Mathematics in Industry 19
The European Consortium for Mathematics in Industry

Magnus Fontes
Michael Günther
Nicole Marheineke *Editors*

Progress in Industrial Mathematics at ECMI 2012

Springer

MATHEMATICS IN INDUSTRY **19**

For further volumes:
http://www.springer.com/series/4650

Magnus Fontes • Michael Günther •
Nicole Marheineke
Editors

Progress in Industrial Mathematics at ECMI 2012

Springer

Editors

Magnus Fontes
Lunds Universitet, Centre for Mathematical Sciences
Lund
Sweden

Michael Günther
Bergische Universität Wuppertal, Applied Math. and Numerical Analysis
Wuppertal
Germany

Nicole Marheineke
Friedrich-Alexander-Universität Erlangen
Department of Mathematics
Erlangen
Germany

ISSN 1612-3956
ISBN 978-3-319-05364-6 ISBN 978-3-319-05365-3 (eBook)
DOI 10.1007/978-3-319-05365-3
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014939511

Mathematics Subject Classification (2010): 34-XX, 35-XX, 49-XX, 60-XX, 62-XX, 65-XX, 68-XX, 70-XX, 74-XX, 76-XX, 87-XX, 80-XX, 81-XX, 86-XX, 91-XX, 92-XX, 94-XX

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

ECMI celebrated its 25th anniversary with its 17th conference held in Lund, Sweden during what turned out to be the most beautiful summer week in the south part of Sweden during 2012. With around 170 attendees, it was a conference on the smaller side but maybe this, together with the festivities, contributed to a very creative and open atmosphere that characterized it.

Showing what Mathematics in Industry can signify, these proceedings reflect many of the topics presented and discussed during the five intense days of the meeting: 23–27 July 2012. The breadth of the conference can also be appreciated by reviewing some of the keynote talks, such as Kuna Huisman's *Decision Making under Uncertainty*, Fahime Nekka's *Information-Loaded Formalism to Assess the Causal Effect of Drug Intake on Therapeutic Outcomes*, Carsten Othmar's *Adjoint Methods for Car Aerodynamics*, or Alistair Fitt's memorable closing lecture on the *Modelling of Disease and Medical Procedures in the Human Eye*, where Alistair, among other things, told us what a truly applied mathematician must be ready to go through to get enough data to test a model. Medical experiments, that at the very least sounded very uncomfortable, performed on your self being part of the criteria for success was without doubt a novelty for many attendees in the auditorium.

In the middle of the conference week, during the half-day anniversary session, Helmut Neunzert gave a talk on the *History of ECMI*, i.e., background, the Mussbach meeting, general impact on education and society. For some in the audience, Helmut's talk slightly lifted the veil to legendary times, heard of, but never experienced, and for others it brought back good memories of grand days when they participated in something strikingly new with a circle of friends.

The 2012 Anile-ECMI prize was awarded during the conference to Franceso Ferranti of Ghent University for his work on *Parameterized Macromodeling and Model Order Reduction of High-Speed Interconnects*.

Many other inspiring talks were given during the week in the many parallel mini-sessions, new problems were identified and started to be attacked during coffee breaks and evening dinner discussions, and new collaborations were initiated. If you were not there, I guess that you have to ask someone who was. Or, you can, of course, read these proceedings reflecting the 25th anniversary conference of ECMI.

Lund, Sweden	Magnus Fontes
Wuppertal, Germany	Michael Günther
Erlangen, Germany	Nicole Marheineke

