data.

A GPS receiver is included in the system to control its position: should the platform go out of the allowed area range, an "unmoored platform" alarm is sent to the base station and to selected cellular phones.

The system software is a new release of the original one, first described by Zappalà [24] and further expanding the enhancements reported by Zappalà [25].

It is coded partly in compiled BASIC and partly in Assembly in a DOS-like environment; remote control software is written in Visual Basic in Windows environment.

Basically, it consists in a "time machine" executing at pre-defined times sequences of macro-commands able to fully control the instruments and the data acquisition and transmission systems.

The sequences can be different for each time and are remotely reprogrammable without suspending system activity.

Acquired data are immediately stored in the local system disk (a solid state one), then formatted and sent via e-mail.

Being the messages sequentially numbered, it is easy for the receiving system to automatically detect a missing message and asking for its retransmission, simply sending an e-mail message or an SMS to the buoy.

2.2.3 The pumping system

The pumping system (shown during refitting operations in Fig. 4) uses a peristaltic pump to drive water from five depths to feed instruments (e.g. multiparametric probes, water samplers, colorimetric analyzers…) into a measurement chamber or in an open flow piping; after the measurements the system can be automatically washed with fresh water contained in a reservoir.
Of course, possible alterations of sample temperature and dissolved oxygen must be taken into account; no influence has been noticed on other parameters,

Figure 4: The water pumping system in the buoy (left) and on the bench (right) during refitting operations.

however, attention must be paid to avoid mixing in calculations *in situ* and pumped measured values (e.g. calculating salinity only using pumped samples).

3 Discussion and conclusions

The TFLAP probe was developed to overcome the lacking of measurements of fluorescence from chlorophyll *a* in available expendable probes and to integrate the third dimension of satellite observation. The acquisition of a huge amount of data concerning chlorophyll *a* estimation, together with temperature, allows the monitoring of marine productivity variation, giving important information to study the state of marine ecosystem.

Worth mentioning is the cost reduction that can be achieved by using the Automatic Multiple Launcher. One of the requirements in operational oceanography is the collection of data allowing the spatial resolution of mesoscale phenomena, which are of the order of 15-20 miles in the Mediterranean Sea.

Since ships are moving faster than 20 knots, the XBTs must be dropped every 20-30 min. Data collection along a section 400-500 miles long with a manual launcher requires the engagement of two-three technicians, having a cost of about 25 USD per hour. The total cost for personnel amounts to 1000-2000 USD for each trip. The cost of a multiple launcher is about 15000 USD, that can be recovered in 10–20 trips by engaging only one person.

Coastal moorings not only offer a low-cost (or better not-too-high-cost) solution for environment monitoring, but also help integrate and calibrate data coming from satellite observations.

The devices here reported can be used for low cost operational oceanography monitoring and can be further implemented with other measuring systems (e.g. water sampler for off-line bacteriological, chemical and nutrient analyses, multispectral instruments, Acoustic Doppler Current Profiler).

Our group focused and is still working on the development of an integrated approach including: for the open waters, the development of low-cost tools and new sensors for the measurement of biological variables; for the coastal waters, the development of automatic measurement systems and their integration with the existing networks, with the ultimate goal of developing automated methods and low cost instruments, that, in a next future, will perform (near) real time biological and chemical measurements.

References

[1] Dickey, T. D. & Bidigare, R. R., Interdisciplinary oceanographic observations: the wave of the future, Scientia Marina, 69(Suppl. 1), 23-42, 2005.

[2] Munk, W., Oceanography before, and after, the advent of satellites, Satellites, Oceanography and Society, edited by David Halpern, 9 Elsevier Science B.V., 2000.

[3] Marcelli, M., Di Maio, A., Donis, D., Mainardi, U. & Manzella, G.M.R.. Development of a new expendable probe for the study of pelagic ecosystems from voluntary observing ships. Ocean Sci., 3, 1-10, 2007.

[4] Carof, A., H., Sauzade D. & Henocque, Y., Arcbleu, an integrated surveillance system for chronic and accidental pollution. Proc. of OES-IEEE OCEANS '94 Conference, III, pp. 298-302, 1994.

[5] Grisard, K., Eight years experience with the Elbe Estuary environmental survey net. Proc. of OES-IEEE OCEANS '94 Conference, I, pp. 38-43, 1994.

[6] Eriksen, C. C., Instrumentation for Physical Oceanography: the last two decades and beyond. NSF APROPOS Workshop – Ailomar, CA 15-17 December 1997.

[7] Griffiths, G., Davis, R., Eriksen, C., Frye, D., Marchand, P. & Dickey, T., Towards new platform technology for sustained observations. Proc. of OceanObs 99, http://www.bom.gov.au/OceanObs99/Papers/Griffiths.pdf

[8] Irish, J. D., Beardsley, R. C., Williams, W. J. & Brink, K. H., Long-term moored observations on Georges Bank as part of the U. S. Globec Northwest Atlantic/Georges Bank program. Proc. of MTS-IEEE OCEANS '99 Conference, I, pp. 273-278, 1999.

[9] Paul, W., Buoy Technology. Marine Technology Society Journal, 35(2), pp. 54-57, 2001.

[10] Seim, H., Werner, F., Nelson, J., Jahnke, R., Mooers, C., Shay, L., Weisberg, R. & Luther, M., SEA-COOS: Southeast Atlantic Coastal Ocean

Observing System. Proc. of MTS-IEEE OCEANS 2002 Conference, I, pp. 547-555, 2002.
[11] Zappalà, G., Caruso, G. & Crisafi, E., Design and use of advanced technology devices for sea water monitoring. Operational Oceanography. Implementation at the European and Regional Scales, eds. Flemming N. C., Vallerga S., Pinardi N., Behrens H.W.A., Manzella G., Prandle D., Stel J.H., Elsevier Oceanography Series, 66, 2002.
[12] Nittis, K., Tziavos, C., Thanos, I., Drakopoulos, P., Cardin, V., Gacic, M., Petihakis, G. & Basana, R., The Mediterranean Moored Multi-sensor Array (M3A): system development and initial results. Annales Geophysicae, 21, pp. 75-87, 2003.
[13] Crisafi, E., Azzaro, F., Zappalà, G. & Magazzù G., Integrated automatic systems for oceanographic research: some applications. Proc. of OES-IEEE OCEANS '94 Conference, I, pp. 455-460, 1994.
[14] Zappalà, G., Crisafi, E., Caruso, G., Azzaro, F. & Magazzù, G., Coastal monitoring by an advanced technology platform. Proc. of Oceanology International 98 Conference Proceedings, I, pp. 69-84, 1998.
[15] Zappalà, G., Alberotanza, L. & Crisafi, E., Assessment of environmental conditions using automatic monitoring systems. Proc. of MTS-IEEE OCEANS '99 Conference, II, pp. 796-800, 1999.
[16] Pinardi, N., Allen, I., Demirov, E., De Mey, P., Korres, G., Laskaratos, A., Le Traon, P., Y., Maillard, C., Manzella, G. & Tziavos, C., The Mediterranean Ocean Forecasting System: first phase of implementation (1998-2001). Annales Geophysicae, 21, pp. 3-20, 2003.
[17] Marcelli, M., Di Maio, A., Piermattei, V., Zappalà, G. & Manzella, G., Development of new technologies for the high variability phenomena data acquisition in the MFSTEP-VOS project. European Operational Oceanography: Present and Future, Proc. of the Fourth International Conference on EuroGOOS, eds. H. Dahlin, N.C. Flemming, P. Marchand, S. E. Petersson, pp. 184-187, 2006.
[18] Marcelli, M., Di Maio, A., Donis, D., Mainardi, U., & Manzella, G., Development of a new expendable probe for the study of pelagic ecosystems from voluntary observing ships, Ocean Sci., 3, pp. 311–320, 2007. Online www.ocean-sci.net/3/3/2007/
[19] Zappalà, G., Manzella G., An automatic multiple launcher for expendable probes. Proc. of the Fourth International Conference on EuroGOOS. Operational Oceanography: Present and Future, eds. H. Dahlin, N.C. Flemming, P. Marchand & S. E. Petersson, European Commission Publication Office, pp. 188-191, 2006.
[20] Zappalà, G., Reseghetti, F. & Manzella, G.M.R., Development of an automatic multiple launcher for expendable probes, Ocean Sci., 3, pp. 173–178, 2007. Online www.ocean-sci.net/3/173/2007/
[21] Zappalà, G., Development of advanced instrumentation for operational oceanography. WIT Transactions on Modelling and Simulation, eds C. A. Brebbia & G. M. Carlomagno, WIT Press, Southampton, 46, pp 841-850, 2007.

[22] Zappalà, G., Caruso, G., & Crisafi, E. The "SAM" integrated system for coastal monitoring. Proc. Of the 4th Int. Conf. On Environmental Problems in Coastal Regions, Coastal Environment IV, ed C.A. Brebbia, WIT Press: Southampton, pp. 341-350, 2002.
[23] Zappalà, G., Caruso, G., Azzaro, F., Crisafi, E. Integrated Environment Monitoring from Coastal Platforms. Proc. of the Sixth International Conference on the Mediterranean Coastal Environment, ed E. Ozhan, MEDCOAST, Middle East Technical University: Ankara, 3: pp 2007-1018, 2003.
[24] Zappalà, G., A software set for environment monitoring networks. Proc. of the Int. Conf. On Development and Application of Computer Techniques to Environmental Studies X. Envirosoft 2004, eds. G. Latini, G. Passerini, & C. A. Brebbia, WIT Press, Southampton, pp. 3-12, 2004.
[25] Zappalà, G., A versatile software-hardware system for environmental data acquisition and transmission. WIT Transactions on Modelling and Simulation, eds C. A. Brebbia & G. M. Carlomagno, WIT Press, Southampton, 48, pp 283-294, 2009.

This page intentionally left blank

Computer-based data acquisition and processing in a technical investigation department

N. J. Durley-Boot & J. R. Maguire
Technical Investigation Department, Lloyd's Register, UK

Abstract

This paper provides a snapshot of the current work of the Technical Investigation Department (TID) of Lloyd's Register EMEA, focusing on how computer-based tools have been integrated to collect, process and analyse data. There are three main parts to the paper: (1) description of the data acquisition and processing systems – both traditional analogue and more modern digital systems; (2) discussion of the analysis techniques – categorised into spreadsheets, signal processing, and engineering design and predictive analysis; (3) example combinations on jobs – in both marine and non-marine industries.

The paper concludes with a look ahead to developments foreseen in the coming years.
Keywords: data, acquisition, processing, technical, investigation.

1 Introduction

Since its formation in 1947, Lloyd's Register EMEA's Technical Investigations Department (TID) has been invited to investigate a wide range of engineering problems across the marine, land-based industrial and off-shore oil and gas industries (Carlton and Bantham [1]). The experience gained from these investigations has been used to develop new and innovative measurement and analysis techniques alongside more orthodox technologies. The department's accumulated experience and contemporary measurement toolkit allows TID to offer services for an extensive range of investigations.

This paper looks at the data acquisition and processing tools and the analysis techniques available. Through a series of case studies, the paper demonstrates how a technical investigation department can combine measurement, analysis,

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110451

experience and professional engineering judgment to deliver high quality results, enabling sound management decisions.

2 Data acquisition and processing

Accurate data acquisition and processing techniques are vital to an investigation department. Investigators must always attempt to identify and understand the prime cause of failures or unexpected performance before deciding on corrective action and this would, in most cases, be impossible without accurate measurement data.

TID has an extensive range of measurement equipment allowing the department to react to a wide range of circumstances and to adapt to clients requirements on an individual basis. A brief overview of some of the principal equipment used by TID is given below.

2.1 Core digital system

National Instrument's NI/CDAQ multi-channel chassis system with various modules and sensors provides the core of the department's digital data acquisition capabilities. This PC based system enables the department to perform a variety of sensor measurements including:

- acceleration;
- velocity;
- displacement;
- frequency;
- pressure;
- force;
- inclination.

Measurements are controlled using LabView SignalExpress and in-house developed MatLab scripts which provide software flexibility through user-defined loggings, analysis and visualisation. More sophisticated processing techniques are dealt with under the "Analysis Techniques" section later.

2.2 Still/video images

The use of low-light digital cameras allows the department to record observations and measurements. These records can then be referenced and referred to when recommending remedial action.

One such application is that of propeller observations. Historically, propeller observations were performed by inserting a viewing window into the hull. This method required a period of dry-docking and specialist work which was expensive to the client.

With improvements in technology TID is now able to offer propeller observations to clients without dramatically affecting the ships' schedule. The use of a borescope coupled with a low-light digital camera means that just a

small hole of the order of 15mm diameter in the ship's hull is required to obtain observations. This technique also allows TID to refer back to the observation when performing analysis, resulting in a more informative investigation. Example images obtained from this method can be seen in Figure 1.

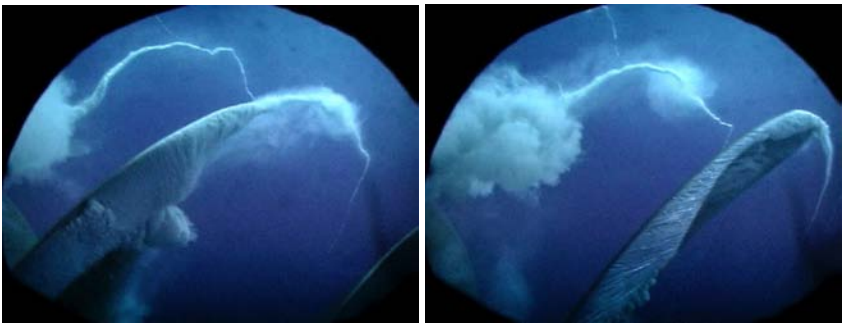


Figure 1: Propeller observations, showing tip cavitation.

The video images can be synchronised with vibration and pressure measurements to aid the understanding of propeller cavitation and its influence on ship structural vibration.

2.3 Acoustic emission systems

Acoustic emission monitoring can be used for leak and crack detection (Rogers [2]). It is sensitive to the propagation of growing cracks and provides information on growth rate under service loading conditions thereby guiding inspection and repair work for cost effective maintenance.

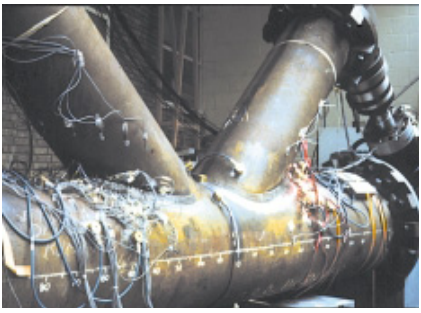


Figure 2: Tubular 'K' joint instrumented with acoustic emission sensors and strain gauges undergoing fatigue testing.

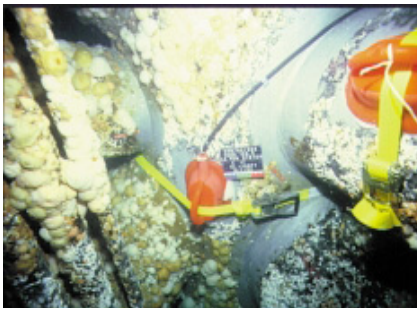


Figure 3: An acoustic emissions sensor attached to a sub-sea node joint.

2.4 Impact testing kit

Hammer test equipment as shown in Figures 4 and 5 can be used to determine the natural frequencies of structures and components. The excitation signal from the hammer, and the response signals from the accelerometers attached to the structure, are acquired in the time domain by digital data acquisition equipment. These signals are then converted by Fast Fourier Transform (FFT) into the frequency domain, which yields the power spectral densities. The power spectrum response can then be divided by the power spectrum excitation to produce the Frequency Response Function (FRF) which contains the natural frequencies of the structure. Mode shapes and damping values can also be extracted from a series of FRF's.

Figure 4: Instrumented sledgehammer.

Figure 5: Seismic accelerometers.

2.5 Telemetry systems

Telemetry systems can be used to transfer data from rotating devices. Two typical applications for such equipment are for measuring propeller blade strains, and for monitoring the operating strain distribution across the teeth of a gear wheel. Figures 6 and 7 shows typical installations in each case.

2.6 Portable measurement devices

TID use a range of portable measurement equipment, allowing investigators to be flexible in their investigations. These devices normally supplement other measurement techniques, and can offer instant data analysis. The results from

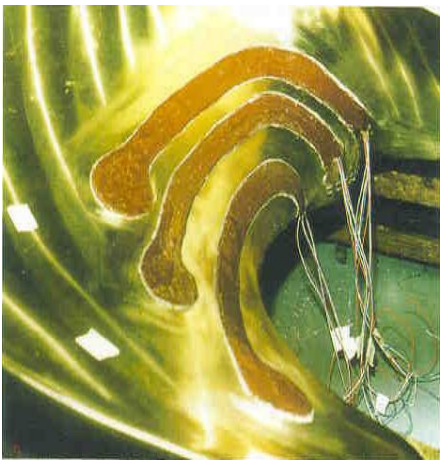


Figure 6: Propeller blade and strain gauge instrumentation.

Figure 7: Telemetry equipment attached to a gear wheel.

these measurements are recorded and later downloaded onto a PC for subsequent analysis if required. Such devices are used for the following measurements:

- sound;
- vibration;
- dye penetrant;
- presence of cracks, using Magnetic Particle Inspection (MPI);
- thickness, using ultrasonics.

2.7 Calibration

Knowledge of the uncertainty of the measurement is vital to the success of the investigation. As a general rule equipment is calibrated in a controlled laboratory environment before and after an investigation. This provides confidence in the measurement data and ensures that equipment has not been damaged or altered during transit.

TID use an externally calibrated vibration table to check the frequency response of accelerometers and velocity transducers. Sensors are calibrated with their associated cabling to remove the effects of resistance between cable lengths. The TID laboratory also has the equipment required to calibrate pressure, eddy current and other types of transducer.

3 Analysis techniques

TID have a range of analysis techniques available, the application dependant on the type of investigation and often adapted to suit the individual requirements of the specific job.

3.1 Spreadsheets

Spreadsheets remain a popular method for manipulating and visualizing data. A commercial spreadsheet application (Microsoft Excel) provides a tool for performing simple statistical analysis, Fourier Transforms and can also offer a platform for more complex programming in Visual Basic.

3.2 Signal processing

TID uses a variety of post-processing tools including standard commercial and custom products. MATLAB is used as the primary software for off-line signal processing. MATLAB is a high-level technical computing language used for algorithm development, data visualization, data analysis, and numeric computation. TID has the capability to develop its own code and are able to perform the following types of analysis:

- statistics;
- time domain;
- frequency domain;
- JTFA (joint time frequency analysis);
- wavelets and filtering;
- 2D and 3D graphical presentation;
- modal analysis.

3.3 Engineering design and predictive analysis

Anticipating and designing problems out at an early stage is an effective risk management strategy and can prove to be extremely cost effective. TID are able to perform analytical studies in conjunction with model tests in order to identify potential problems at the early stages of ship design or with major conversions.

TID undertake domain specific analysis including:

- propeller hydrodynamics;
- ship performance;
- shafting systems analysis;
- diesel engine and performance;
- gear tooth load distribution.

TID's general purpose analysis tools include:

- sound and noise analysis;
- fatigue and crack growth analysis;
- finite element analysis [FEA];
- computational fluid dynamics [CFD].

Computational Fluid Dynamics provides a useful approach for enhancing understanding and helping to identify optimal solutions for various engineering problems. Figure 8 shows the predicted velocity of the inflow to the propeller after the addition of a vortex generator (Figure 9) to the side of the hull. The results were used to minimise cavitation from the propeller and thus reduce ship structural vibration.

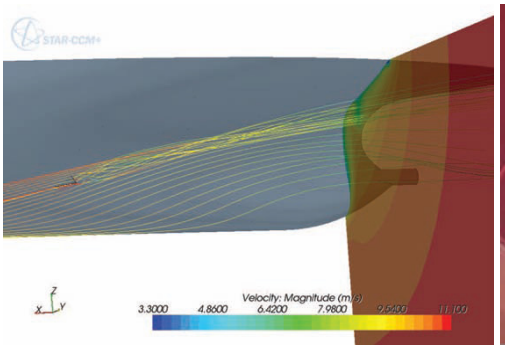


Figure 8: CFD results showing the effect of the vortex generator on inflow to the propeller.

Figure 9: Vortex generator.

4 Example case studies

This section contains three case studies that demonstrate how TID has utilised the computer based data acquisition and processing equipment discussed within this paper. The combination of instrumentation and analysis techniques with sound engineering judgement has provided tailored solutions to specific problems.

4.1 Shaft alignment

When specifying the initial alignment conditions of propulsion shafting, considerations that need to be addressed include:

- the self weight of the shaft;
- the overhung propeller and its associated operational loads;
- gear tooth load distribution;
- main engine main bearing loads;
- the flexibility and loading of the ship's structure.

In this investigation TID were asked to determine the cause of the shaft misalignment of a bulk carrier when loaded, and to make recommendations for future operation (Lloyd's Register [3]). TID measured the existing alignment using a combination of jacking and strain gauge techniques.

Initial measurements revealed that as the vessel's draught increased the sterntube forward bearing became unloaded; the single plummer bearing became overloaded, and the hogging moment imposed on the engine crankshaft increased beyond the design limit. TID recommended a full realignment of the system to restore it to acceptable operating conditions.

Figure 10: Propulsion shafting.

Figure 11: Preparation for jacking.

An analytical model of the propulsion shafting was produced, and used to calculate the theoretical straight line bending moments and the influence coefficients for the load-deflection response at each bearing. The results from the model were used in conjunction with strain gauge measurements (at various axial locations on the shaft) to determine vertical and transverse loads on each of the shaftline bearings. The accuracy of the analytical model was confirmed by jacking the shaft at the sterntube forward bearing and intermediate bearings.

The misalignment of the initial shafting conditions were likely to have been exacerbated by hull structure relaxation of residual stresses on the vessel's maiden voyage. Careful consideration should be given to alignment of shafting systems with a single plumber bearing, to accommodate the large displacement variations between the ballast and loaded condition associated with this type of vessel.

4.2 Gear tooth strain gauge measurements and shaft alignment

A ship owner contacted TID after suffering damage to the main propulsion secondary gearbox assembly on one of their vessels (Lloyd's Register [4]). Subsequent investigations revealed unsatisfactory alignment of shaft to main reduction gearbox. TID were subsequently asked to examine the condition of the

main reduction gear elements and alignment of a sister ship and to conduct load tests while still in service.

Micro strain gauges were installed to the roots of the main wheel and measurements were conducted during sea trials to measure the gear tooth load distribution of the second reduction gear elements (Figures 12 and 13). Strain gauges were also used to measure the gearbox and propulsion shaft bearing loads and were confirmed using jacking measurements. Additionally, the main wheel pin readings were measured and Magnetic Particle Inspection (MPI) of the low speed gear elements was undertaken.

Figure 12: Main gear wheel.

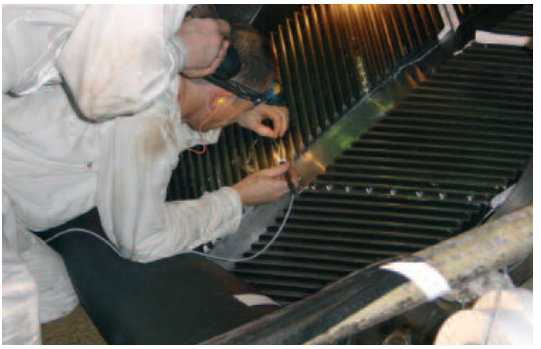


Figure 13: Installation of the micro strain gauges.

Measuring tooth load distribution with strain gauges requires the interpretation of large quantities of measured data. Computer based statistical techniques were employed in order to separate the mean load distributions from the tooth to tooth variations inherent in the meshing process. In order to minimise the required channels of strain gauge instrumentation TID use a Wheatstone bridge arrangement as shown in Figure 14. A typical signal produced by this arrangement as the wheel rotates is shown in Figure 14b. This arrangement allows four or even eight strain values depicting the general tooth root stress distribution, to be obtained from one channel.

Evaluation of the root strain results showed that the load distribution along the length of the secondary reduction gear elements was dominated by base-heavy helix angle mis-match. The low pressure upper pinion was subjected to significant misalignment, which raised the peak stress at the forward end of the forward helix, to levels in excess of the design limits. The MPI conducted on the main wheel and secondary pinions did not show any indication of cracking. The main wheel pin readings showed negligible change since new-build, indicating that distortion of the gear casing had not occurred.

TID concluded that the helix alignment mis-match had been caused by manufacturing errors. It was proposed that immediate adjustments should be made to the pinion bearings to reduce the maximum stress, and that the bearing position should be optimized further at the next scheduled dry-docking.

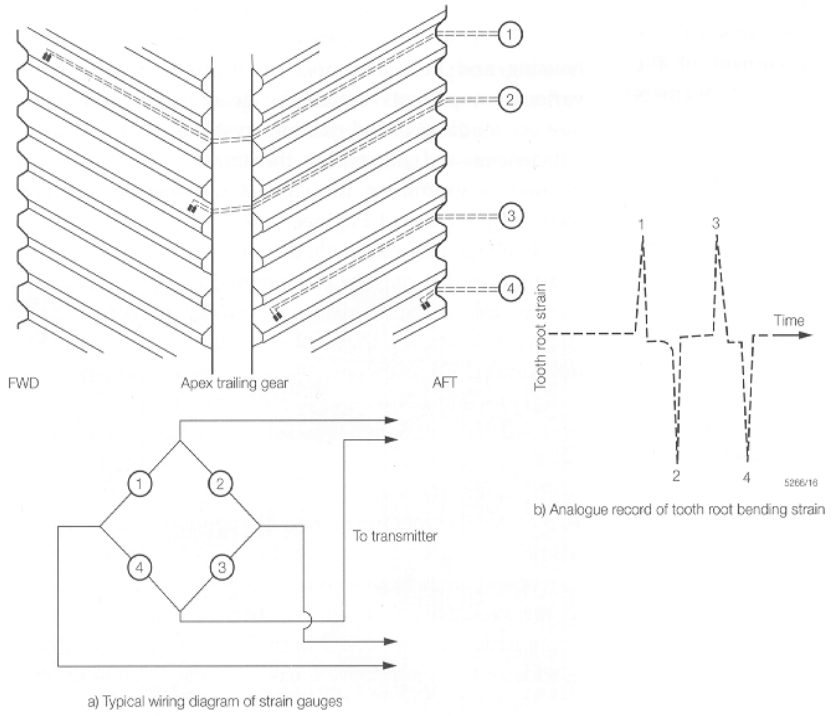


Figure 14: Gear tooth strain measurements.

4.3 Leak testing by acoustic emission methods

In this case TID was asked to consider the feasibility of leak monitoring an offshore oil production manifold, using acoustic emission techniques similar to those previously used for fatigue crack detection and monitoring (Rogers [2]).

The numerous control valves and interconnecting spool-pieces in the manifold represented numerous possibilities for leakage, which might not be evident until appreciable loss has occurred. Environmental regulations and public perception make leak detection by acoustic emissions an attractive option, especially during initial start up and early operation periods.

TID was able to study a manifold to determine the most economic arrangement of sensors to detect leaks in high risk locations. It was concluded that eight acoustic emission sensors, distributed around the central valve block, and one sensor on each of the connecting hubs, would provide the required sensitivity for leak detection.

The acoustic monitoring system sampled the root mean square signal level within octave frequency bands covering the range 10 kHz to 170 kHz. The measurements were relayed to the platform via a hard-wire digital communications link that was also used to supply power to the system. A sensor beacon would be actuated to indicate the existence of a possible leak at which

Figure 15: AE equipment.

Figure 16: Offshore platform (manifold at lower deck level).

point the raw data history would be recovered via an acoustic telemetry link and used to help identify the existence of a leak.

5 Conclusions and look ahead

Engineering failures and unexpected performance can be costly in terms of reduced safety, down time and the expense of repairs. Establishing root causes and taking preventative action are therefore of great importance.

Lloyd's Register EMEA's Technical Investigations Department (TID) has accumulated a wealth of experience and expertise since its formation in 1948. As technology continues to develop and the importance of the environment and efficiency continues to grow, it is inevitable that new technological challenges will arise.

TID continue to combine the experience gained from each investigation with current industry research. This allows the department to utilise new measurement technology and to further develop current analysis techniques. In doing so, TID continues to deliver high quality technical investigations to the marine, land based industrial and offshore oil and gas industries.

Looking ahead it is foreseen that data acquisition hardware will become lighter, cheaper, faster and more reliable. The "bad old days" of TID having to deal with heavy, expensive and cumbersome analogue equipment are behind us.

Processing and analysis techniques are not expected to change dramatically, but faster and more powerful analysis will enable techniques such as FEA, CFD and coupled fluid-structure-interaction (FSI) analysis to become more readily used as part of the investigation process. As always there will be no substitute for good people, adhering to proven procedures, within a supportive team.

References

[1] Carlton, J.S., Bantham I. "The Technical Investigation Department 50 Years of Operation". Lloyd's Register Technical Association, 1998.
[2] Rogers, L.M., "Structural and Engineering Monitoring by Acoustic Emission Methods – Fundamentals and Applications". Lloyd's Register Technical Association, pp. 5-25, 2001.
[3] Technical Matters, Lloyd's Register EMEA, Issue 1, 1997-1998.
[4] Technical Matters, Lloyd's Register EMEA, Issue 3, 2009.

Research on the COD soft-measuring mechanism based on the clustering approach

Y. Z. Feng[1], G. C. Chen[1], D. Feng[2] & M. Zhuo[1]
[1]*Department of Management and Environmental Engineering, Logistic Engineering College, China*
[2]*School of Architecture Liverpool University, UK*

Abstract

By analyzing laws of growth and reproduction of microorganisms in the activated sludge (AS) sewage treatment system, this paper proposes a multi-neural network (NN) COD (chemical oxygen demand) soft-measuring method based on clustering approach for the sewage treatment project. Various reasons which might affect the accuracy of the model are analyzed. Experiments show that the same radiuses of multiple neural network diffusion constant are quite close which means the prediction accuracy is high and the soft-measuring method based on clustering approach is suitable for COD measuring.
Keywords: sewage treatment, clustering approach, neural network, microorganisms.

1 Introduction

Cybenko [1] has proved theoretically that if there are plenty of training data and without restriction of the size of the network, modelling based on NN can always get a satisfactory model structure. But, in the actual industrial process, people often need to face the limited effective process data and due to the real-time requirements, the network structure also cannot be expanded unlimited. Modeling effect normally relies on the good generalization ability of the network. Paper [2] proposed a robust classification method by combining different models which are based on neural networks with fuzzy combination approach. Paper [3] presented a method to improve model prediction accuracy and robustness by adding different models together. Paper [4] discussed Stacked Neural Network Approach for process modeling. The basic idea of these


WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110461

methods is establishing several independent neural network models and then combining the forecast outputs of different networks together through a certain method. Such a structure with multi-models is conductive to the real-time soft measurement with parameters changing in a wide range and meets the technological requirements for accurate description of the dynamic characteristics of the process.

2 Theoretical basis of NN model in the AS process

Shi *et al.* [5] developed an urban sewage treatment plant operation simulation and forecast expert system which shows that there is a strong correspondence between simulated effluent COD (E-COD) value of the aeration tank and its influent COD (I-COD) value. HRT (hydraulic retention time) is a large time-delay process in the sewage treatment, the correspondence misfit caused by time-delay has great influence on model, while the correspondence between effluent and influent COD value is the basis of modeling based on multiple NN. Therefore, it is very important to research on the causes of delay error, analyze ways of effectively reducing the delay error and determine the theoretical basis of multi-neural network modeling.

2.1 Growth and reproduction laws of microorganisms

Bacterial growth can be divided into at least four phases [6]: 1) Lag phase. After cells are seeded into fresh medium, they are living in a new growth environment and will not immediately divide over a period of time and the number of cells basically maintain constant. 2) Exponential phase. In the exponential phase, cells grow and divide with the greatest rate, the number of cells increases logarithmically, different components in bacteria increase regularly according to the corresponding proportion. 3) Stationary phase. Because of the nutrient consumption, increase of metabolites and pH value changes, the environment is not suitable for bacterial growth anymore, cell growth rate progressively decrease to zero. Bacterial population growth changes from exponential phase to stationary phase. Stationary phase owns the highest number of living bacteria and maintain stable. 4) Death phase. Because of the nutrient depletion, toxic metabolite accumulation and pH value changes, the bacterial death rate gradually increase while the living bacteria and their metabolic activity progressively decrease. Cells begin autolyze and senesce.

2.2 Microorganisms proliferation laws in AS

The proliferation of Microorganisms is the inevitable result of AS reaction and organic matter degradation. The proliferation of microorganisms is actually the growth of activated sludge. Activated sludge is a community of microorganisms composed of multiple bacterial colonies with a variety of physiological functions. All kinds of microorganisms work together to finish the biodegradation process of AS. The proliferation laws are complicated but the general trend can still be expressed as Fig. 1.

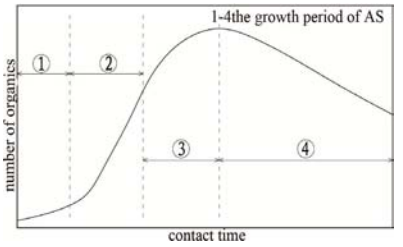


Figure 1: Growth curve of AS.

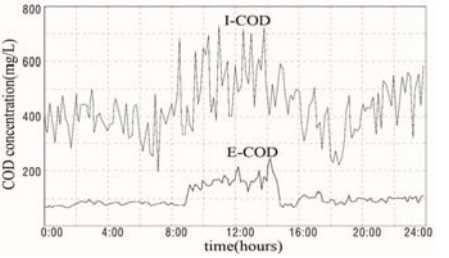


Figure 2: Curve of I-COD and E-COD.

In the practical wastewater treatment process, the conventional AS process mainly depends on the 3 and 4 stage of bacterial population growth phases, while improved wastewater treatment processes based on AS process can use different time interval of the AS growth curve [6]. There are obvious rising and falling phases in the growth curve, which is conducive to divide the treatment process into several stages so as to facilitate the realization of multi-modeling approach.

Proliferation laws of microorganisms in the wastewater treatment process are the interpretation of the process of growth and reproduction of microorganisms.

Through the detailed analysis of all kinds of factors associated with the growth process, researchers can draw a growth curve of activated sludge and decide the application stages of conventional activated sludge process. The growth curve and its corresponding application are the theoretical basis of composite model modeling approach. According to the actual data distribution of sewage treatment plant, laws of the daily routine of residents also should be considered as one of the parameters of the model. The I-COD and E-COD values in Fig. 2 come from a wastewater treatment plant in Chongqing, China [7], with a sampling period of 10 minutes. These data apparently can be divided into four stages in accordance with time-series, including 0:00-8:00, 8:00-14:00, 14:00-20:00, and 20:00-24:00. The fact that there is only a relatively high effluent COD value in 9:00-14:30 on the effluent COD curve indicates that the COD removal ability of the sewage plant can be divided into three relatively stable sections.

Based on the analysis of the 24 hours of operation data on September 1, with the consideration of measurement error, we believe that the wastewater treating capacity of activated sludge microorganisms can change according to certain rules. This is the basis of multiple neural networks.

3 Data sources and processing

The research results of this paper are verified by the data provided in literature [8].

3.1 Normalization processing

If the number of learning samples of a sub-network is small, the estimated value of similar data to a large extent will depend on the number of learning samples of the network and the accurate estimates cannot be obtained because of the large influence of other sub-networks [9]. From the point of view of mechanism of NN, as the NN input nodes, each process variables in the NN is equally important. So, when the output is so sensitive to a certain value of an input variable that even a small deviation of the input variable can cause a large output error, the value of α should be increased. Conversely, the α in the corresponding formula can be assigned a slightly larger value to improve the sensitivity of output to the secondary input variables. In this paper the α is 0.9.

$$X_i' = \alpha \times (X_i - \min(X))/(\max(X) - \min(X)) + 1 - \alpha \tag{1}$$

3.2 Untrusted data processing

Human error comes from manual testing stage or data entry stage. According to 3σ principle, the standard deviation is calculated by the Bessel formula: $S = \sigma = \left[\frac{1}{n-1}\sum v_i^2\right]^{1/2}$, data set 7, 39 and 43 are deleted according to the calculating results. Meanwhile, the COD value in data set 22 is 457.5mg/L, which greatly exceeds the corresponding integrated wastewater discharge standard of China (120mg/L); the effluent COD value in data set 22 is greater than its influent COD value, which is obviously not reasonable, so, both data sets are also removed from the learning samples.

3.3 Analysis of probable influencing factors

The above data are analyzed in the form of multi-dimensional clustering analysis, the sample group affecting effluent quality contains multiple characteristics. Data with similar sample characteristics have the small deviation of the sample performance, which is based on the reliability of all sample data. In fact, there are different kinds of measurement error on field sampling data, which will affect the clustering results. There are main factors and secondary factors in sample groups, if there is a big deviation of the secondary factors, the clustering results will deviate the clustering intervals defined by the main factors. The more characteristics exist in the sampling groups; the far more deviation from the real situation is inevitable, especially when all different secondary factors can affect the results of clustering analysis. Even the membership determination method proposed in literature [9] is a better method than that chosen in literature [10], the problem still cannot be fundamentally solved. This paper intends to try one-dimensional clustering method, namely to cluster only the most important factors. If successful, it is apparently helpful to meet the real-time requirement of on-line running. The first step is to verify and choose the main factors from the data relationship between removal ratio of COD and each auxiliary variable.

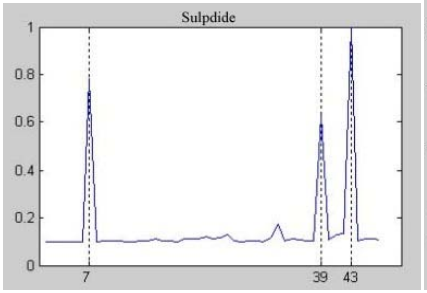


Figure 3: Normalized value of sulphide.

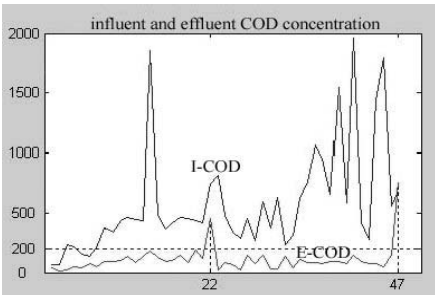
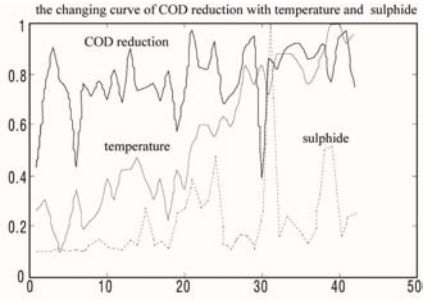
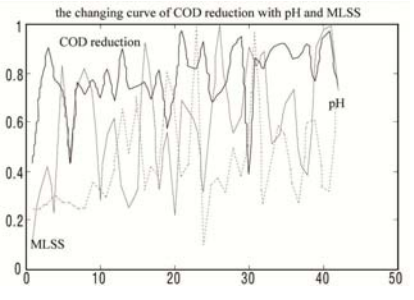
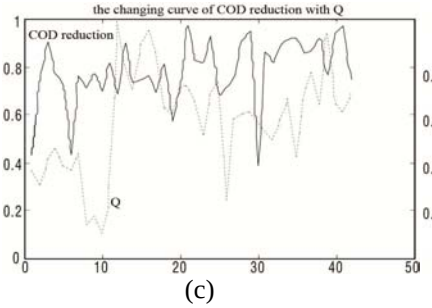


Figure 4: I-COD and E-COD value.

It can be seen from Figs. 5(a) to (d) that the variation trends of T, S, pH, Q and MLSS do not have any correspondence with the removal ratio of COD or its time delay. While from Fig. 5(d), there is an obvious correspondence between COD removal ratio and I-COD values. Furthermore, from the definition of COD removal ratio ($\eta = (COD_{in} - COD_{out})/COD_{in}$), it is quite reasonable to choose I-COD value as the main factor, which is also consistent with Zhang's research results of orthogonal test[11].







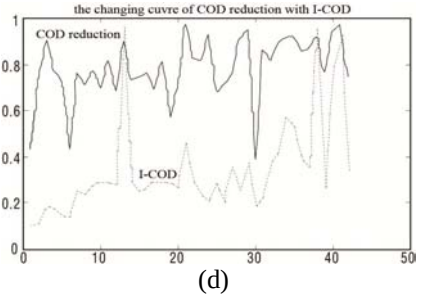


Figure 5: The changing curve of COD reduction with (a) temperature and sulphide, (b) pH and MLSS, (c) Q and (d) I-COD.

3.4 Clustering analysis of historical data

After determining the I-COD value as the main factor, the normalized data of COD data are clustered. The sample data distribution is shown in Fig. 6 in accordance with the order from small to large. In order to improve generalization performance, it is better to resolve the classification problem within the range of 0.1-0.5. The cluster radius r lies in [0, 0.2]. From the convenience of numerical rounded perspective, the value of r can be given as 0.05, 0.1 and 0.2, and the subnet numbers can be 4, 2 and 1. From the perspective of network applications, r=0.05 is a better choice.

Table 1: Data distribution statistics.

Clustering distribution	Data distribution SN of sample group	Sample size
0.1-0.2	1,2,3,4,5,6,28,32 8,9,10,11,12,13,15	8
0.2-0.3	16,17,18,19,20,21,2 4, 25,26,27,30,33,42	20
0.3-0.5	29,31,34,38,40,46, 23,25	8
0.5-1.0	14,36,37,41,44,45	6

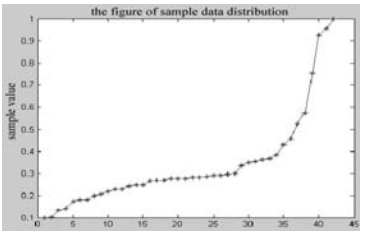


Figure 6: Data distribution.

4 Establishment of multi-NN soft-measuring model

Since the NN constructed in this article is also a distributed architecture, using current research results for reference, this paper takes fuzzy classifier shown in Fig.8 as a tool for network synthesis. The fuzzy clustering unit shown in Fig.7, named as PCA-DRBF, can realize the classification function of input space, in which sample characteristics X is translated as T through a PCA dimension reduction technique, after that, eigenvectors reduced from M dimensions to N dimensions are distributed to N mutually exclusive subspaces. Different from the PCA-DRBF method, this paper proposes an RBF clustering unit shown as Fig. 8, which directly chooses a sample eigenvector, I-COD value, as the classification indicator. From the specific operation concerned, there are many differences

between the two methods. For a new process input data, the fuzzy clustering unit of PCA-DRBF method firstly identifies its membership for each sub-network, and according to the different memberships, outputs of all sub-networks are integrated as the total output of the entire distributed network.

Because the cluster radius r equals 0.05 in this paper, the cluster centers are also fixed as 0.15, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, 0.85 and 0.95. According to the characteristics of samples, the cluster centers in this paper are limited to 0.15, 0.25, 0.4 and 0.75. The respective subordinate network of each new input is determined by Euclidean distance. In addition, to reflect the fuzziness, using the formulation of kernel radius presented for reference [9], this paper defines the kernel radius as 0.005, which means if an input data lies in the interval of [-0.005, 0.005], it can be viewed as belonging to two sub-networks and do not be classified strictly. In the training process, it is determined by the comparison of output values of both sub-networks with test value that which sub-network is more suitable for such an input data. Such a sub-network is retained as retention program. In the forecasting process, the calculating process is a little complicated. Firstly, the correlation degree of input value which lies in the interval [-0.005, 0.005] and the corresponding sample in the retention program will be calculated [12], then the correlation degree will be compared with a set value to determine which subnet will be chosen. Through the above fuzzy approach, the complexity of the wastewater treatment process has been taken into fully consideration.

Assuming that there are N sub-networks $\{RBF_i, i = 1,\ldots,N\}$ in the distributed network shown in Fig. 8, the membership degree set of new input variable x (influent COD value) for each sub-network is $\{\mu_i, i = 1, \ldots,N\}$, then the output of the whole neural network Y is:

$$Y = \sum_{i=1}^{N} \mu_i \bullet f_{RBF_i}(x) \quad (\mu_i = 0 \quad OR \quad \mu_i = 1) \tag{2}$$

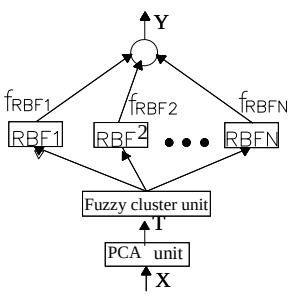


Figure 7: PCA-DRBF architecture.

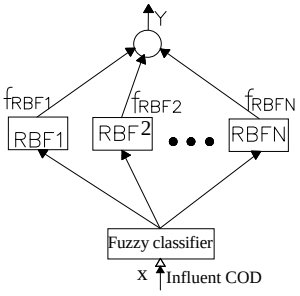


Figure 8: RBF architecture.

The membership degree of input x to each sub-network is obtained by eqn (3) and eqn (4):

$$d_i = \left\| \mathrm{x} - \mathrm{x}_{\mathrm{center_i}} \right\| \tag{3}$$

$$\mu_i = 1, \eta_{j \neq i} = 0, if : 0 \leq d_i \leq 0.005, i = 1, \cdots, N \tag{4}$$

If the input data x lies in the interval [-0.005, 0.005], then, assuming there are two sample groups X_1 and X_2, corresponding to the subnet i and subnet i+1 respectively, based on the calculation of $d = \left\| \mathrm{X} - \mathrm{X}_1 \right\|$ and $d = \left\| \mathrm{X} - \mathrm{X}_2 \right\|$, the finally subnet is chosen by the comparison of the relative size of the two membership degree.

4.1 Evaluation of the model adaptability

Through the clustering analysis, there are only 8 data in the first class. Taking 7 data among the 8 as training sample data and another one as input data, researchers can always find the best fit point because of the convergence performance of RBF network. However, because the sample size is too small, the expansion constant distribution of the network does not show any regularity and cannot be used for accurate forecast. Similar problems lie in the third class and fifth class. Taking relative error (RE) and recognition rate (RR) as criteria of assessing the forecast accuracy of RBF network, the forecast analysis is performed according to the data of second class.

$$RE = (Y - P\_test) / P\_test \times 100\% \tag{5}$$

$$RR = m / n \times 100\% \tag{6}$$

where, m is the number of actual outputs which are satisfied with the condition $\left|(Y - P\_test) / P\_test\right| \times 100\% < \varepsilon$, ε is the forecast precision, ε =10% in this paper. n is the total output number. There are two understandings about parameter Y and P_test. One explanation is that the two parameters represent the desired output and actual output respectively, the other is that the two parameters represent the anti-normalized values of the desired and actual output. This paper adopts the second understanding. Comparing with reference [8], the differences include: (1) calculation accuracy. The relative error of the model proposed in this paper ranges from 0.06% to 23.01%, average relative error is 7.66%, recognition rate is 83.3% (ε =10%). The relative error of the original model ranges from 0.9% to 9.2% and the average relative error is 3.4%, recognition rate is 92.3% (ε =5%). (2) details of models. There are eight hidden layer neurons in the original BP neural network model, training times are not less than 10,000, excitation function of the hidden layer is tansig and learning sample size is 30. When the SSE of the BP model is less than 0.001, the training stops and the predict process begins when the test data are input. The hidden layer of this RBF network provided in this paper are defined in the calculation process, after only 13 times of iteration, the SSE drops to lower than 10^{-28}, the learning

sample size is only 13 groups. The above analysis shows that even the effectiveness of the proposed new model is not as good as the original model, but it is obviously that convergence of the RBF neural network is much better than the original. So, the new neural network is the best choice for online operation.

4.2 Analysis of the results of model prediction

Theoretically, when the diffusion constant of network is 0.25, the prediction results will be highly accurate; while the actual constant is 0.265, with a small deviation from the ideal value, which verify the correctness of the theory analysis of this paper. The new neural model can greatly reduce search range in selecting the optimal value. From the above technical indicator comparison, the forecast precision of the new model is not high enough because of various reasons. Through analysis of these causes, this paper put forward several issues that should be paid attention to in future research.

4.2.1 The lack of sample size

To achieve the desired generalization ability (prediction error is less than a given value ε) requires a large number of samples, which in turn will increase the amount of network calculation. So, there is contradiction between the generalization ability and the amount of calculation of learning. Literature [13] provides an experiential formula: $m \geq d_{vc} \div \varepsilon$, where d_{vc} is the *VC dimension* of the network. It represents the capacity of function class and can be approximate evaluated by the number of independent parameters. According to this experiential formula, the sample size m should equal 50 (ε = 0.1, d_{vc} = 5), while there are only 20 groups used in this case actually. The lack of sample size is one of the reasons causing the lower prediction precision. From the common ideas of researching an issue, we rerun the model by taking cluster radius r as 0.2 so that the sample size increases to 30. The distribution of training and testing samples are shown in Table 2. Data in Table 3 and Fig. 3 shows that the relative error of the model prediction results ranges from 0.16% to 13.27%, the average relative error is 4.67% and the recognition rate is 83.3%. Compared with the initial results above, network performance has been greatly improved. The calculation results indicate that with the increase of sample size, the prediction ability of the model is also improved apparently. Meanwhile, if the cluster radius is 0.2, the diffusion constant needs to be taken 0.371 to ensure the best prediction results. Comparing with the theoretical best value of 0.3, the deviation is increasing, which means the value of cluster radius r (0.2) is not as appropriate as the original value (0.05). In addition, test results also confirm that with the sample size increasing, the number of hidden layer neurons increases synchronously. For BP neural network, the convergence performance declines sharply with the increasing of sample size. Even the operation time of RBF network also increase with the increasing of sample size, its iterations (10^1) is much less than that of BP neural network (10^4), which proves that RBF network has obvious advantages when the sample size is large.

Table 2: Training set and testing set.

Test sets	10, 12, 13, 18, 19, 27
Train sets	1, 2, 3,4 ,5, 6, 8, 9,11,15,16,17,20, 21,23,24,25,26,28,29,30,31,32,33, 34,35,38,40,42,46

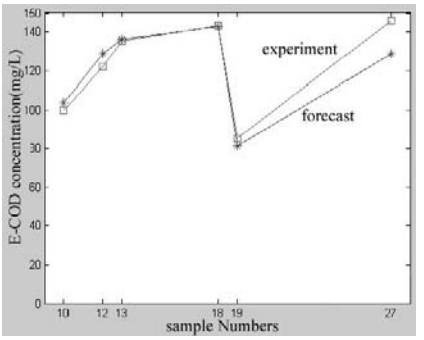


Figure 9: Testing and forecast value.

Table 3: Comparisons of test and prediction results of RBF network (mg/L).

SN	Test results	Prediction results	RE (%)	Desired output	Actual output	RE (%)
10	99.8	103.9	3.95	0.541	0.563	3.84
12	122	128.8	5.26	0.658	0.694	5.14
13	135.5	136.1	0.46	0.729	0.732	0.45
18	143	142.8	0.17	0.768	0.767	0.16
19	85.2	81.2	4.93	0.464	0.443	4.75
27	146	128.9	13.27	0.784	0.694	12.97

4.2.2 Unsatisfactory choice of secondary variables

Through the initial mechanism analysis, the number of auxiliary variables requires 8, which can be reduced by principle component analysis, by the way. However, in order to adapt to the special case of limited sample data and because of the traditional conception that as long as the oxygen supply is not excessive, there is no longer need to take DO into consideration, so, DO, an important control variable, was excluded from the secondary variables. The first deficiency to do so is the loss of the meaning of soft measurement (saving energy consumption), it is inappropriate to exclude the DO factor out of auxiliary variables set because the mechanism analysis shows that the value of DO does have a certain influence on the effluent COD value.

4.2.3 Selection scheme of samples of training set and test set

In the modelling process discussed in this paper, the reason the samples we chose from secondary variables are not maximum or minimum is mainly because the effect of interpolation is better than that of the generalization. But in reality, it is obvious that no guarantee can be provided that the value of input variable is not the maximum or minimum value of current sample source.

4.2.4 Choice of clustering center point and secondary important factors

Two modelling processes are completed in this paper. The deviation between clustering center point and diffusion constant is very small in the first modeling process (0.25 and 0.265) and large in the second modeling process (0.3 and 0.371). The reasons of the existence and change of the deviation include the following aspects. Firstly, the distribution of samples does not fluctuate with uniform amplitude around the clustering center. Secondly, finding the classification formed by the main vector from the five feature vectors will certainly be influenced by the other four vectors. The author assumes that with the sample size increasing, the fluctuation caused by the first reason will decrease and the fluctuation caused by the second reason will increase. Because the more samples, the greater the impact to the classification based on the main feature vector from other feature vectors will be. Choosing the most important feature (i.e. secondary important factor) from the rest feature vectors to balance this effect may solve the problem. This mode of thinking is quite similar with hierarchical clustering method. After the main and secondary factors are chosen from the feature vector set, the selection of sub-network for input variables can be done by fuzzy classifier which is based on the rules of fuzzy control.

4.2.5 Dynamic tracking of the actual process

There are many cases on the research and application of soft measurement, but only a few cases combine prediction model with time series [14, 15]. This is a common problem of current soft measurement technique research and application. As far as wastewater treatment process is concerned, there is no corresponding relationship between the value of input variable detected by operator at a particular moment and the effluent quality of the sewage plant at that time. A time delay which does not be mentioned in many literatures is closely related with HRT in engineering application. The reasons that not many reference literatures point this out are not only because of the sampling period (normally one day) of the data, but also because there is no correlation with time, which means time parameter does not be looked upon as a variable. This adverse effect is more apparent in a distributed multi-network model, because those data with far different time series might be clustered into the same class in the modeling process. Due to the obvious ignorance of time variable, the credibility and reliability of the predicting outcomes provided by such kind of clustering method are questionable. It is obvious that with the pre-treatment of the coarse and fine screen and primary settling tank, the index data of T, pH must have changed after mixing with the original material in the reactor. Such a change process is a complex function of time: the initial change is faster, and then it gradually slows down, accompanying by the potential influences of reaction

process. Neglecting or ignoring the impacts of these factors and other oversimplifications of the process may compromise the credibility of model predictions. The best and feasible way to solve these problems is the implementation of online algorithm based on a small sampling period (such as $\phi = 10m$) of input variables. As a result, effluent quality at time T is a function of feed water quality conditions at time (T-t) and other procedure parameters, which reflects the real process of sewage treatment. Otherwise, soft measurement technique transforms into just some kind of data-based prediction research. If the importance of mechanism analysis to understand the reaction process has been ignored in the modeling, the neural network is obviously not a successful soft instrument. The clustering method needed by multiple network models can use the time-based clustering analysis technique proposed in 2.2 of this paper. If there are historical data of the same time frame of last cycle, they can also be clustered into the same class.

5 The conclusion and prospects

This paper analyzes the microbial basis of multiple neural network modeling based on clustering method in the sewage treatment works. Case analysis shows that the proposed clustering method not only has diffusion constant close to the cluster radius, but also has better prediction accuracy. This paper also analyzes a variety of reasons which may affect the accuracy of model and point out that online real-time algorithm is the future research direction and the focus of subsequent research.

Acknowledgements

This paper is funded by Chongqing Natural Science Foundation (CSTC, 2009BB7175) and National Water Pollution Control and Management Science and Technology Major Projects of China (2008ZX07315-003).

References

[1] Cybenko G. Approximation by superpositions of a sigmoidal function. Math [J]. Control Signal System, 2, pp. 303-314, 1989.
[2] Cho, S. B. & Kim, J.H. Combining multiple neural networks by fuzzy integral for robust classification. *IEEE Trans on System, Man and Cybernetics*, 25(2), pp. 380-384, 1995.
[3] Bates, J.M. & Granger C.W.J. The combination of forecasts. *Operations Research Quarterly*, 20(1), pp. 319-325, 1969.
[4] Sirdhar, D.V., Seagrave, R.C & Bartlett, E.B., Process modeling using stacked neural networks. *AICHE J*, 42(9), pp. 2529-2539, 1996.
[5] Shi Han chang, Diao Huifang, Liu Heng *et al*. Application of treatment plant operation simulation and forecast software. *China Water and Wastewater*, 17(10), pp. 61-63, 2001.

[6] Li Jun, Yang Xiushan, Peng Yongzhen. *Microorganism and water treatment engineering*, Chemical industry press: Beijing, pp.74-98, 2002.
[7] Feng Yuzhao. *Study on hybrid intelligent control for variable rate anoxic biological filter wastewater treatment* [D]. Chongqing: Chongqing University Ph.D. Thesis, pp.5-37, 2004.
[8] Gao Ping ping, *The Study Based on Neural Network of Water Quality Prediction in Sewage Disposal* [D].Chengdu : Southwest Jiaotong University, pp. 11-49, 2005.
[9] Chang Yuqing, Lin Tian, Wang Fuli. Online Learning of Distributed Networks and Its Application in Soft Sensing. *Chinese Journal of Scientific Instrument*, 23(3), pp. 295-297, 2002.
[10] Luo Rongfu, Shao Huihe. Local learning methodology for distributed networks and its application in inferential control. *Acta Automatic Sinica*, 20(6), pp.739-741, 1994.
[11] Zhang Wenyi, Zhong Meiying, Cai Jian-an. Modeling Study of Activated Sludge Process Based on Artificial Neural Network. *Water & Wastewater Engineering*, 28 (6), pp.12-15, 2002.
[12] Wang Xudong, Shao Huihe, Luo Rongfu. Distributed RBF neural networks and its application in soft sensing. *Control Theory and Applications*, 15(4), pp.558-563, 1998.
[13] Yan Pingfan, Zhang Changshui. *Artificial neural network and simulated evolutionary computation*. Tsinghua University Press: Beijing, pp. 32-37, 2002.
[14] Luo Jianxu, Shao Huihe. Developing dynamic soft sensors using multiple neural networks. *Journal of Chemical Industry and Engineering*, 54(12), pp.1770-1773, 2003.
[15] Ran Weili, Qiao Junfei. BOD Soft-Measuring Approach Based on PCA Time-Delay Neural Network. *Transaction of China Electro technical Society*, 19(12), pp.78-82, 2004.

This page intentionally left blank

Surface heat flux determination using a genetic algorithm

L. P. Kanevce[1], G. H. Kanevce[2] & V. B. Mitrevski[1]
[1]*Faculty of Technical Sciences, St. Kliment Ohridski University, Macedonia*
[2]*Macedonian Academy of Sciences and Arts, Macedonia*

Abstract

This paper deals with surface heat flux estimation on the basis of transient temperature measurements using inverse method. As a direct problem the transient heat conduction through a plane slab isolated at one boundary and exposed to the heat flux on the other side is considered. The numerical experiments have been conducted in order to simulate real measurements. The inverse problem consists of heat flux estimation on the basis of experimental temperature response. The objective function, which has to be minimized in order to estimate unknown heat flux, is the least square function of experimental and calculated temperature data. A variant of genetic algorithm has been used for minimization of the objective function. Different convergence histories are presented, compared and discussed. Also, comparison between estimated heat flux change and exact solution and comparison between calculated and experimental transient temperature response are presented.
Keywords: genetic algorithm, heat flux, inverse problems, parameter estimation.

1 Introduction

An inverse approach to parameter estimation in the last few decades has become widely used in various scientific disciplines: like mechanical engineering, heat and mass transfer, fluid mechanics, optimization of the processes, optimal shape defining etc. By inverse method, on the basis of known effects, causes have to be defined, in contrast to the standard, direct method when causes are known and effects have to be determined. Application of inverse method is especially important in investigation of the processes where necessary direct measurements

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110471

are difficult or impracticable and also in finding the optimum regarding different aspects. In fact, the values of parameters which figures in the model of a physical process have to be obtained from the observed data.

Inverse problems of heat and mass transfer are based on temperature and/or heat flux measurements in order to obtain unknown parameter values appearing in the mathematical formulation of a considered problem. This approach is applied in problems of conduction, convection, radiation, and their combination like melting and solidification, in problems of heat and mass transfer with phase change, in drying processes, in composite materials exposed to high temperatures and so on [1]. In this work, the problem of heat flux evaluation on the free surface of the heated body on the basis of transient temperature measurements in the body is analyzed.

2 The objective function

For the parameter estimation problem the objective function is defined as the sum of the squared errors between experimental data and corresponding model predicted values [2, 3]:

$$E(\mathbf{P}) = [\mathbf{Y} - \mathbf{T}(\mathbf{P})]^T [\mathbf{Y} - \mathbf{T}(\mathbf{P})] \tag{1}$$

where $\mathbf{Y}^T = [Y_1, Y_2, \ldots, Y_{imax}]$ is a vector of measured temperatures, $\mathbf{T}^T = [T_1(\mathbf{P}), T_2(\mathbf{P}), \ldots T_{imax}(\mathbf{P})]$ calculated temperatures at times τ_i (i = 1, 2, …, imax), $\mathbf{P}^T = [P_1, P_2, \ldots, P_N]$ vector of unknown parameters, imax total number of measured temperatures, N total number of unknown parameters and superscript T indicates transpose. The parameter estimation problem is solved by minimization of function E. Parameter values at the minimum are solution of this problem. In this paper, for minimization in such a way defined objective function the genetic algorithm has been used.

3 Genetic algorithm

The genetic algorithm [4–6] is categorized as a random search method. During performing calculations the generator of random numbers is called several thousand times. However, the method also contains the certain rules which systematically lead toward solution. Because of the random search characteristic the method has slow convergence but in the same time with its reproduction and selection rules it is stable and reliable.

There are certain variations of this method depending on the application. The basic algorithm, fig. 1, processes the following characteristics:

- An initial population of a given number of candidates is selected by random choice.
- The fitness of each of the members of the population is determined using performance measure for the problem.

- The members of current generation of the population reproduce to create the next generation. The choice of the parents is random but the reproduction should favor the members with better fitness. During reproduction, crossover of the genes results in new members (children) not originally in the previous generation but related to them. If the child is better than the worst of the previous generation it survives and become the member of the new generation.
- With mutation probability some of the children mutate. These mutations introduce new characteristics not in the previous generation and not directly related to the previous generation but they may result in a more suitable child. Mutated child survives if it is better than the worst of the previous generation.
- The process of reproduction repeats until obtaining the member with desired characteristics or until a preset number of generations have been created.

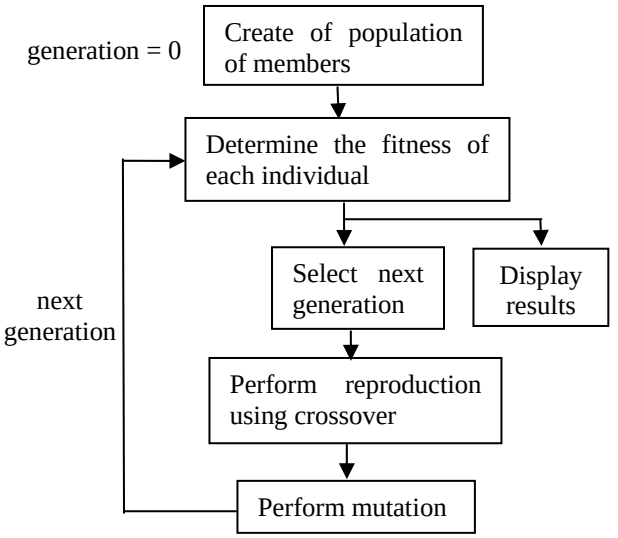


Figure 1: Flowchart of genetic algorithm.

In this paper a variant of the genetic method is applied for minimization of the objective function E of N parameters P_i:

$$E = f(P_1, P_2, ..., P_N) . \quad (2)$$

That means that values of parameters P_i have to be chosen in such a way that the objective function E reaches the global minimum.

The applied variant of genetic algorithm can be described by the following steps.

Initial population The initial population (first iteration population) is typically randomly generated i.e. every member of the initial population is combination of randomly chosen parameters:

$$E^1_j = f(P^1_{1,j}, P^1_{2,j}, \ldots P^1_{N,j}), \quad j = 1, 2, \ldots, N_{pop} \tag{3}$$

Each member of E^1_j parameters P^1_{ij} may be chosen in two ways.

According to the first the random value is generated using Gaussian distribution with variance σ and mean equals to the given value of the parameter in the zero iteration P^0_i:

$$P^1_{i,j} = P^0_i \pm \sqrt{-2\sigma^2 (P_{i,max} - P_{i,min}) \ln R} \tag{4}$$

where R is a random number between 0 and 1. Maximum and minimum values of parameters $P_{i,max}$ and $P_{i,min}$ are defined in advance depending of considered problem.

In the second way uniform distribution is used in the interval $[P_{i,max} - P_{i,min}]$:

$$P^1_{i,j} = (1 - R)P_{i,min} + RP_{i,max} \tag{5}$$

In this work the second approach is used, which enables faster convergence when the initial solution is far from final one.

Selection of parents The selection of parents also may be performed in different ways. In this work, the members of one population are sorted from the worst to the best one depending of values of the function E. The best member of the population has the smallest value of the function E. Then, father and mother are defined by random choice. In addition, in order to provide higher probability of choice for parents with smaller values of E, random numbers are reduced from the smallest to the largest one by special procedure.

Reproduction Using the method of crossover a new child is created, which typically shares many of the characteristics of its parents. The new child may be simple arithmetic mean of its parents. In this work the weighted mean is used and weight is randomly defined:

$$P_{i,child} = (1 - R)P_{i,father} + RP_{i,mother} \tag{6}$$

The value of the random number R in eqn (6) in classical problems usually ranges from 0 to 1, but in this case R ranges from -1 to 2 to provide wider range of combinations of features from parents.

Mutations The algorithm allows for a small chance of mutation with probability defined in advance. Parameters of new child are chosen randomly in the range $[P_{i,min} - P_{i,max}]$.

Elitism The new member is generated by presented processes of selection of parents, reproduction and mutations. This member survives if it is better than the worst member in the parent's generation. The procedure repeats until N_{bred} new survived members are produced. N_{bred} ($\leq N_{pop}$) is defined number depending of considered problem. The new generation is created when N_{bred} members of the previous generation is replaced with N_{bred} members of the new generation. The procedure repeats from generation to generation until minimum of objective function E or maximum number of iterations is reached.

4 The heat conduction model

The one-dimensional transient heat conduction through a plane slab is considered, fig. 2. The left side of the slab (x=0) is isolated and the free surface (x=1) is exposed to the unknown heat flux $q_p(\tau)$.

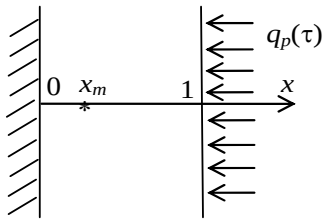


Figure 2: Heat conduction through a plane slab.

The mathematical formulation of this problem in dimensionless form is given by:

$$\frac{\partial T(x,\tau)}{\partial \tau} = \frac{\partial^2 T(x,\tau)}{\partial x^2}, \quad 0 < x < 1, \tau > 0, \tag{7}$$

$$\frac{\partial T(0,\tau)}{\partial x} = 0, \quad x = 0, \tau > 0, \tag{8}$$

$$\frac{\partial T(1,\tau)}{\partial x} = q_p(\tau), \quad x = 1, \tau > 0, \tag{9}$$

$$T(x,0) = 0, \quad \tau = 0, 0 \leq x \leq 1. \tag{10}$$

On the bases on solution of similar problem [3] and introducing certain transformations analytical solution of presented problem is obtained:

$$T(x,\tau) = \int_0^\tau q_p(\tau)\, d\tau + 2\sum_{m=1}^{\infty} e^{-\left(\frac{\beta_m}{2}\right)^2 \tau} \cos\left(\frac{\beta_m}{2} x\right) \cos\left(\frac{\beta_m}{2}\right) \int_0^\tau q_p(\tau) e^{\left(\frac{\beta_m}{2}\right)^2 \tau} d\tau \tag{11}$$

where $\beta_m = m\pi$.

It can be seen that this analytical solution is dependent of temperature flux so, **T(P)** is defined.

In this work transient measured temperatures $\mathbf{Y}^T$ at the location x_m are obtained by simulated experiment i.e. from solution of the heat conduction problem, eqn (11), defined by eqns (7)–(10).

5 Results

5.1 Constant heat flux example

This example considers constant heat flux (unknown parameter) applied at the free surface of the plane slab. The dimensionless value of the heat flux is $q_p = 1$. A simulated temperature sensor is placed at the isolated surface, $x_m = 0$.

For this case, temperature response $Y(0,\tau_i)$, i = 1,…,imax obtained from eqn (11), is shown, fig. 3. In the time interval from 0 to 2, imax (= 501) simulated experimental values Y_i are obtained. In order to analyze influence of measurement errors, the numerical experiment in whish random distributed error with $\sigma = 0.025T_{max} = 0.05$ is added to the exact solution. It is significant error value for this type of problems.

The problem is to estimate value of q_p on the basis of "measured" temperatures by minimization the objective function defined as the sum of the squared errors between model-predicted value and corresponding "experimental" data value, RMS. The problem has been solved using previously described genetic algorithm. The solution has been searched in the interval $-100 \le q_p \le 100$, with initial value $q_{p0} = -100$.

In fig. 4, the various convergence histories are presented. Figs from 3a to 3h represent different convergence histories until value RMS = 0.001 is achieved. It means that the numerical value of the heat flux is estimated with error less than 0.1%. These figures represent change of RMS with number of objective function calls. In this example it takes 7.3 s of CPU time on Pentium 4 (3.06 GHz), for 100 calls. The figures also represent the influence of the number of population

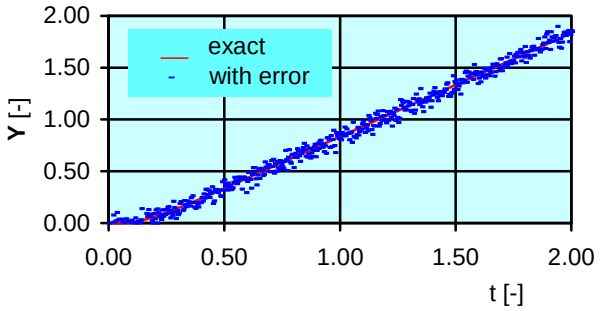


Figure 3: Transient temperature response $Y(0,\tau_i)$, i=1,…,501, for $q_p = 1$.

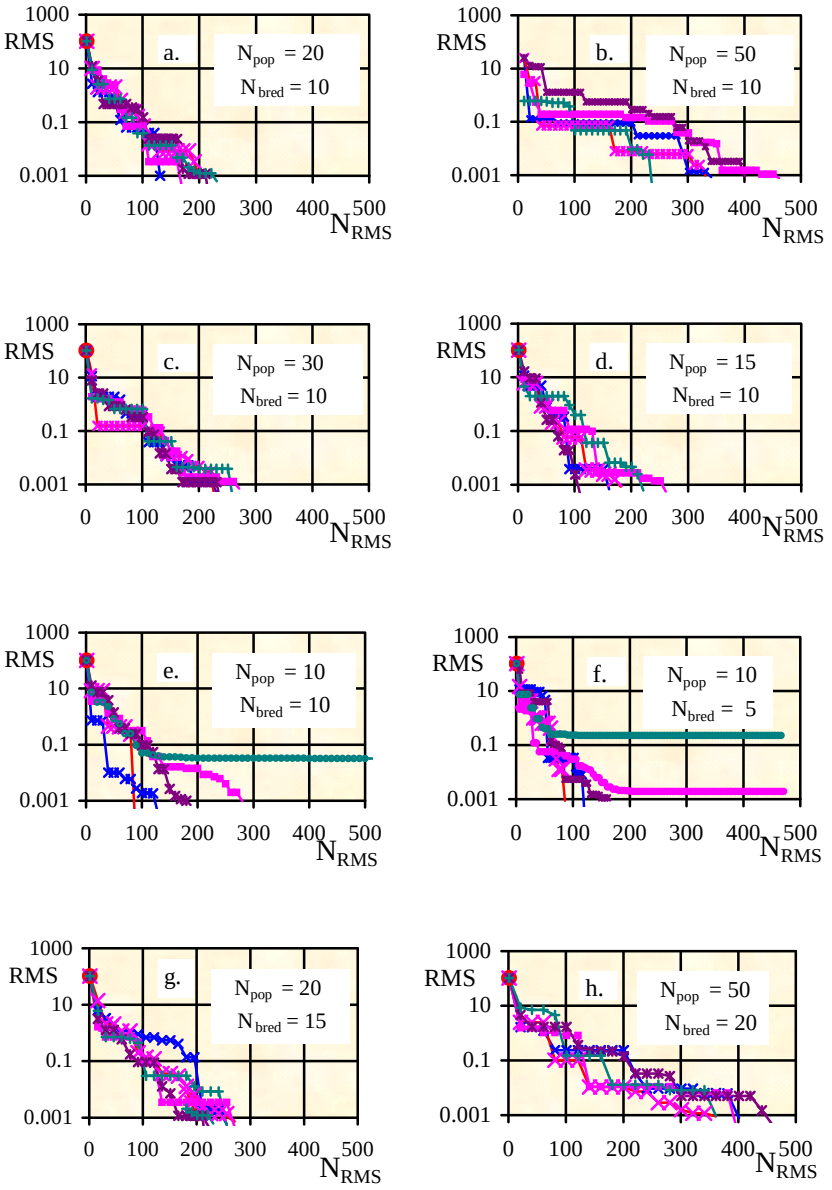


Figure 4: RMS change depending of the number of objective function calls N_{RMS}, different numbers of population members, N_{pop}, and number of generated members in each generation N_{bred} for case $q_p = 1$ and $\sigma = 0$.

members, N_{pop}, and the number of new generated members in each iteration, N_{bred}, to the convergence speed. Because of the statistical nature of the algorithm, each case has been run several times. The most rapid convergence has been obtained for N_{pop} = 20 and N_{bred} = 10, fig. 4a. If the number of population members N_{pop} is changed to 50, 30, 15 and 10, remaining N_{bred} = 10, as it is represented in fig. 4, from 4b to 4e, respectively, the slower convergence is obtained. It can be seen that in the case with N_{pop} = 10 and N_{bred} = 10, fig. 4e, certain runs have not solution. Similar case is with N_{pop} = 10 and N_{bred} = 5, fig. 4f. It leads to conclusion that, in considered problem here, population of 10 members is not big enough to provide new generation diversity and make progress. Fig. 4g and fig. 4h represent combinations with 20 and 50 members with different N_{bred} in respect to previous combinations. This does not lead to improvement in regard to combination with N_{pop} = 20 and N_{bred} = 10. It can be noticed that with increasing the number of the population members, the slower convergence history is obtained, fig. 4b and fig. 4h.

The convergence history in the case with measurement errors is given in fig. 5. It can be seen that, very fast, for about 150-200 objective function calls, RMS = σ = 0.050 is obtained. Error in the heat flux estimation in this case is less than 0.1%.

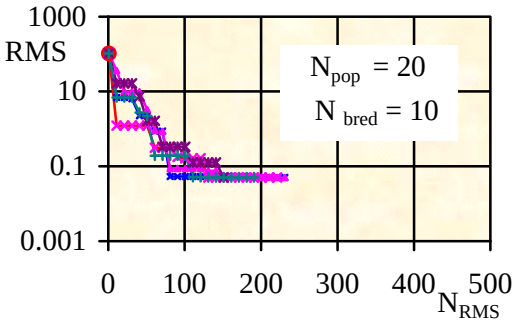


Figure 5: RMS change with objective function calls in the case q_p = 1 and σ = 0.050.

5.2 Transient heat flux example

In this example the heat flux changes in time according to the following exponential relation:

$$q_p(\tau) = a\, e^{-b\tau} \tag{12}$$

where a and b are unknown parameters need to be estimated.

As in the previous case, transient heat flux is determined on the basis of the response of temperature sensor at the location x_m = 0, fig. 2. In the same way, temperature response is obtained using simulated experiment i.e. from eqn (11) with inserted function for transient heat flux eqn (12). 101 temperature values

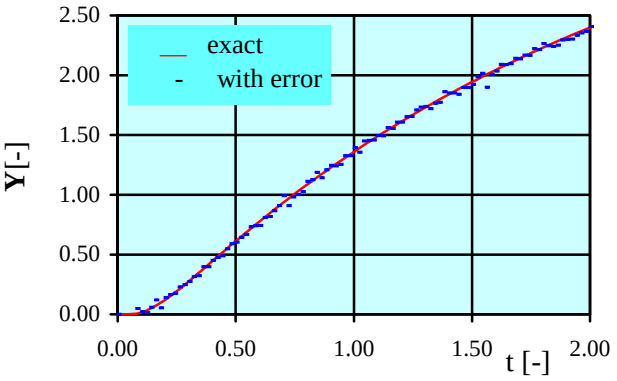


Figure 6: Transient temperature response, $T(0,\tau_i)$, i=1,…,101 in the case of exponential heat flux change, without and with measurement error ($\sigma = 0.025$).

have been calculated in the time interval from 0 to 2. In order to simulate real experiment, random distributed error is added to the exact solution values with $\sigma = 0.014T_{max} = 0{,}025$, fig.6.

Fig. 7 represents convergence history. The solution is obtained for about 4000 objective function calls. In this case unknown parameters a and b can be calculated with error less than 1% and 3% respectively. In this example it takes 1.9 s of CPU time on Pentium 4 (3.06 GHz), for 100 calls.

Model estimated and exact values of the heat flux are compared in fig. 8.

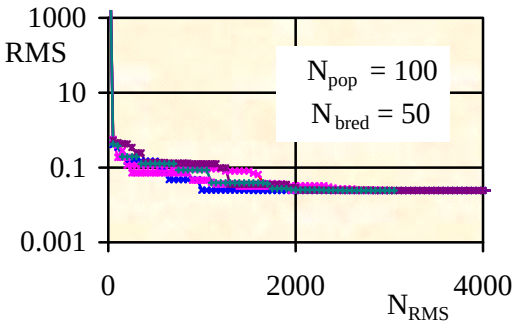


Figure 7: RMS change depending of the number of objective function calls N_{RMS}, for exponential form of heat flux function and measurement error with $\sigma = 0.025$.

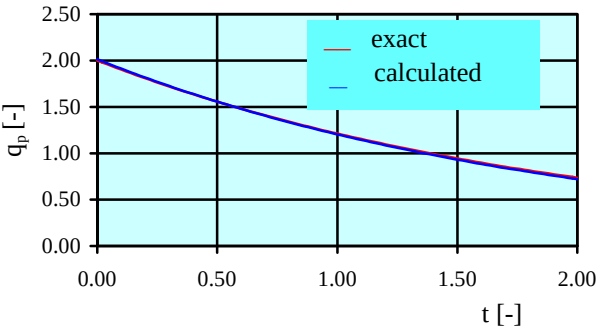


Figure 8: Comparison between estimated heat flux change in the case of exponential flux function and measurement temperature error with $\sigma = 0.025$, and exact solution.

Fig. 9 represents comparison between the temperature calculations with exact parameters and with estimated parameters a and b of the heat flux exponential function.

Increasing measurement error to $\sigma = 0.050$ leads to parameter estimation error for a and b of 1% and 5% respectively.

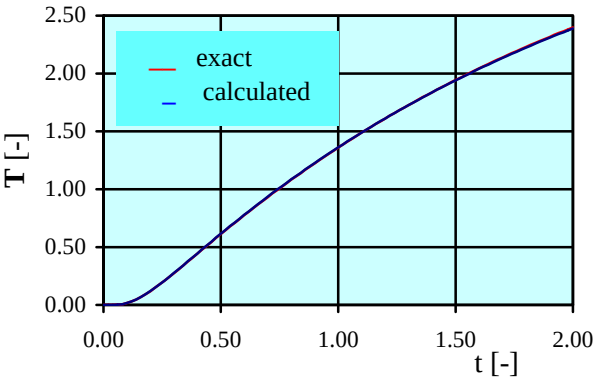


Figure 9: Calculation of transient temperature response $T(0,\tau_i)$ for the case of exponential shape of the heat function with error ($\sigma = 0.025$) in compare to exact solution.

6 Conclusion

In this paper the application of the genetic algorithm to the heat flux estimation on the bases of transient temperature response is presented and analyzed. The

case of constant heat flux (one unknown parameter) and case of exponential transient heat flux (two parameters) are analyzed. The convergence history dependences of different combinations of population and breed number are presented. It can be concluded that the genetic algorithm may be successfully used in both cases. The constant heat flux can be estimated with high accuracy even if measurement is realized with relatively significant error. In the case of transient heat flux, estimation depends of the measurement error.

References

[1] Kanevce, G.H. Kanevce, L. P., Mitrevski, V.B., Dulikravich, G. S. & Orlande, H.R.B., Inverse approaches to drying of thin bodies with significant shrinkage effects, *Journal of heat transfer – transactions of the ASME*, **129(3)**, pp. 379-386, 2007.

[2] Beck, J. V. & Arnold, K. J., *Parameter Estimation in Engineering and Science*, John Wiley & Sons, Inc., New York, p. 370, 1977.

[3] Ozisik M. N. & Orlande H. R. B., *Inverse Heat Transfer: Fundamentals and Applications*, Taylor and Francis, New York, 2000.

[4] Woodbury K. A., Application of Genetic Algorithms and Neural Networks to the Solution of Inverse Heat Conduction Problems, *4th International Conference on Inverse Problems in Engineering* , Ed. H. R. B. Orlande, e-papers, Rio de Janeiro, pp. 73-88, 2002.

[5] Mitchell, Melanie, *An introduction to Genetic Algorithms*, MIT Press, 1999, pp. 2-11, 1999.

[6] Adam Marczyk, *Genetic Algorithms and Evolutionary Computation, What is a genetic algorithm*, Copyright © 2004. http://www.talkorigins.org/faqs /genalg/genalg.html

This page intentionally left blank

New solid-state organic membrane based lead-selective micro-electrode

H. A. Arida[1,2], A. Al-Haddad[3] & M. J. Schöning[4]
[1]*Chemistry Department, Faculty of Science,*
Taif University, Saudi Arabia
[2]*Hot Laboratories Center, Atomic Energy Authority, Egypt*
[3]*Chemistry Department, Faculty of Science, Bahrain University, Bahrain*
[4]*Institute of Nano- and Biotechnologies,*
Aachen University of Applied Sciences, Germany

Abstract

Microfabrication, characterization and analytical application of a new thin-film organic membrane based lead-selective micro-electrode have been elaborated. Prior to the fabrication of the assembly, the gold thin-film substrate has been electrochemically treated using a new technique. The developed micro-electrode based on *tert*-Butylcalix[4]arene-tetrakis(N,N-dimethylthioacetamide) as electroactive sensing material, carboxylated PVC as supporting matrix, 2-nitrophenyl octyl ether as solvent mediator and potassium tetrakiss (4-chlorophenyl) borate as lipophilic additive, respectively, provides a nearly Nernstian response (slope 28±0.5 mV/concentration decade) covering the concentration range $1x10^{-6}$ – $1x10^{-2}$ mole L^{-1} of Pb(II) ions with reasonable selectivity over some tested cations. The merits offered by the new microelectrode include simple fabrication, low cost as well as automation and integration feasibility. Moreover, the suggested microelectrode has been successfully applied for the determination of lead ions in some aqueous samples. These samples were also determined using inductively coupled plasma atomic emission spectroscopy (ICP-AES) for comparison. The proposed electrode offers a good accuracy (the average recovery was 95.5%), high precision (RSD was <3%), fast response time (<30 s.) and long life span (>4 months).
Keywords: all-solid-state microelectrode, thin-film, substrate surface treatment, organic membrane, lead determination.

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110481

1 Introduction

Ion-selective electrodes (ISEs) and potentiometric sensors are the most widely used sensor types for the measurement of toxic heavy metal ions [1, 2]. The development of all-solid-state micro-sensor devices originating from potentiometric sensors has accelerated during the last few years, and this is likely to continue. The realization of such devices seems to be accelerating as micro-scale construction makes it possible to apply principles that would not work in macro-scale analogous. In addition, accurate and reliable analysis using miniaturized chemical sensors is a very useful analytical technique because of the avoidance of laborious and time consuming preliminary sample treatment. Moreover, micro-scale analyses of chemical species have many advantages over conventional methodologies, including high spatial resolution, rapid response, and minimal disturbance of the analyzed substrate [3]. Use of different organic and inorganic sensing materials with versatile properties in fabrication of all-solid-state micro-electrodes makes them suitable for the detection of many chemical species in solution at concentrations lying in the ppm range [4–10]. In realization of such devices, chalcogenide glasses were proven to be very promising ion-selective membranes especially for the detection of heavy metals in solution (Pb^{2+}, Cd^{2+}, Fe^{2+}, Cu^{2+}, Ag^{+}...) [4–6]. However, organic membrane-sensitive layers prepared on transducers, fabricated using different, less or more complicated and expensive, technologies for measurements of potassium [7] and lead [8] have been reported, too. In addition, nanoparticle labels (i.e., gold nanoparticles, silver tags, and semiconductor nanocrystals) have been used in the fabrication of potentiometric micro-sensors for detection of DNA hybridization [9] and carbon dioxide [10]. Miniaturization of solid-electrolyte gas sensors to thin-film micro-devices have been discussed in literature [11]. Microfabrication of chemical sensors and biosensors [12] as well as ISFET-based micro-sensors [13] for environmental monitoring has been reviewed. A Pt-Ir wire-based ISE has been suggested for monitoring the local spatial distribution of magnesium, pH and ionic currents [14]. Moreover, the realization of micro-sensors, based on a “lab-on-a-chip” has also been reviewed [15].

On the other hand, the development of organic membrane based micro-sensors has been recently introduced to overcome the low selectivity of chalcogenide glass and inorganic based thin-film micro-sensors [16–19]. However, there is an additional problem arise that the adhesion of the organic membrane to the thin-film substrate is usually poor, which produces an early degradation of those micro-sensors. To solve this problem, we had recently developed a new approach (Arida Approach) for the organic-based sensors micro-fabrications [17–19]. In this technique, the organic membrane-based sensitive layer has been nebulized in combination with a substrate surface treatment. Using these two steps in combination has distinctly improved the adhesion on the wafer surface. It decreases the leaching out of the ionophore and plasticizer, stabilizes the organic membrane, and consequently increases the micro-electrode’s life-time.

In this paper, we describe the fabrication and characterization of PVC membrane-based micro-sensors using *tert*-Butylcalix[4]arene-tetrakis(N,N-dimethylthioacetamide) as lead ionophore [20], deposited on gold thin-film silicon micro-chips by the nebulization technique in combination with substrate surface treatment. The reliability of the suggested micro-sized lead sensors in comparison to independent standard methods using ICP-AES has been assessed.

2 Experimental

2.1 Chemicals and apparatus

The solvent mediator, 2-nitrophenyl octyl ether and the lipophylic additive potassium tetrakis(4-chlorophenyl) borate were purchased from Sigma-Aldrich (CH-9471 Buchs, Switzerland). The membrane support matrix, high molecular weight (220,000) poly(vinylchloride) carboxylated and the membrane solvent, THF (tetrahydrofurane) were purchased from Riedel-de Haën Chemical Company (Germany). The lead ionophore used was tert-Butylcalix[4]arene-tetrakis(N,N-dimethylthio-acetamide) (15343) purchased from Sigma-Aldrich chemical company. All the standard solutions of cations were prepared from their analytical reagent grade chemicals in deionized water, and then diluted to the desired concentration. Nitrate or chloride salts of the metal used were purchased from Riedel-de Haën. High purity standards (2% HNO_3, Pb 1000 mg kg^{-1}) were used for ICP-AES validation measurements after appropriate dilution with deionized water. Deionized water with conductivity <0.2 μS/cm used in the preparation and dilution of the reagents was produced using a Millipore Deionizer (Millipore, France, Elix 10).

The potentiometric measurements were performed at 20°C with a HANNA microprocessor pH/ion analyzer (Model pH 211) using a thin-film lead micro-electrode in conjunction with a double-junction reference electrode immersed in stirred test solutions. The response characteristics and the selectivity coefficient $K_{Pb,M}^{pot}$ (obtained by separate solution method) of the thin-film lead micro-electrode have been measured using standard methods [21].

The morphology of the substrate surface was studied using a JEOL scanning electron microscope SEM (Model JSM 6390, Japan). A Perkin-Elmer (Optima 2100 DV) inductively coupled plasma atomic emission spectrometer (ICP-AES) instrument connected with an AS 93 Plus autosampler has been used for the standard determination of lead. The 40-MHz free-running generator was operated at a forward power of 1300 W; the outer, intermediate and Ar carrier gas flow rates were 15.0, 0.2 and 0.8 L/min, respectively. The pump flow rate was 1.5 mL/min. The carrier gas flow rate was optimized to obtain maximum signal-to-background ratios.

2.2 Micro-fabrication of the lead solid-state micro-electrode

A cocktail-coating mixture incorporating the Pb^{2+} ionophore was utilized as organic membrane-sensitive layer. The stock solution of this mixture was

prepared by mixing thoroughly 33 mg of powdered PVC, 57 mg of 2-nitrophenyl octyl ether plasticizer, 3 mg of potassium tetrakis (p-chlorophenyl) borate anion excluder and 7 mg of *tert*-Butylcalix[4]arene-tetrakis(N,N-dimethylthioacetamide) lead (II) ionophore in a small beaker. The mixture was then completely dissolved in 5 mL THF [16–19]. The sensor chips with the thin-film gold substrates were fabricated at Research Centre Jülich as described elsewhere [16, 22–24].

Prior to the deposition of ion-sensitive organic membrane, the gold thin-film substrate surface was treated electrochemically to enhance the adhesion to the organic membrane. Here, thin films of Ag precipitates have been deposited on the substrate surfaces from 10^{-3} mol L^{-1} $AgNO_3$ solution for 1 h using a home-made small electro-deposition cell. In combination with this step and based on the previously reported technique [17–19], the cocktail-coating mixture was nebulized on the treated surface of the gold thin-film substrate using a small commercial nebulizer. The nebulized solution was delivered to the substrate in fast pulses with 2 min intervals between the pulses. The deposition time was 30 min. After deposition of the organic membrane-sensitive layer, the sensors were scribed to single chips, glued and contacted via bonding to a printed circuit board (PCB) and encapsulated.

3 Results and discussion

3.1 Physical characterization of the lead micro-electrode

Using the nebulization method, the all-solid-state micro-electrode has been fabricated by depositing the organic membrane sensitive-layer on the treated surface of the silicon micro-chips (10 mm× 10 mm) and tested as lead-sensitive electrode. A video-microscopic picture of the organic membrane-based thin-film micro-electrode after integration into a PCB is shown in Fig. 1. The diameter of the thin-film micro-electrode is about 2 mm.

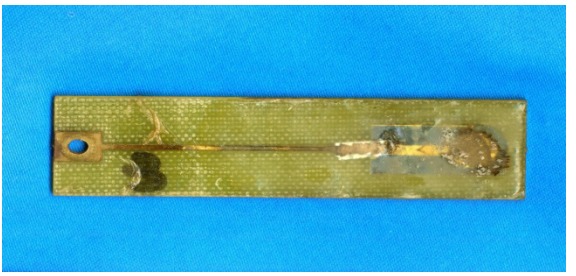


Figure 1: Video-microscopic picture of an all-solid state lead micro-electrode.

Prior to the electrochemical characterization, the morphology studies on the gold thin-film substrate, treated substrate surface and organic membrane deposit have been performed using scanning electron microscopy (SEM) in order to check the film/membrane interface. The micrographs obtained are shown in

Fig. 2. The surfaces of all films do not present any observable defects. While, the untreated film (*a*) appears smooth and luster with poor adhesive properties, the treated substrate surface (*b*) becomes more mountain-like with high roughness and consequently, good adhesion to the organic membrane. The organic membrane film (*c*) is textured, homogenous and uniformly distributed. This significantly enhances the stability and consequently, the life span of the suggested micro-electrode.

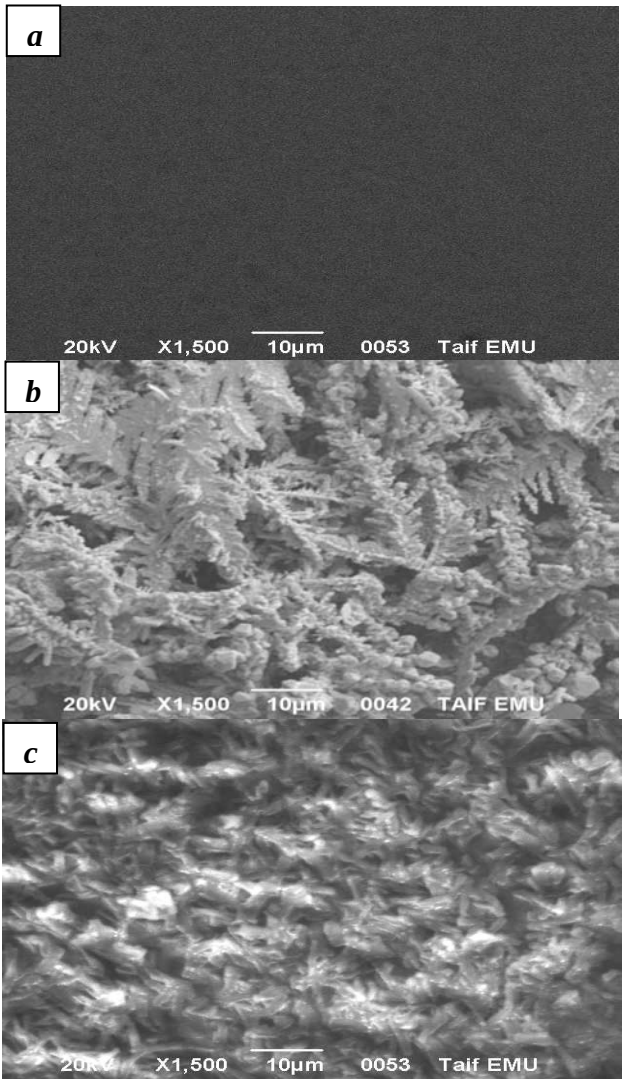


Figure 2: SEM micrograph of the thin-film surfaces; (*a*) untreated gold substrate, (*b*) electrochemical treated and (*c*) nebulized organic membrane-sensitive layer.

3.2 Electrochemical evaluation of the lead micro-electrode

Our goal was to realize sensitive, stable and reliable all-solid-state lead micro-electrodes using the commercialized tert-Butylcalix[4]arene-tetrakis(N,N-dimethylthio-acetamide) lead ionophore. The suggested micro-electrode incorporating this ionophore has been electrochemically evaluated according to IUPAC recommendations in terms of sensitivity, stability, selectivity and reliability. In order to assess the reliability of the suggested micro-electrode, its response properties were compared with those reported for the bulk macro-electrode prepared by the same ionophore [20]. The data obtained are collected in Table 1. As can be seen, the micro-electrode showed a nearly theoretical Nernstian response (slope of 28±0.5 mV/concentration decade) with a straight line between $1x10^{-6}$ and $1x10^{-2}$ mole L^{-1} of Pb(II) ions (see Fig. 3). The limit of detection determined from the intersection of the two extrapolated segments of the calibration, as recommended by IUPAC, was $5x10^{-7}$ mol L^{-1}.

Table 1: Potentiometric performance properties of the thin-film lead micro-electrode and the bulk macro-electrode.

Parameter	All-solid-state thin-film micro-electrode	*Lead bulk macro-electrode
Slope (mV/decade)	28±0.5	28.7
Response time $t_{95\%}$(s)	<30	<10
Linear range (mol L^{-1})	1×10^{-2}-1×10^{-6}	1×10^{-2}-1×10^{-6}
Detection limit (mol L^{-1})	$5x10^{-7}$	$3.2x10^{-7}$
pH	2.2-6.3	3-6
Life span (months)	>4	<2

**Lead bulk macro-electrode [Ref. 20].*

The response time and life span of an electrode are important features for analytical applications. Hence, the potentials of the suggested micro-electrode corresponding to four decade additions of Pb^{2+}, starting from deionized water to 10^{-2} mol L^{-1} have been measured. The results obtained are presented in Fig. 4. The response time ($t_{95\%}$) of the suggested micro-electrode in the whole linear concentration range was about 30 s. The stability and lifetime of the lead micro-electrode were also investigated by repeated calibrations at every two or three days for more than four months. During this period, the response time, slope and linear range are reproducible. Hence, the micro-electrode can be used for at least four months with practically unimpaired performance. Moreover, the suggested micro-electrode properties are almost similar or even somewhat better (life span) than the results observed for the lead macro-electrode based on the same ionophore [20]. The recently developed, nebulization technique of the organic membrane-sensitive layer in combination with the electrochemical treatment of the substrate surface significantly improves the adhesive properties of the membrane to the substrate surface, decreases the leaching out of the ionophore and plasticizer, stabilizes the organic membrane, and consequently increases the micro-electrodes' life-time (4> months).

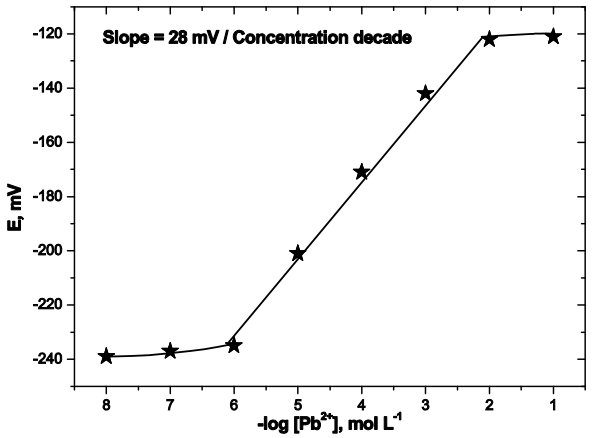


Figure 3: Potentiometric calibration response of the organic membrane based lead microelectrode.

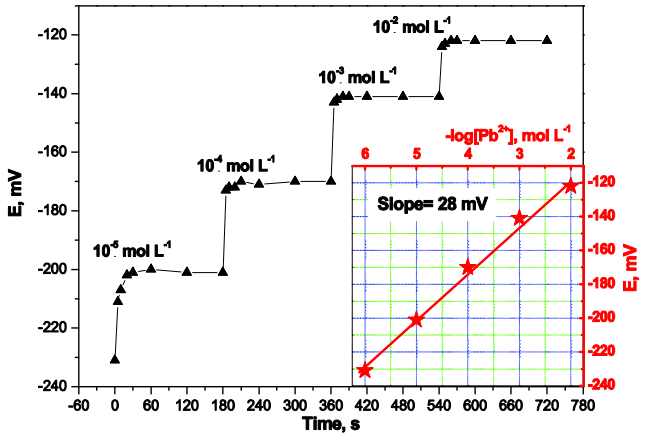


Figure 4: Potentiometric dynamic response of the organic membrane based lead microelectrode.

The dependence of the response of the suggested micro-electrode on the pH value of the test solution was examined at two Pb^{2+} concentrations over the pH range between 1 and 8. As illustrated in Fig. 5, for both 1.0×10^{-3} and 1.0×10^{-2} mol L^{-1} Pb^{2+}, the potential remains constant over the pH range from pH 2.2 to 6.3. As a result, this range can be taken as the working pH range of the suggested micro-electrode. Variation of the potential at pH< 2 could be related to abundant H^{+} ions which can protonate the ionophore in the membrane phase, resulting in its decomplexation of Pb^{2+}. While, decreasing of the potential at higher pH values may be ascribed to the hydrolysis of Pb^{2+} ions, leading to a decrease in its concentration.

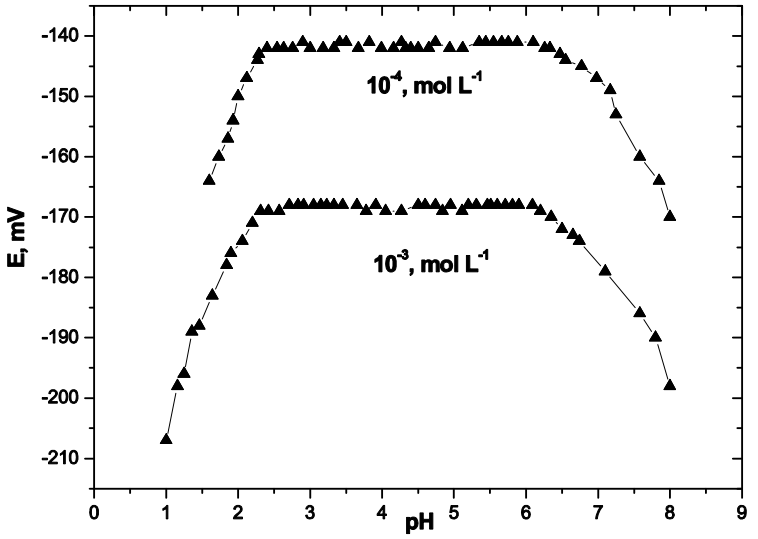


Figure 5: Effect of pH on the potentionetric response of the organic membrane based lead microelectrode.

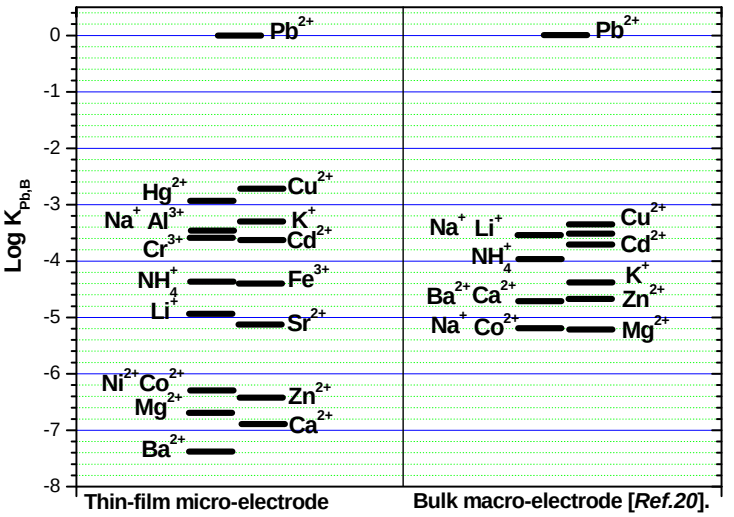


Figure 6: Selectivity coefficients, log $K_{Pb,B}$, for organic lead macro- and micro-electrodes, [0.01 mol L^{-1} of Pb^{+} and interfering cations].

Potentiometric selectivity reflects the relative response of the membrane electrode for the primary ion over other ions, present in solution. This feature is perhaps the most important characteristics, since our goal here is the realization of a highly selective organic membrane-based lead micro-electrode instead of inorganic-based micro-sensors with high cross-sensitivity. The potentiometric response of the lead micro-electrode was studied in presence of some alkali, alkaline earth, transition metals, aluminum and ammonium ions. The potentiometric selectivity coefficients $K_{Pb,B}^{pot}$ were determined using the separate solution method [21]. The results depicted in Fig. 6 are compared with those reported for the bulk lead macro-electrode prepared with the same ionophore [20]. The data revealed that the suggested micro-electrode showed a very high selectivity towards most of the tested ions. This behavior is similar to or even more selective (Ba^{2+}, Ca^{2+}, Zn^{2+}, Mg^{2+}, Co^{2+} and Ni^{2+}) than values obtained from the macro-electrode based on the same ionophore. One possible explanation might be that the response mechanism of the micro-sized electrodes (particularly selectivity) is significantly affected by the size (surface to volume) effect. The enhancement of the micro-electrode' selectivity may be attributed to the fact that these chips have a large size to volume ratio as well as they offer a high density of electronic component and interconnection possibilities in contrast with the macro-electrodes based on the same ionophore.