25 Years of ECMI

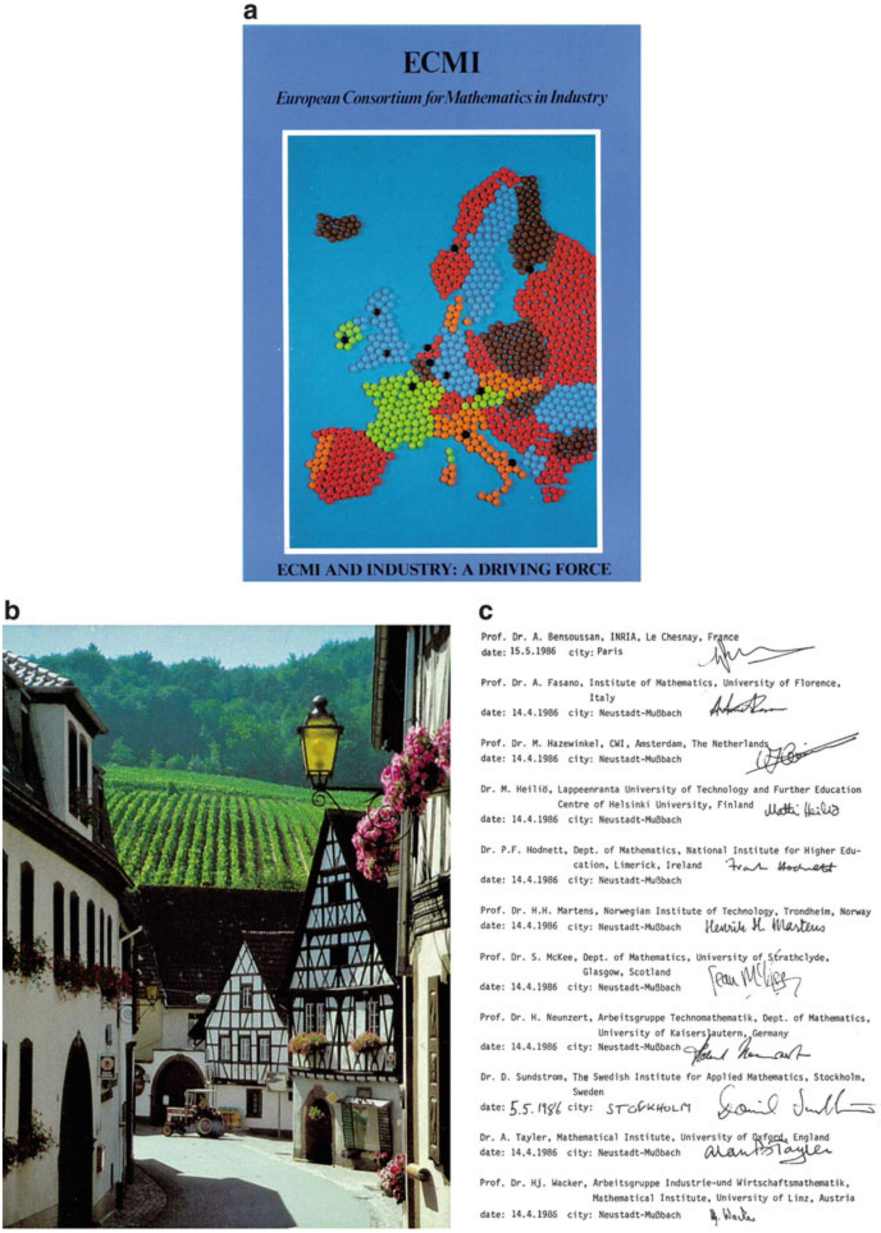
A View Back to Its Childhood

ECMI celebrated its 25th birthday during summer 2012; it is now a pretty young lady (of course: ECMI is a woman!), quite strong and active. As it sometimes happens at a birthday party, one of the grandfathers (ECMI has more than two) talks about the past, a lively childhood, noteworthy escapades. Now, please join me as I look back. Like a human being, an organization's identity is shaped by its history. I will share with you the story of ECMI and mediate on its purpose.

The story begins in Western Europe in 1985. The entire region was dominated by pure mathematicians, interested in algebra, topology, geometry, analysis, etc. The whole region? Well, no. Some small groups resisted, some "black points" on the ivory tower could be found: People, who tried to escape, even cooperated with industry: In Oxford, Linz and Kaiserslautern, Firenze and Bari, Eindhoven and Amsterdam, Trondheim and Lappeenranta, Glasgow and Limerick. Unfortunately, at that time, the Iron Curtain was still shut and the contacts with people from Eastern Europe, who had similar ideas, were rare.

In Autumn 1985, Michiel Hazewinkel and Bob Mattheij called for a symposium in Amsterdam and many came. Most were from the Netherlands, then the UK and Ireland, some came from Italy, Germany and Austria, a few from Scandinavia, and there was even a Slovene and a Pole. In the end, we thought it would be useful to found a European organization, and I invited representatives of the countries to a wine village in the Rhine valley (Mußbach). All invited came, worked a day and a night—with the exception of some wine tasting—and by the end signed a document (Fig. 1).