3.3 Analytical applications of the proposed micro-electrode

To assess the applicability of the micro-electrode to lead measurements, an attempt was made to determine lead in five different samples under optimized conditions**.** These samples were also determined by an independent standard method using ICP-AES, for comparison. The data obtained are summarized in Table 2. As can be seen, there are no significant differences between the results obtained with the proposed micro-electrode and the reference method. In short, the results reveal a high accuracy (recovery 95.5%) and good reproducibility (RSD<3%, n=6).

Table 2: Determination of lead in some aqueous samples employing the micro-electrode in comparison to the reference method ICP-AES.

No.	ICP-AES (mg/L)	*Thin-film micro-electrode (mg/L)	*Recovery %
1	0.21	0.22	104.7
2	5.23	4.88	93.3
3	25.67	23.78	92.6
4	93.01	86.11	92.5
5	210.64	199.57	94.7
Average	recovery		95.5

**The data is a mean of n= measurements.*

4 Conclusion

A new thin-film organic membrane-based lead micro-electrode incorporating *tert*-Butylcalix[4]arene-tetrakis(N,N-dimethylthioacetamide) as electroactive material in a PVC matrix has been realized. The micro-electrode responds to lead ions in a nearly Nernstian behavior and presents a good selectivity and detection limit. The micro-electrode reveals a fast response time and long-term stability. It has been successfully applied for the determination of lead in some aqueous samples. The micro-electrode showed a good accuracy and reproducibility in comparison to the independent ICP-AES method.

Acknowledgement

The authors wish to thank Taif University for financial supports (Project number 766/431/1, 2010).

References

[1] Eric, B. & Ernö, P., Potentiometric sensors for trace-level analysis. *Trends in Analytical Chemistry*, 24(3), pp. 199-207, 2005.
[2] Ru-Qin, Y., Zong-Rang, Z., Guo-Li S., Potentiometric sensors: aspects of the recent development. *Sensors and Actuators B: Chemical*, 65(1-3), pp. 150-153, 2000.
[3] Wróblewski, W., Dybko, A., Malinowska, E., Brzózka, Z., Towards advanced chemical microsensors—an overview. *Talanta*, 63, pp. 33–39, 2004.
[4] Taillades, G., Valls, O., Bratov, A., Dominguez, C., Pradel, A., Ribes, M.,ISE and ISFET microsensors based on a sensitive chalcogenide glass for copper ion detection in solution. *Sensors and Actuators B*, 59, pp. 123–127, 1999.
[5] Guessous , A., Sarradin, J., Papet, P., Elkacemi, K., Belcadi, S., Pradel, A., Ribes, M., Chemical microsensors based on chalcogenide glasses for the detection of cadmium ions in solution. *Sensors and Actuators B*, 53, pp. 13– 18, 1998.
[6] Schöning, M.J. & Kloock, J.P., About 20 Years of Silicon-Based Thin Film Sensors with Chalcogenide Glass Materials for Heavy Metal Analysis: Technological Aspects of Fabrication and Miniaturization- A review. *Electroanalysis*, 19, pp. 2029-2038, 2007.
[7] Zachara, J. & Wróblewski, W., Design and performance of some microflow-cell potentiometric transducers. *Analyst*, 128, pp. 532–536, 2003.
[8] ISILDAK, I., All Solid-State Contact Lead(II) Ion-selective PVC Membrane Electrode Using Dimethylene Bis(4-methylpiperidinedithiocarbamate) Neutral Ionophore. *Turk J Chem*, 24, pp. 389-394, 2000.

[9] Numnuam, A., Chumbimuni-Torres, K.Y., Xiang, Y., Bash, R., Thavarungkul, P., Kanatharana, P., Pretsch, E., Wang, J., Bakker, E., Potentiometric Detection of DNA Hybridization. *J. Am. Chem. Soc.*, 130(2), pp. 410–411, 2008.
[10] Jennifer C.X., Hunter, G.W., Lukco, D., Liu, C.C., Ward, B.J., Novel Carbon Dioxide Microsensor Based on Tin Oxide Nanomaterial Doped With Copper Oxide. *NASA/TM*, 215436, pp. 1- 2, 2008.
[11] Dubbe, A., Fundamentals of solid state ionic micro gas sensors. *Sensors and Actuators B* , 88, pp. 138–148, 2003.
[12] Suzuki, H., Microfabrication of chemical sensors and biosensors for environmental monitoring. *Materials Science and Engineering C*, 12, pp. 55–61, 2000.
[13] Jimenez-Jorquera, C., Orozco, J., Baldi, A., ISFET Based Microsensors for Environmental Monitoring. *Sensors*, 10, pp. 61-83, 2010.
[14] Lamaka, S.V., Karavai, O.V., Bastos, A.C., Zheludkevich, M.L., Ferreira, M.G.S., Monitoring local spatial distribution of Mg^{2+}, pH and ionic currents. *Electrochemistry Communications*, 10, pp. 259–262, 2008.
[15] Kricka, L.J., Miniaturization of analytical systems. *Clinical Chemistry*, 44(9), pp. 2008–2014, 1998.
[16] Arida H., Joachim P., Michael J., Novel organic membrane based thin-film microsensors for the determination of heavy metal cations. *Sensors*, 6, pp. 435-444, 2006.
[17] Arida H., Development of a Long-Term Stable Organic Membrane Based Thin-Film Microsensor using New-Type Substrate Surface Treatment. *Talanta*, 71, pp. 1856-1860, 2007.
[18] [Arida H., Monika T., David R., Michael J., A novel thin-film array based on an organic/inorganic sensor hybrid: Microfabrication, characterization and flow-injection analysis application. *Electroanalysis*, 21(10), pp. 1145–1151, 2009.
[19] Arida H., Mohsen Q., Michael J., Microfabrication, characterization and analytical application of a new thin-film silver microsensor. *Electrochimica Acta*, 54, pp. 3543–3547, 2009.
[20] Malinowska, E., Brzózka, Z., Kasiura, K., Egberink, R. J. M., Reinhoudt, D.N. , Lead selective electrodes based on thioamide functionalized calix[4]arenes as ionophores. *Analytica Chimica Acta*, 289, pp. 253- 258, 1994.
[21] Umezawa, Y., Bühlmann, P., Umezawa, K., Tohda, K., Amemiya, S. , *Pure Applied Chemistry*, 72, pp. 1851, (2000).
[22] Mourzina, Y., Schöning, M.J. , Schubert, J. , Zander, W., Legin, A., Vlasov, Y., Kordos, P., Lüth, H., *Sensors and Actuators B*, 71, pp. 13, 2000.
[23] Kloock, J., Mourzina, Y., Ermolenko, Y., Doll, T., Schubert, J., Schöning, M.J., *Sensors*, 4, 156, 2004.
[24] Kloock, J., Mourzina, Y., Schubert, J., Schöning, M.J., *Sensors*, 2, pp. 356, 2002.

This page intentionally left blank

Measurement of methane pressure by gas-pressure-measurement-tube sets in underground coalbeds

S. Torno[1], E. Alvarez[2], B. Arias[2] & J. Toraño[1]
[1]*School of Mines, University of Oviedo, Spain*
[2]*Hullera Vasco Leonesa S.A. Santa Lucia, Leon, Spain*

Abstract

Sudden methane outbursts which often cause serious accidents with numerous fatalities can be produced in underground coalbed roadways and when driving tunnels for highways or railways passing through carboniferous materials.

Most of the factors which take part in the generation of sudden outbursts are related to methane conditions into the coalbed and specifically to gas pressure and coal permeability.

When excavating roadways or tunnels crossing coalbeds, it is convenient to have easy to use equipment which allows us to know at any moment the measured pressure and its variation in factors such as roadway and tunnel driving velocity, geological faults which lead to fractured formations and natural or induced overstresses.

The designed gas-pressure-measurement-tube set goes into an 8 m deep and 45 mm in diameter borehole drilled in front of the working face. The gas-pressure-measurement-tube set is connected to a flexible rod with a manometer at the end of it.

The obtained measurements are used both in the empirical formulae and in Computational Fluid Mechanics; this allows us to predict the probability of the occurrence of sudden outbursts.

Keywords: gas pressure, methane, outburst, gas-pressure-measurement tube, underground mine.

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110491

1 Introduction

Outbursts in the underground coal mines have been and will be a highly topical question as a consequence of the repercussion which they have on miner safety.

The statistics of fatal accidents are unfortunately widespread and extensive in mines worldwide [1–3].

Alterations which are originated by mining works generate a reorder of tensional state of coalbed till it achieves a new equilibrium, being able to make different phenomena evident during these processes. This affects the stability of coal mine created and consequently the safety of the workers.

Under the denomination of dynamic phenomena, a set of manifestations, in a sudden way, due to existing energetic conditions emitted, are included. As a consequence, a process of projection of coal and rock which can be accompanied by the gas content in the coal is produced. These energetic conditions are generated by the interactions of the following three factors: tensional state of coalbed, presence of gas in the coal and in some cases gravity.

In figure 1, the triangular graphic which related the three factors giving way to the different types of gas emission phenomena is shown. Moreover, it can clearly be seen which properties are more influenced in these factors.

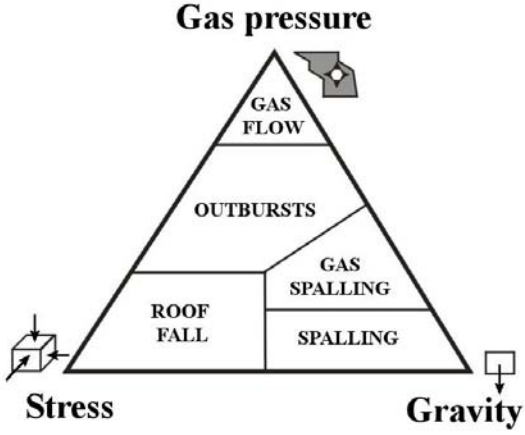


Figure 1: Triangular graphic of gas dynamic phenomena.

The outburst of coal and methane consists of sudden and violent interruption of coal methane in the vacuum created by an explosion [4, 5]. This phenomenon correspond to a modification of the most favorable conditions for the disintegration of the coal close to the face and favours the intense and sudden desorption of methane ([6] and Coal Group of Energy Division of United Nations Economic Commission for Europe (UNECE) [7]).

In the Complementary Technical Instruction developed by the Principado de Asturias (legislation applicable to the location of the mine, S.A. Hullera Vasco Leonesa, and where the experimental measurements are carried out) outbursts of

coal and gas are defined as 'one of the most complex gas dynamic phenomena' with the following characteristics:

-The sudden destruction of face with a common cavity in the coalbed.
-The pneumatic transport of broken coal.
-A higher gas outburst to the normal content of the coal, which is progressively reduced.

In underground coal mine of S.A. Hullera Vasco Leonesa, the outbursts could be classified as a medium type, since from 50 - 400 t of coal are affected. In this study, 300 t of coal were ejected and about 1000 m^3 of gas have been emitted. This classification orders the outbursts in five groups from 0.5 – 10 t to > 1000 t. Most of the factors which take part in the generation of sudden outbursts are related to methane conditions in the coalbed, especially to gas pressure and coal permeability.

In this paper, the necessity to make use of simple equipment which allows us to know at any moment the measured pressure and its variation in factors such as roadway and tunnel driving velocity, geological faults which lead to fractured formations and natural or induced overstresses is manifested.

For experimental measurements a gas-pressure-measurement tube set which goes into an 8 m deep and 45 mm in diameter borehole drilled in front of the working face is designed. This gas-pressure-measurement tube set is connected to a flexible rod with a manometer at the end of it.

A continuous record of the gas pressure measurements in the coalbed allows us to take preventive measures against the outbursts. This avoids the presence of outbursts and reduces the risk of serious accidents for the miners.

2 Risk indexes

There are numerous indexes which quantify the different factors that determine the presence of outburst, some of them are the following: methane concentration in coalbed [8], desorption velocity of methane [9], methane concentration in ventilation, V30 and the German Regulation [10], drill cuttings – Jahns test [11], measurements of gas pressure in the roadway and coalbed [2], the Polish ST index [12], the Russian Ps index, from WostNII Institute in Kuznetsk, the Psi index from the Karaganda coalfield in Kazakhstan and the k and D indexes from China [13].

The nature of these indexes are variable, from normal ones (some of them are applied in other disciplines) to those which are both consider mining experiences and useful.

In both simple indexes and combined indexes, the gas pressure is of greater importance; therefore, it is the most important factor studied in this research.
There are two types of gas pressure measurements in coalbed: measurements in the coalbed and measurements in the heading face. The first measurements which correspond to the phase of coalbed research have been carried out by means of boreholes of greater length where a gas-pressure-measurement tube set which registers the evolution of the gas pressure in long periods of time has been placed.

The pressure changes according to the depth where the results are expressed by means of quotient between the gas pressure and lithostatic pressure. Measurements and control of gas pressure at a few meters.

These techniques associated with exploitation works have been carried out in a mine in Czech Republic, and consist of a daily systematic measurement of the gas pressure in boreholes of 3 m in depth and 42 mm in diameter.

This measurement instrument consists of a tube equipped with a sealing system connected to a manometer or monitoring system.

The measurements are stabilized every 3 minutes. The critical values of this index are specified in each coalbed and are stabilized by numerous sets of measurements. The most common gas pressure value findings are approximately 50 kPa.

In 166 analyzed outbursts, in various coalfields such as Jagorszynski, Kuznetsk, Donetsk or Karaganda, the following conclusions, related to the influence of the gas pressure, have been detected [13]:

-No outbursts have been recorded for gas pressure below 6 MPa.

-For pressure between 0.6 and 0.74 MPa, 79 outbursts were generated with a maximum intensity of 2977 and a medium of 54 t.

-For pressure above 0.74 MPa, 87 outbursts were generated with a maximum intensity of 1500 t and a medium of 145 t.

3 Gas-pressure-measurement tube set

Hullera Vasco Leonesa S.A. Company located in the province of Leon in the North of Spain, is an underground coal mine with three shafts connected between them. The annual coal production is 2 million tons and the proved exploitable coal reserves at the end of 2009 are 45 million tons.

The average coal properties of the mined coalbed are 12.39% Ash, 0.75% Sulfur, 10.41% Volatiles, Calorific value of 7455 Kcal/Kg and Density1.45 kg/m^3 Density.

Stratigraphically, the coal basin is divided into six well-defined formations. The whole group is about 1,500 m thick, formed about 300 million years ago. Pastora Formation shows the best economic prospects and it provides the totality of the present coal production. The Pastora coalbed varies in thickness from 20 to 25 m, its dip is between 40º and 70º and the methane concentration varies between 7 and 10 m^3/t of coal.

The exploitation of this coalbed where the outbursts were produced was exploited by two methods, sublevel caving and shortwall between two levels [14]. The mining method by shortwall and sublevel caving is performed with a particular system of shield supports. The panel width varies according to the thickness of the coalbed. The large coalbed thickness contributes the generation of outbursts [13].

The Mining Company S.A. Hullera Vasco Leonesa has designed a tool in order to measure the gas pressure. Firstly, the tool or gas-pressure-measurement tube was created piece by piece in the SolidWorks software as shown in figure 2, where ‘l’ is the length in meters and ‘d’ is the diameter also in meters.

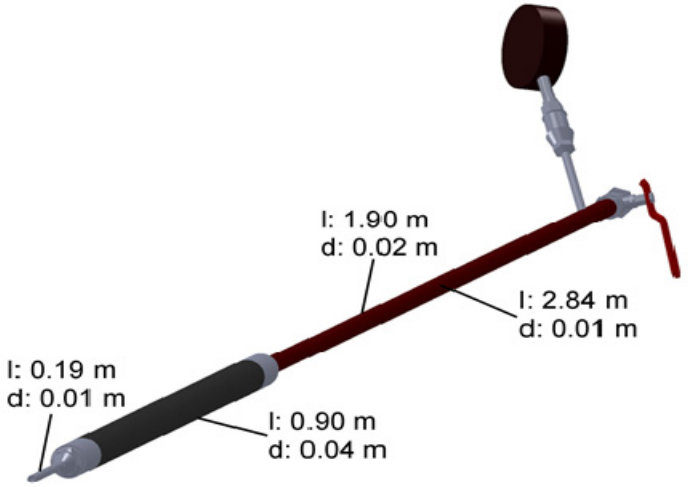


Figure 2: Diagram of gas-pressure-measurement tube designed by SolidWorks.

Secondly, the gas-pressure-measurement tube was manufactured by S.A. Hullera Vasco Leonesa, as it is illustrated in figure 3.

In figure 3 the gas-pressure-measurement tube is illustrated with its final configuration, demonstrating the following parts:

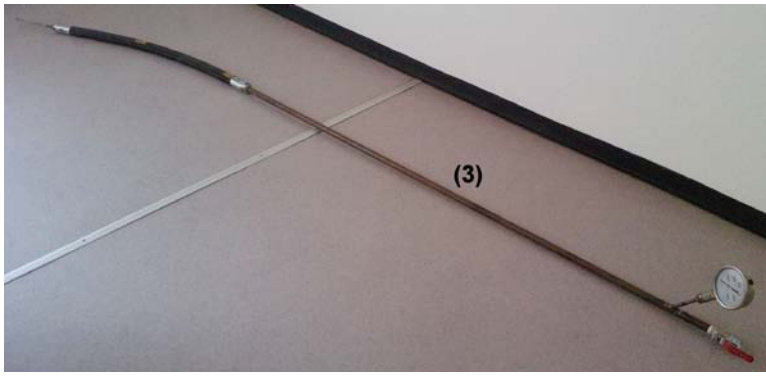
-Measurement tube (1).
-Sealing system (2).
-Borehole rod (3).
-Manometer or monitoring system (4).
-Borehole rod (5) to inject water at a pressure of 1.96 MPa used for the sealing system.

When it is necessary to introduce this tube into deeper boreholes it is lengthened by means of an extensible flexible tube.

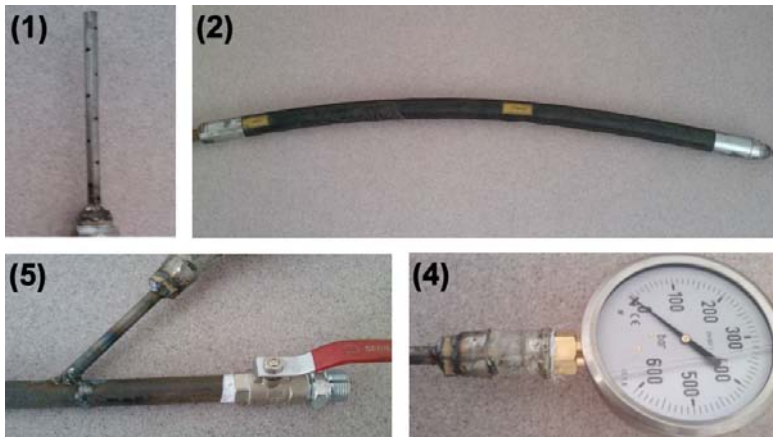
The boreholes used for inserting the gas-pressure-measurement tube sets were carried out at the crown of the heading face. They were 8 m in length, 45 mm in diameter and 10 º angles in upward direction.

The gas-pressure-measurement tubes were put into operation at night, on Friday nights, in order to take gas pressure evolution measurement at heading face until seven o'clock on Monday morning.

When the gas pressure measurements were performed, the heading face was degassed with 6 boreholes at the crown of 3.2 m in length and 65 mm in diameter, and in the middle with 3 boreholes of 5.6 m in length and 105 mm in diameter.



a)



b)

Figure 3: Gas-pressure-measurement tube prior to being introduced into the boreholes.

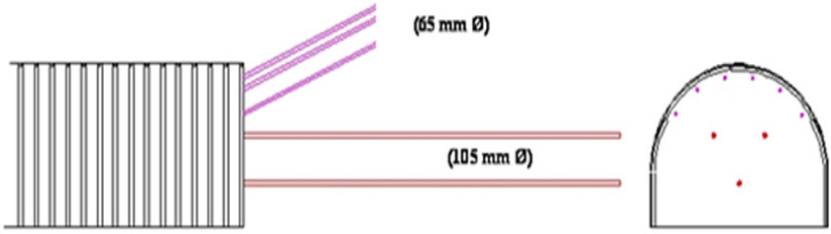


Figure 4: Description of boreholes where the gas-pressure-measurement tube is inserted in order to carry out the measurements.

The horizontal distance that separates the face of heading and face of shortwall-sublevel caving is denominated 'D'. This distance changes as the mine is being exploited and can be divided into 3 groups:

-D is positive when the two faces become closer.
-D is 0 when the two faces coincide.
-D is negative when the faces overlap.

Figure 5 illustrates the results of the evolution over time of gas pressure (kPa) in 3 boreholes in the roof for five distances D = 33, 23, 2, -11.5 and -40m.

It can be seen how, for four hours after inserting the measurement tube (D = 30 m), there is an increase in the gas pressure reaching maximum values of 900 kPa.

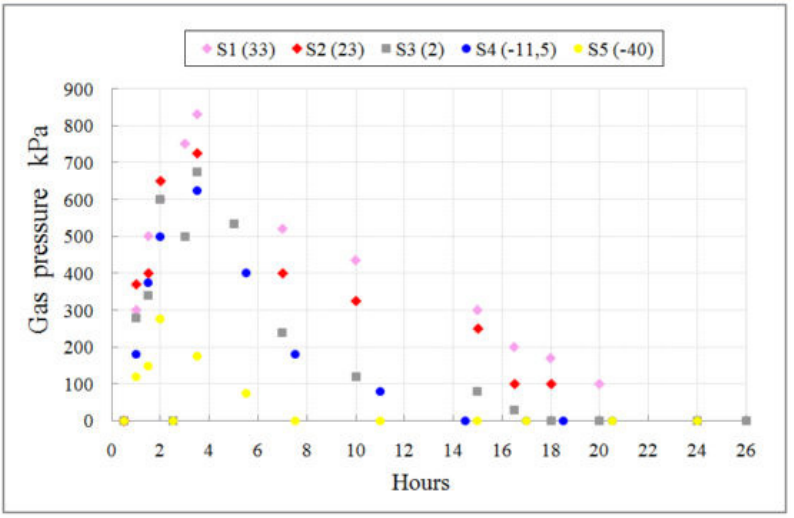


Figure 5: Evolution of the gas pressure in the coalbed in function on distance D and time.

When D decreases, the maximum gas pressure also decreases, reaching values below 700 kPa. Later a faster decrease in this pressure occurs over time until becoming 0.

When the distance D is negative, the maximum values of gas pressure are lower than 500 kPa and when the absolute value of D is higher, the maximum pressure is 300 kPa.

The evolution of the gas pressure shows that there is an outburst-prone zone as a consequence of a gas pressure increase for the distance D (between 25 and 50 m).

For lower values of D, as a result of the decrease in the Protodiakonov index of the heading face (average values of 1.31, maximum values of 2.13 and

minimum values of 0.51), the danger of the occurrence of an outburst diminishes, and the problems related to methane blows increase [13].

4 Conclusions

The measurements and evolution of gas pressure by boreholes in the heading face allow us to predict by means of simple and operative tools, the trend of the generation of outbursts in the coalbed.

It is convenient for the gas-pressure-measurement tube to be motorized in order to have a remote and continuous register of the gas pressure.

When higher elevations of gas pressure, indicated by the gas-pressure-measurement tube are generated, the coalbed should be degassed by means of boreholes in the heading face with an injection of pressure water.

If the pressure increase is excessive, penalization of the works could be necessary.

References

[1] Belin, J. & Vandeloise, R., Résultats des recherches efectúes en Belgique et en France sur les dégagements instantanés dans les mines de charbon. *Annales des Mines de Belgique*. Nº 2 Février, 1969.
[2] Lama, R.D. & Bodziony, J., Management of outburst in underground coal mines *International Journal of Coal Geology*, 35, pp. 83–115, 1998.
[3] Basil Beamish, B.B. & Crosdale, P.J., Instantaneous outbursts in underground coal mines: An overview and association with coal type International. *Journal of Coal Geology*, 35, pp. 27–55, 1998.
[4] Peng, S.S, Coal mine ground control. John Wiley & Sons. New York 1986.
[5] Hargraves, A.J., Update on instantaneous outbursts of coal and gas. Proceeding *of Australian Institute of Mining and. Metallurgy*, 298, pp. 3–17, 1993.
[6] Belin, J., Mesures de prevention des degagement instantanés de methane et de charbon ou de roches. *Organe permanente pour la securité et la salubrité dans les mines de houlle*. Nº 6039/81. Luxembourg, 1983.
[7] ECE, Coal Group of Energy Division of United Nations Economic Commission for Europe. Luxembourg, 1995.
[8] ET0307-2-92. Especificación técnica: Método para determinar la concentración de gas en la capa. Ministerio de Industria del Estado Español.
[9] ET0307-2-92. Especificación técnica: Método para determinar la velocidad de desorción del grisú. Ministerio de Industria del Estado Español.
[10] Renania, Circulares 23.10.87, 6.7.82 y 13.8.81 de la Dirección Superior de Minas del Estado de Renania sobre DI. Dormund. Verlag. H. Bellmann, 1987.
[11] Bräuner, G., Rockburst in coal mines and their prevention. Ed. Balkema. Rotterdam, 1994.

[12] Tarnowski, J., Calculations concerning coal and gas outburst. International Symposium of Workshop on Management, Control of High Gas Emission and Outburst. Ed. Lama. Wollongong, pp. 49-61, 1995.
[13] Lama, R.D., Bodziony, J., Sudden outburst of gas and coal in underground coal mines. Final Report. Project No. C 3034. Australian Coal Association Research Program, 1996.
[14] Toraño, J., Torno, S., Menéndez, M., Gent, M.R., Velasco, J., Models of methane behaviour in auxiliary ventilation of underground coal mining. *International Journal of Coal Geology*, 79, pp. 157-166, 2009.

This page intentionally left blank

Self assembling proteins as matrix for the construction of optical devices

U. Weinert, T. Günther, J. Raff & K. Pollmann
Forschungszentrum Dresden-Rossendorf e.V., Germany

Abstract

Self-assembling biomolecules are widespread in nature and attractive for technical purposes due to their size and highly ordered structures in nanometer range. Surface-layer (S-layer) proteins are one of those self-assembling molecules and their chemical and structural properties make them quite attractive for nanotechnical purposes. They possess a high content of functional groups so a sequential coupling of functional devices is possible and their ability to self assemble in aqueous solutions or on surfaces, e.g. SiO_2 wafers, qualifies them for nanotechnical applications. In this work, first experiments were done in order to construct a sensory device containing S-layer proteins as matrix to bind optical elements and analytes for detection of specific substances. The S-layer proteins were isolated from the *Lysinibacillus sphaericus* strain JG-A12 recovered from a uranium mining waste pile in Germany. As optical elements fluorescent dyes or quantum dots can be used. Three different fluorescent dyes which are able to perform a Fluorescence resonance energy transfer (FRET) were used and coupled to the S-layer proteins. As receptor molecule aptamers were chosen due to their high specifity and stability towards many chemicals. Aptamers are short oligonucleotides which are able to bind specific molecules via their three dimensional structure. In this work, a model aptamer was used that is specific towards human thrombin. The aim was to construct a sensor system which is able to detect specific substances in very low concentration ranges in aqueous solutions.

Keywords: S-layer proteins, fluorescent dyes, aptamers.

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110501

1 Introduction

S-layer proteins represent the outermost cell envelop compound of numerous bacteria and almost all archaea. These structure proteins feature a lot of functions such as protection, shape and molecular sieves. They are able to self assemble in aqueous solutions and on surfaces and form regular sheets exhibiting a paracrystalline nanostructure [1]. Generally, S-layer proteins provide a huge amount of orientated functional groups that can be used for coupling of molecules to the surface, thus introducing a high level of functionality in a small device. The possibility to create highly regular arrays of nanostructured multifunctionalized surfaces, makes S-layer proteins highly attractive for the construction of sensory layers.

Here we present a new concept of biosensors based on the application of S-layers. This biosensor will be able to detect various substances in a very low concentration range, e g. pharmaceuticals in clarified waste water. Figure 4 shows the functionality of the biosensor with following compounds:

(1) S-layers that are used for the nano-structuring and functionalization of surfaces such as SiO_2-wafers or glass.

(2) Aptamers that are coupled to S-layers and work as receptors. Aptamers are oligonucleotides that specifically bind chemical compounds via their three-dimensional structure. In the present study, aptamers that specifically bind thrombin were used as a model.

(3) Two different fluorescence dyes with overlapping excitation and emission spectra. A unit cell of the S-layers has a size of 12.8 nm and comprises four protein monomers. Therefore coupled fluorescence dyes are in a close proximity, thus favouring a fluorescent resonance energy transfer (FRET). FRET means a radiation free energy transfer from one fluorescence dye (donor) to another fluorescence dye (acceptor). The energy transfer causes a decrease in donor fluorescence and an increase in acceptor fluorescence. Such a transfer system is very sensitive against environmental changes and can be easily disturbed e.g. by the binding of analytes to the aptamer.

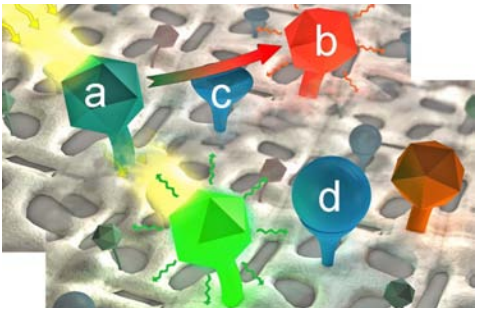


Figure 1: Schematic drawing of a sensory device containing two fluorescent dyes (a and b) which are able to perform a FRET and one aptamer (c) specific for one analyte (d).

The S-layer proteins of *L. sphaericus* JG-A12 that were used for the present study contain 10 mol% NH_2 and 12 mol% COOH groups, consequently providing many residues for functionalization. These groups of the S-layers can be easily coupled with fluorescent dyes by using crosslinking reagents like 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimid (EDC). Additionally, hydroxyl-groups of the protein can be coupled with aptamers. In the present study the coupling of the three different fluorescent dyes Hilyte488, Hilyte555 and TAMRA and the anti-thrombin aptamer was performed and the results analyzed.

2 Materials and methods

2.1 Isolation of S-layer proteins

The isolate *L. sphaericus* JG-A12 was grown in Nutrient Broth medium (6 g/l, Mast Group Ltd., Merseyside, UK) in a 50 L bioreactor over night at pH 7 and 30°C. Extraction of S-layer proteins was performed according to [2]. Determination of protein concentration was determined by Lowry [3] and purity of S-layer proteins was confirmed by SDS-PAGE. Proteins were stored in standard buffer solution at pH 7.5 containing 50 mM Tris, 2 mM $MgCl_2$ and 3 mM NaN_3.

2.2 Functionalization of S-layer proteins

2.2.1 Binding fluorescent dyes to S-layer proteins

S-layer sheets were centrifuged and resolved in 50 mM phosphate buffer at pH 5.6 to give a final concentration of 10 mg/mL S-layer protein polymer solution. Coupling experiments were performed with three different fluorescence dyes. As donor molecule, Hilyte 488 was chosen and used as hydrazide. As acceptor molecules Hilyte555 and TAMRA were tested and used as hydrazide or succimidyl ester (SE), respectively. The carboxyl groups of the S-layer proteins were activated with EDC in a final concentration of 20 mM. The reaction started immediately. Reaction tubes were incubated overnight at 4°C in dark. Afterwards the samples were centrifuged and washed several times to remove remaining fluorescent dyes. The coupling of fluorescent dyes to the proteins was quantified by UV/VIS (Ultrospec 1000, Amersham Pharmacia Biotech, USA). For theses measurements 50 µL of the protein solutions were dissolved in 50 µL of 6 M guanidine solution pH 7 and incubated for one hour in order to get a monomer solution.

2.2.2 Binding anti-thrombin-aptamer to S-layer proteins

Oligonucleotides with the sequence GGTTGGTGTGGTT [4], a TAMRA modification at its 3′ end and a SH-group at its 5′ end were used as anti-thrombin-aptamer. The aptamer was linked to OH-groups of the S-layer protein via its SH-group by using N-[*p*-Maleimidophenyl]isocyanate (PMPI) as crosslinking reagent. The reaction was started by the addition of PMPI, dissolved in DMSO, to a solution of S-layer proteins and aptamers, dissolved in distilled water at pH 7 with a final molar ratio of 1:2.5:500 (protein:aptamer:PMPI).

Reaction was performed overnight at 4°C. Precipitated PMPI was removed by centrifugation and selective resuspension. After reaction, the residual aptamers in the supernatant were removed. The binding of aptamer to S-layer proteins was controlled via UV/VIS spectroscopy and quantification of linked aptamer was done by measuring the uncoupled aptamer in the supernatant.

3 Results and discussion

3.1 Binding fluorescence dyes to S-layer proteins of *L. sphaericus* JG-A12

In a first step, UV/VIS spectroscopy was used to demonstrate that the S-layer protein of *L. sphaericus* JG-A12 can be successfully modified with each of the three fluorescent dyes. TAMRA was linked to the NH_2-groups whereas Hilyte488 and Hilyte555 were linked to the COOH-groups. As an example, the results on S-layer modifications with Hilyte488 or TAMRA and with both dyes together are shown in figure 2. Unfortunately, the spectra of free and protein bound dyes show a slight peakshift and an overlap of donor and acceptor absorption. These facts make it difficult to quantify the total amount of bound fluorescence dyes by UV/VIS spectroscopy. Therefore fluorescence spectroscopy was used to determine the amount of bound fluorescence dyes. If two fluorescence dyes are linked to the protein the total amount of each fluorescence dye can be determined and compared with samples where only one fluorescence dye is linked.

For fluorescence measurements the optimal excitation and emission wavelength of the chosen fluorescence dyes in this experimental setup have to be determined. In result Hilyte488 was excited at a wavelength of 500 nm, TAMRA at a wavelength of 556 nm and Hilyte555 at a wavelength of 553 nm. Fluorescence spectra of protein modified with TAMRA and protein modified with a combination of Hilyte488 and TAMRA show same emission intensities at 580 nm when excited at a wavelength of 556 nm (see figure 3). This fact indicates that comparable amounts of TAMRA were linked to the protein in both samples. An equal approach to determine the amounts of linked Hilyte488 by excitation with light of a wavelength of 500 nm is not possible, as an occurring FRET will cause a decrease of the fluorescence intensity at 530 nm. This effect can be seen in figure 3, whereas the fluorescence intensity at 580 nm simultaneously increases in comparison to the emission spectra of protein modified with TAMRA and excited at a wavelength of 500 nm. Both observations together are a strong indication for a FRET event.

A modification of S-layer protein with Hilyte488 and Hilyte555 at the same time reduces the total amount of bound molecules for both dyes in comparison to samples with only one linked fluorescence dye (data not shown). This can be easily explained by the binding of both dyes via COOH groups so that less binding sites are available for each fluorescence dye. Furthermore an excitation of Hilyte555 at a wavelength of 500 nm shows a clear emission at 565 nm which

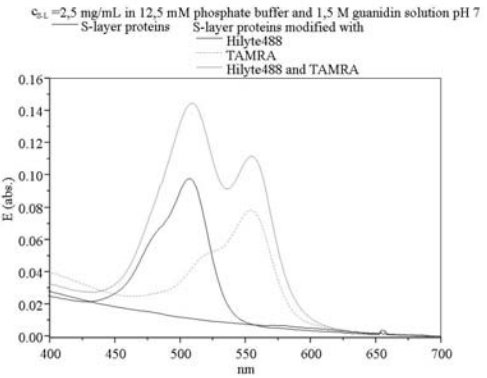


Figure 2: UV/VIS spectra of S-layer proteins (solid black line) modified with Hilyte488 (dashed line), with TAMRA (dotted line) and with both fluorescence dyes (solid grey line).

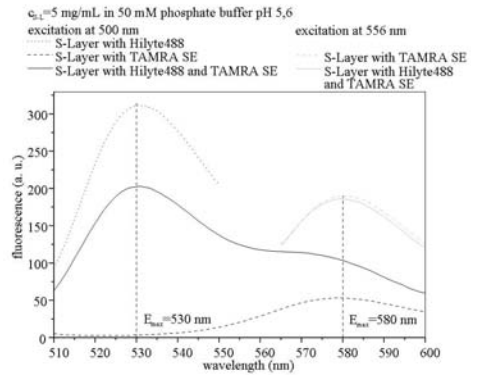


Figure 3: Fluorescence spectra of S-layer proteins modified with Hilyte488, TAMRA or with both fluorescence dyes; excitation wavelength for Hilyte488 was 500 nm and for TAMRA 556 nm.

hinders a potential FRET detection due to the fact that a donor excitation simultaneously excites the acceptor. Therefore further experiments will be done with the Hilyte488 and TAMRA as potential FRET pair.

3.2 Binding anti-thrombin-aptamers to S-layer proteins of *L. sphaericus* JG-A12

As presented in figures 4 and 5, anti-thrombin-aptamers were successfully coupled to S-layer proteins of *L. sphaericus* JG-A12 via PMPI. Modification degree was determined by measuring the absorption maxima of uncoupled anti-thrombin-aptamer that remained in the supernatant after crosslinking. Whereas light microscopic images show that all S-layer polymers are modified with

aptamers, the results of UV/VIS spectroscopy demonstrate that only 20% of the applied S-layer protein monomers were modified with aptamers. Future work will focus on an enhanced modification rate in order to increase the sensitivity of the finally assembled sensor. Additionally S-layer proteins will be stabilized with EDC in order to inhibit degradation after the coupling of aptamer.

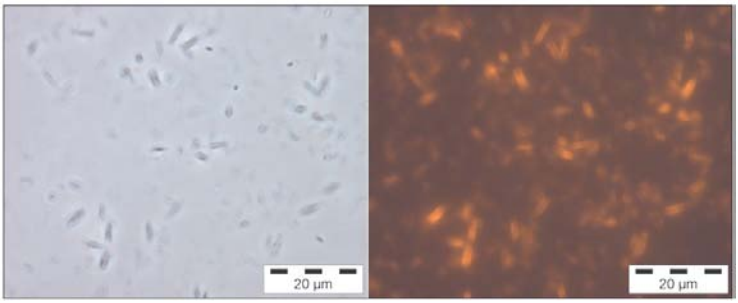


Figure 4: Light microscopic image of S-layer proteins modified with anti-thrombin-aptamer in phase contrast mode and in fluorescence mode using a filter between 540 and 550 nm.

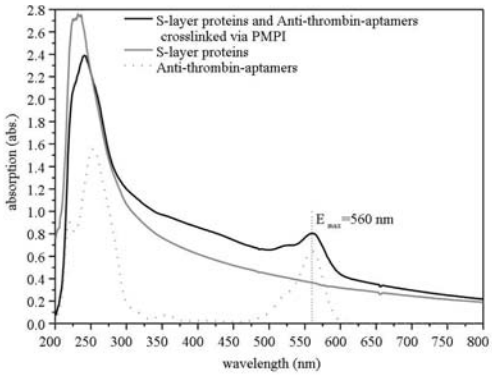


Figure 5: UV/VIS spectra of unmodified S-layer proteins (grey line), free anti-thrombin-aptamer (dotted line) and S-layer proteins modified with anti-thrombin-aptamer (black line).

4 Outlook

For the construction of a biosensor that is composed of S-layers, fluorescence dyes and aptamers, important steps are the stable coupling of the functional molecules to the S-layer proteins without loosing the functionality of those molecules and the structure of the S-layers. The presented results demonstrate that such a functionalization is possible while keeping the structure of the S-layer polymers. In further steps affinity studies of the coupled anti-thrombin-aptamers

will be done in order to verify the intactness of the molecules. In addition, lifetime measurements will be performed in order to prove the occurrence of an energy transfer between the fluorescence molecules.

Acknowledgement

This study was kindly supported by the German Federal Ministry of Education and Research (*BMBF*), grant.01RB0805A.

References

[1] Sará, M. and Sleytr, U. B., S-layer proteins, *Journal of Bacteriology*, **182(4),** pp. 859-868, 2000
[2] Raff, J. et al. *Chem Mater* **15**(1), pp. 240-244, 2003
[3] Lowry, O. H., Rosebrough, N. J., Farr, A. L., Randall, R. J., Protein measurement with the folin phenol reagent. *Journal of Biological Chemistry*, **193(1),** pp. 265-275, 1951
[4] Bock, L. C., Griffin, L. C., Latham, J. A., Vermaas, E. H., Toole, J. J., Selection of Single-Stranded-DNA Molecules That Bind and Inhibit Human Thrombin. *Nature*, **355(6360),** pp. 564-566, 1992

This page intentionally left blank

An algorithm for the calibration of running time calculation on the basis of GPS data

S. de Fabris, G. Longo & G. Medeossi
Dept. of Civil and Environmental Engineering, University of Trieste, Italy

Abstract

Increasing the precision in timetable planning is a key success factor for all infrastructure managers, since it allows us to minimize delay propagation without reducing usable capacity. Since most running time calculation models are based on standard and deterministic parameters an imprecision is implicitly included, which has to be compensated by running time supplements.

At the same time, GPS or even more precise trackings are continuously stored in the event recorders of most European trains. Unfortunately, this large amount of data is normally stored but not used except for failure and maintenance management.

To consider real running time variability in running time calculation, an approach has been developed, which allows us to calibrate a performance factor for each motion phase.

Given the standard motion equation of a train, and a mesoscopic model of the line, the tool uses a simulated annealing optimisation algorithm to find the best regression between calculated and measured instant speed. To increase precision, the motion is divided into four phases: acceleration, braking at stops, braking for speed reductions/signals and cruising. By performing the procedure over a number of train runnings, a distribution of each performance parameter is obtained. Once the infrastructure model is defined and the trackings are imported, the procedure is completely automated.

The approach can be used in both stochastic simulation models and as a basis for advanced timetable planning tools, where stochastic instead of deterministic running times are used. The tool has been tested in the north-eastern part of Italy as input for both running time calculation and microscopic simulation.

Keywords: railway simulation, railway planning, GPS, train performance, calibration.

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110511

1 Introduction

Increasing the precision in timetable planning is a key success factor for all infrastructure managers, since it allows us to minimize delay propagation without reducing usable capacity. When setting up a timetable, it is necessary to estimate running times for different rolling stock using the current infrastructure. Conventional running time calculators solve the motion equation, which is based on a number of empirical parameters. Such parameters have been measured for years for different kinds of rolling stock [1]; therefore it is possible to calculate train speed profile in high detail. However, many influences on running times are not deterministic, such as human behaviour, weather conditions and even different trains of the same series could show different performances. To cope with this variability, recovery times are inserted, implicitly including an imprecision in the representation of train motion.

While a deterministic running time calculation is sufficient to plan timetables, which of course must be deterministic, a more detailed representation is required in micro-simulation and for ex-ante evaluate timetable robustness estimations, which aim at reproducing train behaviour with highest detail.

A *performance factor*, which introduces a stochastic element in the motion equation, has been proposed by some authors and inserted in proven simulation tools. This factor in multiplied by the tractive effort, the speed limit and the braking deceleration during acceleration, cruising and braking respectively.

To estimate the distributions of such parameters, an iterative approach was proposed by the Authors [3]. A software tool which allowed us to compare on-board collected data and running time calculation was developed from scratch and tested on a line in Northern Italy, demonstrating the benefits of calibrated motion equation in stochastic micro-simulation.

The results appeared very promising, although calibration of motion equation was performed manually by the user, graphically comparing simulated and real speed profiles at given points.

To overcome this weakness, and to allow a more precise calibration based on a higher number of records, a new software tool has been developed, which allows an automatic calibration of performance factors.

2 Approach and software tool

The proposed method includes a number of steps that are required to prepare data sets, filter the on-board collected data, calculate running time using deterministic motion equations and then calibrate them. All steps and data are included in a single tool to avoid exporting and importing data from in different software, reducing the number of steps performed by the user.

The software tool has been developed using Java programming language in order to use on different operating systems without compatibility problems, while the Netbeans IDE development environment was chosen for the simple windows drawing procedure and the good built-in debugging and profiling features.

The tool integrates a microscopic infrastructure model, rolling stock and timetable data to calculate running times. Real world data are filtered and then imported. A simulated annealing algorithm is used to calibrate the performance factor over a number of records. The distributions of performance factors are then presented and stored, to be used in the tool or exported to other simulation software supporting stochastic performance factors. The working flow of the approach is shown in Figure 1.

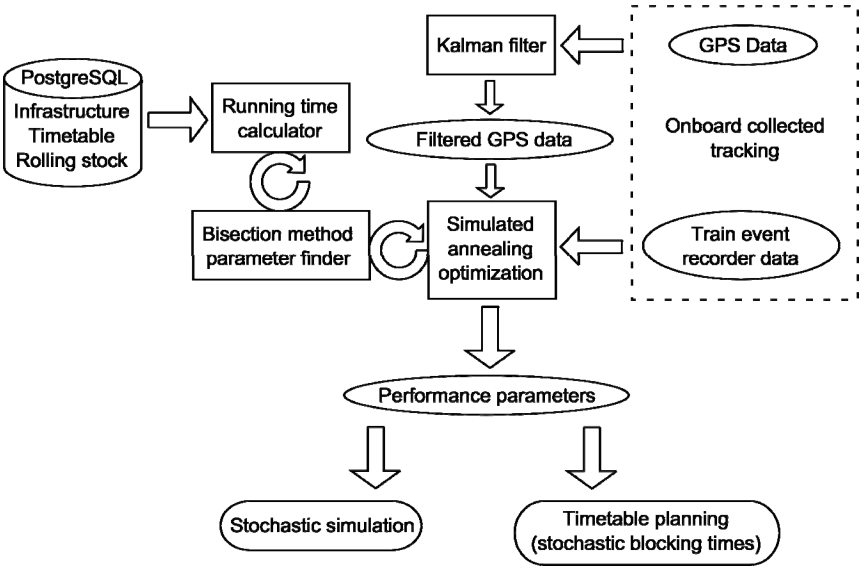


Figure 1: Block diagram of the working flow and software tool.

2.1 Microscopic infrastructure model

The infrastructure, rolling stock and timetables are stored in a PostgreSQL database. Compared to binary files, the database allows a simpler data model, higher exchangeability with other tools, and a smarter management of large amounts of data. Moreover, PostgreSQL represents a freeware, yet performing and reliable alternative to commercial DBMS. Infrastructure is represented according to the RailML standard [5], allowing a simple import and export from the most common simulation tools. The well-known and proven double-vertex graph [6] is used simplify the definition of itineraries.

A microscopic infrastructure model has been chosen rather than a mesoscopic one, which is sufficient for running time calculation, because not only line speed, but also signal aspects and the supervision curve of the protection systems are considered. To achieve the highest precision in the calibration process, the speed profile has to be calculated precisely for the entire train route: therefore the

itineraries within station and the position (and aspects) of each signal are required.

Also timetable data structure is entirely RailML compatible, while rolling stock is defined using all parameters required to solve the motion equation and consider the Italian ATP and ATC (SSC, SCMT, RSC, ETCS L2).

2.2 Data filtering and import

The calibration of motion equation can be performed on the basis of on-board collected GPS tracking or on the archives of digital train event recorders. Train event recorders store very accurate data based on odometers; endowed with a wider range of other data, such as the speed limit calculated by the protection system, the signal aspects the throttle and brake percentage use. These files do not require any filtering process, since tracking data are filtered in real time comparing two odometers and using balises or a discrete calibration of distances.

Significantly more complex to perform is the preparation of GPS data. Trackings are first post-processed using a proven algorithm based on the Kalman filter already in use to reconstruct the trajectories of race cars [8]. The filtered data show a sufficient precision, but do not contain any information on signal aspects, which is required in order to reconstruct the drivers' behaviour. To fill this gap, the recorded speed profile is compared to the planned one (which considers no route conflict), in which the train brakes only for speed reductions and at stops.

Once unplanned brakings are found, the planned profile is recalculated, considering a yellow aspect at the corresponding signal. On lines with discrete ATPs, the braking curve comprehending the approach speed and distance at next signal is computed, since the train is not allowed to re-accelerate before passing the corresponding main signal balise. On lines equipped with continuous ATP, where the train is allowed to accelerate as the man-machine interface shows the green aspect again, the instant when the train re-accelerates is obtained as local minimum of the real speed profile. The planned speed profile is then modified considering braking and then re-accelerating. The weakness of this method is represented by the impossibility to know exactly whether an unplanned braking is due to a "normal" traffic conflict, or to a failure, where the protection system activated the emergency braking.

GPS trackings, filtered and endowed with the corresponding signal aspects, and train event recordings are saved and used as basis for the calibration procedure.

2.3 Performance factors and motion phase

As stated in [3] a single performance factor does not allow a precise representation of the different motion phases. Separated parameters for acceleration, cruising and braking phases where suggested and used in the test case.

Knowing the signal aspects and the planned timetable it is possible to separate braking into three different phases: braking at a halt, for a line speed reduction

and for a restrictive signal aspect. This separation has been decided after a first analysis of real braking curves, which appeared significantly different depending on the motivation of braking. Braking at halts showed lower deceleration compared to braking for signals, and even lower values were recorded at speed limit changes. Moreover, especially when heavy trains brake at a speed reduction often run even lower speeds than allowed or show very variable deceleration to avoid this excessive braking.

As a result, compared to conventional running time calculation, where the braking distance is continuously calculated and stored, three values are computed.

2.4 Calibration algorithm

During calibration, the best fitting set of performance factors for a train ride is calculated. This computation is based on three assumptions:

1) The infrastructure model represents exactly the infrastructure used by the train, in particular concerning signal positions, relative distances and speed limits.
2) At the end of calibration, calculated running time correspond exactly to the measured one. In other words, the set of performance factor must lead to an exact calculation of the running time at the end of the journey.
3) The integration period in running time calculation must correspond to the tracking sampling period. This simplifies the calibration significantly, since to each recorded value corresponds to a calculated one, permitting a simple comparison in order to obtain an indicator of the goodness of the estimated parameters.

The method used to compare the two arrays is the simple mean squared error estimator. The software tool implements the algorithm (1)

$$MSE = \sum_{t=1}^{N} (v_{GPS}(t) - v_C(t))^2 \tag{1}$$

To compute a fixed length speed vector one dependent variable and four independent variables have to be considered. The software tool uses cruising performance as dependent variable for three reasons:

1) It has a small variation and relative high value (always higher than 90% in the test case)
2) Cruising running phase is the longest during a train run, therefore a minimum variation of its value has a great impact on total train running time
3) The value of cruising performance is inversely proportional to train running time.

Assuming 3), the value of the dependant variable value can be found using the bisection method.

2.5 Simulated annealing

In order to find the best performance parameters, an optimisation can be performed. It is necessary to find the minimum mean square error given four independent variables of a not linear function resolving the equation (2).

$$\begin{cases} \min\{\sum_{t=1}^{N}[v_{GPS}(t) - v_C(t)]^2\} \\ v_C = f(a,b,c,d,e) \\ size(v_{GPS}) = size(v_C) = N \end{cases} \tag{2}$$

In the equation a,b,c,d,e are the five performance parameters, v_c is the arrays of the calculated speeds and v_{GPS} is the array of measured speeds.