Bensoussan, by then INRIA president, was not much interested in education—he and France left the ECMI family soon. Sundstrom changed his career, but Sweden soon came with new people. Hodnett, Martens, Tayler, and Wacker participated in the bringing-up of ECMI very much, but died during these 25 years—I will come back to them a bit later. Fasano, Hazewinkel, Heilio, McKee, and I were present at this birthday party in Lund.


a
ECMI
European Consortium for Mathematics in Industry
ECMI AND INDUSTRY: A DRIVING FORCE
b
c

Prof. Dr. A. Bensoussan, INRIA, Le Chesnay, France
date: 15.5.1986 city: Paris

Prof. Dr. A. Fasano, Institute of Mathematics, University of Florence,
Italy
date: 14.4.1986 city: Neustadt-Mußbach

Prof. Dr. M. Hazewinkel, CWI, Amsterdam, The Netherlands
date: 14.4.1986 city: Neustadt-Mußbach

Dr. M. Heiliö, Lappeenranta University of Technology and Further Education
Centre of Helsinki University, Finland Matti Heiliö
date: 14.4.1986 city: Neustadt-Mußbach

Dr. P.F. Hodnett, Dept. of Mathematics, National Institute for Higher Education, Limerick, Ireland Frank Hodnett
date: 14.4.1986 city: Neustadt-Mußbach

Prof. Dr. H.H. Martens, Norwegian Institute of Technology, Trondheim, Norway
date: 14.4.1986 city: Neustadt-Mußbach Henrik H. Martens

Prof. Dr. S. McKee, Dept. of Mathematics, University of Strathclyde,
Glasgow, Scotland
date: 14.4.1986 city: Neustadt-Mußbach

Prof. Dr. H. Neunzert, Arbeitsgruppe Technomathematik, Dept. of Mathematics,
University of Kaiserslautern, Germany
date: 14.4.1986 city: Neustadt-Mußbach

Dr. D. Sundstrom, The Swedish Institute for Applied Mathematics, Stockholm,
Sweden
date: 5.5.1986 city: STOCKHOLM

Dr. A. Tayler, Mathematical Institute, University of Oxford, England
date: 14.4.1986 city: Neustadt-Mußbach Alan B Tayler

Prof. Dr. Hj. Wacker, Arbeitsgruppe Industrie-und Wirtschaftsmathematik,
Mathematical Institute, University of Linz, Austria
date: 14.4.1986 city: Neustadt-Mußbach

Fig. 1 Birth of ECMI. (**a**) The first ECMI-map (The *black pearls* indicate the ECMI-places). (**b**) The cradle of ECMI: Mußbach, Germany. (**c**) The 11 (grand-) fathers of ECMI

Hazewinkel drafted the founding paper, where we agreed on the main goals.

The Charter Goals of ECMI. To promote and further the effective use of mathematics and closely related knowledge and expertise in industrial and management settings. More specifically:

- Research: what is needed by industry and commerce, what is available, and what can be done to fill up the gaps (database)
- Creation, organization, and quality control of a 2-year postgraduate course on industrial and possibly management mathematics
- To encourage joint research ventures among the participating institutions.

Of course, there were many more points about conferences, newsletters, etc. Hazewinkel's preferences are visible—the database was his favorite topic, but never realized. Nevertheless, he was the first ECMI president.

And then we started working, first about topic one, but soon the focus shifted to education. The working group—the founders and some others—met four times during ECMI's first year, for example in Oberwolfach and Oxford. We discussed details, such as an ECMI educational program should contain a course on PDE. But what does "on PDE" really mean? More theory, more modelling, more numerics? We realized how different the mathematical cultures in different European countries were (are?)—each PDE course, in Italy, France, Germany, and the UK was quite special. I myself learned from the Oxford people how important asymptotic analysis is—whether they realized how important good numerics is for solving industrial problems, I am not even sure today. *When computing begins, thinking ends*, formulated my old friend Alistar Fitt, reflecting the Oxford opinion. Anyhow, we learned from each other, we found good compromises, we established a quality control for the educational programs checking spirit, student mobility, the use of languages at different universities.

The first approved centers were Linz, Kaiserslautern, Oxford, Bari, and Eindhoven; Trondheim, Helsinki, Lungby, Milano, and others followed soon.

It's important to note that there were quite different motivations for starting an industrial math program. Why do we do, what we do? It is not at all easy to establish an industrial math group or program. Just to use the name since it is "politically correct" is not fair and it is not at all the idea of ECMI. Often, one has to leave a safe career where you go on generalizing results you discovered during your Ph.D. time. You have to leave the protected area of the university, to expose yourself to the outer world, where your competence is not accepted a priori. You have to enter new mathematical fields, since the problem you find does not fit into your research area. You have to develop new teaching methods like modelling seminars, etc. You have to fight with your colleagues who are most often very conservative. Why do you want to do all that?