This appears as a nonlinear optimization problem: an adequate method to solve it has to be found; moreover, since it is difficult to determine a priori the properties of the target function a robust method must be used. A number of proven algorithms to perform this computation can be found in literature.

In the first tests, using only one parameter for breaking performance, the map represented in figure 2 has been obtained. It represents the value of target function (z-axis) with the variation of acceleration and braking factors. Obviously the value of cruising performance is not represented because it depends on the other two values. It's possible to notice the optimum value point for this train run.

The simulated annealing (SA) optimization method [9] is used in the software tool to find the best performance parameters that reduce the difference between computed and measured train speed. The algorithm begins finding a solution at a random point of the feasible region; after that, many steps are performed in order to find a better solution. It is possible to limit the maximum number of steps to be performed in order to limit computation time.

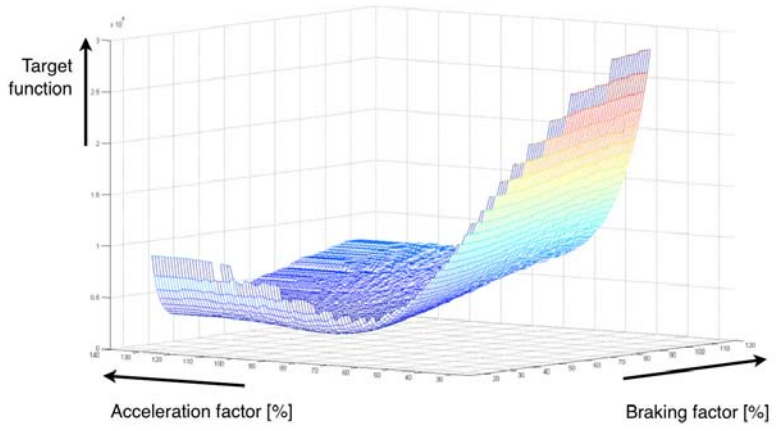


Figure 2: Target function (z-axis) as a function of acceleration and braking factors.

Each step of the SA algorithm replaces the current solution by a random "nearby" solution, chosen with a probability that depends on the difference between the corresponding function values and on a global parameter T (called the temperature), that is gradually decreased during the process. The dependency is such that the current solution changes almost randomly when T is large, but increasingly "downhill" as T goes to zero. The allowance for "uphill" moves saves the method from becoming stuck at local optima.

At the end of the optimization process both performance parameters are saved into a custom file format in order to perform aggregate data analysis. In the same files, the corresponding occupation steps are stored, to enable a precise representation of the real infrastructure usage.

2.6 Multithreading

Simulated annealing algorithm performs a number of experiments looking for value of target function in several points. Since each experiment is a train running time calculation with different performance parameters independent from each other, it's possible to run them concurrently using modern processors multithreading processing capacity. This permits to reduce total computation time, allowing us to use nearly 100% of the CPU power of modern multiprocessor workstations. The experiments to be computed are inserted in a queue, so each thread gets one experiment input data from the queue, computes the value of the target function, compares it with the best value ever found before, generates up to 8 new experiments and adds them in the queue. Synchronization between threads is needed only when accessing the queue to add and remove experiment input data and accessing best target function value. The synchronized code is less than 0.1% of total code lines of the optimization phase.

3 Test and applications

The tool has been tested in North-Eastern Italy, on the Trieste – Venice line. The double track, electrified line is about 130 km long and plays an important role in regional transport and as freight corridor between Eastern Europe and Italy. The line is endowed with SCMT, digital, discrete ATP corresponding to ETCS Level 1. About 100 train runs with the same rolling stock have been computed, obtaining the corresponding parameters sets. Each set represents a driving style, since it contains the way a driver handles in each motion phase.

Comparing the calibrated equation with 5 parameters with the conventional running time calculation, including 7% running time supplement, the difference between the two approaches has been pointed out, showing differences of more than 20% in the expected occupation steps.

In Figure 2, the speed limit (black), the planned speed profile (red) the real (green) and the calibrated (purple) ones are compared, clearly showing the differences among them. In particular, the goodness of the fitted equation compared to the planned one is remarkable.

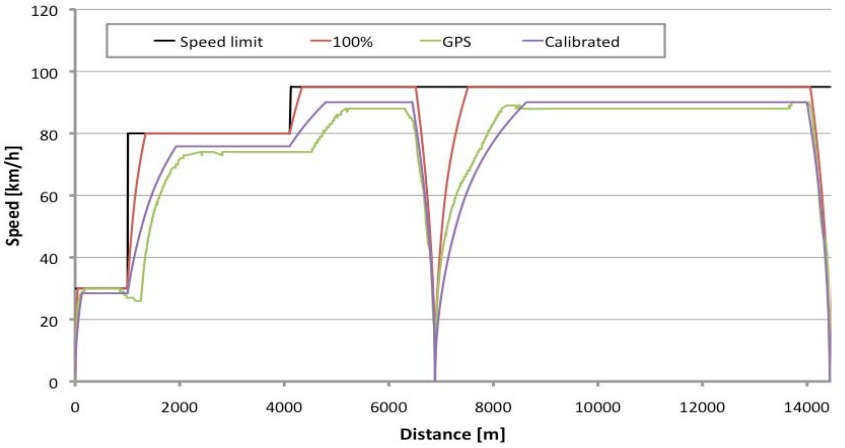


Figure 3: Speed limit, simulated speed profile with 100% performance, calibrated and real speed profiles.

3.1 Future applications

The obtained driving styles (calibration sets) over a statistically significant number of records can be used to compute the stochastic behaviour of a single train or a calibrated simulation of the interactions among trains. The second option would lead to a precise micro-simulation that could be either asynchronous or synchronous.

In the first option a representation of the blocking times over a number of realistic operational days are obtained, as tested on the Trieste-Venice line. The resulting blocking times can be stored and depicted in a timetable graph using transparencies to represent each run. As a result, a wider occupation staircase is obtained, which colour intensity is proportional to the probability to have the infrastructure used. While planning a timetable, the possibility to view the realistic infrastructure usage supports the definition of buffer times, which can be chosen on the basis of the real variability instead of empiric rules.

4 Conclusions and outlook

In this work, an approach for thee accurate calibration of motion equation, which introduces a quantitative representation of real driving styles has been presented. The method has been implemented on a software tool, using standard representation of input data and taking advantage of powerful DBMS and multithreading to perform onerous computations in a competitive time.

The case study has shown a significant improvement in the stochastic micro simulation output, demonstrating the importance of a precise parameter calibration especially regarding train braking and acceleration. A further

improvement will come from the calibration of the motion equation parameters to fit the DIS data.

The approach appears very useful for the calibration of motion equation within micro-simulation tools, while a promising application is represented by the implementation of stochastic blocking time staircases instead of deterministic ones in timetable planning software.

References

[1] Wende, D. "Fahrdynamik des Schienenverkehrs". Wiesbaden, Teubner Verlag, 2003.
[2] Hansen, I.A., Pachl, J., "Railway Timetable & Traffic" Hamburg, Eurailpress, 2008.
[3] Medeossi, G. "Capacity and Reliability on railway networks, a simulative approach" University of Trieste, 2010-07-23.
[4] de Fabris, S., G. Longo, et al. Automated analysis of train event recorder data to improve micro-simulation models. In: J. Allan, E. Arias, C. A. Brebbia et al., Computers in Railways XI, WIT Press, Southampton, 575-585, 2008.
[5] Nash, A., Huerlimann, D., Schuette, J., Krauss, V.P., "RailML – A standard data interface for railroad applications", In: Allan, J., Hill, R.J., Brebbia, C.A., Sciutto, G., Sone, S. (eds.), Computers in Railways IX, WIT Press, Southampton, 45-54, 2004.
[6] M. Montigel. Representation of Track Topologies with Double Vertex Graphs. In T.K.S. Murthy, F.E. Young, S. Lehman, W.R. Smith, editor, Computers in Railway, volume 2, Washington D.C., 1992. Computational Mechanics Publications.
[7] Huerlimann, D., Nash, A., "Railway simulation using Opentrack", In: Allan, J., Hill, R.J., Brebbia, C.A., Sciutto, G., Sone, S. (eds.), Computers in Railways IX, WIT Press, Southampton, 45-54, 2004.
[8] Robert Grower Brown, Patrick Y.C. Hwang, "Introduction to Random Signals and Applied Kalman Filtering" Second Edition, John Wiley & Sons, NavtechGPS, 1997.
[9] Hillier, F. and Lieberman, G., Introduction to Operations Research (8th ed.), McGraw-Hill Science/Engineering/Math, 2005.

This page intentionally left blank

Section 8
Multiscale modelling

This page intentionally left blank

Simulation of intergranular crack nucleation and evolution in polycrystal metals on mesoscale

G. Shi & G. Renaud
Structures and Materials Performance Laboratory,
Institute for Aerospace Research, National Research Council, Canada

Abstract

This paper presents research on the development of a simulation method for the study of the mechanisms of intergranular crack nucleation and propagation in polycrystal metals on the mesoscale. Microstructural geometry models were built randomly using Voronoi techniques. Based on these grain structure geometry models, two-dimensional grain structure finite element models were created using a Patran Command Language (PCL) program. Techniques for the implementation of the cohesive elements between grain boundaries were developed in PCL for the generation of two-dimensional cohesive models. Simulations on intergranular crack nucleation and evolution along grain boundaries using two-dimensional finite element cohesive models were carried out on the mesoscale level. Several aspects that affect the crack nucleation and propagation were studied, which included random grain geometries, grain boundary misorientations, grain boundary peak strength, grain boundary fracture energy, grain properties, and grain plasticity. The simulations demonstrated that the cohesive model is a useful and efficient modeling tool for the study of the intergranular crack nucleation and evolution on the mesoscale level. The simulation results showed that the factors studied have large impacts on intergranular crack nucleation and evolution based on the current model capabilities and conditions.
Keywords: intergranular crack nucleation and evolution, cohesive zone model, polycrystal metals, mesoscale.


WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110521

1 Introduction

A major challenge in life cycle prediction and management of aircraft structural components is the lack of information on the crack nucleation and short crack propagation stages. However, for many airframe materials, the life of an aircraft component could be almost completely exhausted within these two phases. Therefore, studies on smaller-scale levels, such as the microscopic scale or the mesoscale, are strongly needed to better understand the physical nature of the crack nucleation and propagation.

A fundamental research project was launched at Institute for Aerospace Research in National Research Council Canada aimed at developing a coupled atomic-meso-macroscopic modeling strategy for the simulation of crack nucleation and propagation in aircraft components. To this end, the development of modelling capabilities at each length scale was essential. Research work on the mesoscale level is presented in this paper. The simulation methods, developed capabilities, and the studied mechanisms of intergranular crack along grain boundaries (GB) are discussed.

2 Grain geometry model and finite element model

To study the fracture mechanism on the mesoscale, grain geometry models need to be properly built in order to further simulate the crack nucleation and propagation at this scale level. In this study, the microstructural geometry models were statistically created by using Voronoi techniques [1]. A general-purpose two-dimensional microstructure model simulation software, MicroSimu, was used to accomplish the geometry generation with a clean database ready for further finite element mesh generation.

Based on the grain structure geometry model, a Patran Command Language (PCL) program was developed to automatically generate two-dimensional grain structure finite element (FE) models. Using the developed PCL program, the material properties and orientation angle in each grain were statistically assigned to represent the heterogeneity of the grain structure.

3 Cohesive zone model (CZM)

Cohesive zone modeling has gained considerable attention over the past decade, as it represents a powerful and efficient technique for fracture studies [2–11]. Initially, cohesive zone models were developed to study the perfectly brittle materials [2], and later, models were extended to describe the damage that occurs in the plastic zone ahead of a crack [3]. In recent years, cohesive models were widely used in different areas for different applications, such as simulations of delamination formation and propagation in composite materials and structures [4], debonding interface separation simulations in multimedia systems [5], simulations for bonded structures [6], dynamic fracture of homogeneous materials [7], and simulations of grain boundary separation between grains on the mesoscale [8–10].

The major advantage of the cohesive zone models is that they can predict the formation of damage without the need to pre-define any initial damage in the model. Moreover, cohesive zone formulations can be easily implemented in finite element codes using cohesive elements [4].

3.1 Traction-separation based CZMs

There are two types of cohesive zone models, the continuum-based constitutive model and the traction-separation based constitutive model. If the interface thickness is negligibly small (or zero), the constitutive response of the cohesive layer can be defined directly in terms of traction versus separation. If the cohesive layer has finite thickness and if macroscopic properties of the adhesive material are available, the response can be modeled using conventional material models. For the simulation of grain boundary characteristics, the traction-separation based cohesive models can be used to define the grain boundary behavior.

One of the existing traction-separation models assumes an initial linear elastic behavior, followed by the formation and evolution of damage, as shown in Figure 1. The nominal traction stress vector, $\vec{S}$, consists of two components in two-dimensional problems, S_n and S_t, which represent the normal and shear traction, respectively. The corresponding separation vector is denoted by $\vec{\delta}(\delta_n, \delta_t)$.

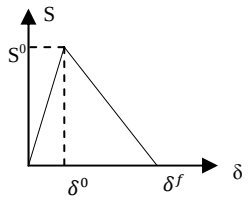


Figure 1: Linear damage evolution.

In Figure 1, S^0 denotes the maximum traction (cohesive strength), δ^0 is the separation corresponding to the maximum traction, at which the crack nucleates, and δ^f is the separation at which the final failure occurs.

3.2 Damage nucleation criteria [12]

3.2.1 Maximum nominal stress criterion

Damage is assumed to nucleate when the maximum nominal traction stress ratio reaches the value of one. This can be written as below for a two-dimensional problem.

$$\max\left\{\frac{<S_n>}{S_n^0}, \frac{S_t}{S_t^0}\right\} = 1.0 \qquad (1)$$

where S_n^0 and S_t^0 represent the peak values of the nominal stress when the deformation is either purely normal to the interface or purely in the shear direction. The symbol < > used in equation (1) represents the Macaulay bracket, which means:

$$< S_n > = \begin{cases} S_n, & S_n > 0 \\ 0, & S_n \leq 0 \end{cases} \tag{2}$$

The use of the Macaulay bracket is to signify that a pure compressive deformation or stress state does not start damage.

3.2.2 Quadratic nominal stress criterion

Damage is assumed to nucleate when a quadratic interaction function involving the nominal stress ratios reaches the value of one.

$$\left\{\frac{<S_n>}{S_n^0}\right\}^2 + \left\{\frac{S_t}{S_t^0}\right\}^2 = 1.0 \tag{3}$$

3.3 Damage evolution laws [12]

3.3.1 Evolution based on effective displacements

The damage evolution law describes the rate at which the material stiffness is degraded once the damage nucleation criterion is reached. If based on displacements, the damage evolution law can be written as

$$S_n = \begin{cases} (1-D)\overline{S_n} & \overline{S_n} \geq 0 \\ \overline{S_n} & \text{otherwise} \end{cases} \tag{4}$$

$$S_t = (1-D)\overline{S_t} \tag{5}$$

where $\overline{S_n}$ and $\overline{S_t}$ are the stress components predicted by the elastic traction-separation behavior for the current strain damage. D is a damage variable representing the overall damage in the material.

For the linear softening, the evolution of the damage variable D reduces based on the effective displacement: $\delta = \sqrt{< \delta_n >^2 + \delta_t^2}$,

$$D = \frac{\delta^f(\delta^{max} - \delta^0)}{\delta^{max}(\delta^f - \delta^0)} \tag{6}$$

where δ^{max} refers to the maximum value of the effective displacement attained during the loading history, δ^0 and δ^f represent the displacements at damage nucleation and final failure, respectively.

3.3.2 Evolution based on fracture energy

The damage evolution can also be defined based on the fracture energy that is dissipated as a result of the damage process.

The energy-based damage evolution law can be expressed by a power law fracture criterion, which states that failure under mixed-mode conditions is governed by a power law of the energies required to cause failure in the individual (normal or shear) mode. For the two-dimensional problem, it is given by

$$\left\{\frac{G_n}{G_n^c}\right\}^{\alpha} + \left\{\frac{G_t}{G_t^c}\right\}^{\alpha} = 1.0 \tag{7}$$

where G_n and G_t refer to the work done by the tractions and its conjugate relative displacement in normal and shear directions, respectively. G_n^c and G_t^c are the critical fracture energies required to cause failure in the normal and shear directions, respectively.

4 An example of CZM applications

The application of cohesive zone models was implemented for the simulation of crack nucleation and propagation under monotonic loading on the mesoscale. Several samples of grain geometry were statistically created using the Voronoi tessellation method. The sample size was 500x500μm filled by 100 grains. The grain geometry models were imported into the MSC/Patran preprocessor and the finite element models were created with fine meshes in each grain. The grain orientation and material properties for each grain were randomly assigned to capture the heterogeneity using the developed PCL program. Cohesive elements between grain boundaries were automatically inserted for the generation of two-dimensional cohesive models using the developed PCL program, which included the automatic insertion of cohesive elements between grain boundaries, calculation of grain boundary misorientations based on random angles in each grain, properties of the cohesive model, and criteria for crack nucleation, and evolution. The cohesive zone model used in the study was composed of a bi-linear traction-separation relationship, a stress threshold for crack nucleation and an energy based damage evolution law. The general-purpose commercial finite element software ABAQUS was employed to conduct the simulations in this study [12]. The grain properties and grain boundary (GB) properties used in this study are listed in Table 1 [8].

Table 1: Grain and grain boundary (GB) properties.

Grain properties (orthotropic, elastic-plastic)		GB properties
E_{11} = 72,000 MPa	σ_{y1} = 505 MPa	G_n^c = 0.25 N/mm
E_{22} = 42,000 MPa	σ_{y2} = 450 MPa	S_n^0= 500 MPa
E_{12} = 26,900 MPa	σ_{y12} = 400 MPa	k_n = 2.5 e7 MPa
μ = 0.33		

In the orthotropic models, the GB parameters, such as G and S, were varied with the misorientation angles across the boundaries [8], [9].

$$G(\theta) = G_{ave} + \Delta G * Cos(4\theta) \tag{8}$$

$$S(\theta) = S_{ave} + \Delta S * Cos(4\theta) \tag{9}$$

where $\theta = \beta_1 - \beta_2 \qquad 0^0 \le \beta \le 180^0$

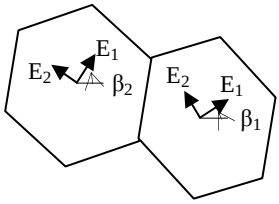


Figure 2: Demonstration of grain orientation.

The boundary conditions and load applied onto the samples are illustrated in Figure 3. Three sides were pinned and one side was loaded by a displacement, 0.005mm (5 μm), which was equivalent to 1% strain.

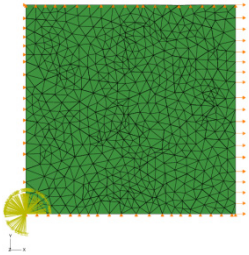


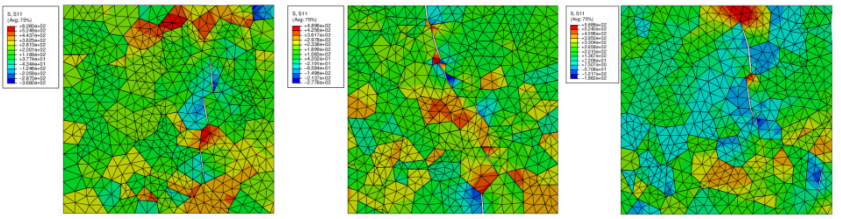
Figure 3: Boundary and load conditions on the sample.

5 Parametric studies

Using the mesoscale models, simulations were carried out to study the effects of several sets of parameters on the crack nucleation and evolution in the polycrystalline samples. The studied parameters and results are described in the next sections.

5.1 Variations of random grain geometry

In this section, the effects of random grain geometry on the crack nucleation and evolution are examined. The grain geometry was randomly created using Voronoi tessellation method. By fixing the overall size of the sample and the grain number (100), samples with different grain geometries (grain size, grain shape, aspect ratio) were generated. Figure 4 demonstrates the crack paths and the stress distributions in the three samples with different grain geometries. From this figure, it can be seen that the crack patterns and stress distributions are quite different. Figure 5 shows the damage dissipation energy versus applied strain for the three samples shown in Figure 4. The figure indicates that the crack evolution patterns, i.e. the energies dissipated by the damage, are quite different.



(i) (ii) (iii)

Figure 4: Crack path and stress with three different grain geometries.

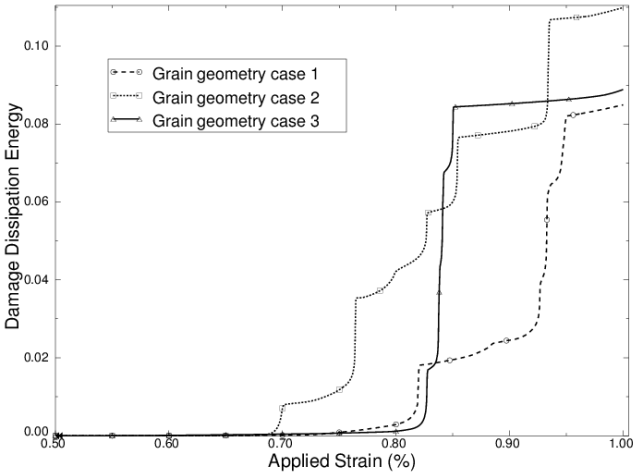
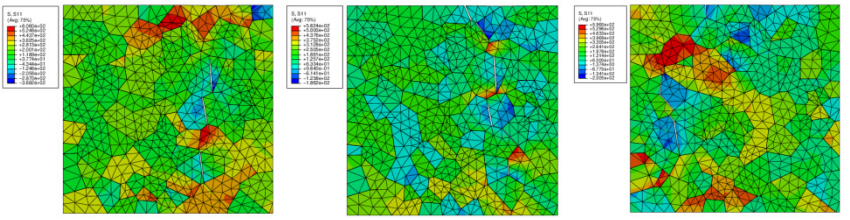


Figure 5: Variation of damage dissipation energy with different grain geometries.

5.2 Effects of random GB misorientations

Since the grain orientation angles were randomly assigned to each grain, the grain boundary misorientations between the conjunct grains were obtained randomly. By keeping the grain geometry the same, three cases with different grain orientation angles and GB misorientations were created to examine their effects on the crack nucleation and evolution behavior. Figure 6 shows the stress distributions and crack paths for the three cases. It can be seen that the crack locations and their evolution paths are totally different when the grain boundary misorientations are differently distributed. Figure 7 shows the variation of damage dissipation energy with the randomly distributed grain boundary misorientations. This figure demonstrates that the grain boundary misorientation has significant effects on the crack nucleation and evolution. In the figure, it is seen from the dissipated damage energy that crack nucleates earlier in case 2 than in case 1 and 3. Relate to Fig. 6(i), (ii), (iii).



(i) (ii) (iii)

Figure 6: Stress distribution and crack path with different GB misorientation (same geometry).

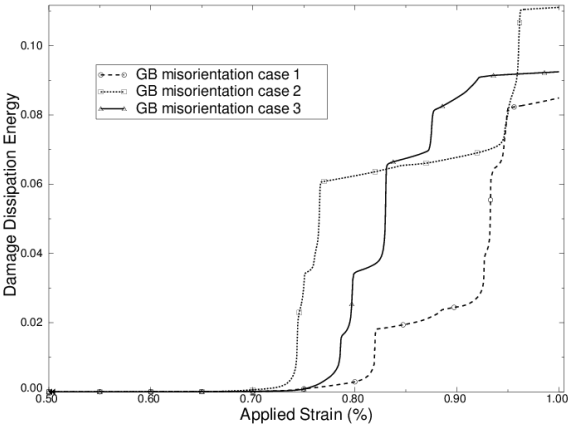


Figure 7: Variation of damage dissipation energy with different GB misorientations.

5.3 Effects of grain boundary cohesive parameters

In this section, the grain boundary properties, GB strength, and GB fracture energy, were studied for the investigation of their effects on the intergranular crack nucleation and evolution.

5.3.1 GB strength (peak tractions)

By fixing the GB fracture energy, the GB peak tractions were varied by ±20%, ±10% and 0%. The mean value of the GB strength was assumed to be 500MPa. Simulations were carried out until the crack nucleation and evolution process were observed in the damage energy dissipated during the load increments. Figure 8 is a plot of the variation of damage dissipation energy with GB peak tractions. It can be seen that the GB strength had significant effects on the crack nucleation and evolution process. The lower the GB strength, the earlier the crack nucleates and the larger the damage dissipation energy can be generated under the same applied load.

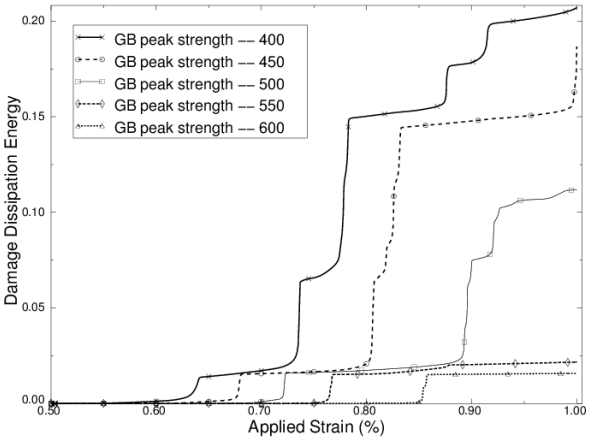


Figure 8: Variation of damage dissipation energy with GB strength.

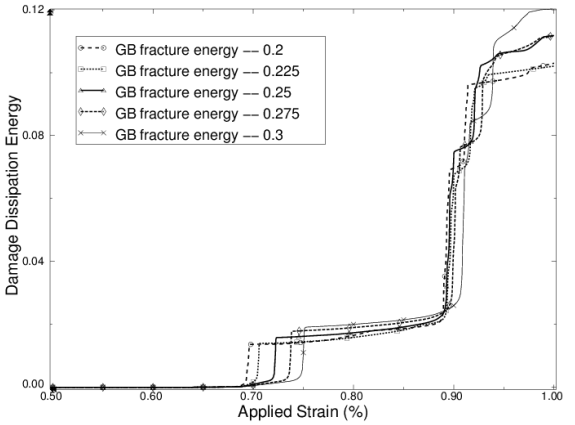


Figure 9: Variation of damage dissipation energy with GB fracture energy.

5.3.2 GB fracture energy

Fixing GB strength at its mean value, the GB fracture energy was varied by ±20%, ±10% and 0% of its mean value, which was assumed to be 0.25 N/mm. Figure 9 shows the variation of damage dissipation energy with GB fracture energy. From this figure, it is noted that the cracks nucleate at the same applied load level for all GB fracture energy cases because of the constant GB strength. However, the damage evolution processes under different GB fracture energies are quite different. The lower the GB fracture energy the earlier and faster the crack propagates after the crack nucleation occurred. Higher GB fracture energy leads to larger damage dissipation energy.

5.4 Effects of grain material properties

The effects of grain material properties on the intergranular crack nucleation and evolution along the grain boundary was evaluated by keeping the grain boundary properties, such as the GB stiffness, the GB strength, the GB fracture energy and the GB misorientation, constant, while the grain properties were varied to isotropic elastic, isotropic elastic-plastic, orthotropic elastic, orthotropic elastic-plastic. Figure 10 shows the variation of damage dissipation energy with the grain material properties. From this figure, it can be seen that for the orthotropic case, cracks along the grain boundary nucleated at the same level for both elastic and elastic-plastic materials, but much more damage dissipation energy was generated in elastic material than in plastic material. This is because part of the external work was transformed into plastic dissipation energy. The similarity can be seen for the isotropic cases in the figure. It can also be seen that, by comparing the isotropic and orthotropic cases, cracks nucleated much earlier and the damage dissipation energy was much larger in the isotropic cases than in the orthotropic cases.

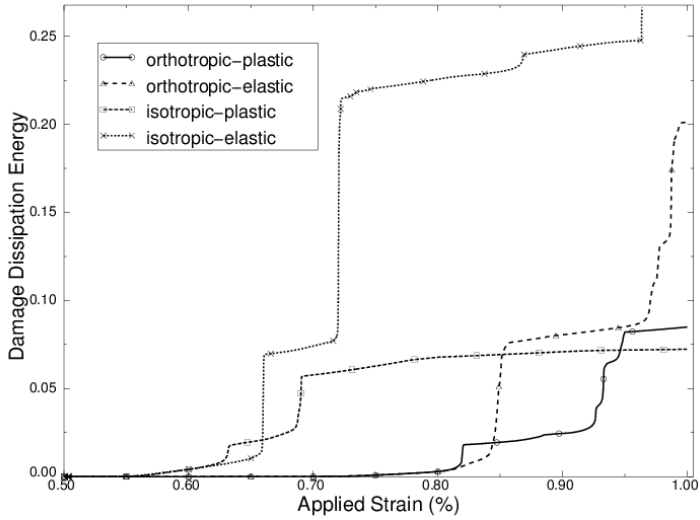


Figure 10: Variation of damage dissipation energy with grain properties.

5.5 Effects of the competition between grain plasticity and GB strength

Although the current study focuses on the intergranular crack along grain boundaries, the grain plasticity is still a factor that affects the crack characteristics on grain boundary. This can be shown by keeping the GB peak strength constant at 500MPa, while varying the reference yield stresses in grains with orthotropic elastic-plastic properties. Figure 11 shows the variation of damage dissipation energy with grain plasticity. From the figure, it can be seen that when the reference yield stress became higher, the damage dissipation

energies were larger. This may be because less external work was converted plasticity dissipation energy, and more was used to generate the crack surface along grain boundaries. When the reference yield stress became much lower than the GB peak strength, the damage energy dissipated along grain boundaries was much lower, and most of the external work was converted into plasticity deformation in the grains.

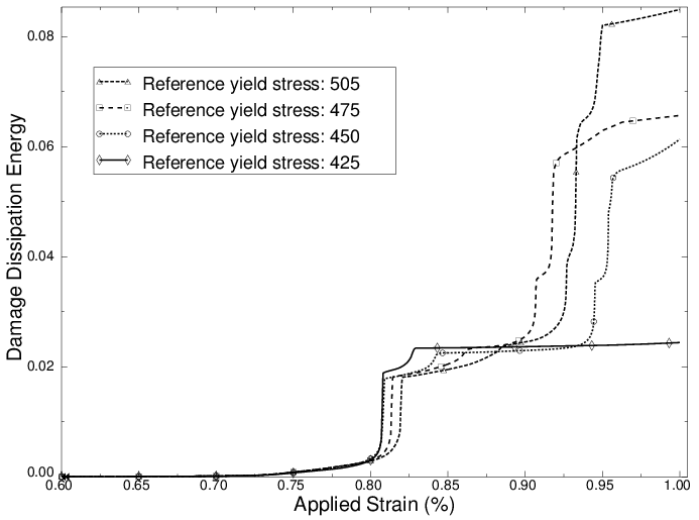


Figure 11: Variation of damage dissipation energy with grain plasticity (the units for reference yield stresses are MPa).

6 Summary

In the present work, capabilities for the construction of grain structure finite element cohesive zone models were developed. Modeling and simulation of intergranular crack nucleation and propagation in polycrystal metals on the mesoscale were conducted. The intergranular fracture characteristics were investigated through parametric studies. The simulation results showed that grain boundary cohesive properties, such as the grain boundary peak strength and the grain boundary fracture energy, directly affected the intergranular crack nucleation and evolution. The lower the grain boundary strength, the earlier the cracks nucleated. The lower the grain boundary fracture energy, the earlier and faster the cracks propagated after nucleation occurred. Different grain geometries and grain boundary misorientations resulted in different crack nucleation and evolution patterns. Moreover, the grain material properties, and the competition between grain plasticity and GB strength also have influences on the crack nucleation and evolution along grain boundaries.

Acknowledgements

This research was supported by the Technology Investment Funding (TIF) from the Department of National Defence Canada.

References

[1] Zhao, Y. And Tryon, R., "Automatic 3-D simulation and micro-stress distribution of polycrystalline metallic materials," Computer Methods in Applied Mechanics and Engineering, Vol. 193, pp. 3919-3934, 2004.
[2] Barenblatt, G.I., "The formation of equilibrium cracks during brittle fracture. General ideas and hypothesis, Axially-symmetric cracks", Journal of Applied Math. & Mech. (PMM), Vol. 23, No. 3, pp. 622-636, 1959.
[3] Dugolale, D., "Yielding of steel sheets containing slits." J. of Mech. Phys. Solids, Vol. 8, pp. 100-104, 1960.
[4] Fan, C., Jar, P.-Y. B and Cheng, J.J.R., "Cohesive zone with continuum damage properties for simulation of delamination development in fibre composites and failure of adhesive joints", Engineering Fracture Mechanics, Vol. 75, pp. 3866-3880, 2008.
[5] Chandra, N., Li, H., Shet, C., Ghonem, H., "Some issues in the application of cohesive zone models for metal-ceramic interfaces", International Journal of Solids and Structures, Vol. 39, pp.2827-2855, 2002.
[6] Moura, M.F.S.F. D., Goncalves, J.P.M., Chousal, J.A.G. and Campilho, R.D.S.G., "Cohesive and continuum mixed-mode damage models applied to the simulation of the mechanical behavior of bonded joints," International Journal of Adhesion & Adhesive, Vol. 28, pp. 419-426, 2008.
[7] Ruiz, G., Pandolfi, A. and Ortiz, M., "Three-dimensional cohesive modeling of dynamic mixed-mode fracture", International Journal for Numerical Methods in Engineering, Vol.52, pp.97-120, 2001.
[8] Iesulauro, E., Ingraffea, A. R., Arwade, S., and Wawrzynek, P. A., "Simulation of Grain Boundary Decohesion and Crack Initiation in Aluminum Microstructure Models," Fatigue and Fracture Mechanics: 33rd Volume, ASTM STP 1417, W. G. Reuter and R. S. Piascik, Eds., American Society for Testing and Materials, West Conshohocken, PA, 2002.
[9] Luther, Torsten and Konke, Carsten, "Polycrystal models for the analysis of intergranular crack growth in metallic materials", Journal of Engineering Fracture Mechanics, Vol. 76, pp. 2332-2343, 2009.
[10] Glaessgen, E., Seather, E., Phillips, D., and Yamakov, V., "Multiscale modelling for grain-boundary fracture: cohesive zone models parameterized from atomistic simulations", Proceedings of the 47th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics & Materials Conference, AIAA 2006-1674, Newport, Rhode Island, 2006.
[11] Li, H. and Chandra, N., "Analysis of crack growth and crack-tip plasticity in ductile materials using cohesive zone models", International Journal of Plasticity, Vol. 19, pp. 849-882, 2003.
[12] ABAQUS, ABAQUS reference user manual, version 6.8.

A multiscale finite element simulation of piezoelectric materials using realistic crystal morphology

Y. Uetsuji[1], H. Kuramae[2], K. Tsuchiya[3] & M. Kamlah[4]
[1]*Department of Mechanical Engineering, Osaka Institute of Technology, Japan*
[2]*Department of Technology Management, Osaka Institute of Technology, Japan*
[3]*Department of Precision Engineering, Tokai University, Japan*
[4]*Institute for Materials Research II, Karlsruhe Institute of Technology, Germany*

Abstract

This paper presents the full components of macroscopic homogenized material properties and the microscopic localized response obtained through a multiscale finite element simulation using realistic crystal morphology. Crystal morphology analysis was performed to reveal microstructure and texture of a polycrystalline piezoelectric material. The insulative specimen of piezoelectric material was coated with a conductive layer of amorphous osmium to remove an electric charge, and crystal orientations were measured by means of electron backscatter diffraction. Then the obtained crystal orientations were applied to a multiscale finite element simulation based on homogenization theory.
Keywords: piezoelectric material, EBSD, crystal morphology, multiscale finite element simulation, homogenization theory.

1 Introduction

Piezoelectric materials have been used in actuators or sensors as a component of various electronic and mechanical devices. Generally these materials consist of many crystal grains and domains at a microscopic scale. Since each domain shows strongly anisotropic mechanical and electrical behaviours according to

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110531

asymmetrical crystal structure such as tetragonal and rhombohedron ones, the macroscopic properties of polycrystalline piezoelectric materials have a large dependence of the microscopic crystal morphology. Therefore, it is important to understand the microstructural features. Scanning electron microscope (SEM) and electron backscatter diffraction (EBSD) technique [1, 2] is well-established and useful to characterize the crystal morphology in polycrystalline materials. Electron beam is irradiated to a specimen in SEM, and then crystal orientations of sub-micron order grains and domains can be analyzed by EBSD. SEM·EBSD technique is hitherto applied to various materials, which range from metals [3, 4] to functional materials [5]. Especially if we focus on piezoelectric materials, a lot of published literatures have reported their microstructures and textures for lead titanate $PbTiO_3$ [6], barium titanate $BaTiO_3$ [7] and PZT $Pb(Zr, Ti)O_3$ [8]. All of them indicate that the microstructures have a significant role in the overall material properties. However, there are few researches to analyze the quantitative relationship between the microstructure and the macrostructure.

On the other hand, we employ a multiscale finite element simulation based on homogenization theory [9–11] and attempt to estimate the relationship of SEM·EBSD-measured realistic microstructure with material properties and behaviours of piezoelectric materials. In case of our previous study on $BaTiO_3$ [12], crystal orientation maps with high resolution couldn't be obtained because of the electrification of specimen. Additionally, in our previous investigation on PZT [13], the computations were carried out by using uncertain material properties to microstructural grains because single crystals are nonexistent and their material properties were unknown. In this paper, we retry to analyze the microstructure of $BaTiO_3$ by SEM·EBSD measurement. With the aim of a high resolution crystal orientation map, a conductive layer of amorphous osmium is utilized to remove an electric charge on the specimens. Then we employ a multiscale finite element simulation and investigate on the macroscopic homogenized material properties developed by SEM·EBSD-measured realistic microstructure and the microscopic localized behaviours in response to macroscopic external load.

2 Crystal morphology analysis

A $BaTiO_3$ polycrystalline piezoelectric ceramics (Murata Manufacturing Co. Ltd.) was used for the SEM·EBSD specimen. As shown in figure 1, the specimen was a circular disk of 15 mm diameter and 1 mm thickness, and it was electrically poled along the thickness direction. The crystal structure was examined by means of X-ray diffraction (XRD). Figure 2 shows XRD spectrum of $BaTiO_3$ piezoelectric material. It has a perovskite tetragonal structure and the lattice constants of *a* and *c* axes are 0.3995 nm and 0.4034 nm, respectively.

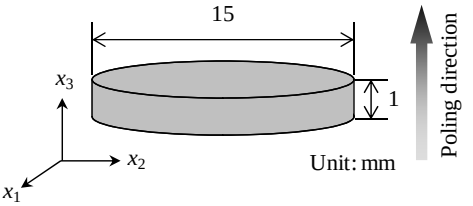


Figure 1: Geometry of SEM·EBSD specimen.

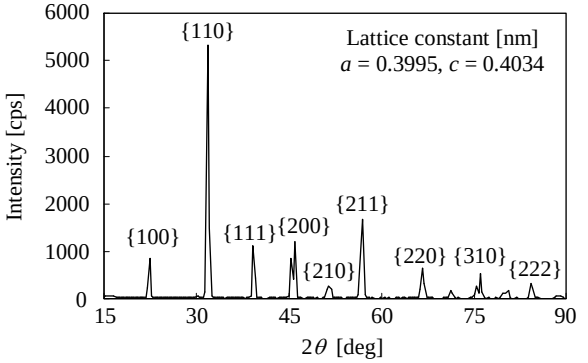


Figure 2: XRD spectrum of $BaTiO_3$ piezoelectric ceramics.

The observed surface was mechanically polished for 20 minutes using 3 μm diamond particles (Marumoto Struers Co., DP-Spray) with a polishing sheet (Marumoto Struers Co., DP-Mol). And then, it was chemically polished at pH 3.5 for 20 minutes using colloidal particles (Marumoto Struers Co., OP-A) with a polishing sheet (DP-Chem). The surface of the insulative specimen was coated with the electrical conductive and amorphous osmium layer to defend the electrification due to electron beam. We employed an osmium coater (Meiwafosis Co., LTD. Neoc-ST), and set the discharging time in osmium gas to 10 seconds. The estimated thickness of osmium layer is 0.5 nm/sec. In addition, a silver paste (Fujikura Kasei Co. Ltd., type D-500) was applied to the surface except for the SEM-EBSD measurement region to leak the accumulated charge.

Figure 3 shows crystal morphology images obtained by means of EBSD (Oxford Instruments plc, Link ISIS C.7272) implemented in SEM (JEOL Datum Ltd., JSM-5410). The scanning interval was set to 0.64 μm that is smaller than one over ten of the average grain size. As shown in figure 4, crystal orientations in four neighboring regions of 81.3×63.5 μm^2, which was the obtainable maximum region at once, were measured successively. Their four-divided crystal orientation maps were spliced to keep their exact spatial arrangement. As a

result, a crystal orientation map of 127.6×90.8 μm^2 was obtained as shown in figure 4. The mean diameter of grain was estimated to approximately 6.71 μm.

Figures 3(a) and 4 are the collar map of crystal orientations, and figure 3(b) is the grayscale map of pattern q1uality (PQ) value defined by eqn. (1).

$$PQ = \sum_{i=1}^{3} h_i / 3\sigma_h \tag{1}$$

where h_i means the peak height of the Hough transformed i-th Kikuchi's band. σ_h is the standard deviation of Hough transform. The PQ value represents quantitatively the fitting condition of Kikuchi patterns with the target crystal

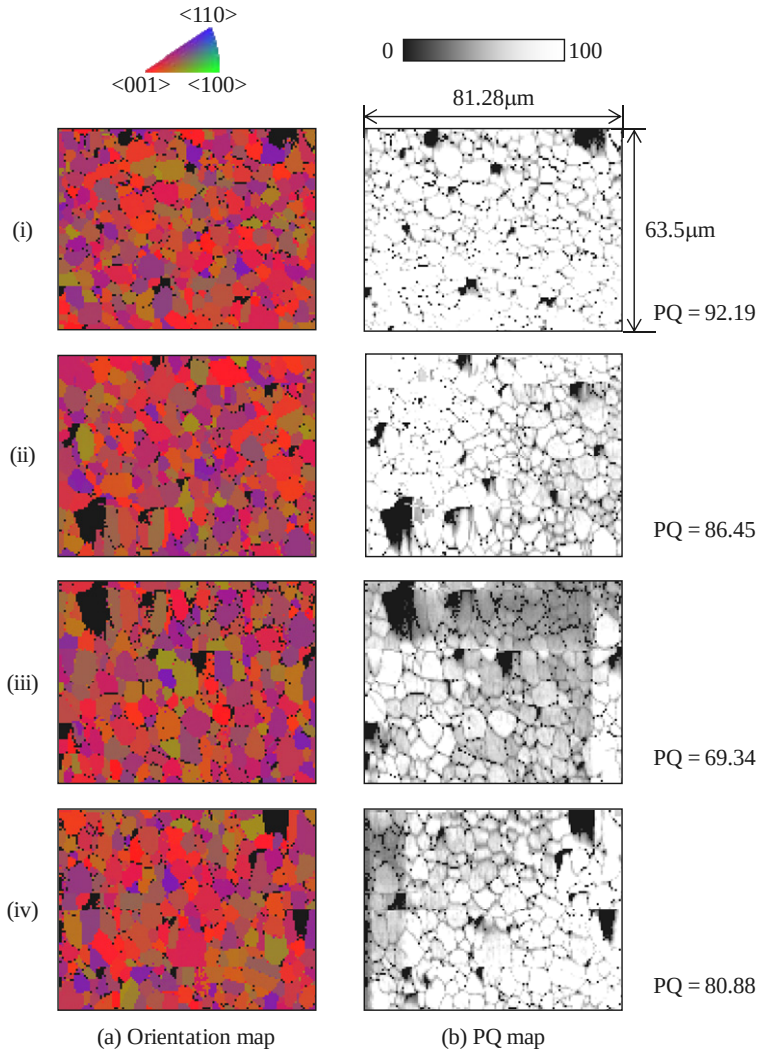


Figure 3: Crystal orientation and PQ maps of $BaTiO_3$ piezoelectric ceramics.

structure. On grain boundaries, where Kikuchi pattern may be composed of two or more orientations from adjacent grains, the PQ parameter yielded low value. Additionally, the reduction of PQ value appeared at the overlapping regions of scanning electron beam as demonstrated in (iii) of figure 3.

The crystal orienting degree in eqn. (2) was employed for an indicator to catch the orientation of domains toward an optional direction.

$$S^{c}_{y_1'} = \frac{\sum_{i=1}^{n} \left| \boldsymbol{e}_c^{\,i} \cdot \boldsymbol{e}_{y_1'} \right|}{n} \tag{2}$$

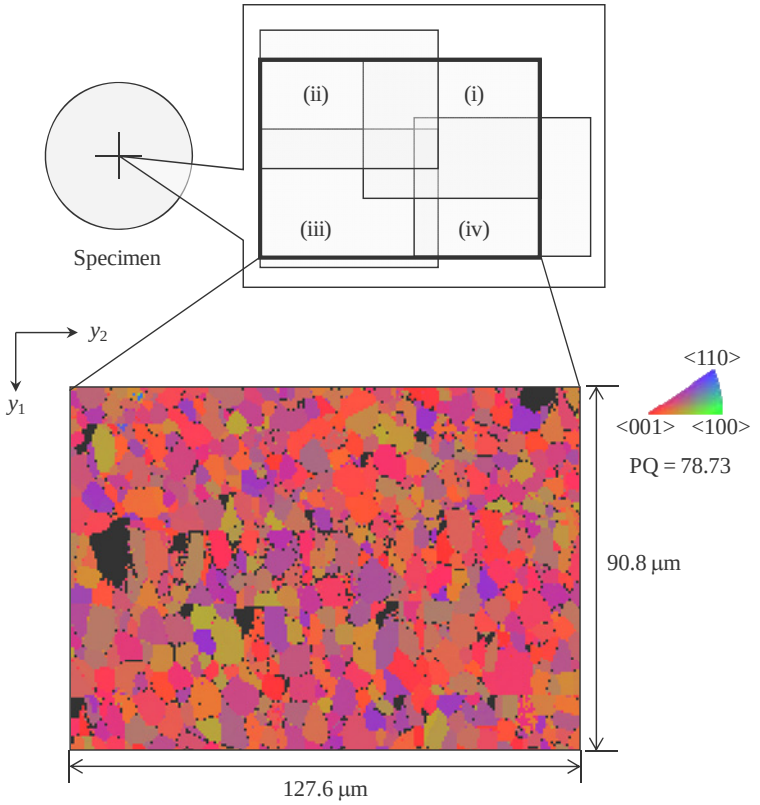


Figure 4: Spliced crystal orientation map of $BaTiO_3$ piezoelectric ceramics.

where $\boldsymbol{e}_c^{\,i}$ means the basis vector of c axis of the crystallographic coordinate system in microstructural coordinate system at the i–th measuring point. $\boldsymbol{e}_{y_1'}$ is the basis vector of an optional direction y_1' existing in the y_1– y_2 plane. And n is the total number of measuring points. The crystal orienting degree becomes zero when all points orient to the normal direction of the y_1– y_2 plane and it becomes

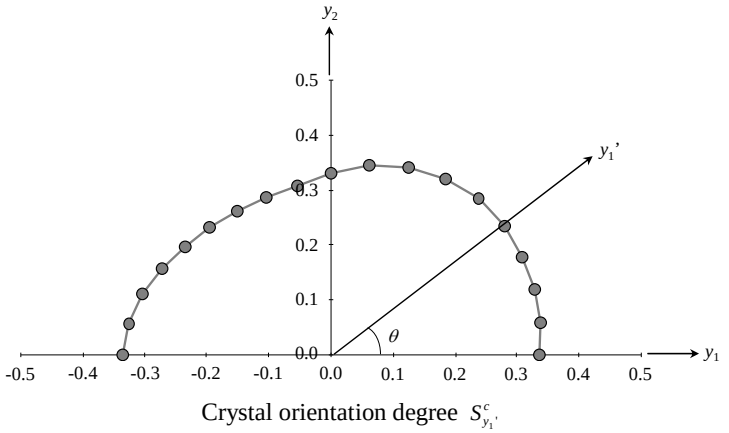


Figure 5: Inplane crystal orienting degree of microstructure.

one when all points orient to y_1' direction. Figure 5 shows the crystal orienting degree for the crystal orientation map in figure 4. These values were calculated through changing the angle θ between y_1 and y_1' from 0˚ to 180˚. The crystal orienting degree indicated a multi-directionally uniform orientation in the y_1– y_2 plane.

3 Multiscale finite element simulation

A three-dimensional multiscale finite element simulation based on homogenization theory was employed to investigate the relationship between SEM·EBSD-measured realistic microstructure and macrostructure. Figure 6 illustrates the scheme of multiscale modelling for polycrystalline piezoelectric materials. The details of formulations and computational approach were presented in the published literatures [10, 11]. A macrostructure is enough large compared with microstructures and it can be assumed that microstructures have the periodicity at microscopic scale. In that case, the multiscale simulation based on homogenization theory enables us estimate the macrostructural homogenized material properties from heterogeneous microstructures. And it also enable us evaluate the microscopic mechanical and electrical behaviours in response to the external load applied to macrostructure.

Figure 7 shows the finite element modelling of SEM·EBSD-measured realistic microstructure. The crystal orientations in 90.8×90.8 μm^2, which consisted of approximately 233 grains, were sampled from the central region of 127.6×90.8 μm^2 in figure 4. A plate-shaped regular cubic mesh, whose number of elements corresponds exactly to one of measuring points, was employed for a microstructural finite element model. The crystal orientation of each measuring point was mapping independently into one element. Our previous investigation

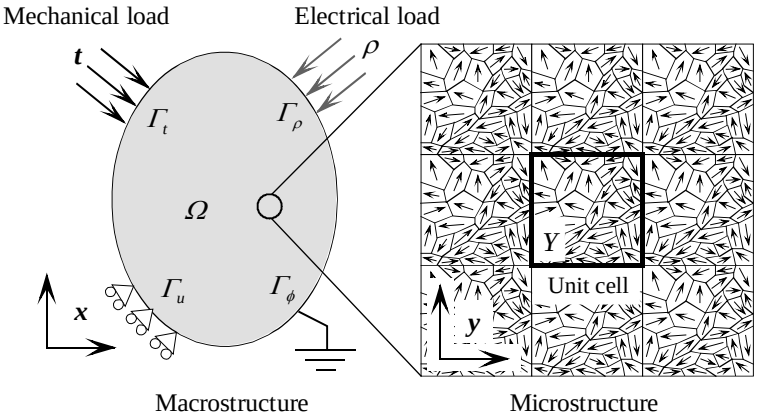


Figure 6: Multiscale modelling of polycrystalline piezoelectric materials.