It may be interesting to detail some of the motivations of some ECMI founders. Since those who are still active and have been present in Lund can speak for themselves, I will take a closer look at three very important grandfathers, who passed away during the first 10 years of ECMI, but influenced it very much: Hansjörg Wacker, Henrik Martens, and Alan Tayler (Fig. 2).

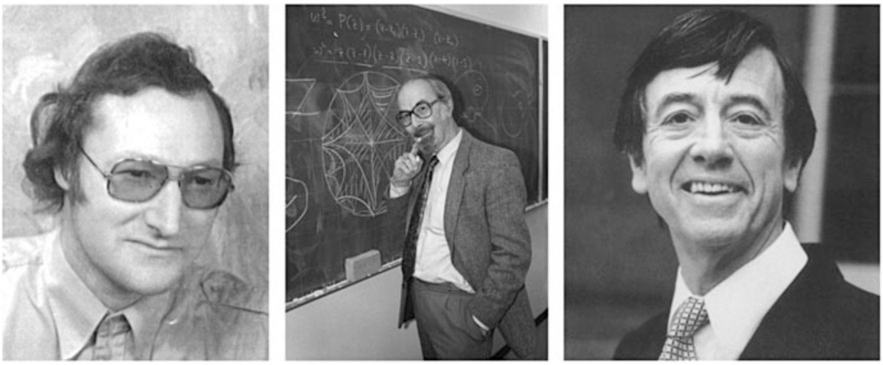


Fig. 2 Hansjörg Wacker, Henrik Martens and Alan Tayler

Hansjörg Wacker, ECMI President 1989–1990. Hansjörg Wacker (1939–1991) was educated at the Technical University of Munich (TUM) and became full professor at Johannes Kepler University in Linz in 1973.

His competence was in numerical analysis, more specifically continuation methods. Why did he turn to industrial mathematics during the 1970s? According to his student and later his successor Heinz Engl, it was his most important goal to encourage mathematicians to come out of the ivory tower and to start cooperating with industry, thus increasing the chances of math students for interesting employments in industry. This was, indeed, his main concern: To help students, to give them better chances. Many of us in Germany or Austria had studied 10 to 15 years after World War II: There were little chances for math graduates. We expected to become high school teachers—what else? One had to convince the slowly recovering industry that it was worthwhile to hire mathematicians. In Linz, there was—and still is—a big steel company that could be convinced—and Hansjörg succeeded. He was a man to visit companies, to form student groups who work on problems discovered there, to inspire young colleagues. He was also very active in helping foreign students—we, in Linz and Kaiserslautern, had the idea to help foreigners in order to correct the image of the past (Fig. 3).

It was only consequent that ECMI established a Wacker prize for young students. The first winner was Joachim Weickert, today a leading image processing expert in Germany, who got one of the most prestigious German prizes, the Leibniz-prize, 3 years ago. One of the editors of this volume, Nicole Marheineke, also got the Wacker prize in 2002.

Henrik Martens, ECMI President 1993. Henrik Martens (1927–1993) was born in a little town on the Norwegian coast and escaped at age 15 from the German occupation. He worked as a radio telegraphist on a ship, settled in New York in 1949, and worked as a technical assistant at Bell Labs, studying at the same time to graduate as an electrical engineer at 29. While at Bell Labs, he began to study mathematics and got his Ph.D. in 1962 at the Courant Institute. He was then a very pure mathematician, specializing in compact Riemannian manifolds. In 1968,

Fig. 3 Hansjörg Wacker with Ferryanto from Indonesia. Ferryanto, who studied in Kaiserslautern and Linz, is now head of quality control at Ford in the USA

he returned to Trondheim and was appointed as professor at NTH. At Trondheim he began to build a bridge between pure mathematics and engineering. I met him 1985 in Trondheim, we became close friends, vividly discussing mathematics, politics, literature, philosophy while hiking in the Norwegian mountains. He died in 1993, when returning from an ECMI board meeting at Como. His last activity was to see Leonardo da Vinci's "The Lord's Last Supper."

In 1986 he explained his motivation to start with industrial mathematics in a paper called "The task ahead": "We are witnessing an increasing invasion of mathematics and mathematicians into the engineering environment. How do we prepare our students for such a task? If we want to understand the issue, it seems reasonable first to make some effort to understand how mathematics enters engineering as a discipline, as a profession, and as an educational task.

"As a discipline [. . .]: My point, however, was to emphasize that the mathematics that is relevant for contemporary technological problems extends far beyond the boundaries of traditional applied mathematics, and it seems to me that all this—whatever mathematics is useful to technology—deserves the (new) heading industrial mathematics".

"As a profession: The primary goal is to solve practical problems, not to prove theorems [. . .]. We must convince ourselves that it is indispensable to simplify and systematize the results obtained in a discipline in order to permit the most easiest access for those who cultivate different (other) disciplines. [. . .] Modern technology is a source of potentially interesting problems that it would be a serious mistake to turn ones back on them."

"As an educational task: It implies that we must find ways to train the students in the important and difficult art of modelling. It implies that they must get experience in team work and communication. Above all, it implies that they should be exposed to a mathematical environment where these topics are cultivated and regarded as important, for it is through the social mechanisms of the environment that attitudes are transmitted. It is hardly possible to teach mathematics to users without a realistic understanding of how mathematics is used. I think it can be put quite simply: To seek

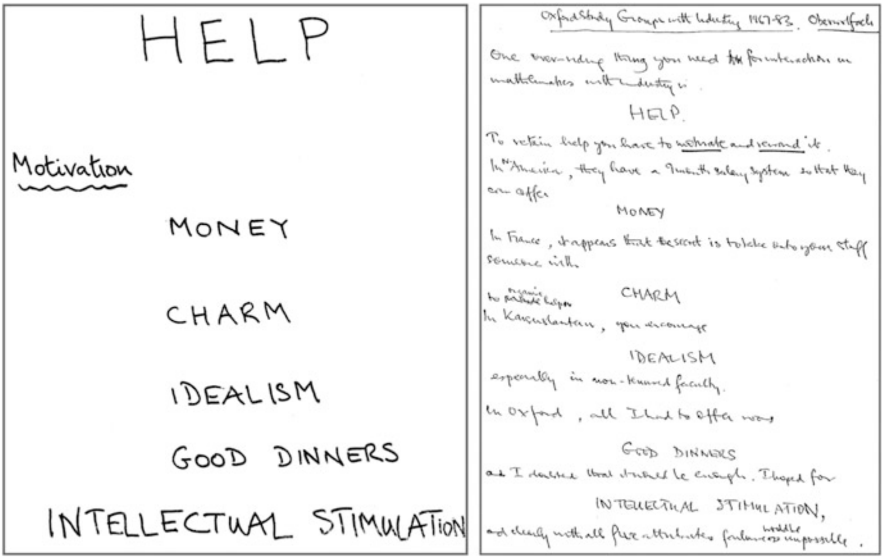
Fig. 4 "The first recorded instance of industrial mathematics occurred more than 2,000 years ago in Syracuse when Archimedes ran naked through the streets yelling Eureka."; "We may well see the emergence of industrial mathematics as member of a new technological profession, solidly rooted in the mathematical sciences, but with its own professional profile and goals." (Martens)

out whatever mathematics is relevant to technology, and make it available to the engineer."

This is a whole program—Henrik's motivation was to give (even pure) mathematics the role in technology it earns. He himself was an educated engineer who turned into a pure mathematician. His dream was to bring the two sides of his thinking finally together. I remember his ironic smile when showing him this picture, especially about the emergence (in a literal sense) of the first industrial mathematician, cf. Fig. 4.

Alan Tayler. Alan Tayler (1931–1995) initiated the study groups with industry in Oxford in 1967, 20 years before ECMI. He was the first and a real pioneer for industrial mathematics in Europe. Why did he leave the ivory tower at such a prestigious university as Oxford, and a very comfortable one at that?

In 1988 he wrote: "At this time there was much discussion worldwide about the role of applied mathematics, and in Britain the Royal society produced a report on the future of applied mathematics. This report was strongly influenced by the Cambridge group and one of its proposals was to require the applied mathematician to carry our experiments as part of research work, thus ensuring that theory and application did not diverge. As a young lecturer, with some experience of very bright mathematics students, this seemed to be totally inappropriate for Oxford where high intellect often was paired in the same individual with practical incompetence, and where graduate students waited to use their mathematics as soon as possible. Nevertheless, I recognized the need for their problems, and mine, to be real ones and not extensions to text book exercises. I looked therefore at the problems of interest to faculty members in physics, engineering and chemistry, but found my colleagues too busy or poorly motivated to communicate with me, an applied mathematician with little knowledge of their specialist topic and rather dubious mathematical skills. They would ask me to solve an equation, perform a mathematical manipulation, explain the pitfalls of the use of complex variables, but would not give me their



HELP

Motivation

MONEY

CHARM

IDEALISM

GOOD DINNERS

INTELLECTUAL STIMULATION

Oxford Study Groups with Industry 1967-83. Oberwolfach

One over-riding thing you need ~~for~~ for interaction in mathematics with industry is

HELP.