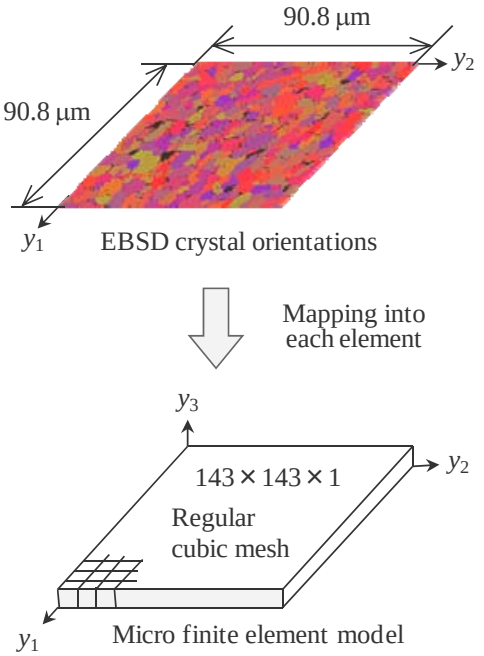
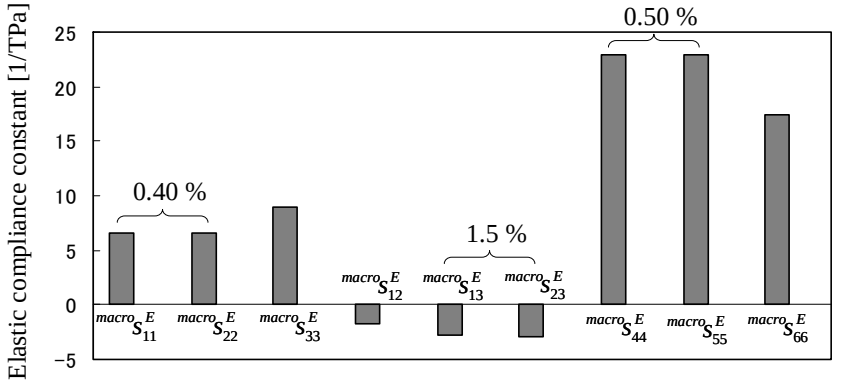


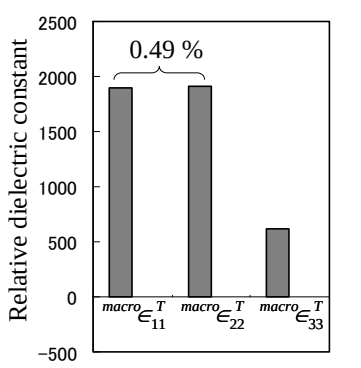
Figure 7: FE modelling of SEM·EBSD-measured realistic microstructure.

for PZT [13] indicated that the dependence on the sampling area of crystal orientations falls off and the relative error of macroscopic homogenized material properties becomes 1% or less if the number of grains is beyond approximately 200. Consequently, the crystal orientation map of 90.8×90.8 μm^2 can be regards as a representative volume element of microstructure since it includes 233 grains.

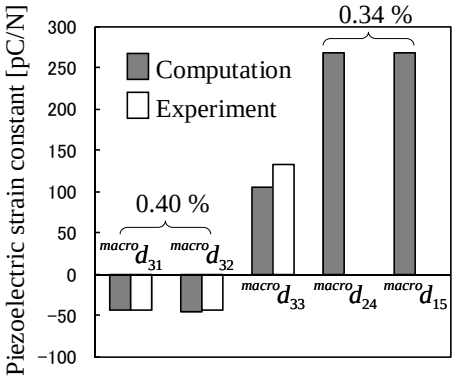
Figure 8 shows the full components of macroscopic homogenized material properties obtained through the multiscale simulation using the SEM·EBSD-measured realistic microstructure in figure 7. The material properties of $BaTiO_3$ single crystal [14] were utilized for each element on the assumption that each grain is poled fully and it consists of a single domain. Figure 8 (a), (b) and (c) indicate the elastic compliance constants, relative dielectric constants and piezoelectric strain constants, respectively. Figure 8 (c) includes the experimental values described in a brochure provided by Murata Manufacturing Co. Ltd. The computational values of piezoelectric strain constants have some correspondence to experimental values although ${}^{macro}d_{33}$ has relatively-large error more than 20%. If we focus on the transverse isotropy, good agreements can be confirmed between the components relating to inplane first and second axes.



(a) Elastic compliance constants



(b) Relative dielectric constants

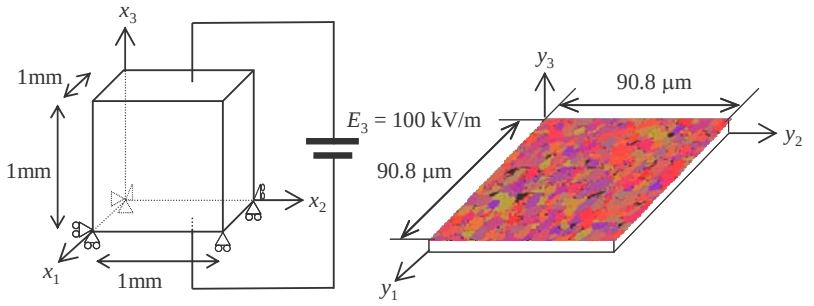


(c) Piezoelectric strain constants

Figure 8: The full components of macroscopic homogenized material properties obtained through the multiscale finite element simulation using SEM·EBSD-measured realistic microstructure.

And now, we discuss the microscopic localized behaviours of SEM·EBSD-measured realistic microstructure in response to a macroscopic external load. Figure 9 (a) demonstrates the macrostructure and its boundary conditions, and (b) shows the microstructure. While the above-mentioned 143×143×1-divided mesh was used for microstructure, a one-element mesh was employed for macrostructure because of a linear displacement and electrical potential field. The macroscopic homogenized material properties in figure 8 were introduced into macrostructure, and the free deformation of a macrostructure was analyzed under a uniform electric field along macrostructural x_3 axis. Then material behaviours of the SEM·EBSD-measured realistic microstructure were evaluated in response to the macroscopic external load.

The difference between single crystal and polycrystal is with and without the interference of material behaviours among neighbouring grains. Namely, grains show the various deformations according to their orientations and they have a large influence on others in polycrystalline microstructure. The interference of mechanical deformation under electrical loads appears obviously in piezoelectric energy. Therefore, we picked out the almost [001]-orientating grains from many grains in the microstructure as an example of computations, and investigated their piezoelectric energy. To be more precise, firstly we calculated the angle between crystallographic c axis, which is the direction of spontaneous polarization in case of perovskite tetragonal structure, and microstructural y_3 axis for the microstructural finite element model. The elements whose angle between both axes is less than 5 degree were picked out as almost [001]-orienting grains and their piezoelectric energy was calculated from mechanical and electrical states cased by macrostructural external load. Figure 10 shows the relation between the piezoelectric energy and the average of orientation gap, which was calculated from neighbouring eight elements, in the realistic microstructure.



(a) Macrostructure (b) Microstructure

Figure 9: The macrostructure, its boundary conditions and the microstructure used for the estimation of microscopic localized behaviours.

Additionally, the piezoelectric energy is compared with two specified states of [001]-orienting single crystal. One is a free deformation state and the other is a perfect-constraint state omnidirectionally under the same electric field with the

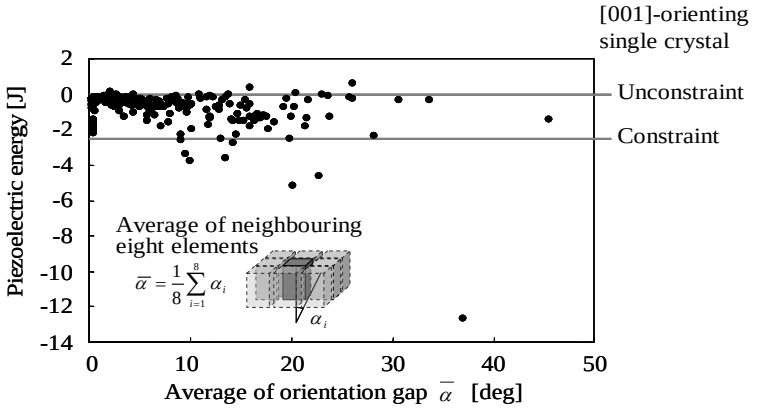


Figure 10: The relation between the piezoelectric energy and the average of orientation gap in the realistic microstructure.

polycrystal. This figure indicates that the piezoelectric energy of some grains reaches beyond one of perfect-constraint single crystal when the average of orientation gap is over 10 degree. However, there is no strong correlation between low piezoelectric energy and the orientation gap.

4 Conclusions

The realistic crystal morphology of a $BaTiO_3$ polycrystalline piezoelectric material was obtained by SEM·EBSD measurement utilizing an amorphous osmium coating for the prevention of electrification. Then we employed a multiscale finite element simulation and investigated the effect of realistic crystal morphology on material properties and behaviours. As a computational result, the macroscopic homogenized material properties correspond approximately to experimental values and they satisfy the transverse isotropy. In a realistic polycrystal under electric field, there are some grains whose mechanical deformation is interfered more strongly by neighbouring grains compared with an omnidirectional-constraint state.

Acknowledgements

One of the authors (Y. Uetsuji) was financially supported by a Grant-in-Aid for Young Scientists (B) (No. 22760087) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

[1] Venables, J. & Harland, C., Electron back scattering patterns – A new technique for obtaining crystallographic information in the scanning electron microscope. *Philosophical Magazine*, **27**, pp.1193-1200, 1973.

[2] Dingley, D. & Randel, V., Microstructure determination by electron back-scatter diffraction. *Journal of Material Science*, 27, pp.4545-4566, 1992.
[3] Wu, X., Pan, X. & Stubbins, J.F., Analysis of notch strengthening of 316L stainless steel with and without irradiation-induced hardening using EBSD and FEM. *Journal of Nuclear Materials*, **361**, pp.228-238, 2007.
[4] Calcagnotto, M., Ponge, D., Demir, E. & Raabe, D., Orientation gradients and geometrically necessary dislocations in ultrafine grained dual-phase steels studied by 2D and 3D EBSD. *Materials Science and Engineering: A*, **527**, pp.2738-2746, 2010.
[5] Yasutomi, Y. & Takigawa, Y., Evaluation of crystallographic orientation analyses in ceramics by electron back scattering patterns (EBSP) method. *Bulletin of the Ceramic Society of Japan*, **37**, pp.84-86, 2002.
[6] Yang, L.C., Dumler, I. & Wayman, C.M., Studies of herringbone domain structures in lead titanate by electron back-scattering patterns. *Materials Chemistry and Physics*, **36**, pp.282-288, 1994.
[7] Koblischka-Veneva, A. & Mücklich, F., Orientation imaging microscopy applied to $BaTiO_3$ ceramics. *Crystal Engineering*, **5**, pp.235-242, 2002.
[8] Tai, C.W., Baba-kishi, K.Z. & Wong, K.H., Microtexture characterization of PZT ceramics and thin films by electron microscopy. *Micron*, **33**, pp.581-586, 2002.
[9] Gudes, J.M. & Kikuchi, N., Preprocessing and postprocessing for materials based on the homogenization method with adaptive finite element methods. *Computer Methods in Applied Mechanics and Engineering*, **83**, pp.143-198, 1990.
[10] Uetsuji, Y., Nakamura, Y., Ueda, S. & Nakamachi, E., Numerical investigation on ferroelectric properties of piezoelectric materials. *Modelling and Simulation in Materials Science and Engineering*, **12**, pp.S303-S317, 2004.
[11] Kuramae, H., Nishioka, H., Uetsuji, Y. & Nakamachi, E., Development and performance evaluation of parallel iterative method. *Transactions of the Japan Society for Computational Engineering and Science*, Paper No.20070033, 2007.
[12] Uetsuji, Y., Yoshida, T., Yamakawa, T., Tsuchiya, K., Ueda, S. & Nakamachi, E., Evaluation of ferroelectric properties of piezoelectric ceramics based on crystallographic homogenization method and crystal orientation analysis by SEM･EBSD technique. *JSME International Journal Series A*, **49**, pp.209-215, 2006.
[13] Uetsuji, Y., Satou, Y., Nagakura, H., Nishioka, H., Kukamae, H. & Tsuchiya, K., Crystal morphology analysis of piezoelectric ceramics using electron backscatter diffraction method and its application to multiscale finite element analysis. *Journal of Computational Science and Technology*, **2**, pp.568-577, 2008.
[14] Jaffe, B., *Piezoelectric ceramics*, Academic Press: London and New York, p.74, 1971.

This page intentionally left blank

Section 9
Ballistics

This page intentionally left blank

Coupled numerical-experimental study of an armour perforation by the armour-piercing projectiles

B. Zduniak, A. Morka & T. Niezgoda
Department of Mechanics and Applied Computer Science, Military University of Technology, Poland

Abstract

This paper concerns an analysis of target penetration by a selected armour-piercing (AP) projectile: 7.62x54R with steel core. Numerical and experimental research was carried out. The aim of this work was a comparison of the results obtained in real conditions of ballistic test and computer simulation. In this study, two three-dimensional targets and the core of the projectile were built. The structure of the projectile is complex, but steel core plays the main role in the perforation process. Then the numerical model of the projectile was reduced to describe only steel core dynamics. The 3D Element Free Galerkin method is applied to solve the problem under consideration. The algorithm implemented in the Ls-Dyna code was used. Space discretization of the analyzed problem was prepared by means of the HyperWorks Software (HyperMesh module). The total amount of the elements reaches 500 000 in this model. The Johnson-Cook constitutive model is applied to describe the behaviour of the metallic parts: steel layers and the projectile's core. The experimental results were obtained using a high speed video-camera. The target penetrations by the projectile were recorded. The processing of the data obtained from a high speed camera was carried out by means of the TEMA Software. In this paper, a good correlation between the numerical and experimental results was obtained. A lot of interesting mechanical effects observed during the experiment were analyzed.

Keywords: penetration, perforation, numerical model, constitutive model, armour.

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110541

1 Introduction

The paper describes a coupled numerical-experimental study of an armour perforation by a given type of projectile. An analysis of an AP 7.62 x 54 R type projectile with steel core was carried out. The structure of the projectile is complex, but steel core plays the main role in perforation process. The numerical model of projectile was reduced to describe only steel core dynamics. Two types of targets are used in modeling. The numerical simulations were performed using the Element Free Galerkin Method (EFG) implemented in LS-DYNA code. Three dimensional numerical models for each version of impact were developed. The initial stage of the problem is presented in Figure 1.

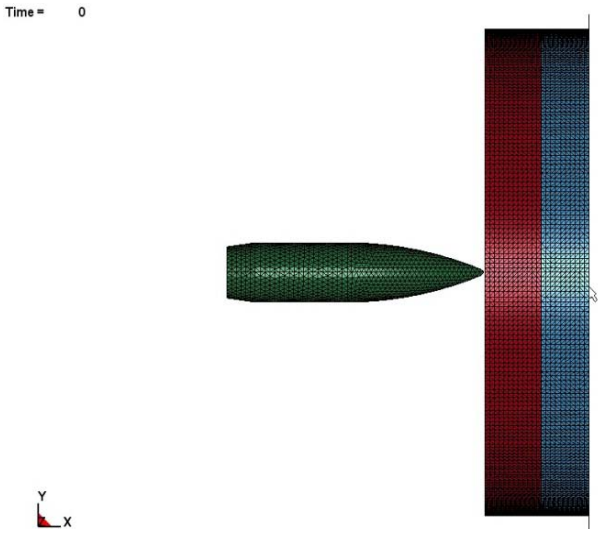


Figure 1: The initial stage of the problem. The target block with perpendicular impact of the 7.62 type projectile on target.

Three-dimensional models of the projectile and target were developed with a strain rate and temperature dependant on material constitutive relations. An impact of the 7.62 x 54R mm armour-piercing projectile on a 6 mm thick plate of steel and 5mm thick plate of aluminum was modelled as the first type of target. The second type was made of four 5mm thick aluminum plates. The target block with perpendicular impact of the 7.62 type projectile on target was assumed as reference. The perpendicular impacts are the most dangerous cases. The 7.62x54Rmm projectile belongs to level III of the ballistic protection according to the STANAG 4569 standard. That type of projectile consists of a soft metal jacket, incendiary material and a hard steel core which is the crucial element for penetration effectiveness. The impact velocity equals 854 m/s. The geometry of this kind of projectile is shown in Figure 2.

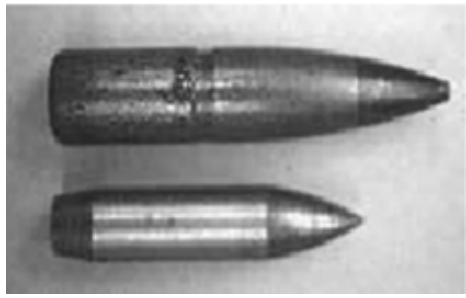


Figure 2: The geometry of projectile [1].

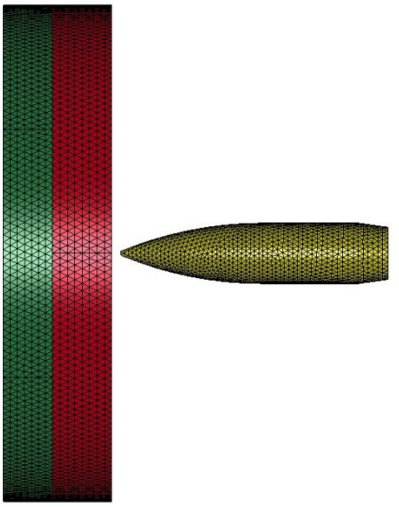
The experimental results were obtained by using a high speed video-camera. A scheme of the experiment is shown in Figure 3.

Figure 3: Scheme of the experimental test of impact.

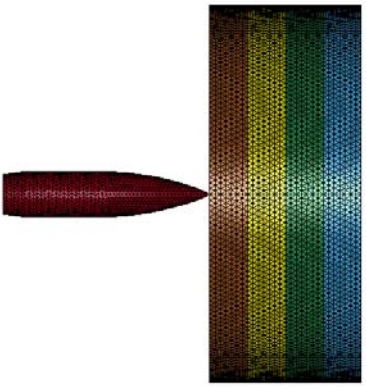
The target penetration by projectile 7.62x54 R type AP was recorded. The processing of the data obtained from the high speed camera was carried out by means of the TEMA Software. The aim of this work was a comparison between the results obtained in real conditions of ballistic test and computer simulation.

2 Numerical model

For the purpose of this study numerical models of projectile and targets were constructed. Target can be built using three types of configurations: with steel plate and aluminum plate or four aluminum plates. Numerical models for panel's configurations: steel/aluminum plates, 4 aluminum plates are presented in Figure 4 (a), (b).



(a)



(b)

Figure 4: A 3D view of the mesh and geometry of the panel's configuration, (a) steel/ aluminum, (b) 4 aluminum plates.

Thickness of steel (ARMOX 500) plate equals 6 millimeters, thickness of aluminum (PA11) is 5 millimeters, diameters of all plates are 50 millimeters.

Figure 1 represents meshes of the numerical model consisting of three parts: the steel target backed by an aluminum plate and projectile's steel core. The typical node to node distance was equal to about 1mm in all cases of targets and projectiles. The total number of elements per single case exceeded 500k. The Johnson-Cook [2] constitutive model was chosen for armox500 and PA11. The data for constitutive model was derived from work [3]. The excessive deformations often observed in the perforation/penetration issues caused the choice of a meshless method as the method for the problem solution. The Element Free Galerkin (EFG) [4] method implemented in the LS-DYNA solver was selected [5, 6]. EFG only uses a set of nodal points describing geometry of

the body, no mesh in the classic sense is needed to define the problem. The initial condition was reduced to the given projectile velocity, 854 m/s. The boundary condition was assumed as the target plate fixed at its edge a 5 millimeters thick ring.

3 Experimental setup

This experimental work was carried out at a firing ground. The experimental setup is shown in Figure 3. There was a ballistic tunnel, in which firing takes place. A ballistic barrel was used for firing. 7.62x54R type AP of projectiles was used. All of projectiles were turned, because fuse of projectile induces a flash, what makes recording a film impossible. A turned projectile is shown in Figure 5.

Figure 5: A turned projectile.

The target plate is fixed at its edge with a 5 millimeters thick ring. A high speed camera was put in the ballistic tunnel. It had a very resistant cover, made of a transparent polycarbonate.

4 Results and discussion

Computer simulations were performed for the selected cases. A high performance computing system based on the cluster architecture was used. This allowed us to assign 4 to 8 CPUs per single job, limiting the total computing time to a reasonable level.

The first analysed variant was the steel-aluminum target. Armox500 type of steel was used and aluminum was of PA11 type.

This case underwent an entire perforation, in both numerical and experimental cases. It is presented below on a plot of kinetic energy, where ending kinetic energy value after target's perforation does not equal zero.

Figure 6 shows the change in the kinetic energy of the projectile in time for the first case.

Figure 7 shows the change in the velocity of projectile in time.

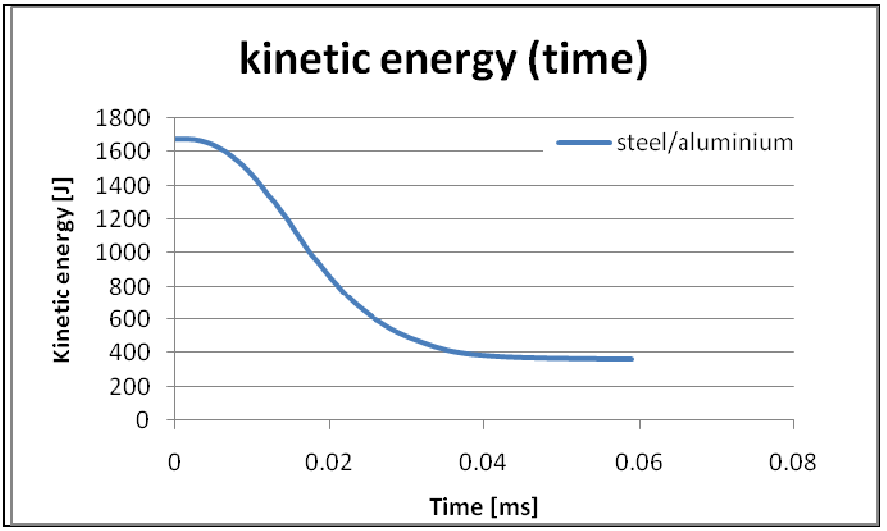


Figure 6: Change in the kinetic energy of projectile in time for steel/aluminum target for numerical simulation.

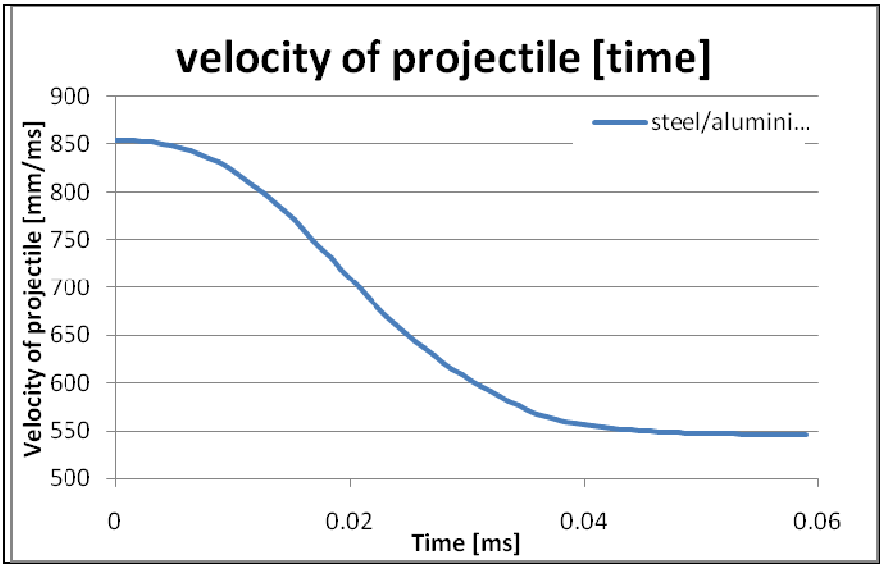


Figure 7: The change in the velocity of projectile in time (numerical result).

The velocity of projectile in numerical model, after perforation of the target, equals about 550 m/s.

In experimental result, the velocity of the projectile after perforation of the target is about 576 m/s. This velocity is obtained from velocity's measurements on a short-circuit sensor. The velocity measurement obtained from a film analysis in Tema is impossible, because dust does not allow us to trace the projectile's path.

However, we can observe similar projectile's velocity values for both experimental and numerical cases.

Figure 8 presents a comparison of numerical model and experimental protective panel at the same time of a perforation process. This picture shows a very similar behavior of both targets. It means, that a damage model of numerical simulation is similar to fracture in real material, and that the used constitutive models and material parameters are compliant.

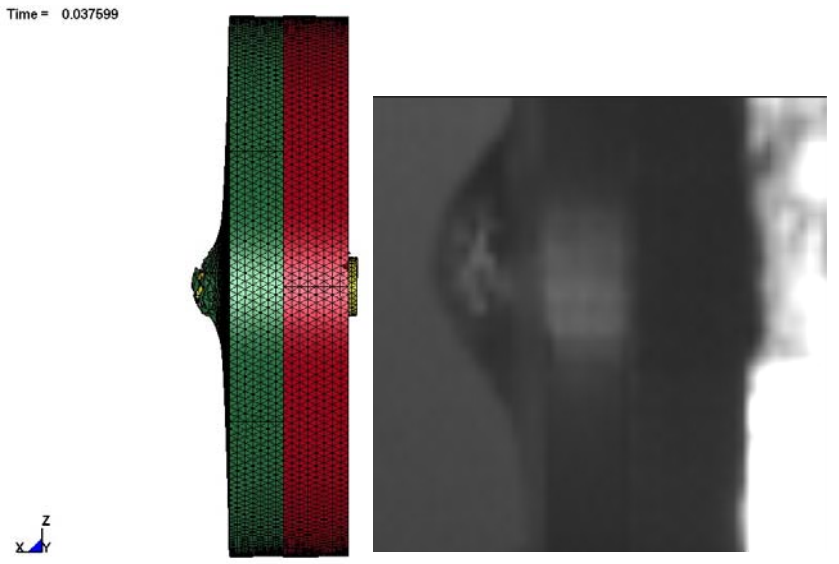


Figure 8: Comparison of numerical model and experimental protective panel at the same time of a perforation process for steel/aluminum target.

The second analysed variant is a target consisting of 4 aluminum plates. A type of aluminum was PA11.

This case underwent an entire perforation, in both numerical and experimental cases, which is shown in Figure 9. Figure 9 presents a change in the kinetic energy of the projectile in time for the second case for numerical model.
Figure 10 shows a change in the velocity of projectile in time.

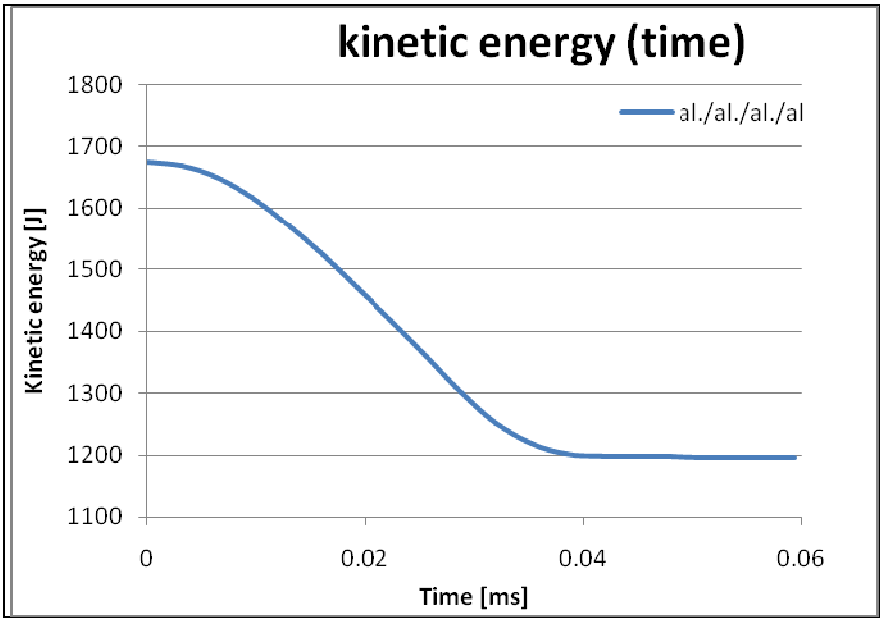


Figure 9: Change in the kinetic energy of the projectile in time for the second case for numerical model.

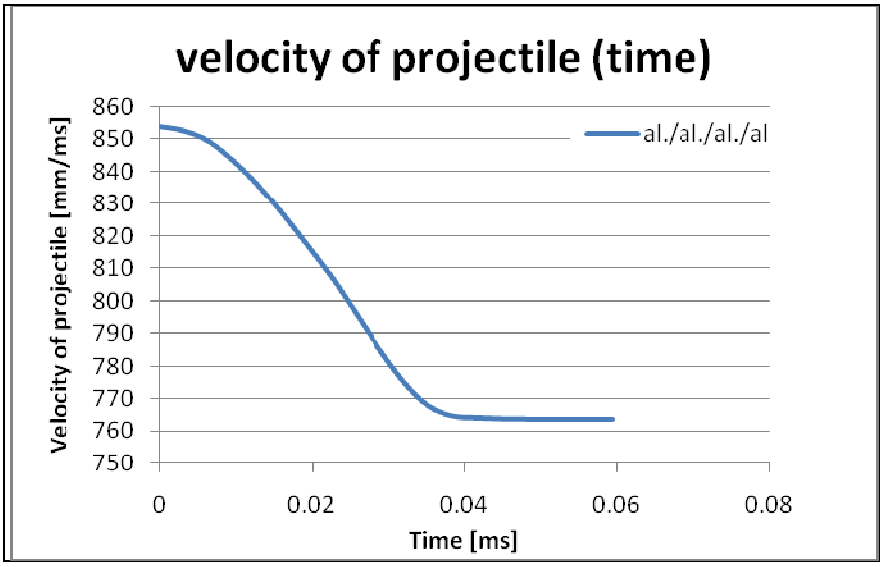


Figure 10: Change in velocity of projectile in time for second numerical model.

The velocity of projectile in numerical model, after perforation of the target, equals about 762 m/s.

In experimental result, the velocity of the projectile after perforation of the target is about 706 m/s. This velocity is obtained from velocity's measurements on a short-circuit sensor. For this velocity value can have an influence on external factors, e.g. shock wave, fragments simulating projectile while using the short-circuit sensor. This may cause 7% error between final experimental and numerical results.

A velocity obtained from Tema equals 709 m/s.

Figure 11 shows a comparison between the numerical model and experimental protective panel at the same time of a perforation process. This picture also shows a very similar behavior of both targets.

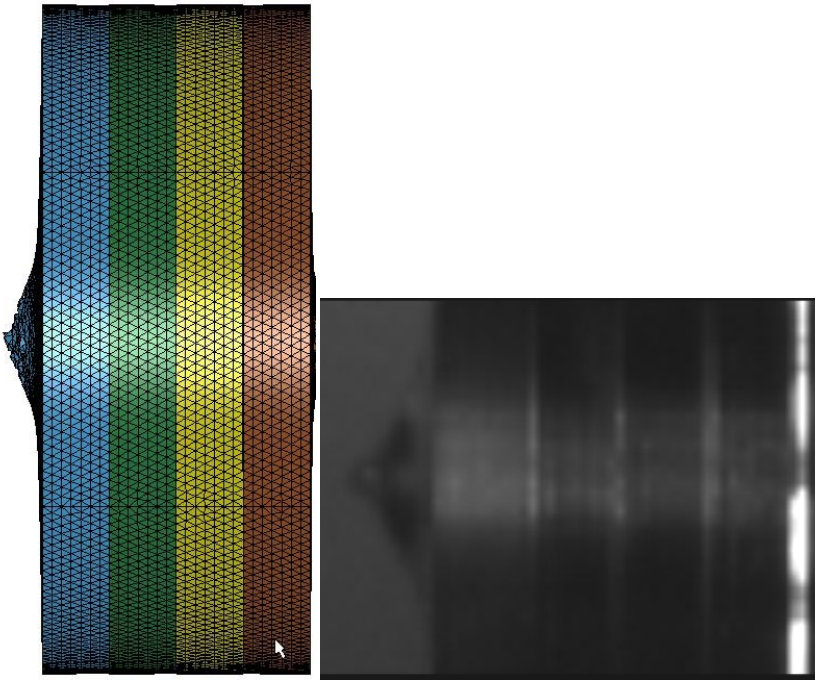


Figure 11: Comparison of the numerical model and experimental protective panel at the same time of a perforation process.

5 Conclusions

The presented analysis has provided only preliminary information about the comparison between the numerical and experimental model of the perforation problem. There is insufficient data about the mechanical properties of the steel,

which the projectile is made of. This makes a quantitative assessment of the results difficult in many cases.

The most important task of this article is to show, that using numerical models allows us to reduce the cost of experimental research, which is very expensive. The correlation obtained in this job is enough to know if this target gives in perforation or not. A similar behaviour of penetration targets is also achieved. Plate materials were destroyed in similar ways. In this paper, a good correlation of the numerical and experimental results is received. Further work for finding a good damage parameters for steel projectile will be carried out.

References

[1] Paul J. Hogg, "Composites for ballistic applications", Department of Materials Queen Mary, University of London.
[2] G. R. Johnson, W. H. Cook, A constitutive model and data for metals subjected to large strains, high strain rates and high temperatures, Engineering Fracture Mechanics, Volume 21, Issue 1, p. 31-48, 1985.
[3] A. Morka, B. Jackowska, T. Niezgoda, Numerical study of the shape effect In the ceramic based ballistic panels, Journal of Kones Powertrain and Transport, Vol.16, No.4, 2009.
[4] C. Kambur, Assessment of Mesh-free Methods in LS-DYNA: Modeling of Barriers in Crash Simulation, 2004.
[5] J. O. Hallquist, LS-DYNA Theory Manual, Livermore Software Technology Corporation, 2006.
[6] LS-DYNA Keyword User's Manual, version 971, Livermore Software Technology Corporation, 2007.

Numerical analysis of missile impact being shot by rocket propelled grenades with rod armour

T. Niezgoda, R. Panowicz, K. Sybilski & W. Barnat
Department of Mechanics and Applied Computer Science
Military University of Technology, Poland

Abstract

During the operation of the UN and NATO forces in Africa and Asia it appeared that the equipment held and utilized by these organizations was not proof against the missiles being shot by rocket propelled grenades. The missiles utilized by the terrorist organizations belong to the first generation. They have a head with a shaped charge jet permitting them to pierce up to 300 mm RHA steel.

There are four protection methods against this hazard: reactive armour, thick proper armour, rod armour and active defence system. The reactive armour cannot be applied on each kind of vehicle. The thick, proper armour limits other important features of vehicles. The active defence systems continue to develop, except the Russian system "Arena". Therefore, the fast development of rod armour appeared. The main task of this armour type is to disturb the symmetry of the shaped charge jet as a result of the shock that is effective of the deflagration and not the forming of cumulative jet.

The finite element model of the missile with the cumulative head and its influence with two types of rod armour will be presented in this article. The cases of the missile impact with circular and square section for different typical speeds of missile will be considered.

Keywords: mechanics, finite element method, rods armour, RPG.

1 Introduction

As a result of armed conflicts observation all over the world and actions carried on by Polish army divisions within a framework of stabilization missions, we can notice a change in the way of fighting and the usage of available ammunition. It appears that currently the biggest threat is posed by two types of ammunition.

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110551

The first are all kinds of landmines and improvised explosive device (IED), whose basic destructive factor is a pressure wave formed due to the explosion. The destruction effect is multiplied by all kinds of shrapnel (marbles, nails, screws etc.) placed inside or on the explosive device. The second group of the most threatening ammunition included are bullets with accumulation heads. They were worked out to destroy military armoured vehicles. Their popularity is connected, before all, with high penetration (reaching 900 mm of reinforced steel RHA [1]), low production cost and the simplicity of use.

High penetration of projectiles with an accumulation head is a big problem for the protection against this kind of ammunition. One of the possible protection methods against this threat is the use of rod armour. A perfect protection of this type should not allow the creation of accumulation stream, that is the striking power of a projectile with accumulation head. The principle of operation of this kind of armour consists in neutralizing the accumulation insert and the explosive material contained inside the projectile by a strong deformation of these elements. The result is an eventual creation of accumulation stream with a lower efficiency, in an extreme case – the lack of it. In this situation the destruction force is comparable with a detonation of an explosive material contained within the projectile and the reaction of fragments created by the explosion.

In the following article the results of a simulation of an impact of a projectile with an accumulation head type PG-7G in two different kinds of rod armours will be presented. Two impacts under different angles in respect to the projectile axis will be considered.

For the analysis LS-Dyna software was used containing an open implementation of finite elements method.

2 Construction and principle of operation of projectile with an accumulation head

A projectile with an accumulation head is made of three principle elements: head, rocket engine and flight stabilizer (fig. 1) [2, 3]. The head contains the most important parts of the projectile: accumulation insert, explosive material, a shield and a fuse. During the shooting the projectile is fired outside the barrel towards the target and accelerates to the maximum speed reaching 300 m/s. In the moment of impact on the target the fuse is generating an electrical impulse exciting through the detonator of the explosive material. The effect of exciting the detonator is the detonation of the explosive material and creation of the accumulation stream.

The condition of achieving the maximum destruction power of the accumulation stream is a proper, very accurate workmanship of the device and the accumulation insert. In the case of both of these elements the inaccuracy of workmanship should not exceed 0,05 mm, and the radius deviation should not be bigger than 0,03 mm. Together with the increase of the array's inaccuracy its effectiveness decreases. In an extreme case, with large workmanship errors, an accumulation stream will not be created and the projectile's effect will be

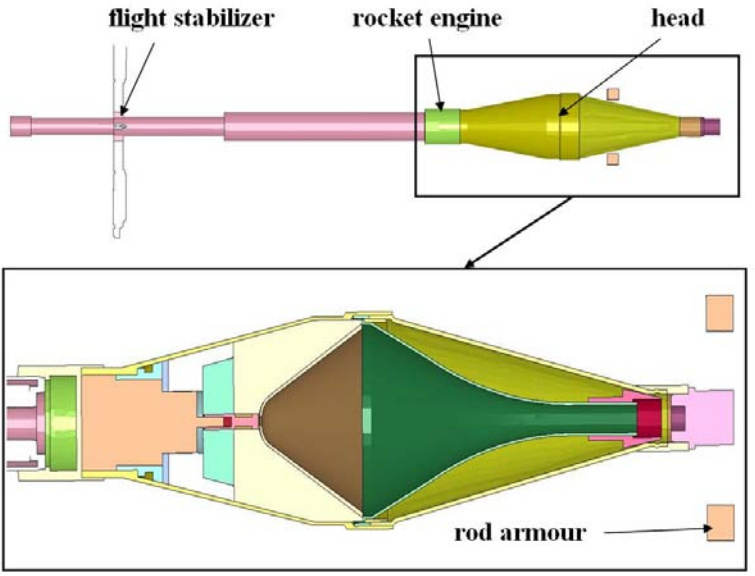


Figure 1: An intersection of the projectile PG-7G head.

comparable with the effects of an explosion of an explosive material with a mass of the charge contained in the head. This feature is used in rod armours. It is assumed, that with certain probability amounting to less than 50%, it will not come into contact between the fuse of the incoming projectile and the elements of the armour. As a result an accumulation stream will not be created.

Results from the carried out analysis indicate that in most cases the projectile hits the rods with the elements of the projectile's casing, which causes deformation of the accumulation insert and crumbling of the explosive material, reducing considerably the effectiveness of this weapon.

3 MES model construction of PG-7G projectile

To carry out the projectile PG-7G type impact simulation in different kinds of rod armours the elaboration of a numerical model was necessary. The principal restriction during its construction was the maximal number of used finite elements. Together with the increase of the finite elements number the quantity of unknown in the dynamical equation increases, and with this the time needed to solve both this equation as well as the whole problem prolongs. To make the limitation of the number of finite elements possible, introducing certain simplification of the projectile's geometry was necessary (fig. 2), that do not influence the behaviour of the array as a whole.

In the model five main simplifications were introduced:

1. Shoulders were removed and a hole in the tube between the nozzles and the fins.
2. The flight stabilizers were replaced by a cylinder with an identical mass.
3. Holes fastening the fins were removed.
4. Nozzles were removed.
5. A part of the fuse protruding outside the projectile was replaced with one cylinder.

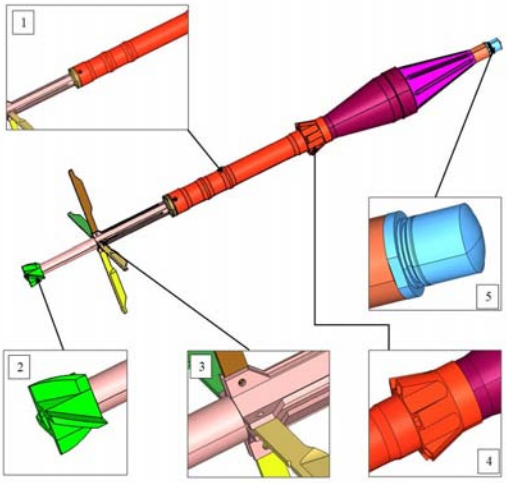


Figure 2: Projectile's geometry simplifications.

The numerical model constructed and used for the calculations is composed of 19 parts. Particular parts were attributed to material properties, that characterize the behaviour of real materials used in the analysed projectile. For the description of copper a simplified Johnson-Cook model was used. The behaviour of other materials was reflected by the use of a multilinear model (Piecewise_Linear_Plastic). This model allows to image materials characteristics by introducing an experimental curve displaying the dependence of stress from strain.

Between all of the parts of the model a contact was defined, that is usually used in dynamical problems for example in analysis of crash type. In these analyses, similarly as in described case, we deal with great deformations and elements displacement speeds. Parts of the projectile element are connected to each other with a screw thread. Between these elements of the model a contact, simulating this connection, was defined. The chosen contact type enables inserting a conditioned separation of connected elements.

During the simulation imaging of projectile PG-7G type impact into an obstruction made out of two angle sections, being a fragment of the rod armour structure, was endeavoured (fig. 3). In the first case rods with a circular intersection with a diameter of 15 mm were considered, and in the second – with

a square intersection (side length 14 mm). The rods dimensions were matched so that the mass of the armour in both solutions was close. In both cases it was accepted, that the total length of a singular rod amounted to 500 mm. It was assumed, that the projectile hits centrally between the rods, at the half of their length.

The simulation was carried out for three initial speeds of the projectile V_z. In the first version the projectile moved with the minimal speed that it acquires during firing (100 m/s), in the second with maximal speed (300 m/s). The third case was put into practise for an intermediate speed amounting to 200 m/s.

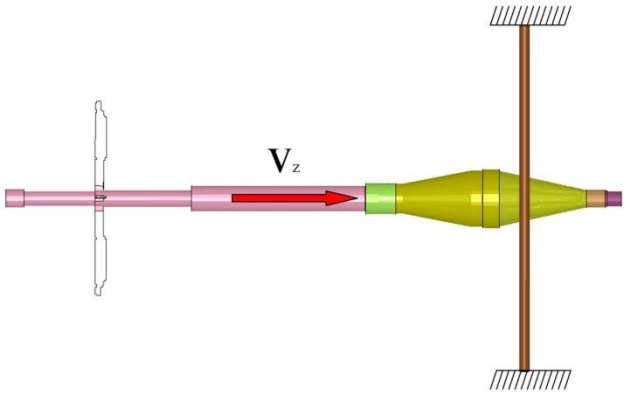


Figure 3: Initial-boundary conditions.

Numerical calculations were carried out in LS-Dyna system. This system is used, among others, for the analysis of fast-changing processes with the finite elements method. Based on the information contained in the model it builds a dynamical movement equation in the following form [4]:

$$M\ddot{q} + C\dot{q} + Kq = f \tag{1}$$

This equation is solved then with the direct integration method (so called Explicite), that finds a wide use in the analysis of highly nonlinear phenomena (impacts, stamping, explosion wave influence on a structure). In calculations of this type we have to manage with great strains and strain speeds.

4 Numerical analysis results

In figures 4 and 5 the strain process of the projectile during the impact into the rod with a square intersection for the initial speeds amounting to 100 and 300 m/s are presented. In the initial stage we can observe the crushing of the ballistic cap after the impact into the head's body. This impact causes strong folding of the body that in consequence results in significant damage of the explosive

material. In the final stage it comes to the burring of the body and passage of the projectile between the rods. In attached figures one can notice that with the increase of the initial speed the way the projectile's elements behaviour changes. The head's body burring proceeds more on the outside of the projectile, the rods strain less, but the accumulation insert is strained more.

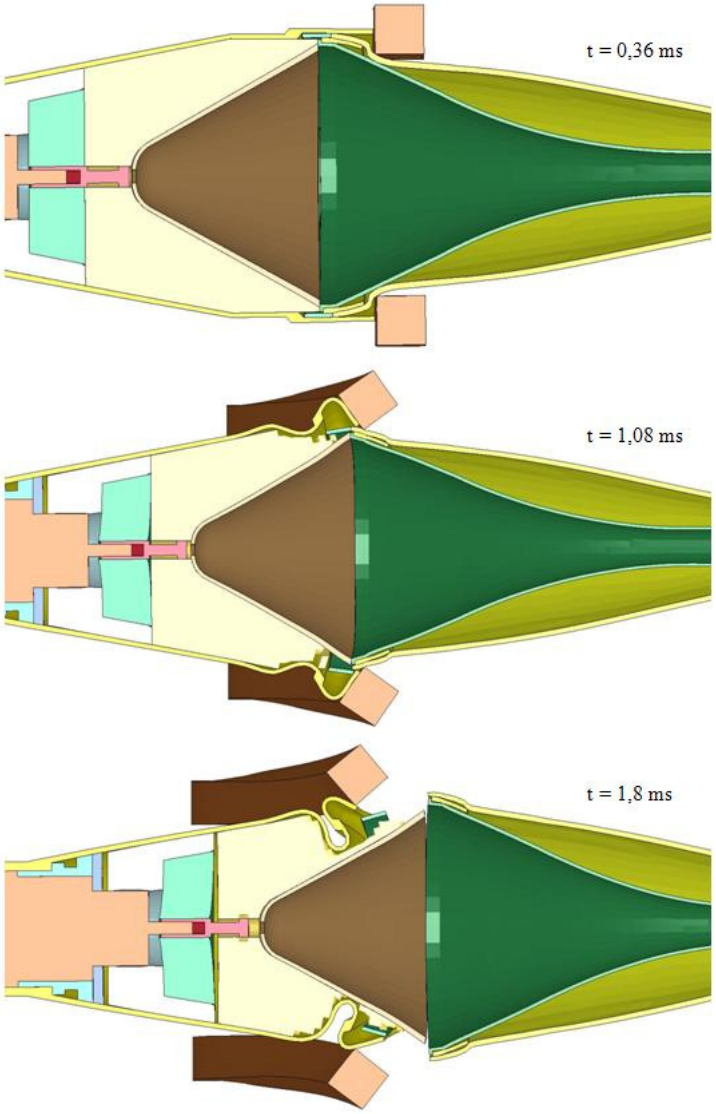


Figure 4: The course of the projectile's deformation for the impact speed 100 m/s.

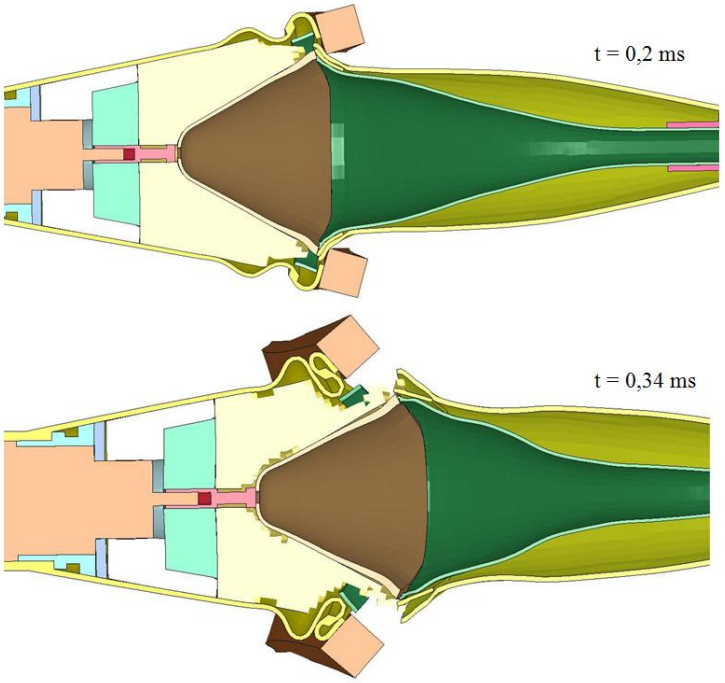


Figure 5: The course of the projectile's deformation for the impact speed 300 m/s.

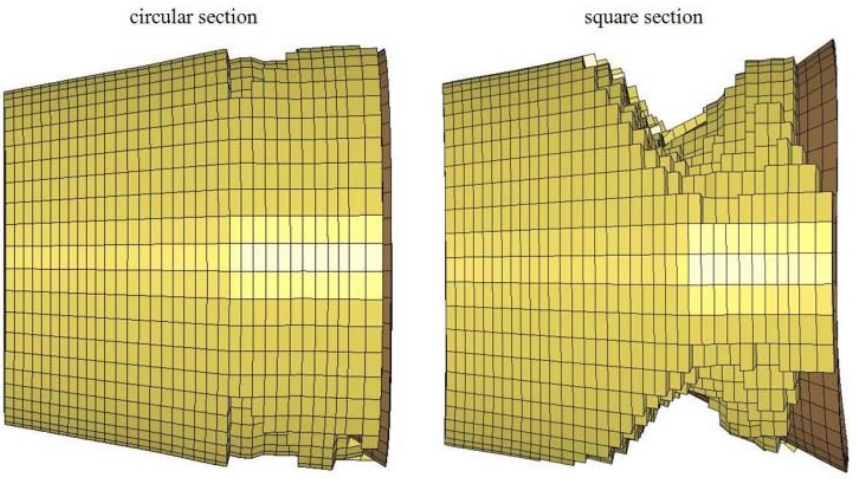


Figure 6: Maximal deformation of the accumulation insert and the destruction of the explosive material for the initial speed of the projectile flight 100 m/s.

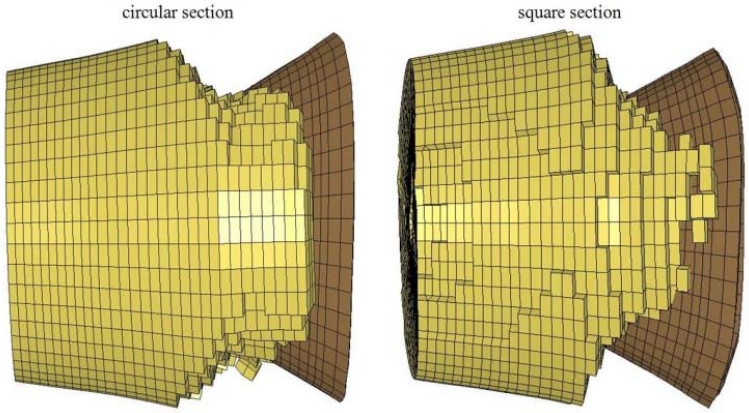


Figure 7: Maximal deformation of the accumulation insert and the destruction of the explosive material for the initial speed of the projectile flight 200 m/s.

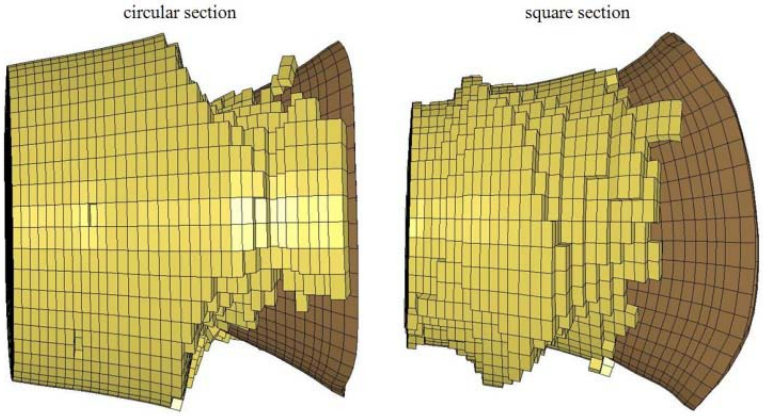


Figure 8: Maximal deformation of the accumulation insert and the destruction of the explosive material for the initial speed of the projectile flight 300 m/s.

In figures 6, 7 and 8 the maximal destruction of the explosive material and the accumulation insert recorded during the whole simulation, were presented. On the basis of the figures we can observe that the destruction is greater in case of the impact into the rods with a square intersection. Additionally, together with the increase of the initial speed of the projectile, the explosive material is crushed more uniformly on its whole length.

5 Conclusions

From the point of view of protecting the soldiers and the equipment the most important thing is the fact of the creation, or the lack of it, of the accumulation stream. The condition of the explosive material and the accumulation insert at the moment of the fuse actuation decide about the creation of it.

The analysis carried out showed that with the change of the rods crosswise intersection shape (from circular to square) maintaining the same mass of the whole armour, the efficiency of the projectile decreases. The key elements of the projectile, such as the explosive material and the accumulation insert, are deformed more and at the same time have lower ability of creating an accumulation stream which is the greatest destructive power of this type of weapon.

The speed of the projectile during impact with the armour is not without meaning. The increase of the speed causes, before all, the damage of the explosive material on a greater surface.

References

[1] Wiśniewski A., *Pancerze budowa, projektowanie i badanie*, WNT, Warszawa, 2001.
[2] www.ruag.com.
[3] Military Parade, Russia, Moscow, 2001.
[4] J.O.Hallaquist: LS-Dyna. Theoretical manual, Livermore Software Technology Corporation, California, 2005.

This page intentionally left blank

On the truss-type structures backing the ceramic tiles in the ballistic panels subjected to the impact of hard steel projectiles

A. Morka & P. Dziewulski
[1]*Military University of Technology, Warsaw, Poland*

Abstract

The objective of this paper is a feasibility study of a new concept for ceramic backing in multilayer ballistic panels. The truss structures were applied as a backing layer. They were based on a diamond crystal structure. The analysis of length/diameter ratio of the bars was performed. It was aimed at the achievement of the required mechanical properties maximizing the ballistic resistance and minimizing panel surface density. The panel structure is considered as a composition of Al_2O_3 ceramic tile backed by a plate in a truss form made of aluminum alloy or steel. These results were compared with classic multilayer solutions based on solid aluminum alloy backing plates. The study was carried out for a normal impact of the AP (armor-piercing) 7.62x51 projectile type with tungsten carbide (WC) core. A method of computer simulation was applied to study the problem. The Finite Element Method (FEM) implemented in the LS-DYNA code was used. The full 3D models of the projectile and target were developed with strain rate and temperature-dependent material constitutive relations. The Johnson-Cook constitutive model with Gruneisen Equation of State (EOS) was applied to describe the behavior of metal parts: aluminum alloy and projectile's core. However, the Johnson-Holmquist model was used to describe the ceramic material. The Boundary Conditions (BC) were defined by supporting the panel at its back edges. The obtained results show the alternative solution to the classic solid plates supporting the ceramic layers in ballistic panels. It was identified that the main deformation mechanism in truss-type backing components is buckling if the L/D ratio goes up. A virtual prototyping technique could be applied to manufacture the developed truss structure.

Keywords: computational mechanics, finite element method, ballistic protection, multilayer armour, ceramic armour systems, truss structure.

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110561

1 Introduction

Modern ballistic protection systems, especially for light-weight armored vehicles, are based on the multilayer armour concept, fig. 1 [11]. The main task to resist the projectile is typically given a ceramic layer. However, the ceramic material is known as very brittle so its strength in tension or bending loads is very low. Therefore, the next layer behind should compensate for this disadvantage of the ceramics. It is known as a backing effect. The solid materials like aluminum or polymer composite are considered as the backing plate in classic solutions of ballistic panels. The surface density of the panel is a crucial parameter at given effectiveness, and then its minimization is especially important. There is no alternative to ceramics in case of Armour Piercing (AP) projectiles containing hard cores made of steel or tungsten alloys. Then the backing plate can be considered to be replaced with a lighter material. One possibility is a choice of a truss structure. However, selection of the best structure is not a simple task. So we turned to nature in looking for a hint. The answer came from space. It could be lonsdaleite; in other words hexagonal diamond. Its crystal structure is an allotrope of carbon with hexagonal lattice. It is formed in nature when graphite-containing meteorites strike the Earth. The great heat and pressure of the impact transforms the graphite into diamond, but retains graphite's hexagonal crystal lattice. Lonsdaleite was first identified from the Canyon Diablo meteorite, where it occurs as microscopic crystals associated with diamond. It was first discovered in nature in 1967 [4]. Specific gravity of 3.2 to 3.3, and Mohs hardness of 7–8 [5]. The Mohs hardness of diamond is 10, and the lower hardness of lonsdaleite is chiefly attributed to impurities and imperfections in the naturally-occurring material. Numerical analysis showed that pure sample has been found to be 58% harder than diamond [6]. Lonsdaleite is expected to be 58% harder than diamond and to resist indentation pressures of 152 GPa, whereas diamond would break at 97 GPa [6].

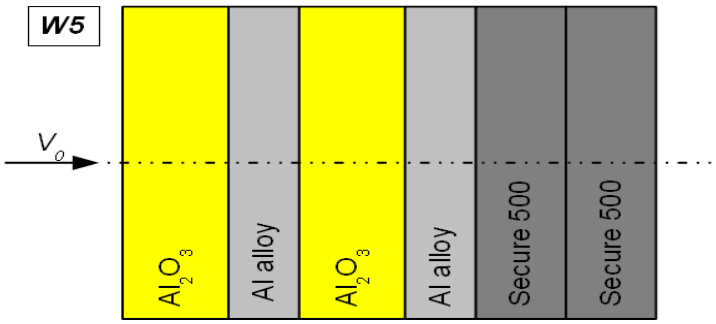


Figure 1: An example of the multilayer armour concept [11].

The authors do not intend to apply such rare and strange material as the backing plate in ballistic panels. The truss structure in demand could be based on the lonsdaleite's crystal lattice instead. The main idea is that the very high

hardness of this crystal may partially be a result of its specific crystal structure. It is clear that the real truss structure cannot reproduce the atomic forces of interaction, where the force is proportional to the square of displacement. In contrast, for macroscopic materials, the Hook's law yields only the linear relation between force and displacement. However, it is still interesting to know the role of the geometric factor in overall structure strength. The truss structure formed as a plate is presented in fig. 2 with different points of view. The characteristic size of the elementary cell is around 0.45mm.

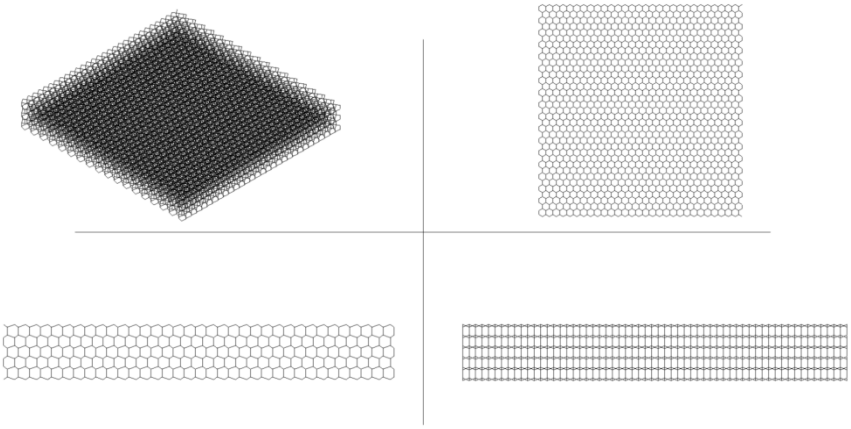


Figure 2: A truss structure formed based on an idea of the lonsdaleite's crystal lattice.

2 Description of the problem

An assessment of the effectiveness of the truss-type backing plate is performed by comparison with other standard solutions. Several cases are considered: (1) the absence of the backing plate, (2) rigid backing plate with friction, (3) solid aluminium plate supporting the ceramics and, finally, (4) truss structure covered by two thin aluminium alloy sheets applied as the backing of the ceramic tile. The test problem was selected as a normal impact of the AP (armor-piercing) 7.62x51mm projectile type with WC core on 9mm thick hexagonal ceramic tile supported by one of the above-mentioned ways. The dimensions of the projectile and targets were depicted in fig 3. The structure of this projectile type is complex, but the tungsten carbide core plays the main role in perforation process. Then the problem was simplified to describe only WC core dynamics. The experimental tests were conducted for 10mm thick Al_2O_3 ceramic tile backed by PA11 aluminum alloy disc with thickness of 10mm. The final result is shown in fig. 4, where the deformed backing plate is visible. It is important to note that such configuration of the ballistic panel is effective enough to stop the 7.62x51mm AP projectile.

The quantitative results assessment should be based on the independently-measured parameters. They are supposed to include the time history of the projectile kinetic energy (PKE) and residual length of the projectile (RLP). However, the qualitative evaluation is going to be based on the projectile/target deformation at a selected moment of time after impact.

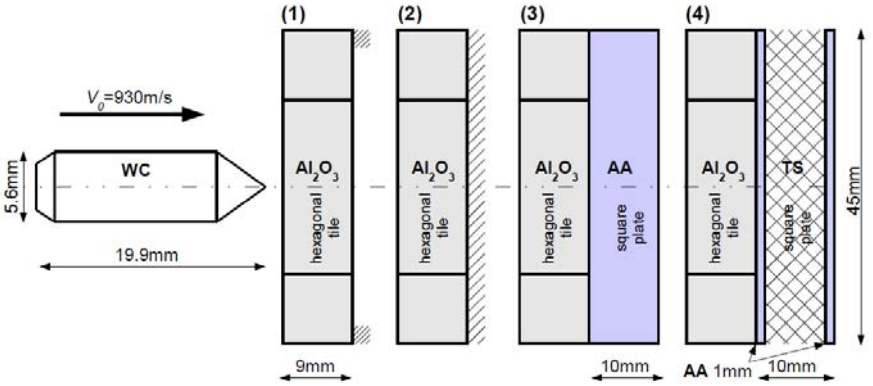


Figure 3: The dimensions of the projectile and the target configurations in the test problem.

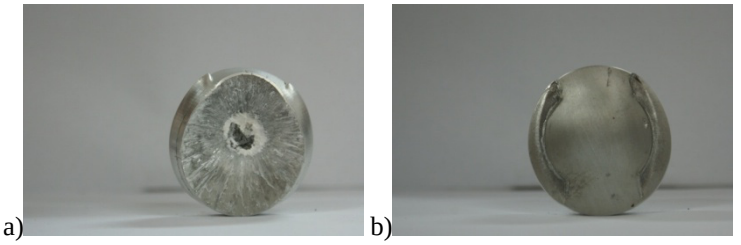


Figure 4: The result of the experimental impact test for the system of 10/10 mm thick Al_2O_3 ceramic tile backed by disc of PA11 aluminum alloy, (a) front, (b) back view. The projectile was the 7.62x51mm AP with WC core, initial velocity: 921 m/s.

3 Description of the numerical model

A method of computer simulation was applied to study the problem. The Finite Element Method (FEM) implemented in the LS-DYNA commercial code was used with explicit (central difference) time integration algorithm [3].

The full 3D models of the projectile and target were developed with a strain rate and temperature-dependent material constitutive relations. The Johnson-Cook constitutive model with Gruneisen Equation of State (EOS) was applied to describe the behavior of the metallic parts: aluminum alloy and projectile's core, table 2. However, the Johnson-Holmquist (JH2) model was used to describe the ceramic material Al_2O_3, table 1. Besides, the truss components were modeled as a bilinear elastic-plastic material with the typical elastic parameters for aluminum alloy (AA) or steel supplied with yield stress and hardening modulus: 350MPa, 7GPa and 1.25GPa, 20GPa respectively for AA and steel. Generally, the failure model based on the effective plastic strain was applied, but the JH2 describes the damage evolution and even completely failed ceramics. In this case, the given effective plastic threshold (150%) was used to limit the excessive finite elements deformations which can lead to numerical errors like negative volume.

The Boundary Conditions (BC) were defined by supporting the panel at its back edges at a distance of 4.5mm. It was numerically realized by frictional contact with a rigid body and a friction coefficient of 0.5 was assumed. The initial conditions were limited to the given initial projectile velocity, 930 m/s.

A spatial problem discretization was conformed to available computing resources. The projectile (WC core) mesh was built of the tetrahedronal elements with one integration point sized from 0.1mm at the sharp head to 0.5mm elsewhere. Similar mesh topology was selected for the hexagonal ceramic tile, but its density gradually grows in the direction of the location of the impact point. Characteristic single element length varies from 1 to 0.5 mm. The solid components of the backing plate were divided into constant stress brick finite elements with a typical size of 0.5mm. Finally, the numerical representation of the truss structure was built with an application of 1D beam elements. A total number of beam elements exceeded 500k with a single beam length of 0.1-0.15mm. Summarizing the mesh configuration of the studied problem, it comes up in one million elements.

Table 1: Johnson–Holmquist constitutive model constants for alumina [9].

Parameter	Units	High purity Al_2O_3
JH-2		[7]
ρ	kg/m³	3840
A		0.88
B		0.45
C		0.007
m		0.6
n		0.64
T	GPa	0.462
HEL	GPa	7.0
D_1		0.0125
D_2		0.7
FS		1,5
EOS		
k_1	GPa	231
k_2	GPa	-160
k_3	GPa	2774

Table 2: Johnson–Cook constitutive model, failure model and Gruneisen Equation of State data for WC and aluminum alloy materials [9].

Parameter	Units	93%WC6%Co	AA2024-T3
Johnson -Cook		[9]	[10]
A	GPa	3	0.369
B	GPa	89	0.684
C	-	0	0.0083
m	-	1	1.7
n	-	0.65	0.73
Gruneisen Equation of State			
c	m/s	5210	5328
S_1	-	1.14	1.338
S_2	-	0	0
S_3	-	0	0
Γ_0	-	1	2
a	-	0	0.48
Failure			
D_1 (JC)	-	0.03	0.5
PC (spall)	GPa	2.7	1.67

4 Analysis of the results

The truss type backing plate effectiveness was studied according to the assumptions mentioned earlier. The numerical results of the computer simulations of the impact test problem were mutually compared. The qualitative evaluation was performed by comparison of the projectile/target deformation at selected moment of time after impact. Figs. 5 and 6 present these results where the targets were crossed and half of them were removed from view for better interpretation. Two characteristic stages of the panel perforation process were chosen: 10 μs when the projectile reaches the middle surface of the ceramic tile – the most resisting stage during impact, and 20 μs when the ceramic layer is completely defeated and the projectile starts to perforate the backing plate – this stage describes the progressive loss of the panel's resisting properties. The careful examination of these images lets us form the conclusion that the backing effect is very important in overall ballistic panel effectiveness and the perfect stiff support gives the best result, solid backing plates are seemed to be more realistic, preserving quite an acceptable level of functionality. And finally, the truss based backing structures appear quite efficient but are unbeatable if compared with the surface mass density.

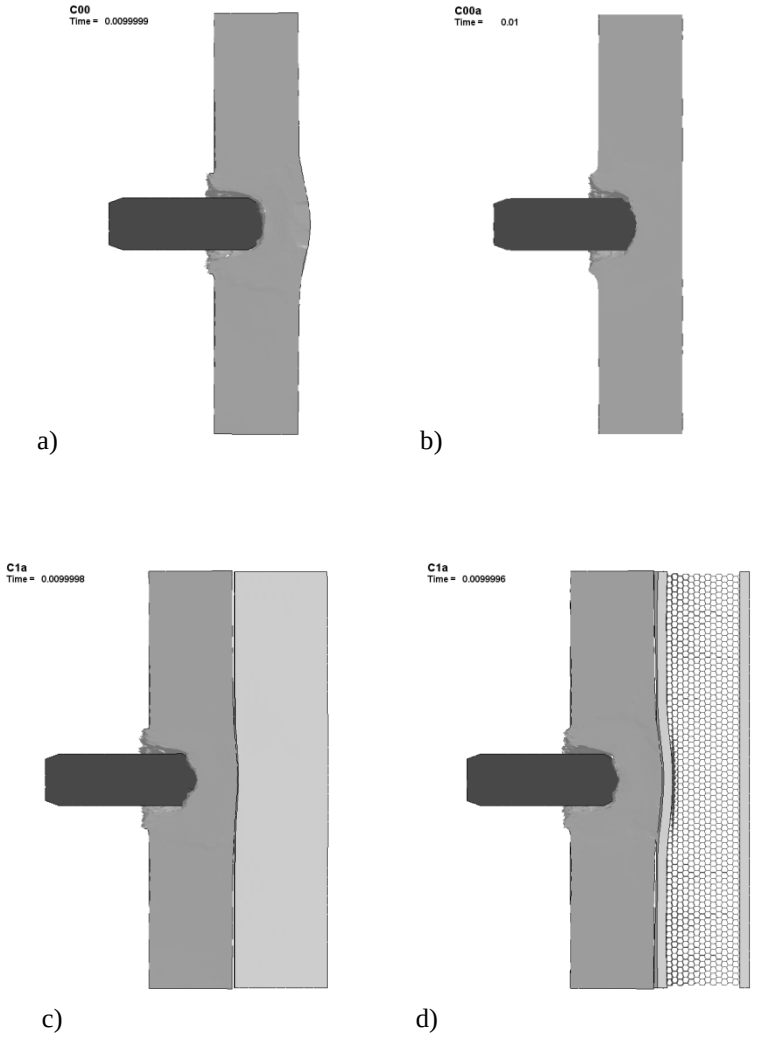


Figure 5: The images of the projectile/target deformation at 10 μs after impact for studied cases where the following were applied as the backing plate: (a) none, (b) rigid body, (c) solid AA, (d) truss structure.

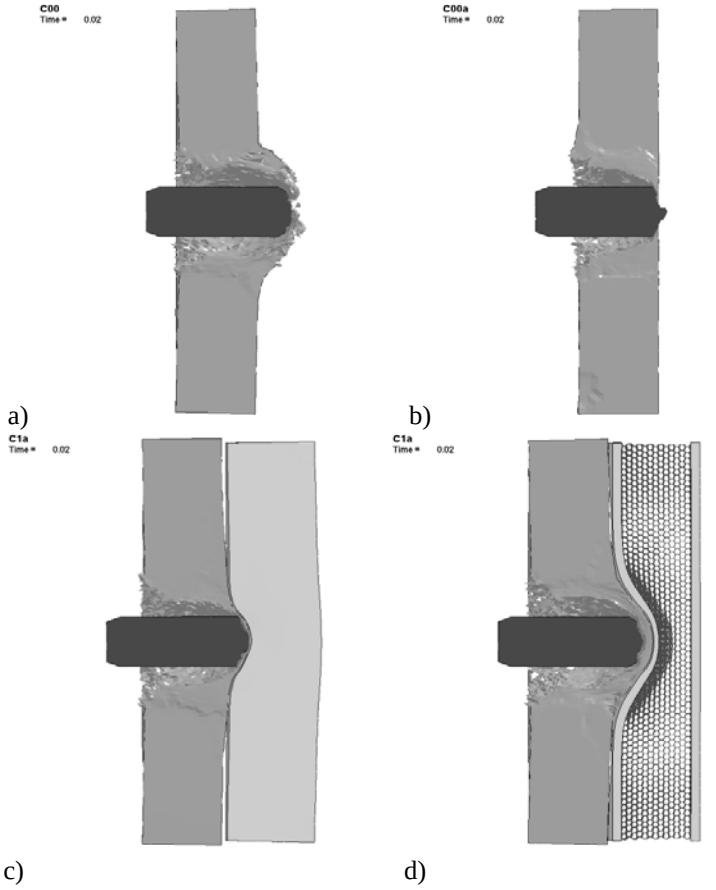


Figure 6: The images of the projectile/target deformation at 20 μs after impact for studied cases where the following were applied as the backing plate: (a) none, (b) rigid body, (c) solid AA, (d) truss structure.

The quantitative assessment was carried out by comparison of the specific problem parameters. These are the time history of the projectile kinetic energy (PKE), fig. 7, as well as the residual length of the projectile (RLP), table 3. The rigid support of the ceramic layer ensures the most efficient dissipation of the projectile kinetic energy and leads to its shortest residual length. Unfortunately, it is an unrealistic case and substitute solutions have to be considered as the solid or truss-type plates backing the ceramic tiles. Both of them mutually confronted proved sufficient and effective in providing the global panel resistance. Additionally, the mass effectiveness presented in table 3 emphasises the advantages of the truss-type backing solutions.

Table 3: The Residual Length of the Projectile (RLP) and the minimal identified PKE as functions of the single truss characteristics (L/D, material type).

Backing type:	**L/D**	**Panel surface density [kg/m^2]**	**Minimum PKE, *t*=50 μs [J]**	**Residual Length of the Projectile [mm]**
none	-	34.2	1721	16.5
rigid body	-	34.2	637	14.2
AA	-	61.2	390	16.0
AA truss structure	3	44.7	1159	16.3
	5	41.5	1360	16.5
	9	40.2	1443	16.5
steel truss structure	3	54.2	604	16.5

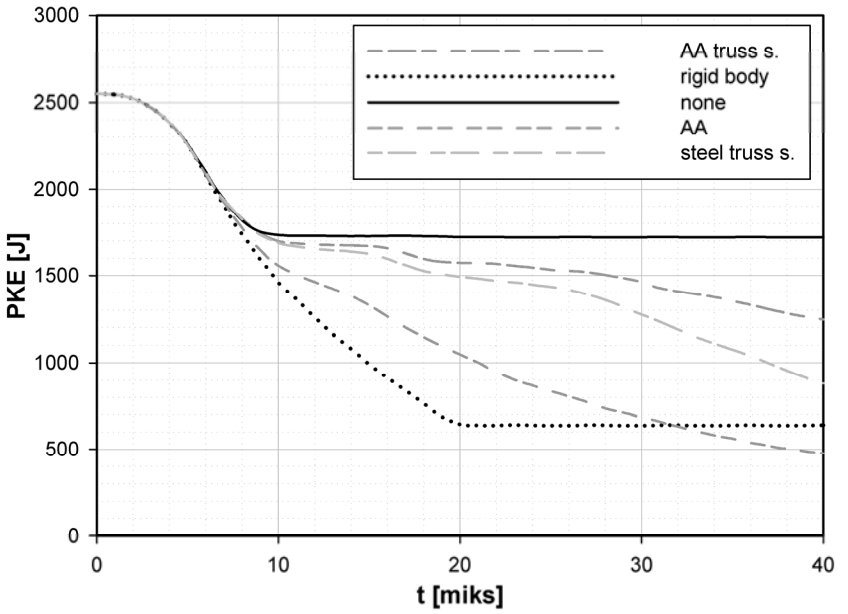


Figure 7: The time histories of the projectile kinetic energy (PKE) for studied cases.

5 Conclusions

The obtained results show that the satisfactory level of the ballistic protection may be accomplished by application of the truss type backing plates with preserving very low surface density of the panel. The hexagonal crystal lattice of

the lonsdaleite was used as an idea to build the studied truss-type plate successfully. It may be interesting to compare the efficiency of this structure with other concepts like honeycomb or Kelvin structure (tetradecahedron). The improvement in backing effectiveness was observed with scaling down the single truss ratio L/D. It was identified that the main deformation mechanism in truss-type backing components is buckling if the L/D ratio goes up. A virtual prototyping technique could be applied to manufacture the developed truss structure to confirm its properties in experimental impact tests.

Acknowledgement

The paper is supported by Grant No. O R00 0056 07, financed in year 2009-2011 by the Ministry of Science and Higher Education, Poland.

References

[1] Johnson, G.R., Cook, W.H., A Constitutive Model and Data for Metals Subjected to Large Strains, High Strain Rates and High Temperatures. *Proc. of the 7th International Symposium on Ballistics*: Hague, Netherlands, 1983.
[2] Kury, W.J., Breithaupt, D., Tarver, M.C., Detonation Waves in trinitrotoluene. *Shock Waves*, **9(2)**, pp. 227-237, 1999.
[3] Hallquist, J.O. (compiled), *LS-Dyna Theory Manual*, Livermore Software Technology Corporation (LSTC), 2006.
[4] Frondel, C., Marvin U.B., Lonsdaleite, a new hexagonal polymorph of diamond. *Nature*, **214**, pp. 587–589, 1967.
[5] The mineral and locality database, www.mindat.org/min-2431.html
[6] Pan, Z., Sun, H., Zhang, Y., Chen, Ch., Harder than Diamond: Superior Indentation Strength of Wurtzite BN and Lonsdaleite. *Physical Review Letters*, **102(102)**, 2009.
[7] Tasdemirci, A., Hall I.W., Numerical and experimental studies of damage generation in multi-layer composite materials at high strain rates. *International Journal of Impact Engineering*, **34**, pp. 189–204, 2007.
[8] *LS-DYNA Keyword User's Manual*, version 971, Livermore Software Technology Corporation (LSTC), 2007.
[9] Holmquist, T.J., Johnson G.R. & Gooch, W.A., *Modeling the 14.5 mm BS41 projectile for ballistic impact computations,* Computational Ballistics II, Transaction: Modelling and Simulation volume 40, pp. 61-75, 2005.
[10] Panov, V., Modelling of behaviour of metals at high strain rates, Cranfield University, PhD Thesis, 2005.
[11] Morka, A., Zduniak, B., Niezgoda, T., *Numerical Study of the Armour Structure for the V+ Ballistic Protection Level according to STANAG 4569*, transl. from Polish, Technical Report PBN 03-433/2010/WAT, AMZ-Kutno, Poland, 2010.

The influence of conical composite filling on energy absorption during the progressive fracture process

W. Barnat, T. Niezgoda, R. Panowicz & K. Sybilski
Department of Mechanics and Applied Computer Science, Military University of Technology, Poland

Abstract

Energy absorbing composite elements, which can be used for the protection of people, vehicles and structures, have been widely investigated in recent years. Different kinds of materials (metals, composites, foams) and structures have been investigated. Many articles present the glass, carbon and aramid composites investigation results. The sleeve and conical structures are described. The impact energy in these elements is absorbed during the progressive fracture process. Glass composites are very popular due to their low cost.

This paper deals with numerical and experimental research on the possibilities of energy absorption by a conical element made from a glass composite in comparison with a conical composite element filled with polyurethane foam. One element and a small composite panel are investigated and compared.

Keywords: blast wave, energy absorbing element and panel.

1 Introduction

The problem of energy dissipation by composite absorbing elements is considered from an aspect of the local stability loss or a progressive destruction [1, 2]. The work, carried out by the destruction of an energy-absorbing element, causes a significant limitation of shock load effects of the construction, for example an impact of the landing platform or an airship with the ground, or the influence of a pressure wave created by an explosion [3].

The greatest absorption energy in relation to a mass unit is possessed by the elements made of composite [4].

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110571

The aim of this work is the estimation of the influence of the used filling of composite elements on the absorbed energy by the energy-absorbing element in a form of a sleeve filled with a polyurethane foam. Preliminary, quasistatic examinations were carried out on the INSTRON machine, and then they were verified in testing ground conditions.

Cylinder composite elements are characterized by the higher swelling strength initiating destruction processes than the elements with a different geometry (e.g. scones) [5]. In case of these elements, a destruction initiator should be used to protect against catastrophic element destruction. The element destruction in this mode shows that the absorbed energy is considerably smaller than in the case of a progressive destruction.

2 Description of examined objects

In this work selected results from numerical-experimental research of following energy-absorbing objects were presented: a composite sleeve, a composite sleeve filled with a polyurethane foam, the system of three sleeves with a destruction initiator, the system of three filled sleeves with a destruction initiator. Additionally, in the work experimental results of loading an energy-absorbing panel, made out of sleeves, with a pressure wave originating from an explosion of an equivalent of 4 kg of TNT are presented.

In quasi static research the INSTRON testing machine was used. Examined elements were squeezed kinematically with a speed of 0.01 m/s. During the examination the displacement and the destructive force were recorded.
In the calculations the method of finite elements, implemented in MSC Dytran program, was used. The destruction process was described with the Hashin model.

Composite sleeves (fig. 1 a, d) with an internal diameter of 40 mm, height 50 mm and thickness of walls equal to 3 mm were made from a glass mat with following properties: E_{11}=18.5 GPa, ν_{12}=0.158, G_{12}=3.48 GPa, and next part of them (fig. 1b, c) were filled with an air-foam – foamed polyvinyl chloride PCHW-1. The foam was characterized by following material properties: ρ = 115 kg/m^3, E_r = 92.4 MPa, E_c = 80.6 MPa, R_m = 1.82 MPa, R_c = 0.911 MPa, ν = 0.26.

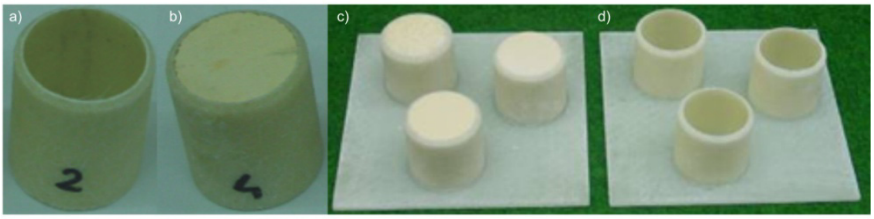


Figure 1: Composite sleeves investigated experimentally.

Their behaviour was described by the generalized Hooke's law [6], that can be used for the description of composite material properties. Physical (of the

material's model) and geometrical nonlinearities (great displacements and deformations) were taken into account. Calculations were made in the Dytran program containing an implementation of finite elements method.

The numerical model loading (figs. 2 and 3) was put into practice similarly as in carried out experiments by a kinematical input function. Analysed models were loaded by a stiff plate, described by a material of MATRIG type. The superficial contact type was used. It was defined between the bottom plate, energy-absorbing elements, and the striking plate.

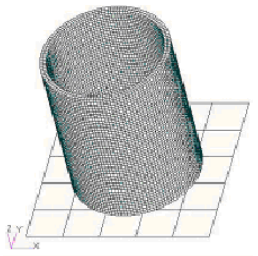


Figure 2: Numerical model of the sleeve.

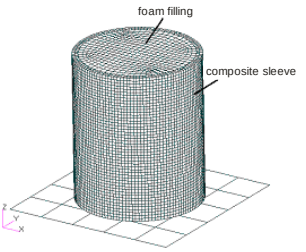


Figure 3: Numerical model of the sleeve with a polymer filler.

3 Research results for the composite sleeve

In figure 4, photos from experimental examinations undertaken are presented, and in figure 5 the change of destructive force value from the upper plate displacement in the contact zone is depicted. Analyzing this graph a really great stiffness of the examined object in the initial, linear-elastic zone can be noticed. Then the destruction of examined object was performed under a loading with upsetting force of the order of 23 kN.

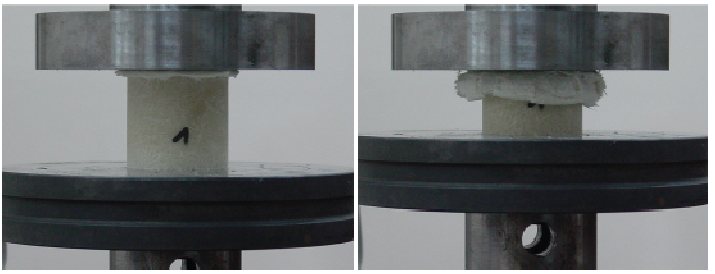


Figure 4: Failure mode of the composite sleeve.

For the same arrangement numerical simulations were made. Deformations of the composite sleeve numerical model are shown in fig. 6. The relative absorption energy was estimated to be about 42 kJ/kg.

A sleeve, subjected to the experimental examinations, behaved similarly as the numerical model. The way of the composite sleeve destruction was shown in fig. 6 and in fig. 7 the change of a value of a destructive force in dependence on the upper plate displacement.

Upsetting force graphs, acquired numerically and from experiments, have the same nature, which testifies the correctness of the model's assumptions and the model's material selection.

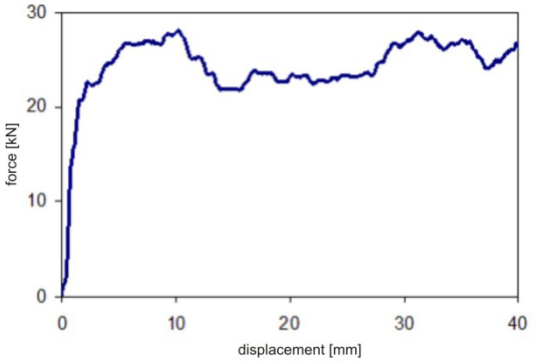


Figure 5: Experimentally obtained compressive force diagram for the composite sleeve.

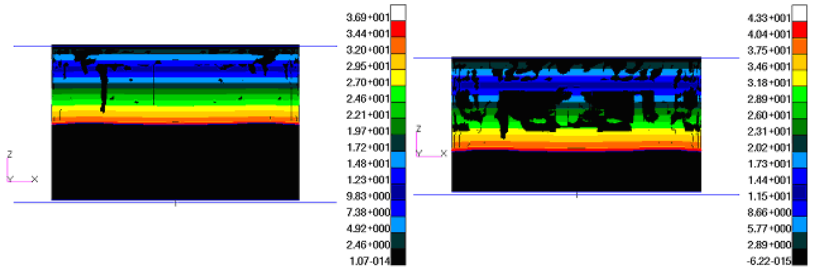


Figure 6: Deformation mode of the composite sleeve numerical model.

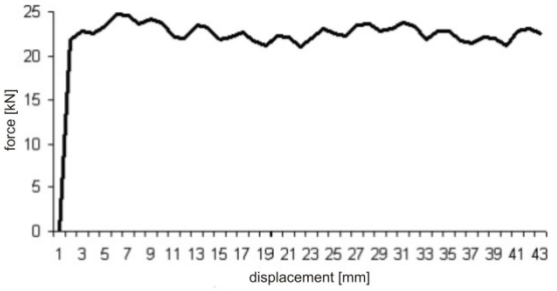


Figure 7: Numerically obtained compressive force diagram for the composite sleeve.

4 Research results for the composite sleeve with filling

As a result of the carried out numerical experiment, a graph of upsetting forces in relation to the displacement was acquired (fig. 8). The use of a filler caused the composite sleeve to be destroyed in a different way than in the former case. The filler precludes the sleeve's burring inside of the model. Upsetting force graph has got a complex character. On the initial part of the graph a clear leap of the upsetting force value up to circa 37 kN is seen. Next its value (thanks to the catastrophic sleeve destruction) falls down and attains the value of 7.5 kN. In the next stage of the load the upsetting force value increases. The numerical model of the composite sleeve is being destroyed in a similar was as the real object. The deformation manner of the composite sleeve is presented in fig. 9.

Based on the numerical calculations it was found that the average upsetting force amounts to 20 kN. After taking into account the route that is passed by stiff plates, the upsetting force work was equal to 800 J. After taking into account the examined element's mass it was found that the relative absorption energy for the composite sleeve with a filling amounts to 24.4 kJ/kg.

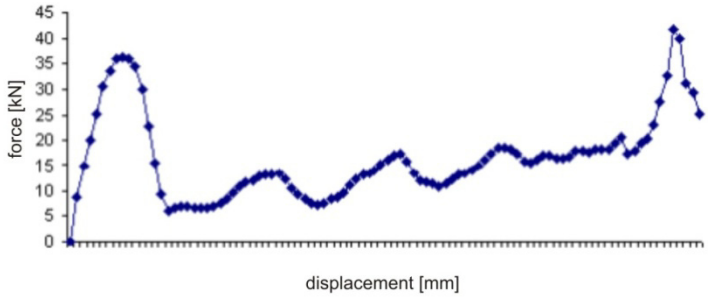


Figure 8: Numerically obtained compressive force diagram for the composite sleeve with a polymer filler.

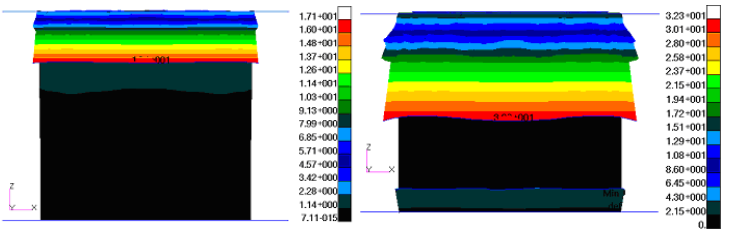


Figure 9: Deformation mode of the numerical model of the composite sleeve with a polymer filler.

The real object behaves similarly as the numerical models. The destruction force graph is presented in fig. 10. In the first stage, the progressive destruction

of few upper layers took place. Next catastrophic destruction of the sample's part was recorded, characterized by the fall of the loading force and the appearance of a crack on the sample, outside the contact zone. Then the upsetting force is rising gently.

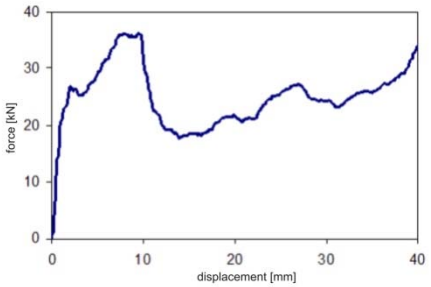


Figure 10: F(Δl) graph from the epoxide composite tube reinforced with a glass mat, filled with foam compression.

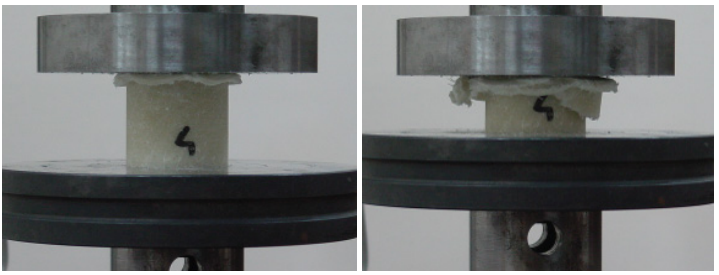


Figure 11: Deformation mode of the numerical model of the composite sleeve with a polymer filler.

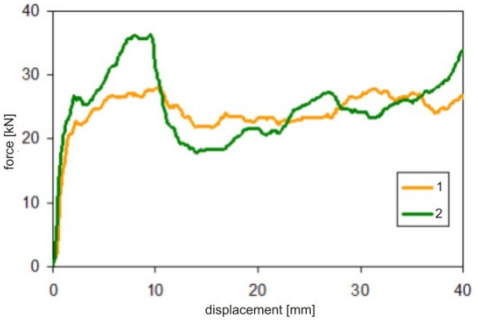


Figure 12: Comparison of vertical reaction value increase graphs obtained experimentally 1 – sleeve 2 – sleeve with filling.

The matching of the most important data is presented in Table 1.

The use of a filler causes the increase of the absorption energy. The use of a filler also has got its disadvantages. When poorly selected it can cause a deterioration of the element's energy-absorbing properties. In the case of a sleeve with a foam filling a necessity to increase the minimal thickness of walls from circa 2 up to circa 3 mm exists. Too thin wall doesn't withstand the foam's pressure and yields to catastrophic destruction (the sleeve's crosswise cracking). Whereas a too thick wall can preclude the destruction of the energy-absorbing element, precluding at the same time the stroke's energy absorption.

Table 1: Comparison of obtained results.

Sample type	Thickness [mm]	D_{w2} [mm]	H [mm]	P_{max} [kN]	$P_{śr}$ [kN]	$\frac{P_{śr}}{P_{max}}$	EA [J]	WEA [kJ/kg]	M [g]
Sleeve	3	39.3	50	28.1	23,1	0.822	1155	44.3	26.1
Sleeve with a filling	3	39.3	50	36.3	25	0.689	1250	37.7	33.2

5 Energy-absorbing panels examination – sleeves

Results obtained from the experimental examinations for hollow sleeves are presented in figs. 13 and 14. The first figure presents the process of an element's destruction, whereas the second depicts the upsetting force graph. The gentle force rising caused by the use of cracking initiators (in a form of phases) for composite tubes is characteristic for the graph. An average upsetting force was 74.3 kN, at the mass of a set amounting to 66.1 g.

The sleeves were being destroyed in a progressive way, mainly through the material's delamination.

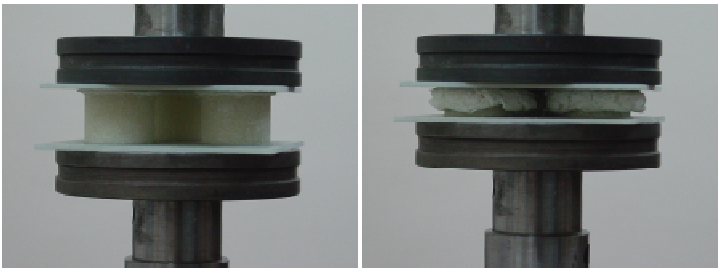


Figure 13: Energy-absorbing panel way of destruction in a form of three composite sleeves.

Absorption energy amounted to 2972 kJ, and the relative absorption energy amounted to 45 kJ/kg.

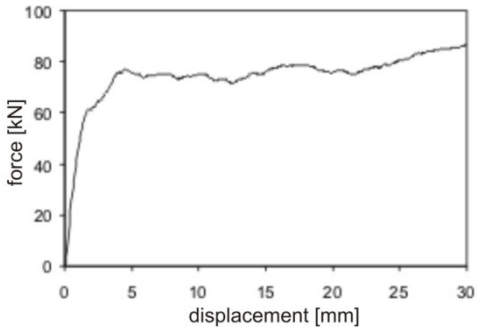


Figure 14: P-Δl dependences for the third energy-absorbing object.

The final character of destruction is presented in fig. 15.

Figure 15: Energy-absorbing panel way of destruction in a form of three composite sleeves.

6 Energy-absorbing panels examination – sleeves with filling

Results obtained from the experimental examinations are presented in fig. 16. An air-foam caused a limitation of free delamination of composite elements of energy-absorbing sleeves, which resulted in their catastrophic destruction [7]. With the use of thick-walled elements (wall thickness 3 mm) progressive destruction of the element was also acquired but not in all cases. A part of the sleeves with filling examinations ended in the element's catastrophic destruction as presented on the right side of the fig. 16. Most likely it is connected with a great dispersion of mechanical properties of composite elements. To avoid this effect completely, in case of elements filled with foam, the walls thickness should be further increased.

An upsetting force character is presented in fig. 17. Characteristic for the graph is the appearance of the force refraction for the 5 mm displacement. Such force drop is probably caused by catastrophic destruction of one of the tubes (fig. 18). The mixed destruction process of energy-absorbing elements and the limitation of the possibility of delamination of energy-absorbing elements by the air-foam caused an increase in the upsetting force value. The protective panel

was destroyed with an average force of the order of 90 kN. The maximum force that was measured amounted to, in this case, 105.7 kN. In comparison with the former examinations of an object the upsetting force value was 1.13 times greater.

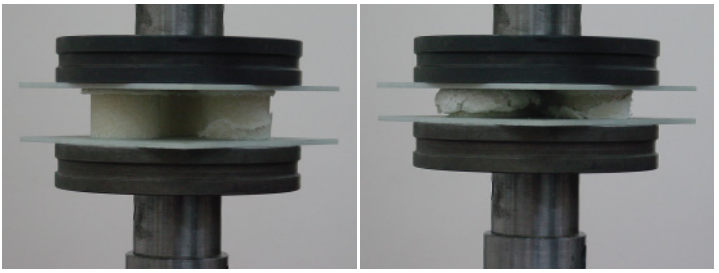


Figure 16: The way of energy-absorbing panel destruction in a form of three composite sleeves filled with air-foam.

The absorption energy, for this case, amounted to 3600 kJ, whereas relative absorption energy - 42,6 kJ/kg. The statement of the results for panels is presented in Table 2.

Table 2: Comparison of obtained results.

Structure	Thickness crank [mm]	Dw2 [mm]	H [mm]	P_{max} [kN]	$P_{śr}$ [kN]	$\frac{P_{śr}}{P_{max}}$	EA [J]	WEA [kJ/kg]	M [g]
3 Sleeves	3	39.3	40	86.7	74.3	0.857	2972	45.0	66.1
3 Sleeves with filling	3	39.3	40	105.7	90	0.851	3600	42.6	84.5

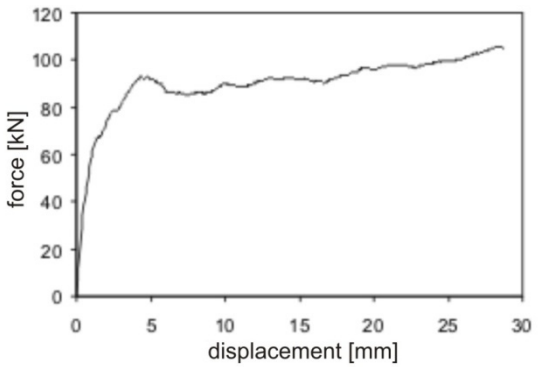


Figure 17: P-Δl dependence for the second energy-absorbing object.

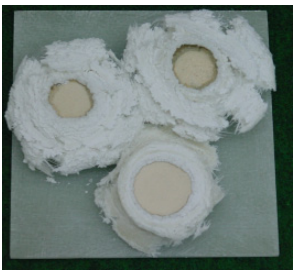


Figure 18: The way of energy-absorbing panel destruction in a form of three composite sleeves with a foam filling.

7 Testing ground examinations

Examinations presented above showed the usability of the application of the filling in energy-absorbing elements. That is why testing ground examinations for the panel, presented in fig. 19, with the dimensions 750 x 750 mm, were made, composed out of 121 composite sleeves. The panel was loaded with a blast wave originating from an explosion of an equivalent of 4 kg of TNT. The charge was positioned centrally over the panel at the distance of 40 cm from the panel's top surface.

Comparison examinations were made of a system with a steel plate and the energy-absorbing panel made out of sleeves. After the experiment (fig. 19), in both considered cases, the protected surface and the surface loaded with the pressure wave were scanned. For this task a 3D scanner and software were used. The comparison of deformations is presented in figs. 20 and 21.

The deformation of upper plate, loaded by the blast wave is visible (fig. 20) in the sleeves fastening places and the differences in the final deformation of the considered systems (figs. 20 and 21). In the case of a plate protected for the panel with the sleeves, maximum durable deflection amounts to 15.4 mm, and for the second case – 66.3 mm. This means decreasing of the maximal durable deflection over 4 times in case of an element, where the sleeves were used.

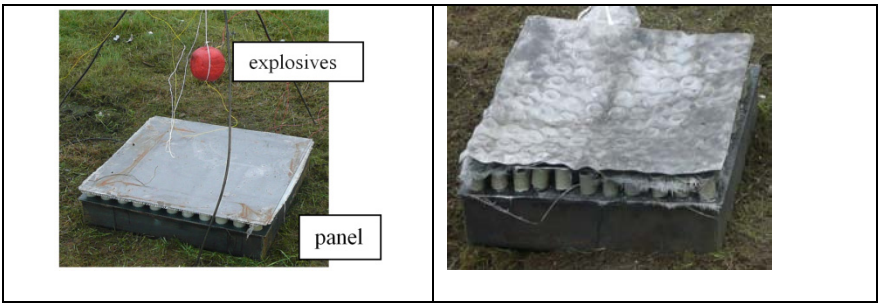


Figure 19: Energy-absorbing panel before and after the experiment.

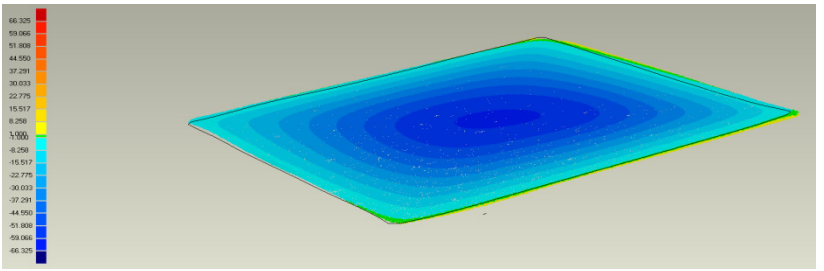


Figure 20: Steel plate deformation burdened with a pressure wave originating from a detonation of an equivalent of 4 kg of TNT.

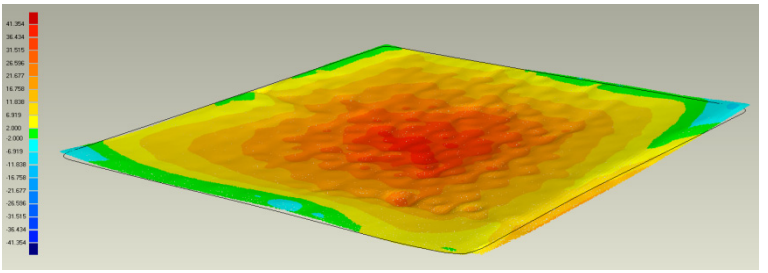


Figure 21: Durable deformation of the plate burdened with a blast wave originating from the detonation.

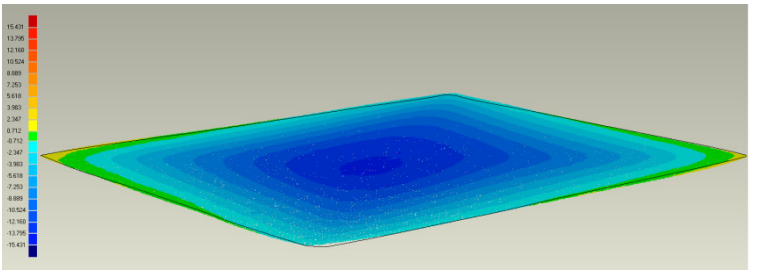


Figure 22: Durable deformation of the protected plate.

8 Conclusions

In the work results of numerical examinations with an experimental verification of two energy-absorbing objects are presented. Quasistatic examinations were carried out on the testing machine INSTRON, and dynamical tests on testing ground with the use of explosive materials.

The use of a filling in a form of a foam caused an increase in the absorption energy at the relatively low increase of the protective panel mass. This is quite

essential in situations, where we have to deal with a limited amount of space, and the energy-absorbing element's mass isn't the main limitation of the designed construction.

Based on the preliminary evaluation of acquired results it was found out that adequate selection of the filler's material should further increase absorbed by the elements energy needed to destroy the examined construction. The authors plan to use for this purpose optimization procedures connected with programs for calculations with finite elements method.

References

[1] Niezgoda T., Barnat W., *Numeryczna Analiza Wpływu Kształtu Podstawowych Struktur Kompozytowych na Energię Zniszczenia*, III Sympozjum Mechaniki Zniszczenia Materiałów i Konstrukcji Augustów, 1 – 4 czerwca 2005.

[2] Nagel G., Thambiratnam D. (2003), *Use of thin-walled frusta energy absorbers in protection of structures under impact loading.* Design and Analysis of Protective Structures against impart/Impulsive/Shock Loads, Queensland.

[3] Niezgoda T., Barnat W.: Analiza pasa bariery drogowej wzmocnionej elementami kompozytowymi w zastosowaniu do poprawy energochłonności elementów infrastruktury Górnictwo Odkrywkowe 5-6/2006.

[4] Barnat W., Niezgoda T.,: *Badania energochłonności elementów podatnych w aspekcie zastosowanych materiałów* Journal of Kones Powertrain and Transport vol 14 No 1/2007.

[5] Niezgoda T., Barnat W., *Numeryczna Analiza Wpływu Kształtu Podstawowych Struktur Kompozytowych na Energię Zniszczenia*, III Sympozjum Mechaniki Zniszczenia Materiałów i Konstrukcji Augustów, 1 – 4 czerwca 2005.

[6] MSC Dytran, Example Problem Manual, Version 3.0 MSC 1996.

[7] Niezgoda T., Ochelski S., Barnat W. Doświadczalne badanie wpływu rodzaju wypełnienia podstawowych struktur kompozytowych na energię zniszczenia, Acta mechanica et automatica 1/2007.

Section 10
Railway transport

This page intentionally left blank

Stochastic traffic generator for Monte Carlo load flow simulation

E. Pilo de la Fuente[1,2], J. Jiménez-Octavio[1], R. Rodriguez-Pecharroman[1] & A. Lopez Lopez[1]
[1]*Universidad Pontificia Comillas de Madrid, Spain*
[2]*Multitest09, Madrid, Spain*

Abstract

This paper presents a stochastic traffic generator in which (i) the driving manner is assumed to be stochastic, (ii) random unexpected stops are considered and (iii) the stop time in the stations is also supposed to be stochastic. This simulation tool provides a large number of different driving scenarios to be used for a Monte Carlo load flow analysis. Traffic simulations algorithms are described in detail. Reference to a section of the Madrid–Barcelona high-speed line illustrates the potential of such a tool.

Keywords: power supply, train driving, stochastic simulation.

1 Introduction

One of the inputs required for designing the electrification in a railway system is the power to be supplied to the rolling stock in each time step, which is in itself closely related to the considered operation. Accordingly, compared to other power systems, railways are quite different: power consumptions are moving loads and change very quickly. The power consumption depends on the manner in which the trains are driven (the speed along the trip, how fast it accelerates or it brakes, etc.), the characteristics of the train (mass, running resistance, motors, etc.) and the path (slopes, speed limits, etc). Due to its complexity, computation of instant power consumption of the trains is normally performed by software tools.

When traction power consumptions have to be estimated, the characteristics of the train and the path are typically well defined. However, rail lines are normally not equipped with automatic driving systems and each driver drives his

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110581

own way (within mandatory safety limitations). Thus, in theory, there is an infinite number of possible ways of driving, with very different power consumption spatial distributions. The driving type in which the trip is done as fast as possible, considering both the infrastructure speed and the train operational limits, is known as minimal time driving (MTD) [1].

When performing power supply system design, MTD is normally considered because: (i) it is very simple to define the manner in which trains are driven and (ii) it is mistakenly believed to be the most demanding driving mode.

This paper discusses why MTD mode does not correspond to the worst driving mode for the railways electrification and presents a simulation algorithm which combines MTD simulations and stochasticity in the operation. The final purpose of this simulation is to carry out stochastic analysis of power supply systems.

Section 2 describes the power supply system, its design and why MTD does not necessarily correspond to the worst operating conditions. Section 3 presents the simulation procedure that is proposed in this paper. Section 4 presents some results obtained with this simulation tool. Finally, Section 5 summarizes the conclusions of this work.

2 Power supply system

2.1 General structure

Figure 1 shows the general structure of power-supply systems of AC electrified railways, when they are fed from a three-phase network. Even if this paper is focused on this case, similar conclusion can be reached in other systems.

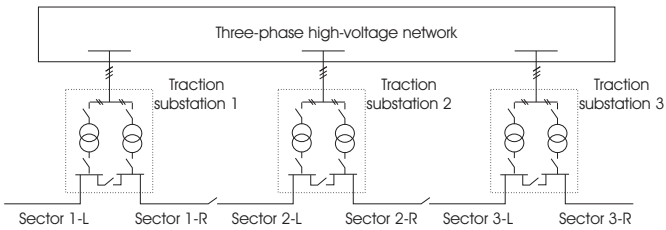


Figure 1: Structure of the power supply system.

As shown, the electrical system is divided in electrically-isolated sectors, which are fed from the three-phase network through a traction substation [2].

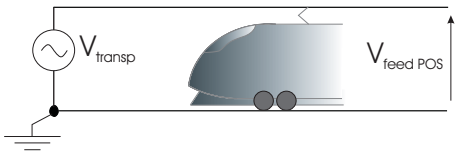


Figure 2: Mono-voltage system configuration.

These substations are connected between two of the three phases of the high-voltage network. Each of these sectors can use either mono-voltage system (1x25kV) or bi-voltage system (2x25kV). In mono-voltage system, the feeding conductors are set to the specified voltage level (see Figure 2).

In bi-voltage systems, a higher voltage is set between feeding conductors [3, 4]. This voltage is reduced by using autotransformers distributed along the catenary (see Figure 3). In these systems, the term cell normally refers to the portion of catenary located between two consecutive autotransformers. Typical values for cell lengths are 10–15 km.

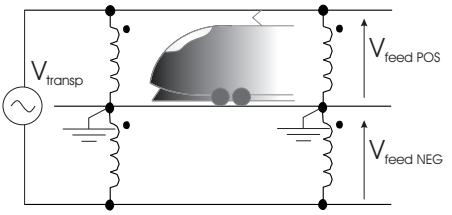


Figure 3: Bi-voltage system configuration.

It has been proven that bi-voltage systems can be represented as if they were mo-voltage [5, 6]. For that reason, in this paper only mono-voltage systems are considered in the discussions, without loss of generality. Actually, as this paper is focused on driving simulation, the power supply system considered for the discussions is not a critical aspect.

2.2 Electro-technical dimensioning criteria and driving mode

Railways power supply systems are dimensioned to be able to supply the power required by the trains, usually assuming a long-term estimation of the traffic needs. From an electro-technical point of view, the power supply system is considered to be able to supply the required power if the following restrictions are fulfilled:

- Voltage in the catenary has to be within the range specified by UIC-600 standard, which specifies the upper and lower limits depending of its duration.
- Currents circulating along the catenary and through transformers (and autotransformers) have to be lower than the rated values, in order to ensure no overheating will occur. These limits are often expressed as power limits instead of current limits.
- Power supplied by the three-phase network through the traction substations is also frequently limited when the network is too weak, in order in order to ensure a proper operation of the network.

Figure 4 shows the train power consumptions of two different driving modes as a function of the train position. In the figure, driving mode 1 and 2 could correspond respectively to a MTD and to a less aggressive driving (lower accelerations and somehow lower speeds). Electrical sections have been

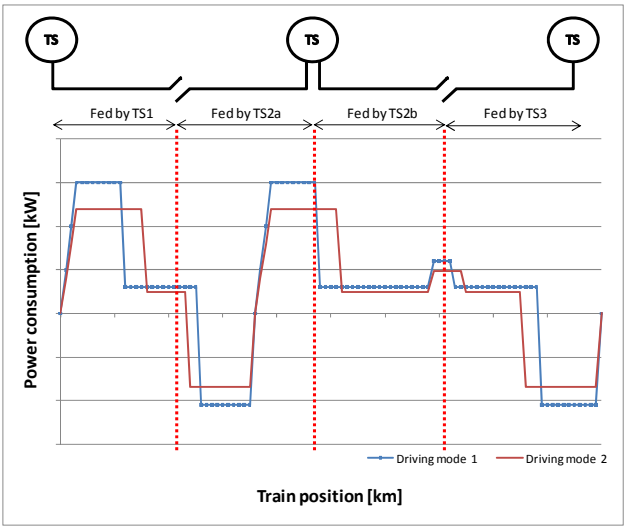


Figure 4: Power consumption comparison of two different driving modes.

represented and each of them is fed with one power transformer (TS1, TS2a, TS2b and TS3), as they are electrically isolated.

The voltage $\mathbf{V}(t,ts)$ at train t pantograph, fed from a substation transformer ts can be calculated using Eq. (1):

$$\mathbf{V}(t,ts) = \mathbf{V}_{no-load}(ts) - \sum_{i \in ts} \mathbf{Z}(ts) \cdot \mathbf{I}(i) - \sum_{i \in ts} \mathbf{Z}'_{CAT} \cdot \mathbf{I}(i) \cdot D(i,ts) \quad (1)$$

where $\mathbf{V}_{no-load}(ts)$ is the non-load voltage in the substation transformer ts, $\mathbf{Z}(ts)$ is the Thevenin equivalent impedance of substation transformer ts, $\mathbf{I}(i)$ is the current consumed by a train i fed by substation transformer ts, $\mathbf{Z}'_{CAT}$ is the per length unit impedance of the catenary and $D(i,ts)$ is the distance between train t and substation transformer .

As shown in Eq. (1), the second term is proportional to currents (which are proportional to power consumptions) but the third one is proportional both to currents and to distances to the substation. When considering the driving modes represented in Figure 4, it cannot be established that MTD (driving mode 1) leads to lower voltages than driving mode 2, even if peak power consumptions are higher in MTD, because distances from power peaks to substations are just different in both cases.

As the topology of each electrical section is radial, the maximum instant current circulating in the catenary $\mathbf{I}_{C,\max}(ts)$ corresponds normally to the total current supplied by the substation, which is proportional to the supplied power as

expressed in Eq. (2) (even if exceptions to that rule may occur when regenerated power is high).

$$\mathbf{I}_{C,\max}\left(ts\right)=\sum_{i\in ts}\frac{\mathbf{S}^{*}\left(i\right)}{\mathbf{V}^{*}\left(i,ts\right)} \tag{2}$$

In order to evaluate if currents are under the rated values, normally quadratic mean currents are calculated from the instant values.

Also for currents, MTD do not necessarily correspond to the higher values due to two different effects. First, if time is kept constant between two consecutive trains, the distance between two consecutive trains is always higher in MTD than in other driving modes. In other words, the same substation may have to supply power to eventually more trains. Second, as power consumption spatial distributions differ from one driving mode to other, the substations that supply the power (and thus the current) may be different.

Finally, the instant power $\mathbf{S}_{total}\left(ts\right)$ supplied by the substations transformer ts is calculated in Eq. (3)

$$\mathbf{S}_{total}\left(ts\right)=\mathbf{S}_{loss}\left(ts\right)+\sum_{i\in ts}\mathbf{S}\left(i\right) \tag{3}$$

where $\mathbf{S}_{loss}\left(ts\right)$ represents the power losses supplied by the substation transformer ts and includes losses in the catenary, in the transformer itself and in the three-phase network.

For the same reasons mentioned for currents, MTD is not necessarily the most demanding driving mode for the power supply limits of each substation.

3 Simulation procedure

3.1 Overview

In order to design the power supply system of a railway network the traffic (including train movement) and the power supply are often supposed to be uncoupled. The design is then done iteratively in two different steps, as shown in Figure 5.

Firstly the traffic is simulated for each discrete time step and a traffic scenario is obtained, usually assuming MTD. A traffic scenario is composed by the list of locations and power consumptions of each train at each time step. Then an electrical simulator, typically solving a power flow for each time step, is used to determine all the voltages, currents and power flows and to determine if the power supply works properly. The design is modified and the electrical simulation is repeated until the power supply works properly.

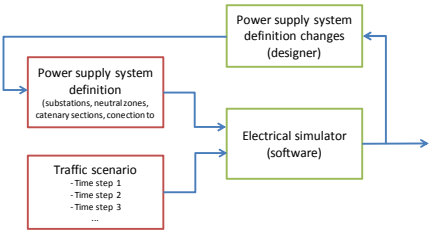


Figure 5: Power supply design process considering deterministic traffic scenario.

Once determined that MTD is not necessarily the most demanding driving mode (neither for voltages, nor for currents, nor for power flows), it is proposed to modify the design process so that the power supply is evaluated with a set of stochastic scenarios (see Figure 6).

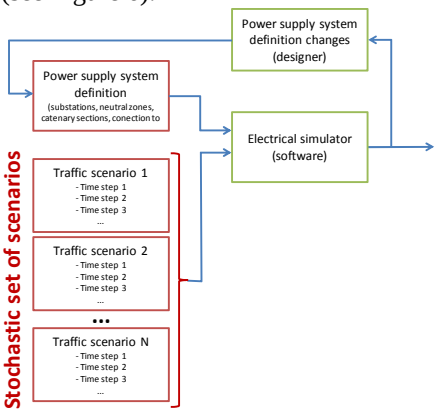


Figure 6: Power supply design process considering stochastic traffic scenarios.

The stochastic traffic generator presented in this paper is intended to provide such a set of scenarios to run a Monte Carlo load flow. To produce these scenarios, well-known MTD simulations have been enriched with the following add-ins: (i) stochastic speed profiles, (ii) random speed reductions and (iii) stochastic stop times.

3.2 Stochastic traffic generation

Traffic generation is usually performed in two steps, which are described in the following parts of this section: (i) stochastic minimal time single train driving (single train movement and power consumption calculations) and (ii) stochastic traffic mesh generation.

3.2.1 Stochastic minimal time driving

The stochastic MTD driving simulation calculates the movement of the train between two consecutive stops (position, speed and acceleration) as well as the power consumption o regeneration, for each time step. To determine the movement, Newton's second law (see Eq. (4)) is integrated:

$$F_t - F_w - F_c - F_r = kM\frac{dv}{dt} \tag{4}$$

where F_t is the traction force, F_w is the resistant force due to the slope of the railway line, F_c is the resistant force due to the curvature of the railway line, F_r is the running resistance force, M is the mass of the train, k is a factor that is commonly used to consider the rotating inertia, and v is the speed of the train.

The resistance force due to the slope of the line is proportional to the mass of the train and the slope of the line. The resistance force due to the curvature is also proportional to the mass of the train and inverse to the curvature radius. The running resistance force is normally approximated by a quadratic expression function of the speed. Finally, the traction force F_t is decided for each time step in order to fulfil the speed limits (rolling stock limits and infrastructure limits), maximum acceleration/deceleration or other constraints defining the driving mode.

Once the movement of the train has been solved, the electrical consumed power P_e and regenerated power P_r is determined using Eqs. (5) and (6) respectively:

$$P_e = \frac{F_t \cdot v}{\eta_{e,m}} + P_{anc} \tag{5}$$

$$P_r = F_t \cdot v \cdot \eta_{m,e} + P_{anc} \tag{6}$$

where $\eta_{e,m}$ is the efficiency of the electrical to mechanical power conversion, $\eta_{m,e}$ is the efficiency of the mechanical to electrical power conversion and P_{anc} is the power consumed by ancillary services (air conditioning, pressure equipment, lights, etc.).

The simulation of each train movement between two consecutive stations is done assuming that starting time and starting station are the temporal and spatial origins. For that reason, once all the train movements have been calculated, the traffic mesh has to be built-up by shifting the individual movements to their space and time origin. To assign time origins, train frequencies and stop times in the stations have to be specified. Once the traffic mesh has been constructed, for each time step all the active trains are identified and the traffic scenario can be constructed.

As discussed previously, actual traffic can be very different depending on the driving, on the traffic contingencies or even on the number of passengers. For that reason, the following aspects have been modelled:

- Depending on the operational conditions, train driving can be faster or slower, but maximum speed limits have always to be observed. Thus, the speed limit of each section has been assumed to follow a normal probability distribution, which is truncated at the maximum speed limit (see Figure 7).
- In dense traffic situations, due to track occupancy some trains may interfere with other trains, by forcing them to reduce their speed or even making them to stop. In many cases these situations can be identified and eliminated when signalling is designed, but when power supply is being designed no information is normally available. To model this kind of situations, in some cases (randomly selected) an additional speed reduction is added (see Figure 7): their position is assumed to be uniformly distributed along the line and their established speed has been assumed to take discrete values.

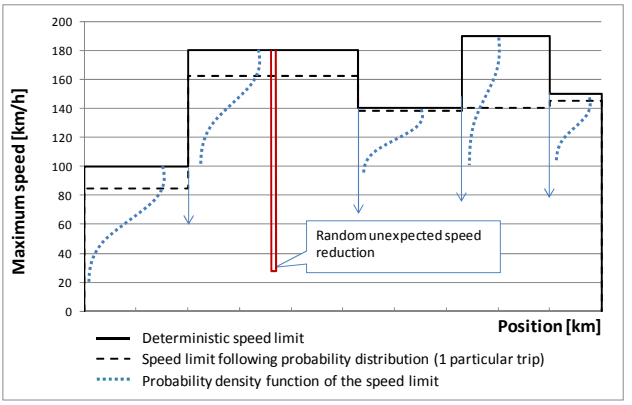


Figure 7: Stochastic speed limits for a single train.

- While in peak hours some stations can be overcrowded and stop time higher, in off-peak hours stop time can be reduced. To model this, the stop time has been considered to follow a normal probability distribution (see Figure 8), whose mean is the rated time stop, truncated in order to have always positive times.

In all these aspects, an additional work has to be done to analyze other probability distributions and to compare them with reality.

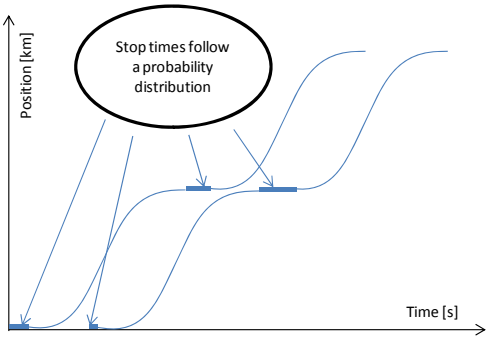


Figure 8: Stochastic stop times when building up the traffic mesh.

4 Study case

In order to show the potential of the presented stochastic traffic generator, a 436 km long section (from Madrid to Lérida) of the Spanish high-speed line Madrid–Barcelona has been used (its cross-section is shown in Figure 9).

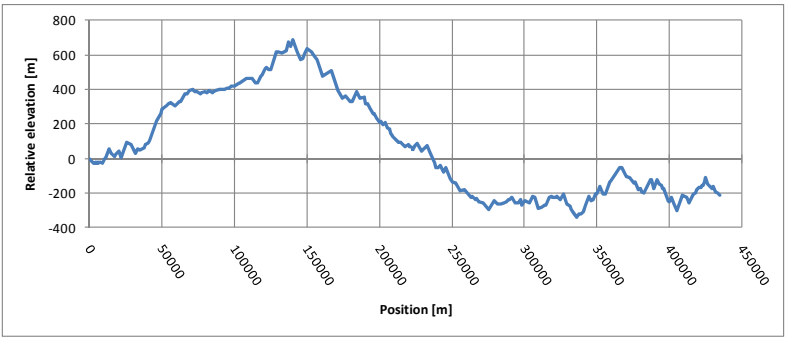


Figure 9: Cross-section of considered part of high-speed line Madrid–Barcelona.

Generic high-speed trains (410 t heavy, 168 meter long, 850 kW ancillary services power consumption, without regeneration, 300 km/h maximum speed, 280 kN maximum traction force) have been considered. 3 min intervals have been presumed.

Figure 10 shows, for every position, the power that should be supplied by a substation to a standard section 35 km long. It has been calculated as the sum of the electrical power required by all the trains within a 35km moving window, assuming MTD. This kind of graphics gives a good idea about how does power consumption accumulate along the railway and which parts of the line are more demanding.

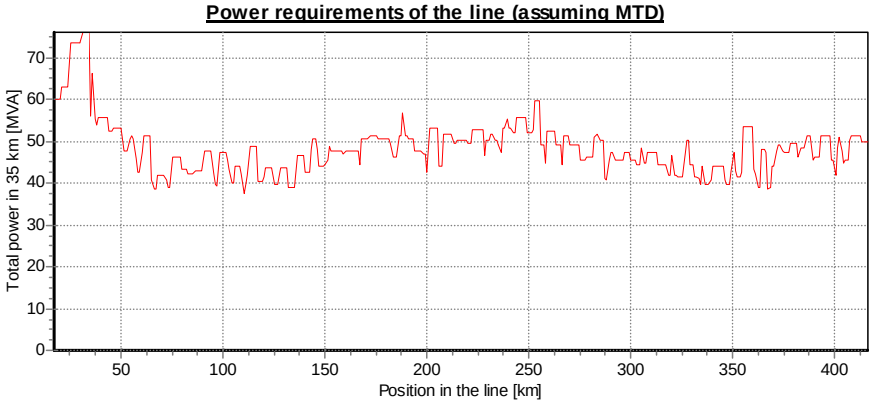


Figure 10: Power requirements considering MTD traffic.

Figure 11 shows the same kind of graphics, but obtained from one of the many scenarios obtained by the stochastic traffic generator. Actual speed limits have been taken as mean values of the normal distributions, with a typical 10% standard deviation. 3 minutes stop times have been consider, with 30 s standard deviation. Finally, 20% steps have been considered for unexpected stops, which occur with a 5% probability.

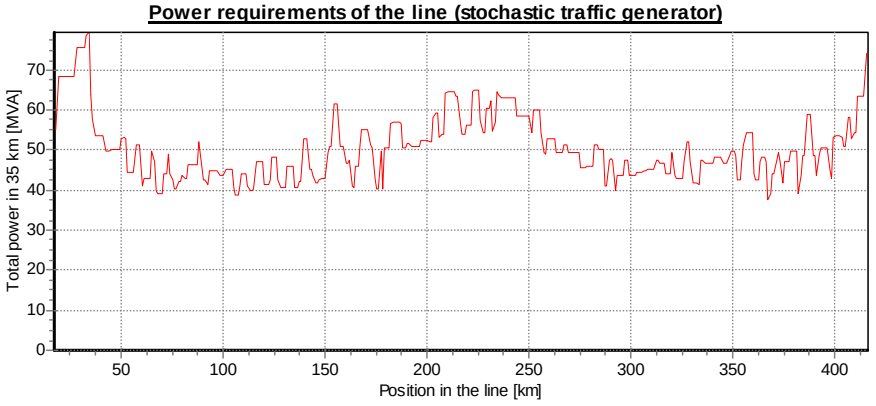


Figure 11: Power requirements considering stochastic traffic (one of the scenarios).

Comparison of Figures 10 and 11 shows that power requirement in both cases are slightly different. Therefore, considering several operation types would be very enriching when carrying out the power supply design.

5 Conclusions

This paper has presented a stochastic traffic generator in which (i) the driving manner is assumed to be stochastic, (ii) random unexpected stops are considered and (iii) the stop time in the stations is also supposed to be stochastic. This simulation tool provides a large number of different driving scenarios to be used for a Monte Carlo load flow analysis for the design of the power supply system.

This stochastic traffic generator has been used to generate traffic scenarios for a section of the Spanish high-speed line Madrid–Barcelona. The power requirement analysis shows that using this kind of tools could lead to more accurate power supply system designs, as more operating conditions can be analyzed.

Future developments should include refining probability distribution functions and validating them with measurements. In addition, the integration of this tool with a power flow tool would ease the penetration of such tools in the industry.

References

[1] P. Lukaszewicz, "Energy Consumption and Running Time for Trains: modelling of running resistance and driver behaviour based on full scale testing" PhD Thesis. Royal Institute of Technology (KTH). Stockholm, 2001

[2] E. Pilo, L. Rouco, A. Fernández and A. Hernández-Velilla, "A simulation tool for the design of the electrical supply system of high-speed railway lines," in IEEE PES Summer Meeting 2000, Seattle (USA)

[3] P. H. Hsi, S. L. Chen and R. J. Li, "Simulating on-line dynamic voltages of multiple trains under real operating conditions for AC railways," IEEE Transactions on Power Systems, vol. 14, pp. 452-459, 1999

[4] R. J. Hill and I. H. Cevik, "On-line simulation of voltage regulation in autotransformer-fed AC electric railroad traction networks," IEEE Transactions on Vehicular Technology, vol. 42, pp. 365-372, 1993

[5] E. Pilo, L. Rouco and A. Fernández, "A reduced representation of 2/spl times/25 kV electrical systems for high-speed railways," in IEEE/ASME Joint Rail Conference, Chicago, 2003, pp. 199-205

[6] G. Varju, "Simplified method for calculating the equivalent impedance of an AT system," Innotech Ltd (Innovation Park of Technical University of Budapest), Budapest, July 1996

This page intentionally left blank

The rescheduling system for trains, drivers and crew of high speed railways

H. Shimizu[1], H. Tanabe[2] & M. Matsumoto[2]
[1]*JR East Japan Information Systems Company, Japan*
[2]*East Japan Railway Company, Japan*

Abstract

We are developing the new rescheduling system for drivers and crew that synchronizes the train forecasted plan. The East Japan Railway Company has five Bullet Train lines (SHINKANSEN) operating about 800 trains per day, and dispatching nearly 250 drivers and 350 crew per day in order to operate the trains. To keep the operation plan and management of SHINKANSEN properly, we have the system named COSMOS (Computerized Safety Maintenance and Operation systems of SHINKANSEN). The drivers and crew rostering schedule are made with the drivers and crew rescheduling system that is one of the sub-systems of COSMOS Transportation Planning System. Each driver and crew works on their trains according to a rostering scheduling decided upon beforehand normally. However, sometimes trains would be unable to operate on schedule when the weather is bad or car troubles happen. In such cases, the rostering scheduling must be redesigned. The influence reaches the two or more crews' rostering scheduling by changing only one crew member. It is very difficult to make this rescheduling because a delay of the train changes minute by minute. To make the crews' rostering scheduling change adequately, we have developed a new system. This system displays some contradictions of the rescheduled plan of drivers and crew if a train delay happened. Dispatchers reschedule the plan to solve these contradictions.

We reduce the delay time of trains by utilizing this system and will improve our services for customers.

Keywords: train traffic rescheduling, drivers rostering rescheduling, crew rostering rescheduling, forecast, real-time rescheduling.

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110591

1 Introduction

East Japan Railway Company (JR-East) has five Bullet Train (SHINKANSEN) lines – Tohoku line, Joetsu line, Hokuriku line, Yamagata line, and the Akita line.

Our network of SHINKANSEN is shown in Figure 1, and features of each line are shown in Table 1. These lines all start from Tokyo and directly connect to five areas in east Japan.

A part of the section of the Yamagata line – between Fukushima and Shinjo and a part of section of the Akita line – between Morioka and Akita are low speed sections. In these sections, SHINKANSEN trains and local trains are operated in coexistence. Trains of Yamagata and Akita line are combined with the train of Tohoku line between Tokyo and Fukushima or between Tokyo Morioka.

To respond to a variety of passenger needs in the SHINKANSEN transportation, JR-East has a variety of types of cars. The types of vehicles in April, 2010 are shown in Table 2. The transportation scale, the number of trains per day, car rostering, driver rostering and crew rostering are shown in Table 3.

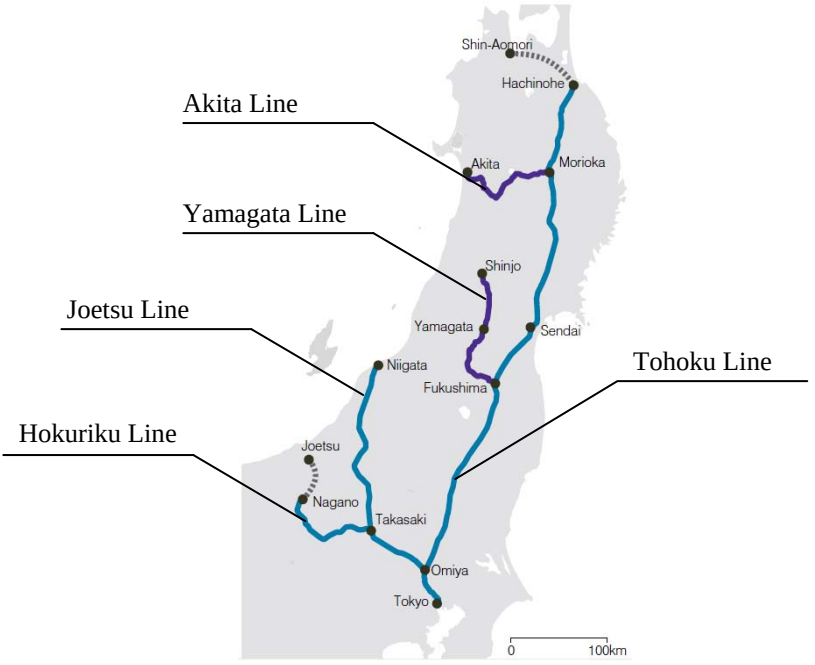


Figure 1: Network of JR-East SHINKANSEN [3].

Table 1: Features of JR-East SHINKANSEN (As of March 2010).

Line name	Distance of line	Name of trains	Number a day of trains
Tohoku Line	631.9 km (Tokyo – Hachinohe)	Yamabiko	89
		Hayate	37
		Nasuno	38
Joetsu Line	333.9 km (Tokyo – Niigata)	Toki	56
		Tanigawa	39
Hokuriku Line	222.4 km (Tokyo – Nagano)	Asama	56
Yamagata Line	421.4 km (Tokyo – Shinjo)	Tsubasa	33
Akita Line	662.6 km (Tokyo – Akita)	Komachi	32

Table 2: Kinds of vehicles of JR-East SHINKANSEN (As of March 2010).

Type	Name of vehicle	Line chiefly operated	Car type	Partner of combining
200	K	Joetsu	10 cars, Flat	None
E1	M	Joetsu	12 cars, Duplex	Enable
E2	J	Tohoku, Joetsu	10 cars, Flat	R
	N	Hokuriku	8 cars, Flat	Enable
E3	R	Akita	6 cars, Flat	J
	LR	Yamagata	7 cars, Flat	P
E4	P	Tohoku, Joetsu	8 cars, Duplex	P, LR
E5	Unsigned	Tohoku	10 cars, Flat	Undecided

Table 3: The transportation scale of JR-East SHINKANSEN.

Trains	About 1,000
Car rostering	About 200
Driver rostering	About 250
Crew rostering	About 350

2 The transportation feature in JR-East SHINKANSEN

There are following three large transportation features in JR-East SHINKANSEN.

(1) All trains of five lines and share limited track capacity between Tokyo and Oomiya.
(2) To operate many trains five lines in limited track capacity, some of them are combined with another.
(3) There is a difference at the transportation stability level between SHINKANSEN line and conventional line, and a conventional line train delay might influence SHINKANSEN trains.

There is the possibility of influencing all areas and all trains even if a delay occurs in a small area. Therefore, we need to reschedule the plan of train operation as soon as possible.

For example, in conventional sections between Fukushima and Yamagata of Yamagata line, a delay of trains and operation suspension of trains occur owing to rainfall or snowfall and so on.

If a delay of trains of Yamagata occurs, it influences combined trains of Tohoku. In this case, we reschedule a splitting and combining plan of trains of Yamagata line and Tohoku line. It is a simple solution that all splitting and combining plans are cancelled and each train is operated by itself. But there are the following problems.

(1) An unscheduled driver and crew is needed.
(2) It is difficult that we add trains to the section between Tokyo and Omiya limited line capacity.

To avoid these problems, we change combined partner trains one by one. It is called "DAN-OCHI" in Japanese.

One of the examples of this rescheduling is shown in figure 2.

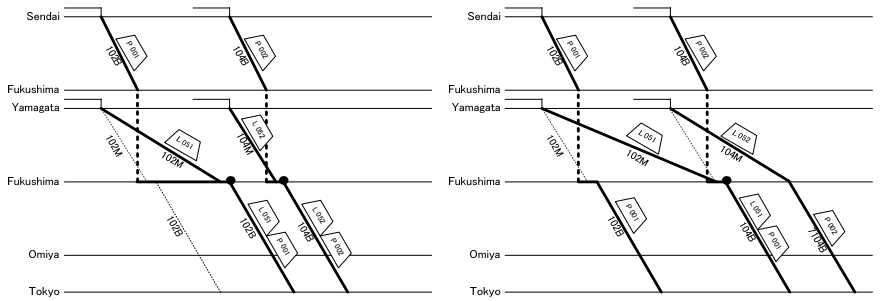


Figure 2: One of examples of combining plan rescheduling.

3 COSMOS overview

3.1 General overview

In an operation of JR-East SHINKANSEN, we need to reschedule the train operation plan as soon as possible. Therefore, we have some systems called "COSMOS" for supporting our judgement of rescheduling. COSMOS has eight systems – Transportation Plan System, Operation Control System, Rolling Stock Control System, Facility Control System, Maintenance Work Control System,

Railway Yard Work Management System, Facility Monitoring System and Electric Power Control System. We are highly managing and controlling information of transportation of SHINKANSEN by cooperatively using these eight systems. Systems overview is shown in Figure 3.

In this thesis, the sub-system that the drivers and crew use in the rescheduling system which is one of sub-systems of the Transportation Plan System, cooperates with the train traffic rescheduling system which is one of sub-systems of the Operation Control System is described. The overview of two systems is shown in Figure 4. This development overview of two systems was reported in our thesis [1]. This drivers and crew rescheduling system came into practical use beginning in March, 2009.

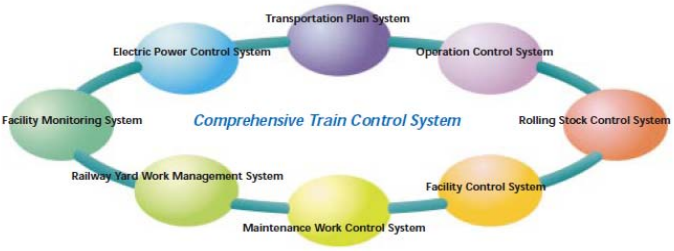


Figure 3: Overview of COSMOS systems.

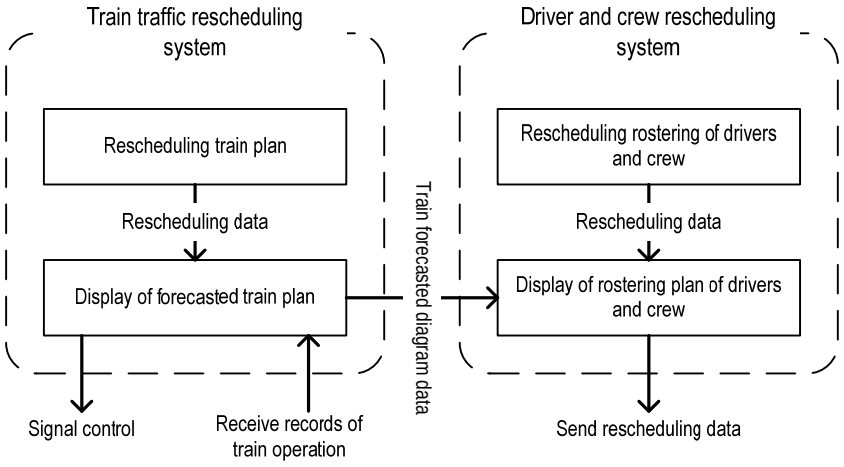


Figure 4: Cooperation between rescheduling functions.

3.2 Effect of improvement

We remade and began to use the COSMOS Transportation Planning System and Operation Control System in May, 2008. In this renewal, we redesigned some

input window and input flow of train traffic rescheduling system. Consequently, we achieved the effect shown in Table 4. The average input time has shortened two minutes or more before and after the function improvement. Because it is necessary to input a lot of train traffic rescheduling, an effect of this time crunch is large.

Table 4: The average input time.

	Number of input times		Average input time	
	Before	After	Before	After
Window 1 (Run alone)	2	1	1min 38sec	54sec
Window 2 (Change combined partner)	4	2	2min 36sec	1min 52sec
Window 3 (Cancel to combine)	7	3	4min 37sec	2min 27sec
Example input of rescheduling shown in Figure 2	8	4	5min 52sec	3min 40sec

4 Case study

4.1 The train traffic rescheduling system

This chapter introduces the train traffic rescheduling system which is one of the sub-systems of the COSMOS Operation Control System.

Figure 5: Input window of system.

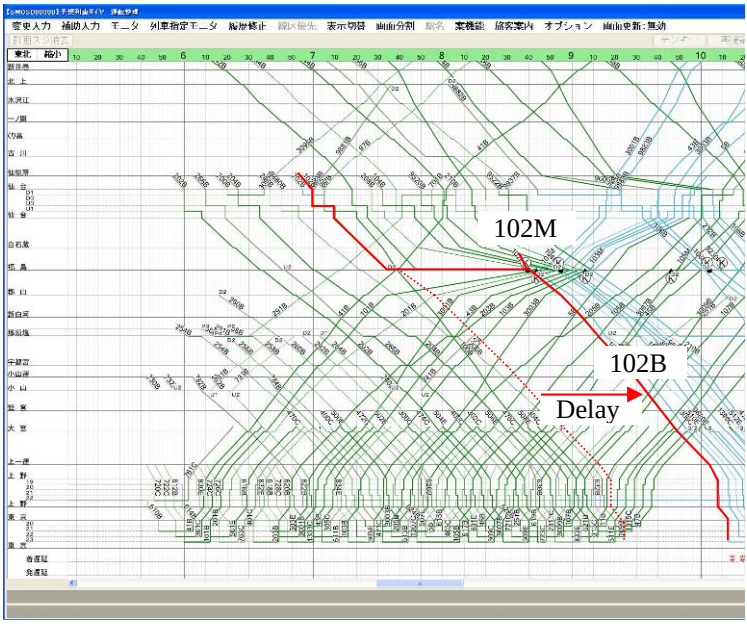


Figure 6: Train forecasted **diagram before** rescheduling.

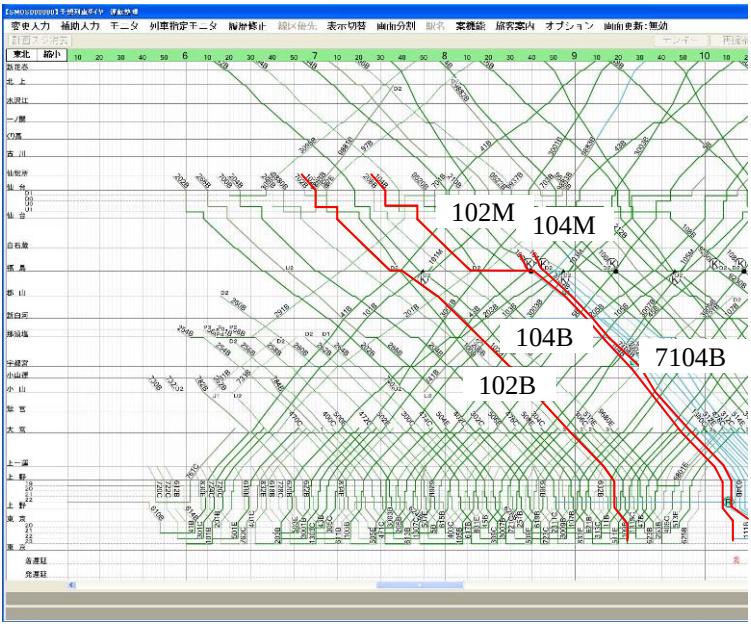


Figure 7: Train forecasted **diagram after** rescheduling.

If a delay occurs in a conventional section, we need to reschedule the train operation plan as soon as possible. For example, when there is a delay of train number “102M” between Yamagata and Fukushima, we reschedule train traffic using the input window of the system which is shown in Figure 5, and a train forecasted diagram before rescheduling is shown in Figure 6. A train forecasted diagram after rescheduling is shown in Figure 7.

Train “102M” is combined with “102B” at Fukushima. “102M” is assigned car “L 051”. In this case, we change the combined partner of “102M” from “102B” to “104B”, and we change combined partner of “104M” from “104B” to none. “104M” is operated alone as train number “7104B” from Fukushima to Tokyo.

4.2 The drivers and crew rescheduling system

When train traffic is rescheduled, the driver and crew rostering plan is influenced, and we have to reschedule the rostering plan using the drivers and crew rescheduling system.

This chapter introduces the drivers and crew rescheduling system which is one of the sub-systems of COSMOS Transportation Plan System. In this sub-system, the forecasted driver and crew rostering diagrams and some contradictions are displayed in a window. This forecasted rostering scheduling is created based on a train forecasted diagram data. Dispatchers monitor this window and input the changing data of rescheduling to the system.

For example, when there is a delay of train number “102M” between Yamagata and Fukushima, we reschedule drivers rostering plan using the input window which is shown in Figure 8, and the drivers forecasted diagram before rescheduling is shown in Figure 9-left figure. The drivers’ forecasted diagram after rescheduling is shown in Figure 9-right figure.

Figure 8: Input window of system.

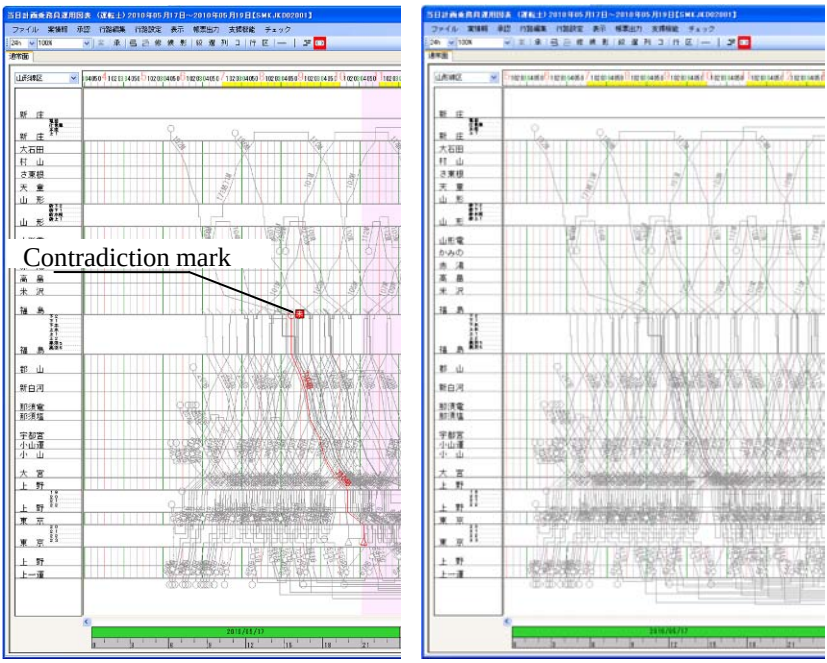


Figure 9: Drivers forecasted diagram before and after rescheduling.

5 Conclusion

In this thesis, the features in JR-East SHINKANSEN and supporting systems for train rescheduling and drivers rescheduling are shown.

When the train traffic rescheduling system and the drivers and crew rescheduling system are not used, the planners and the dispatchers have planned by using the paper diagram by their knowledge, experience, and capability.

After these systems have started to be used, we can quickly reschedule train traffic and drivers and crew rostering. In addition, these systems led to the revolution of the drastic business rule of rescheduling of drivers and crew rostering. Dispatchers changed from a person at the center to a local person.

These systems are contributing to the stability improvement of transportation. JR-East thinks our supporting systems to be important and will develop a further upgrade with JR East Japan information systems Company.

We have now been developing the proposal system for SHINKANSEN using constraint programming [2] as one of the upgrades of the train rescheduling system for practical use.

After this proposal system is put to practical use, we expect that we will achieve a higher effect by cooperating with the rescheduling system for drivers and crew.

References

[1] Hiroyuki SHIMIZU, Hitoshi TANABE, Satoshi HONDA, Kazutoshi YASURA "The new Shinkansen rescheduling system for drivers and crew", Proceedings of Computer System Design and Operation in the Railway and Other Transit Systems – Computers in Railways X, pp. 227-234, (2006)
[2] Hiroyuki SHIMIZU, Hitoshi TANABE, Masashi YAMAMOTO, "The proposal system for Shinkansen using Constraint Programming", Proceedings of World Congress on Railway Research 2008, O.1.3.2.3, (2008)
[3] East Japan Railway Company "2009 Annual Report", http://www.jreast.co.jp/e/investor/ar/2009/index.html

Railway disruption recovery: lessons from airlines

I. Evans
Constraint Technologies International, Australia

Abstract

CTI is combining its products for situation awareness and disruption recovery in airlines with its railway timetabling software to develop a tool for situation awareness and disruption recovery in a railway environment.

Analysis has shown that many concepts are similar, but that the extra complications caused by interactions between train paths require extra visualisation options such as train diagrams and spatial network displays, along with extra constraint checking to identify and prevent conflicts.

It has been found that these extra requirements can fit within the same overall framework as is used for airlines. The extra visualisation options for rail were then seen to in turn provide value for airlines, as the format used for train diagrams is useful to visualise crew connections, and spatial network displays are useful to visualise air corridors. The extra constraint checking required for rail can also be useful for airlines to model flow restrictions placed on congested runways and air corridors.

Keywords: disruption management, situational awareness, recovery optimization, airlines, railways, fleet, crew, passengers.

1 Introduction

The disruption management process in airlines has been the subject of a large amount of academic study summarised by Kohl et al. [1] and Clausen et al. [2] and is expedited by commercial products offered by several vendors [3, 4] that provide integrated fleet disruption management taking into account scheduled maintenance, crew, passengers and cargo. The situation for other modes of transport such as heavy and light rail and buses is much less mature, yet there are many similarities in the problems such that much can be learned and borrowed

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110601

from the airline work. The similarities are most pronounced for scheduled services rather than on-demand services, so that this paper will confine discussion to these.

2 Terminology

The disruption management procedure for airlines has been described using a range of terminology, and thus it is necessary to define the terms used in this paper. Here we use the following terms for the three main components:

- *Situational Awareness*, where controllers of an airline's operations are provided with accurate and timely information about all aspects affecting the operation of a schedule.
- *Recovery Scenario Evaluation*, where a controller can evaluate the effect of actions that might be taken to recover from disruptions without committing to them.
- *Publishing of Schedule Updates*, which publishes changes in the schedule to recover from the disruption.

We will also define the term *Recovery Optimisation* to describe a tool for use during Recovery Scenario Evaluation that automatically produces on request one or more suggestions for recovery scenarios that will minimise the amount of disruption. Much of the academic literature for both airlines and railways concentrates on such tools that provide a single optimal solution, however our experience with airlines has shown that it is not possible to provide a single solution that will be acceptable in practice for all disruptions. There are two reasons for this:

- It is difficult to state a single objective function to be minimised that correctly encapsulates the tradeoffs that are required in all disruption scenarios.
- There is often extra information that is not available electronically when the optimisation is run. An example from airlines is that the time taken for a mechanical fault to be rectified can only be estimated, so that the ability to produce a range of provisional solutions for comparison that are based on different estimates is a very useful part of the decision making process.

Each of these components will be discussed separately in the following sections.

3 Operator versus network disruption management

One confusing aspect of the usage of the term disruption management in the transport context is that disruption management occurs at two different levels.

With airlines, disruption management is performed by each airline for its own fleet, and there is a separate level of disruption management that occurs at the level of Air Traffic Control which is concerned with airspace and airport resource usage by all airlines. In this paper we will use the term *operator disruption management* to refer to the first of these, and *network disruption management* to refer to the second.

A similar situation occurs for rail where *above-rail* operators (also known as train operators) are concerned with disruptions to the operations of their fleet, while *below-rail* operators (also known as infrastructure managers) are concerned with managing the provision of paths in the event of disruptions to infrastructure availability or the schedules of the above-rail operators. The distinction between these two types of disruption management in rail is less clear where an operator is both the above- and below-rail operator, such as commonly occurs in dedicated metro systems.

The two levels of disruption management are interlinked, but there are well defined systems that allow information to be exchanged between the systems to enable good decisions to be made in the two separate systems. In the airline context, the Air Traffic Control system allocates movement slots to limit movements, and an airline can then make adjustments to its schedule to make best use of its available movement slots. A similar situation occurs with paths requested by an above-rail operator and allocated by a below-rail operator.

In this paper we are suggesting that it is best to keep these two levels of disruption management separate, even for cases where an operator is responsible for both levels. There are several reasons for this:

- The lower level of detail that needs to be considered at the level of operator disruption management enables options to be seen more clearly without having to concentrate on infrastructure details.
- The timeframes involved in operator disruption management are longer than those involved in network disruption management. The extra time available to make these decisions enables better strategic decisions to be made at the higher level.
- Operator disruption management should be concerned with more detailed checks for non-network resources – for example crews, passenger overcrowding and scheduled maintenance for particular vehicles are not considered in the context of network disruption management.
- The systems that control the below-rail infrastructure are complex, and since development and maintenance of these systems is costly and time-consuming it is sensible to only include essential functionality at this level.
- Separate software used to assist with disruption management currently exists for each level, so it is much easier to take advantage of existing functionality if this separation is maintained.

When seen from this point of view, it can then be seen that much commercial software currently exists for both levels of disruption management for air transport [3–6], but most commercial software for railways concentrates on network disruption management or only handles a limited number of resources in the context of operator disruption management [7, 8]. It will be further seen that there are many similarities between operator disruption management for airlines and other modes, and this paper will concentrate on showing the possibilities for the use of this technology by above rail operators.

4 Situational awareness

Situational awareness is the most critical factor in the process of disruption management. Tools that make it easy for a controller to quickly comprehend the essential features of a disruption increase both the speed and quality of the recovery process.

A vital component of situational awareness software is the ability to show events in real-time. This requires an event-driven architecture that processes messages as they arrive and immediately updates all visualisations. Such architectures have been in use in an airline context at CTI for 20 years, and we are now extending this technology for use with rail situational awareness.

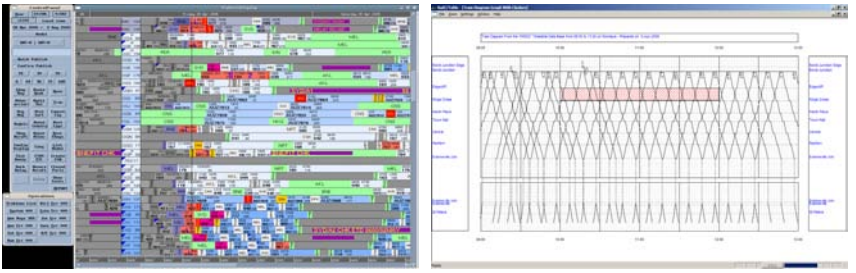
To understand the full picture of the current situation it is necessary to have the following information:

- The current position and availability of all vehicles.
- Current information about any network restrictions. For airlines, this information includes arrival and departure slot availability at airports. For railways this can include information about available paths including time restrictions caused by factors such as timed track possessions, and other factors such as temporary speed restrictions.
- Forecast information to predict whether any current deviations from the nominal schedule are likely to naturally correct from the built-in recovery times in the schedule, or whether recovery actions need to be taken to prevent problem escalation.
- Information about planned crew duties, rest breaks, connections and location of standby crew. It is important to also be able to check maximum duty and minimum rest break rules to determine if delays will make the currently planned duties illegal. Crew information is particularly important in long-distance networks.
- Information about passenger numbers and itineraries (and/or cargo for freight operations), or estimates if these are not accurately known. This is vital for determination of the impact of a disruption, and particularly so in a railway context since overcrowding has a significant effect on dwell times, and thus can affect running times significantly.
- Information about scheduled maintenance times and locations for vehicles. This is important if repositioning a vehicle for planned maintenance would be very expensive, as is normally the case for airlines and long-distance rail lines.

A range of visualisation formats have been used for both airline and railway disruption management, and as different formats have different strengths for different classes of disruption the most desirable option is to have all formats available so that the controller can choose the most suitable for any given situation. It is also interesting to note that most of the formats are also applicable to other transport types, although the amount of usage of the different formats is likely to differ between modes.

4.1 Gantt chart

Gantt charts (Figure 1(a)) provide easy visualisation of resource usage over time.



(a) Gantt chart (b) Train diagram

Figure 1: Main visualisation displays for airlines and rail.

For airline operations, it is typical to show the actual or planned legs for each vehicle as the prime visualisation tool. This display is typically annotated to show the origin and destination of each leg as well as the flight number, with colour coding used to show problems such as late departures or arrivals, crew or passenger connection problems, ATC slot problems, insufficient turnaround times at airports, aircraft unavailability and problems meeting scheduled maintenance locations or times.

For rail operations, this can be used in a similar way to show the trips planned for each vehicle while highlighting any problems. In order to avoid excessive clutter, only trip segments between major stations or junctions would be shown at this level, with the stopping pattern information being available by clicking on each trip segment.

A Gantt chart has the disadvantage of not highlighting the interactions between vehicles, which is less of a problem in an airline context but is a significant factor for rail. Thus for rail it is likely that the prime visualisation tool would instead be a Time-Distance Chart.

4.2 Time-distance chart

Time-distance charts (Figure 1(b), commonly known for railways as Train Diagrams or sometimes Service Planning Diagrams) are an excellent method of visualising interactions between services and also any network limitations such as track possessions.

These can be updated in real time to visualise actual delays and estimated arrival times, and can also have overlaid crew connection information to assist with visualisation of likely crew impacts.

This type of visualisation can also be useful in an airline context to visualise connections between services for vehicles and crew in the context of point-to-point networks with an approximately linear topography such as the east coast of Australia.

4.3 Platform occupancy diagram

Diagrams showing occupancy of platforms over a given time such as that shown in Figure 2(a) are commonly used in rail planning, and are also very useful on the day of operation. These are also known as Station Activity Diagrams. This type of visualisation is also well suited to showing airport gate usage.

(a) Platform occupancy diagram

(b) Real time timetable

Figure 2: Other visualisation displays.

One of the critical factors when dealing with metro rail systems in peak periods is the requirement to ensure that the number of passengers arriving at each station can be transported using the available services. In many metro systems, some overcrowding occurs even in the absence of any cancellations. If trains are cancelled, it is vital to ensure that the capacity of subsequent trains on the same line is sufficient to recover from the cancellation, as otherwise the resulting increase in dwell times caused by extreme overcrowding is likely to lead to a decrease in the line capacity and thus an escalating problem. One way of visualising whether this is occurring would be to include estimated passenger figures on the platform occupancy diagram.

4.4 Timetable

A standard means of visualising a railway schedule is the timetable view. Additions showing the actual and estimated times along with any cancellations or changes to stopping patterns as shown in Figure 2(b) are another way of visualising the current situation on the day of operations that is immediately familiar.

4.5 Spatial displays

Network displays such as track and signal diagrams shown in Figure 3(a) are used extensively in railways by below-rail operators, but are in general too detailed for higher level planning. Simplified network diagrams may be useful for above-rail operators to visualise alternative routings in complex networks with several alternatives, however. A similar concept may also be useful for airline operators where bad weather can cause flow restrictions through certain areas.

A geospatial display such as that in Figure 3(b) could be useful where it is necessary to take into account options that can use other modes of transport – e.g. the use of buses for transport of passengers in the event of a major network disruption.

(a) Network diagram (b) Geospatial display

Figure 3: Spatial visualisation displays.

4.6 Summary information

Summary information can be used to evaluate the current situation and to help to decide whether measures need to be taken to aid recovery from a disruption. If a timetable is robust, recovery from small delays is likely to be automatic through the recovery time built in at certain stops. One of the key functions of situational awareness is to understand when the situation is such that this is unlikely to happen and that other measures need to be taken to expedite recovery.

Table 1: Key performance indicators.

Indicator	One week ago	So far current day	Forecast to midnight	Total last month	Forecast to end of month
Delays > 5 mins	53	30	51	1510	1456
Delays > 10 mins	32	20	32	923	912
Delays > 15 mins	15	10	17	478	456
Delays > 30 mins	6	2	5	204	193
Delays > 60 mins	3	0	1	56	45
Cancellations	4	0	2	98	87
Pax delayed ('1000)	6.7	4.2	6.6	202	195
Delay pax-mins ('1000)	132	79	121	3962	3765
Cancelled maintenance	1	0	0	11	9
Standby crew used	4	2	3	120	105
Crew overtime hours	25	10	24	753	742

Table 2: Remaining problems.

Type	Count
Insufficient turn time 1-5 mins	5
Insufficient turn time 6-10 mins	0
Insufficient turn time 11-15 mins	0
Insufficient turn time > 30 mins	0
Leg port mismatches	0
Missed crew connections	1
Missed scheduled maintenance	1

In both airlines and railways, periodic on-time performance statistics are published which materially affect the profitability of the operator. For airlines, this effect is often mainly on reputation, although there are financial repercussions if Air Traffic Control slots are consistently missed. For railways, there are often penalty clauses such that operators lose revenue if performance drops below certain values. In such situations, it is useful to be able to track the performance during the current reporting period and to forecast the likely numbers at the end of the period. This information can then be used when choosing recovery options – for example to choose to increase the overtime payments to crew in order to avoid excessive service cancellation.

Summary information can then be shown in real time in a manner similar to Table 1, depending on the performance figures important to the operator.

As it is typically not possible to solve all problems for the rest of the day immediately, it is also important to show the number of remaining problems in a format such as shown in Table 2.

5 Recovery scenario evaluation

Recovery Scenario Evaluation involves a controller making a number of provisional changes in one or more private "what-if?" scenarios to recover from the disruption, and then evaluating the expected outcome from these changes, often by means of comparison of the summary statistics.

Since these changes are being made in a real-time environment, they are stored in a format such that updates to legs that have not been changed in the scenario are reflected as they happen. Updates to legs that have been changed, however, can result in the scenario being updated if it is no longer possible to make the provisional changes. For example, if a scenario plans to swap the lines of flying for two aircraft at port ABC starting from flights 123 and 124, then if either of these flights actually depart from port ABC then the swap is no longer possible and will automatically be removed from the scenario.

The changes that are typically performed in an airline context are:

- Delaying all legs that are to be flown by an aircraft, but reducing the turnaround time between future legs in order to make up time. This is only appropriate for short delays.
- Swapping the lines of flying for one aircraft with another at a port so that an aircraft that has already arrived but has a long scheduled turnaround time can be used to fly the leg that has been planned for an aircraft that is late. This is a good choice when the late aircraft is due to arrive before the scheduled departure of the other aircraft.
- Cancelling a set of legs in a loop (often a return trip). This is appropriate only if the load on the legs is low and the frequency on that route is high, as the passenger disruption would then be limited, or if flow limitations at an airport make it essential.
- Adding or removing a stop for a multi-leg flight. This is not common for airlines, and only makes sense in certain network geographies with several airports are located approximately in a line. The most common use of adding a stop is for situations with strong headwinds where an extra stop is required to take on more fuel.
- Delaying, cancelling or moving scheduled vehicle maintenance activities where this is possible. This can have a cost impact since often a maintenance crew will have been organised at a particular location and time to perform this maintenance.
- Adding another service to cater for stranded passengers.

For an airline, the main criteria against which the impact is assessed against are:

- The on-time performance statistics of the airline, which are published regularly and can be seen as a competitive advantage or disadvantage.
- The total number of affected passengers.

- The total passenger delay in passenger-minutes.
- The requirement for standby crew or overtime.
- The number of disrupted scheduled vehicle maintenance activities.

Similar actions are possible in the rail context as outlined in Norio et al. [9], but the extra constraints associated with railways would result in a greater emphasis on adding or removing stops, with this being further broken down into the following main options:

- Changing the stopping pattern on a trip section, e.g. from all-stop to express or limited express or vice-versa.
- Terminating a trip at an intermediate station and connecting it to another trip that has been changed to start at that station.
- Changing the departing or meeting orders of trains, e.g. to allow a rapid train to overtake a local train.

For a railway, the main criteria against which the impact is assessed against could be:

- The on-time performance statistics of the railway, which may be tied to penalties if they drop below specified levels.
- The total number of affected passengers.
- The total passenger delay in passenger-minutes.
- The number of trips with severe overcrowding.
- The requirement for standby crew or overtime.

Note that for rail lines where actual passenger numbers are not well known, passenger statistics would have to be estimated based on expected passenger flows. For this reason, these measures are currently mainly used in an informal way, even though they more accurately reflect the perceived passenger service quality, and the extent to which they are used depends on the experience of the traffic controllers.

6 Publishing of schedule updates

When an acceptable provisional recovery scenario has been produced, the set of associated changes is published as a single action to ensure that consistency is maintained.

In an airline context, publishing normally involves sending one or more teletype messages to the worldwide SITA or AFS [10] networks, which are then directed to any interested airlines. Situational Awareness tools used by these airlines will then process the received messages to update the current status of the airline's flights or connecting flights that the airline is interested in.

In the rail context, standards for messaging are less mature due to many railways operating on isolated networks.

In Europe, draft messaging standards have been specified in response to Interoperability Directives 96/48/EC (High-Speed), 2001/16/EC (Conventional Rail) and 2004/50 (alignment of High-Speed and Conventional Rail Directives and extension of the scope). Implementation of these standards is progressing across the network.

In the UK, the SIRI and TransXChange standards have been developed for publishing of real-time public transport information and are in widespread use, although TransXChange is aimed more suited to buses than railways.

7 Recovery optimisation

The recovery optimisation problem has been the subject of much study for airlines and for rail.

The problems studied for airlines consider a much greater range of resources, with many considering fleet, scheduled maintenance, airport slots and full passenger itinerary modelling [11], and several also incorporating crew [12].

For railways, most work on recovery optimisation concentrates on the fleet and incorporates quite detailed consideration of the network paths along with some consideration of passenger connections [13], while passenger overcrowding and crew are only considered for simple problems [14] [15].

As crew connection problems can have a large impact on the feasibility of solutions for complex or long-haul networks, and the passenger service quality can have a large impact when comparing the quality of the solution and also the feasibility (due to the effect of overcrowding on dwell times), we consider that extensions of airline recovery optimisation formulations to allow for the more complex constraints of railway paths would be extremely beneficial.

8 Conclusion

There are many advantages in considering disruption management in railways at a high level, as this enables consideration of factors such as passenger overcrowding and crew connections. When considered this way, there are many similarities that allow the use of the mature technology used for disruption recovery in airlines to enable more effective disruption recovery for rail.

References

[1] Kohl, N., Larsen, A., Larsen, J., Rossd, A. & Tiourine, S., Airline disruption management – perspectives, experiences and outlook. *Journal of Air Transport Management*, **13(3)**, pp. 149–162, 2007.

[2] Clausen, J., Larsen, A. & Larsen, J., Disruption management in the airline industry – concepts, models and methods. *University of Denmark, DTU*, 2005.

[3] TPAC Operations product. Constraint Technologies International, http://www.constrainttechnologies.com/software/TPAC_Operations/.

[4] iFlight Operations product. IBS Software, http://www.ibsplc.com/fleet-scheduling-and-management-software.html.

[5] OSYRIS software suite. Barco, http://www.barco.com/en/productcategory/29.

[6] Air Traffic Management products. SELEX Systems Integration, http://www.selex-si-uk.com/markets/air_traffic_management/atm_products.aspx.
[7] RailEdge Traffic Control & Dispatch Systems. GE Transport, http://www.getransportation.com/na/en/trafficcontroldispatchsystems.html.
[8] Rail Operations System. Siemens Mobility, http://www.mobility.siemens.com/mobility/en/pub/interurban_mobility/rail_solutions/rail_automation/operations_control_systems.htm.
[9] Norio, T., Yoshiaki, T., Noriyuki, T., Chikara, H. & Kunimitsu, M., Train rescheduling algorithm which minimizes passengers' dissatisfaction. *Lecture Notes in Computer Science*, Springer: Berlin and Heidelberg, volume 3533, pp. 829–838, 2005.
[10] AMC and European Fixed Service (AFS). EUROCONTROL, http://www.eurocontrol.int/amc/public/standard_page/aeronautical_fixed_service.html.
[11] Artigues, C., Bourreau, E., Afsar, H.M., Briant, O. & Boudia, M., Disruption management for commercial airlines: Methods and results for the ROADEF 2009 challenge. *Hyper Articles en Ligne*, **hal-00492035**, 2010.
[12] Petersen, J.D., Solveling, G., Johnson, E.L., Clarke, J.P. & Shebalov, S., An optimization approach to airline integrated recovery. *AGIFORS Airline Operations 2010*, 2010.
[13] Tornquist, J., Computer-based decision support for railway traffic scheduling and dispatching: A review of models and algorithms. *Proc. of the 5th Workshop on Algorithmic Methods and Models for Optimization of Railways*, Dagstuhl Research Online Publication Server: Sweden, 2006.
[14] Shaw-Ching Chang, Y.C.C., From timetabling to train regulation – a new train operation model. *Information and Software Technology*, **47**, pp. 575–585, 2005.
[15] Walker, C.G., Snowdon, J.N. & Ryan, D.M., Simultaneous disruption recovery of a train timetable and crew roster in real time. *Computers & Operations Research*, **32(8)**, pp. 149–2094, 2005.

Kaban: a tool for analysis of railway capacity

J. Ekman
SICS AB, Sweden

Abstract

The capacity analysis tool Kaban aims at being efficient for examining if planned infrastructure meets the expected need for railway capacity. Infrastructure, safety rules, signalling and traffic are all modelled in great detail in Kaban and hence the tool is a useful support for signalling design. The tool is also useful for finding out which routing and what train order suits existing or planned track layout. The idea of Kaban is that traffic patterns can be modelled as discrete event systems, with the minimum cycle time as a capacity measure. This measure answers the question if a certain timetable is possible at a station and tells how much buffer time there is. Kaban also presents results on what is critical for the capacity, aiming at explaining how to adjust to increase capacity. The GUI of Kaban displays the infrastructure and train paths and takes care of the user interaction. The development of Kaban is supported by the Swedish Transport Administration (Trafikverket).

1 Introduction

Kaban aims at being a user-friendly capacity analysis tool, efficient to work with and producing relevant and dependable results. Kaban aims in particular at supporting signalling design, but the tool is useful also for other phases of building and utilising the railway. The computations of capacity in Kaban are analytic. The tool is based on the methods presented in [1] and [2].

1.1 The goal of capacity analysis

In this paper the point of view is that analysis of railway capacity is conducted with the aim of supporting decisions on how to most efficiently build and utilise

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110611

railway infrastructure. The basis for the development of the capacity analysis tool Kaban is that

- capacity analysis results of high quality make a substantial difference for the efficiency in building and utilising the railway; and
- there is a lot to gain by more efficiently building and utilising the railway.

To build railway infrastructure that efficiently meets the present and future traffic demands and to efficiently utilise the existing railway network are difficult tasks. Nevertheless, those are important tasks and the efficiency of the railway industry relies on how well those tasks are performed. In the railway building phase it is important with analysis results feasible for:

- suggesting how to adjust the infrastructure to increase performance
- making clear advantages and drawbacks of infrastructure proposals
- making a decision on an infrastructure investment.

1.2 Capacity analysis efficiency

Reducing the time and work to answer specific analysis questions is an important step to obtain capacity analysis of higher overall quality. It enables for the analysis to cover a larger number of cases of infrastructure and traffic. Thereby favouring a more complete understanding of the infrastructure utilisation possibilities. The result is that advantages and drawbacks of different proposals are better understood. Possibly the risk of mistakes or unforeseen consequences are also reduced. In this context a capacity analysis tool needs to be supportive for entering data. Since a lot of data is needed for analysing capacity it is often the case that entering all this data is time consuming. This is especially the case if infrastructure and train conflicts are considered at a detailed level.

Transparency of the capacity analysis is helpful for understanding and trusting the results. Such a transparency is provided by presenting how the results are derived. If computations are presented as divided in parts, where each part is conceived as reasonable, then the result can be trusted. A benefit of capacity analysis of high quality in the building phase is the potential to use those results for track utilisation at stations and for planning traffic.

2 Capacity analysis in Kaban

2.1 Basic ideas

Kaban is based on analytic methods for estimating capacity, as presented in [2]. That is, Kaban is not a simulation tool. The models of railway operation of Kaban are static and not probabilistic. Kaban is designed for analysing capacity at complex stations with a lot of conflicting movements. These are cases where it is crucial how safety rules are interpreted and implemented in the interlocking system.

Kaban implements the Swedish safety rules in detail and offers the possibility to specify certain kinds of deviations from the rules.

The static and analytic character of Kaban makes it in a sense model traffic in an ideal world. Some results can be thought of as theoretical maximum capacity, to be used as measures for comparing different solutions. Other results, to become realistic, needs to be calibrated by asserting buffer times. There are also results, though, that has a static nature of their own. Static are some requirements on vehicle positions in the moment a train route is locked. Such requirements are capacity results in their own, they are significant as they are the basis for other capacity analysis results. The advantages with Kaban's models of railway operation are that:

- the assumptions are precisely defined and we know exactly what the results mean
- fast computations are made possible and this allows for considering large number of infrastructure proposals and modifications within available time
- precise measures of capacity are provided for measuring the benefits of infrastructure modifications, comparing infrastructure proposals and for deciding on which proposal to choose.

It is possible to further develop Kabans, still being analytic, to become a tool based on probabilistic models. For instance using probabilistic models for running times. But such a development may be hard to enable without the drawback of slowing down the computations. A development more likely to be advantageous is the search for the minimum cycle time for all traffic pattern satisfying some specification.

2.2 The concepts waiting point and cleared path

The essential concepts on which the method of Kaban relies are *waiting point* and *cleared path*. Detailed description and exemplification are given also in [2]. A waiting point is a point at which a trains stops to wait for prior trains to pass. Waiting points of trains are input to the capacity analysis. A waiting points is associated with the end of a train route. A train path is divided into parts called *cleared paths* by the waiting points. A train path here means the path on which the train runs through the analysed railway section. Cleared paths shall not to be confused with train routes and waiting points may not be confused with stops in general.

2.3 Conflicts and cycle time computation

The model of train operation of Kaban is a discrete event system (see [3]). As events we chose starting times of movements on cleared paths. In the continuation we will just use movements to refer to those movements on cleared paths. Considering two events separately from all other events the following principles

will assure that the minimum duration in between two events is uniquely defined

- a train moves unhindered by other trains on each of its cleared paths
- stop durations are fixed, but waiting for other trains at waiting points is not included in those durations
- it is uniquely determined where and how two trains are in conflict.

Notice that the first two points uniquely determine the time from the start of a movement to any point along the way to the end. The last point gives what in addition is needed to know for uniquely determining the minimum duration between events. Another way to formulate the first point is to say that a movement may start no earlier than that it will be unhindered by prior movements. This means that the speed of trains is not at all affected by the occurrence of other trains. The interaction between trains is manifested as conditions on time for trains to enter the analysed railway section and time to wait at waiting points. The first point does not mean that all of the train routes of a cleared path must be locked before the movement on the cleared path starts.

As is shown in [2] max-plus algebra provides notation for defining and characterising the cycle time of a traffic pattern in Kaban. The cycle time of a traffic pattern in Kaban is expressed as the *minimum cycle mean* for a discrete event system. This suits the use of Howard's algorithm [4], a fast algorithm for computing the minimum cycle mean. The approach to model train operation as a discrete event system and to use max-plus algebra to express cycle time is well studied, see for instance [5]. Although such approaches are basically the same, they may differ in what the analysis aims at and precisely how the modelling is done.

2.4 An example of a cycle time computation in Kaban

The figures 1 and 2 show the drawing area of the Kaban GUI in which a fictive railway section is pictured. The two figures also show a traffic pattern with just

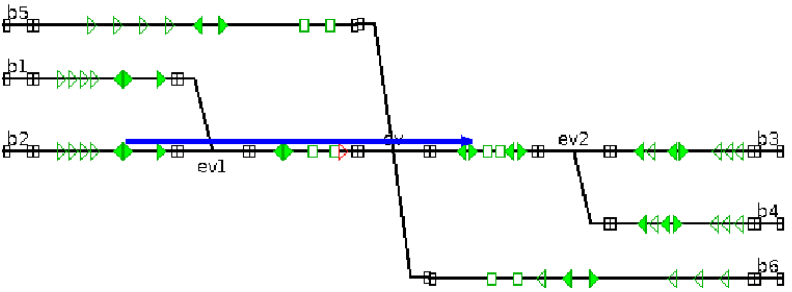


Figure 1: Movement *R1* entering the railway section.

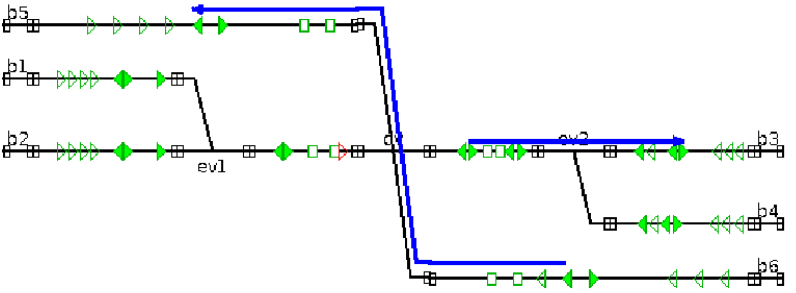


Figure 2: Movement *R2* leaving and movement *V* passing the railway section.

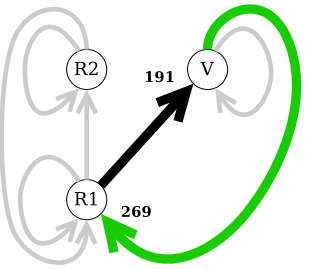


Figure 3: The conflict graph for the traffic pattern *R1-V-R2*.

two trains. One of the trains, say train *TR*, is divided in two movements, *R1* and *R2*, by a waiting point. The other train, say train *T*V, has no waiting point and consists of just one movement *V*. Figure 1 shows us the movement *R1* and figure 2 shows us the two movements *V* and *R2*. Assume that, for this example all data needed for the analysis capacity has been given to the system. Among the given data is the traffic pattern given by ordering the movements, here given as *R1-V-R2*.

The first step in the capacity analysis is to make clear which the relevant conflicts between movements are. In Kaban, the conflicts of a traffic pattern is represented by a *conflict graph*. Figure 3 shows the conflict graph for the traffic pattern *R1-V-R2*. The conflict graph has events, i.e. movement starts, as nodes and conflicts between movements as *arcs*. Subsequently we will refer to conflicts between movements and minimum durations between two events as *arcs* and *arc weights*, respectively. Since the analysis is concerned with a repeated traffic pattern there are two kinds of arcs, those of conflicts of movements in the same period in the traffic pattern and those of movements in successive periods. The first kind is depicted and referred to as *straight arcs* and the second kind is depicted and referred to as *bowed arcs*.

Straight arcs:	
R1→R2	262.2
R2→R1	191.1
Bowed arcs:	
R1→R1	270.0
R2→R1	82.8
R2→R2	240.0
V→R1	269.4
V→V	270.0
Critical Graph:	
V→R1	269
R1→V	191
Cycle Time: 460	

Figure 4: Some Kaban results from the cycle time computation.

Let us now study the traffic pattern *R1-V-R2* as given by the conflict graph. *R2* starts from a waiting point and there it waits for other trains until *R2* can go on unhindered. But the graph tells us that movement *R2* and *V* are not in conflict, so what does *R2* wait for? The arcs ending in *R2* tell us that. There are two such arcs: *R1→R2* and *R2→R2*. The arc *R1→R2* concerns one and the same vehicle, the vehicle of train *TR* and therefore the time of the start of *R2* depends on the time of the start of *R1*. The other arc, *R2→R2*, says that the time of the start of *R2* also depends on the movement *R2* of the previous period. Hence, train *TR* in one period waits at the waiting point for the train *TR* of the previous period to move ahead far enough not to interfere with movement *R2* of the waiting train. Therefore, in this example, the waiting point makes a difference.

The *critical graph* is the subgraph of the conflict graph that determines the cycle time. In figure 3 the critical graph for the traffic pattern *R1-V-R2* is given by the bold arcs. We have here chosen to show the arc weights only in the critical graph. The weight of the arc *R1→V* means that *V* can start earliest 191 seconds after the start of *R1* and the weight of the arc *V→R1* means that *R1* can start earliest 269 seconds after the start of *V*. Hence, the minimum time for one period, i.e. the cycle time, is $191 + 269 = 460$ seconds. Figure 4 show us how Kaban presents the results concerning conflict graph, critical graph and cycle time for the studied example.

2.5 Analysis results for separate conflicts

As is shown in the previous section Kaban estimates capacity for repeated traffic patterns on a railway section and gives as main results, cycle time, critical sequences of movements and conflict graph. The cycle time estimates the feasibility of timetable options. For instance by answering the questions if it possible to

run a given timetable on a railway section and how much buffer time there is. The conflict graph tells us which the relevant conflicts are and can be used to answer questions about slack time for a chosen sequence of movements and with respect to the cycle time or a given timetable. The critical sequences of movements are of interest for robustness sensitivity analysis and for finding out how to reduce the cycle time further.

Since the cycle time can be reduced only by reducing the arc weight of arcs in the critical graph, we need to study the computation of those arc weights for understanding what the options are for reducing the cycle time. For this reason Kaban describes, for each arc $A \rightarrow B$ of the analysed traffic

- how is the arc weight computed
- the critical train route *RB* of the succeeding movement *B*
- the position that vehicle *A* must pass before *RB* is locked and the position that *B* must not pass before *RB* is locked.

Concerning the first point Kaban considers capacity limits for the surroundings of the analysed area and hence the weight of an arc is either a consequence of that capacity limit or an internal limitation. Moreover vehicles has a brake reaction time and the signalling system also has some reaction time which makes the final arc weight become a sum of some terms. That is what the first point refers to.

Regarding the second and third points let us once again consider the traffic pattern $R1-V-R2$ described by figures 1 and 2. The arc weight of $V \rightarrow R1$ is the minimum duration from the start of *V* till the start of *R1*. *R1* consists of more than one train routes, each one locked in one piece. It is not only the locking of the first train route of *R1* that is critical for the arc weight of $V \rightarrow R1$, it may be any of the train routes of *R1*. The reason for this is the assumption that *R1* is an unhindered movement all the way to the waiting point. Hence, we need to know which is the critical train route of *R1*, w.r.t. $V \rightarrow R1$ to reduce the arc weight of $V \rightarrow R1$. Given that the capacity limitation is internal and given the critical route, then the arc weight depends on requirements of the vehicle positions at the moment in time the critical train route is locked. The safety regulation says which position *V* must pass before locking the critical train of *R1*. The assumption that *R1* is unhindered does not allow the vehicle of *R1* to adjust its speed to the fact that some train route ahead is not yet locked. Hence, this assumption will determine the position the vehicle of *R1* must not pass before locking the critical train route of *R1*.

2.6 Support for entering data

An extensive amount of data is required for careful analysis of capacity and a tool with a supportive user interface will save a lot of time for the users. Capacity analysis in Kaban is based on data on

- infrastructure
- vehicles
- train routes and protection of train routes
- trains and traffic.

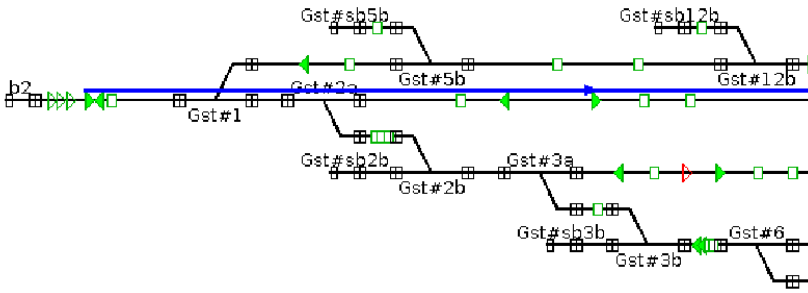


Figure 5: A station and a train. The station, Gamlestad, is based on BIS-data.

All needed data can be entered via the Kaban GUI. Some of the features of Kaban are

- visualisation of infrastructure, train routes and movements
- import of infrastructure data
- construction and modification of infrastructure data
- automatic generation of train routes including protection.

Infrastructure data includes, points, track circuits, speed limits, various signals and buffer stops. It is possible to import infrastructure data on a special XML-format generated from BIS, the infrastructure database of Banverket (the Swedish rail administration). Kaban does not automatically draw a visually nice picture of the track layout, but the Kaban GUI has got drawing modification support for concisely picturing the real track layout. Figure 5 shows an example on a Kaban drawing obtained by modifying BIS-imported data.

The vehicle data includes length, brake reaction time, acceleration and deceleration. The computations in Kaban is based on constant acceleration and deceleration. Train routes are automatically generated for all pairs of successive main signals. Protection for train routes are generated according to the Swedish safety regulation, where the first valid object is chosen in all directions needing protection. For any train using any train route the user may adjust the chosen protection. Those adjustments appear as allowance of non-valid objects to protect the train route and by disallowance of otherwise valid protection. There is several reasons for enabling such adjustments:

- to modify the train route protection is to adopt to older safety regulation, which may still be in use at some sites
- to give the user the possibility to override the safety regulation for some *almost* acceptable protective objects
- to increase capacity by cancelling conflicts which in turn is done by forbidding some protective objects for certain train routes used by certain trains.

2.7 Comparison with other tools for capacity analysis

The research in the area of capacity analysis is extensive and so are the number of tools. For surveys on methods and tools see for instance [6–8] and [8]. In contrast to Kaban, many of the tools simulate railway traffic and many of the methods aim at line capacity. Another tool that is based on the same kind of cycle time computations as Kaban, is presented in [9]. This tool used for evaluating timetable performance. The user-friendliness of capacity analysis tools ought to be measured by the time needed to obtain results of sufficient quality to support decisions on how to build railway or how to utilise existing railway.

3 Conclusions

This paper views *capacity analysis* as a means to support efficient construction and utilisation of railway infrastructure. Different methods and tools may contribute in different ways to help getting a solid understanding of which economic investments will make the most for satisfying the growing railway transportation needs. In this context time efficiency is crucial. Saving time and work is a decisive factor for increasing the quality of analysis results, balancing operational aspects and further developing capacity analysis, by equipping it with new powerful and relevant results. In this context analytic, static methods of analysing capacity has the potential to contribute with valuable results such as

- fast computations that enable a lot of cases for infrastructure and traffic to be part of the analysis
- analysis of trains behaving like we would like them to behave
- providing measures of capacity for comparing different solutions and enabling automatic search for the best solution
- offering results with a precise and clear meaning.

The paper presents Kaban aiming at being a user-friendly capacity analysis tool. Kaban is based on analytic, non-probabilistic methods and the results of Kaban aim at helping users to get a good understanding of the capacity advantages and drawbacks of proposals infrastructure and traffic. A main result is the cycle time of a repeated traffic pattern, useful for estimating feasibility of time table options. Other results aim at explaining how to adjust the infrastructure to increase capacity. The Swedish safety regulation is the basis for Kaban's careful model of train conflicts. Hence, Kaban suits capacity analysis in the signalling design phase of railway construction.

References

[1] Forsgren, M., Computation of capacity on railway networks. SICS Technical Report T2003:12, Swedish Institute of Computer Science, 2003. FoU-rapport, Banverket.

[2] Ekman, J., Capacity estimation of new infrastructure based on a discrete event model of train paths. *Computers in Railways*, WIT Press, www.witpress.com, volume IX, pp. 539–548, 2004.
[3] Baccelli, F., Cohen, G., Olsder, G. & Quadrat, J.P., *Synchronization and Linearity: An Algebra for Discrete Event Systems*. Wiley: Chicester, 1992.
[4] Cochet-Terrasson, J., Cohen, G., Gaubert, S., Mc Gettrick, M. & Quadrat, J.P., Numerical computation of spectral elements in max-plus algebra. *IFAC Conference on System Structure and Control*, Nantes, France, 1998.
[5] Goverde, R. & Soto y Koelemeijer, G., *Performance Evaluation of Periodic Timetables: Theory and Algorithms*. Number S2000/2 in TRAIL Studies in Transportation Science, Delft University Press: Delft, The Netherlands, 2000.
[6] Barber, F., Abril, M., Salido, M., Ingolotti, L., Tormos, P. & Lova, A., Survey of automated systems for railway management. Technical Report DSIC-II/01/07, Department of Computer Systems and Computation, Technical University of Valencia, 2007.
[7] Abril, M., Barber, F., Ingolotti, L., Salido, M.A., Tormos, P. & Lova, A., An assessment of railway capacity. *Transportation Research*, **Part E(44)**, 2008.
[8] Kontaxi, E. & Ricci, S., Techniques and methodologies for railway capacity analysis: comparative studies and integration perspectives. *Rail Zurich 2009 3rd International Seminar on Railway Operations Modelling and Analysis*, 2009.
[9] Goverde, R. & Odijik, M., Performance evaluation of network timetables using peter. *Computers in Railways*, WIT Press: Southampton, volume VIII, 2002.

Two step method for rail potential rise analysis in a DC traction power supply system

S. Chung, H. Kim & H. Jeong
Electric Power Supply Research Division,
Korea Railroad Research Institute, Korea

Abstract

An algorithm for analyzing the rail potentials in a DC traction power supply system is proposed. The distinctive feature of this algorithm is that it uses the node voltage method twice in the rail potential analysis. To calculate the rail potentials, the leakage resistance of the running rail must be included in the equivalent network which makes it difficult to apply the node voltage method. The mesh current method has the drawback that the initial mesh currents are difficult to estimate, which is necessary for an iterative network solution. In this algorithm, the rail potentials are obtained by applying the node voltage method twice. In the first stage, the injection currents to the negative rail are obtained from a load-flow study. In the next stage, a network consisting of the negative rail and the injection currents is constructed. The leakage resistance to ground is added to the network a, and the rail potentials of the network are analyzed. A computer load flow analysis software package was developed to verify the validity of the algorithm. The results from the software are compared with the EMTP-RV circuit analysis results, and the results are nearly identical.
Keywords: rail potential, simulation, DC traction power.

1 Introduction

In most DC powered electric railways, running rails are also used as the negative return path to the rectifier negative bus in the substation. In this system, the potentials of the running rails rise with increasing load current. The increased rail potential causes concern for human safety, due to the possibly hazardous excessive touch voltage and step voltages [1]. For this reason, the maximum allowable rail potential is limited by IEC standard 62128 [3]. The rail potential

WIT Transactions on Modelling and Simulation, Vol 51, © 2011 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM110621

rise can also generate stray currents, which causes concern for equipment safety due to the possibility of electro-chemical corrosion of the metal [1].

In this paper, a computer algorithm to analyze the rail potential rise is presented. The distinctive feature of this algorithm is that it uses the node voltage method twice in the rail potential analysis. In the node voltage network analysis method, every node has to be directly connected to the reference node. However, if the leakage resistance of the running rail to ground is included, it becomes difficult to make an equivalent network in which every node is directly connected to the reference node. Under these conditions, the mesh current method may be used to solve the network. However, the mesh current network method has some drawbacks in this situation because it is difficult to set the initial mesh currents for iterative network analysis.

In the proposed algorithm, node voltage network analysis is conducted twice. First, the load flow is studied to obtain the injection currents to the substations and trains. The leakage resistance of the running rail is not modeled explicitly in this step, which makes every node in the equivalent circuit directly connected to the reference node. In the next step, a different node voltage equivalent circuit is constructed, with only the longitudinal resistance of the negative rail and the injection currents. The leakage resistance of the running rails is added to the equivalent circuit. A set of node equations is constructed to analyze the circuit and solve for the node voltages, corresponding to the potentials of the running rail at the corresponding locations. In conclusion, by conducting node voltage analysis twice, we can avoid using the mesh current method, which is difficult to use in traction power supply system analysis. The leakage current can be calculated from the rail potentials. Computer load flow analysis software using the algorithm was developed. A test run was conducted on a test system. The result is compared with the results of EMTP-RV circuit analysis, where the potentials of the running rail were calculated in one step. The comparison shows that the two results are nearly identical.

2 Review of rail potential analysis

A simple railway DC power feeding system is shown in Fig. 1, to illustrate the method for analyzing negative rail potentials [2]. The rail potentials at the rectifier station end and the train end are:

$$V_{GL} = \frac{R_L}{R_L + R_S} \cdot V_N \tag{1}$$

$$V_{GS} = -\frac{R_S}{R_L + R_S} \cdot V_N \tag{2}$$

$$V_N = I_N \cdot R_N \tag{3}$$

R_S and R_L are the effective resistance to ground at the rectifier station end and the train end, respectively. They are lumped resistances obtained by converting the distributed rail-earth conductance of the negative rail to a lumped pie

resistance circuit. R_N is the longitudinal resistance of the running rail. I_N is the load current in the running rail. V_N is the longitudinal voltage drop in the running rail.

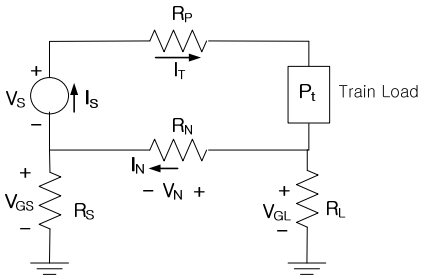


Figure 1: Basic model for rail potential analysis.

3 Proposed computer algorithm for rail potential analysis

Our purpose is to obtain the rail potentials, denoted by V_{GS} and V_{GL} in Fig. 1. To analyze the circuit in Fig. 1, the mesh current network shown in Fig. 2 has to be analyzed iteratively, which requires initial mesh currents. However, the initial value of i_2 in Fig. 2 is difficult to estimate. In the proposed algorithm, the node voltage analysis method is applied twice. In the first step, the leakage resistance of the running rail does not explicitly appear in the node voltage equivalent network, as shown in Fig. 3. In Fig. 3, R_N' is the parallel resistance of R_N and (R_S+R_L). P_t is the train load which is given. The circuit is analyzed iteratively to find the injection currents at the DC source nodes and the train nodes, which are I_s and I_t in Fig. 3. Fig. 2 and Fig. 3 are equivalent in terms of I_s and I_t. In the next step, a different node voltage network is drawn, as shown in Fig. 4, with only the longitudinal resistance of the running rail and the injection currents. In Fig. 4, I_s and I_t are the injection currents obtained in step 1. The leakage resistance of the running rails is added to the equivalent circuit. The circuit in Fig. 4 is also equivalent to Fig. 2. if only the circuit elements in Fig. 4 are considered.

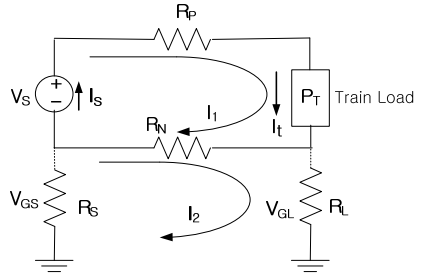


Figure 2: Mesh equivalent network of a simple DC traction power supply system.

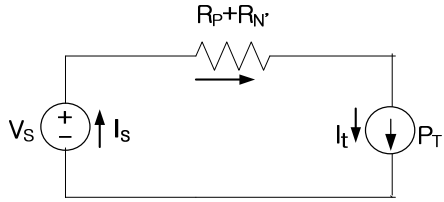


Figure 3: Node network in step 1 for the system in Fig. 2.

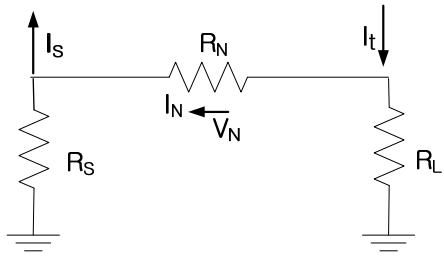


Figure 4: Node network in step 2 for the system in Fig. 2.

The logic flow diagram is shown in Fig. 5. In step 3 of the logic flow diagram, the node equation is constructed in matrix form, as shown in equation (4). Equation (4) is set for the circuit shown in Fig. 3.

$$[G][V]=[I] \tag{4}$$

where [G]: the conductance matrix of the network
[V]: the node voltage vector
[I]: the node injection current vector

In eq. (4), the train load is given in terms of power instead of current or resistance. Given the required power of the train, the corresponding element in the [I] vector is obtained from eq. (5).

$$I_t = -\frac{P_t}{V_t} \tag{5}$$

where P_t: the required power of the train
I_t: the injection current at the train node
V_t: the train node voltage (assumed value)

Because I_t is a function of V_t, equation (4) is nonlinear, and must be solved iteratively. In step 4 of the logic flow diagram in Fig. 5, a new network is constructed, and a new set of node equations is built, as shown in eq.(6).

$$[G]_{new}[V]_{new}=[I]_{new} \tag{6}$$

where $[G]_{new}$: the conductance matrix of the new network
$[V]_{new}$: the new network node voltage vector
$[I]_{new}$: the injection current vector, with elements corresponding to the source current or train load current

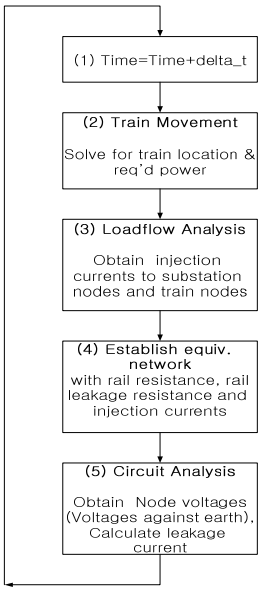


Figure 5: Logic flow diagram for the rail potential analysis.

Unlike equation (4), eq. (6) does not require an iterative solution because the elements of $[I]_{new}$ are given as constants. All elements of $[V]_{new}$ represent the running rail potential at every node location.

4 Test run

Computer load flow software for rail potential analysis was developed to implement the algorithm proposed above. It was applied to the test system, and the result was compared with the result of EMTP-RV analysis for the test system. Fig. 6 shows the test system, and the symbols in Fig. 6 are explained in Tables 1 and 2. In Fig. 6, m stands for 'milli-'. The line parameters and the train

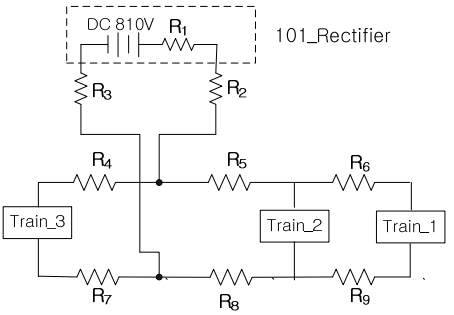


Figure 6: Test system.

Table 1: Resistances in the test system.

Symbol	Resistance [mΩ]	Remark
R_1	22.5	Internal resistance of DC source
R_2	0.35	Cable resistance between source and positive feeding line
R_3	0.35	Cable resistance between source and running rail
R_4	10.4412	Resistance of positive feeding line
R_5	4.9728	
R_6	9.7272	
R_7	9.4717	Resistance of running rail
R_8	4.5110	
R_9	8.8240	

Table 2: Rectifier station and trains in the test system.

Source/ Loads	Locations [m]	Req'd Pwr [kw]	Injection Current [A] After load flow analysis	Remarks
Rectifier Station	2670		1030.9493	Source Resistance = 0.0225Ω
Train_1	1427	200	256.0867	
Train_2	3262	-627.2	-805.4106	Regenerating
Train_3	4420	1184.3	1580.2787	

locations are shown in Table 1. In the test system, the rectifier produces 810V DC with no load, and three trains are drawing power, as shown in Table 2. The rail potentials of the test system are obtained from the developed load flow software. The results are shown in Fig. 7. The same rail potentials were analyzed with EMTP-RV s/w. Fig. 8 shows the EMTP-RV model of the test system. However, EMTP-RV cannot conduct a DC load flow study, therefore, instead of the power required for the trains, the injection currents from the load flow software are inserted as a DC current source at the train location. This can be validated once the power consumption of each train is identical to its required

```
 --------  Dump Rail_potential   ---------
no, type location   volt[v]   src_type   power[kw]   Name
 1   t     1427     -2.471888   c          200.00     train_2
 2   p     2670     -4.898177   n          811.18     101Station
 3   t     3262     -1.404086   c         -627.20     train_1
 4   t     4420     12.537014   c         1184.30     train_0
```

Figure 7: Results from simulation s/w.

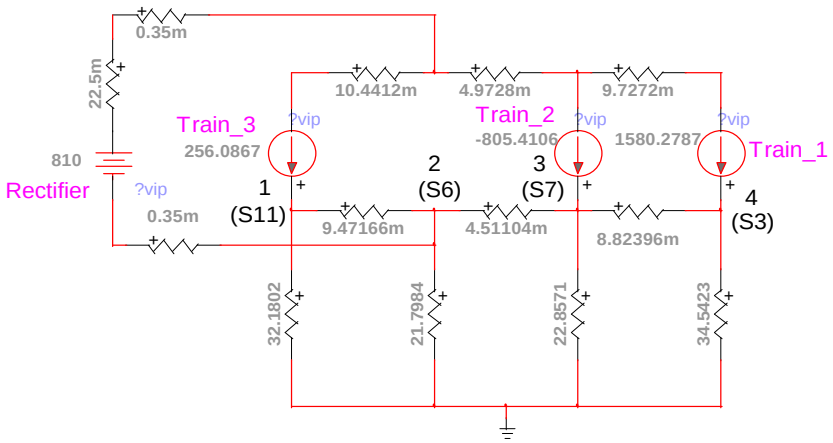


Figure 8: EMTP-RV model of the test system.

power. Fig. 9 shows the resulting rail potentials from the EMTP-RV analysis. The rail potentials from the previous two studies are summarized in Table 3, and the two results are identical. Next, the power that each train consumed according to the EMTP-RV analysis is compared with its required power to validate the

Hide Node Voltages

Node Voltages

	Node		Real (V)	Imaginary (V)	Module (V)	Phase (degrees)
	s4	1	+0.8047409780E+03	+0.0000000000E+00	+0.8047409780E+03	+0.0000000000E+00
	s14	2	-0.5259022044E+01	+0.0000000000E+00	+0.5259022044E+01	+0.1800000000E+03
Node_4	s3	3	+0.1253702638E+02	+0.0000000000E+00	+0.1253702638E+02	+0.0000000000E+00
	s17	4	+0.7619587097E+03	+0.0000000000E+00	+0.7619587097E+03	+0.0000000000E+00
Node_3	s7	5	-0.1404087030E+01	+0.0000000000E+00	+0.1404087030E+01	+0.1800000000E+03
	s9	6	+0.7773303967E+03	+0.0000000000E+00	+0.7773303967E+03	+0.0000000000E+00
Node_1	s11	7	-0.2471894153E+01	+0.0000000000E+00	+0.2471894153E+01	+0.1800000000E+03
	s16	8	+0.7785098083E+03	+0.0000000000E+00	+0.7785098083E+03	+0.0000000000E+00
	s1	9	+0.7815444950E+03	+0.0000000000E+00	+0.7815444950E+03	+0.0000000000E+00
	s5	10	+0.7811836608E+03	+0.0000000000E+00	+0.7811836608E+03	+0.0000000000E+00
Node_2	s6	11	-0.4898187864E+01	+0.0000000000E+00	+0.4898187864E+01	+0.1800000000E+03

Figure 9: EMTP-RV analysis results.

Table 3: Rail potential comparison between the results of two studies.

	Rail potential [V]			Remarks
Node No	From the simulation s/w	From EMTP_RV analysis	% error	
1	-2.47189	-2.47189	0.00	Train_3
2	-4.89818	-4.89819	0.00	DC Source
3	-1.40409	-1.40409	0.00	Train_2
4	12.53701	12.53703	0.00	Train_1

replacement of the required powers in the test system with the DC current sources in the EMTP model. Fig. 10 shows the power consumed by each train and by the rectifier, which are compared to the required powers in Table 4. The two results are again identical. In Table 4, the output power from the rectifier is 835.09kW, while Fig. 7 shows a rectifier output power of 811.18 kW, because the rectifier output power in the load flow study includes the i^2r loss in the 0.225Ω source resistance. If we add the i^2r loss to 811.18 kW, we obtain 835.09 kW, which is the rectifier output power before its internal resistance. In Fig. 8, 1–4 are the node numbers of the train and rectifier nodes in the developed load flow software, and S3, S6, S7 and S11 are the node numbers of the train and rectifier nodes in the EMTP-RV model. Fig. 7 and Fig. 8 show the rail potential analysis results from the EMTP-RV analysis and the simulation software developed using the presented algorithm, respectively.

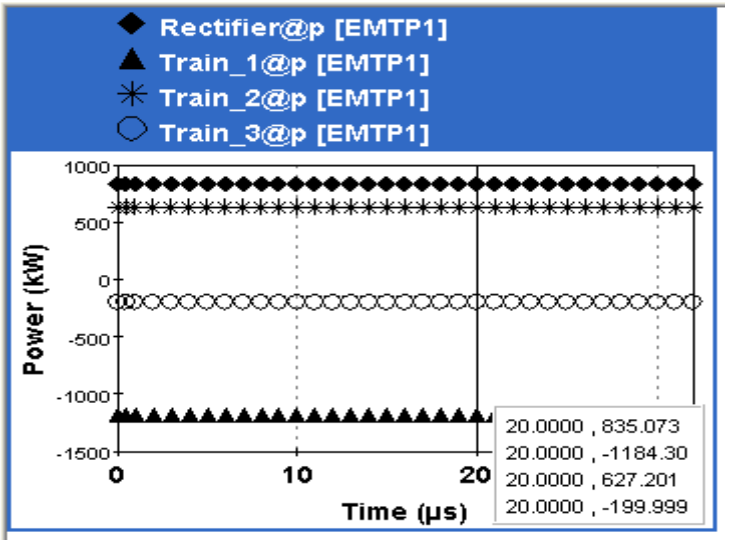


Figure 10: Power generation in the trains & DC source.

Table 4: Comparison of the power consumption results of the two studies.

	Power Consumption [kW]		
Node	Require power in the test system	Results from the EMTP_RV analysis	% error
Train 3	200.0	200.0	0.00
DC Source	-835.09 (before internal resistance of the rectifier)	-835.07	0.00
Train_2	-627.2	-627.2	0.00
Train_1	1184.3	1184.3	0.00

5 Conclusion

The rail potentials in a DC traction power supply system can be analyzed by applying the node voltage analysis method twice consecutively with the proposed algorithm. This algorithm was validated in section 4. The advantage of the algorithm is that we can avoid the mesh current method in the rail potential analysis, which is difficult to use because assumed initial mesh currents must be used in the iterative circuit analysis. The mesh current method also requires more than twice the computer memory required for the node analysis method, because the number of meshes is more than twice the number of nodes.

References

[1] Jian Guo Yu B.Sc.: 'Computer Analysis of touch voltages and stray currents for DC railways', thesis for Ph.D. University of Birmingham, July, 1992.
[2] Pham, K.D., Thomas, R.S. and Stinger, W.E. 'Operational and safety considerations in designing a light rail DC traction electrification system', Proc. 2003 IEEE/ASME Joint, PP.171-189.
[3] IEC62128-1: 'Railway applications – Fixed installations – Part 1: Protective provisions relating to electrical safety and earthing'.

This page intentionally left blank

Author Index

This page intentionally left blank

WIT PRESS ...for scientists by scientists

Computational Methods and Experimental Measurements XIV

Edited by: ***C.A. BREBBIA****, Wessex Institute of Technology, UK and* ***G.M. CARLOMAGNO****, University of Naples, Italy*

As computer models become more reliable and able to represent more realistic problems, detailed data is leading to the development of appropriate new types of experiments.

Experimental measurements need to be conditioned to the requirements of the computational models. Hence it is important that scientists working on experiments communicate with researchers developing computer codes, as well as those carrying out measurements on prototypes. The orderly and progressive concurrent development of all these fields is essential for the progress of engineering sciences.

This book resulted from the latest in a series of successful meetings that provide a unique forum for the review of the latest work on the interaction between computational methods and experimental measurements. The topics include: Computational and Experimental Methods; Experimental and Computational Analysis; Direct, Indirect and In-Situ Measurements; Detection and Signal Processing; Data Processing; Fluid Flow; Heat Transfer and Thermal Processes; Material Characterization; Structural and Stress Analysis; Industrial Applications; Forest Fires.

WIT Transactions on Modelling and Simulation, Vol 48
ISBN: 978-1-84564-187-0 eISBN: 978-1-84564-364-5
Published 2009 / 672pp / £255.00