To retain help you have to motivate and reward it.

In N. America, they have a 9 month salary system so that they can offer

MONEY

In France, it appears that the secret is to take onto your staff someone with

CHARM

to ~~persuade~~ organise helpers

In Kaiserslautern, you encourage

IDEALISM

especially in non-tenured faculty.

In Oxford, all I had to offer was

GOOD DINNERS

and I doubted that it would be enough. I hoped for

INTELLECTUAL STIMULATION,

and clearly with all five attributes failure ~~was~~ would be impossible.

Fig. 5 Alan Tayler's arguments for an industrial math group

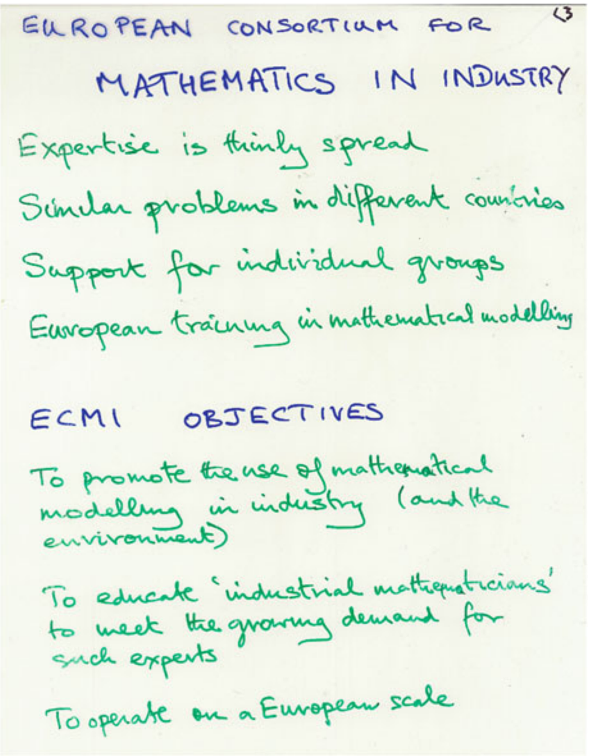
problems, which is all very understandable. So I had to look elsewhere and decided to approach research workers in industry and government laboratories."

I believe that this is a perfect description of what many of us experienced—and may still experience. No, our engineering colleagues will not give us their problems—we have to find them ourselves. And he did not want to repeat textbook exercises, they were boring. He wanted intellectual stimulation, new and exciting problems. And, in fact, at an Oberwolfach conference in 1983, he showed in a transparency how he was asking for help in establishing an industrial math group and promising rewards for that help. These are the rewards in his own handwriting (Fig. 5).

For him, mathematics was not the same as for Hansjörg and Henrik. He spoke mainly about modelling, of asymptotic analysis. Numerics and algebraic geometry were not the objects of his interest. A bit later, he explained, why we need a European consortium—and it still holds (Fig. 6). Alan was extremely convincing, he spread enthusiasm, was very friendly and inspiring. He brought Oxford into ECMI and shaped ECMI very much.

The Three Main Reasons for ECMI. Now we have seen three main reasons for industrial mathematics and for ECMI, personalized in three outstanding persons:

- Improving the chances for students,
- Promoting mathematics in society,
- Intellectual stimulation.



EUROPEAN CONSORTIUM FOR

MATHEMATICS IN INDUSTRY

Expertise is thinly spread

Similar problems in different countries

Support for individual groups

European training in mathematical modelling

ECMI OBJECTIVES

To promote the use of mathematical modelling in industry (and the environment)

To educate 'industrial mathematicians' to meet the growing demand for such experts

To operate on a European scale

Fig. 6 Alan Tayler's view on ECMI

Fig. 7 Further ECMI-supporters: Sten Ackermans (1936–1995), Marcello Anile (1948–2007), and Frank Hodnett (1939–2011) with Helmut Neunzert and others in Limerick (1991)

These reasons are still valid today, 25 years after ECMI's birth. All ECMIsts, now and in the future, should follow them. We see: To start industrial mathematics at your university is not an easy way to political correctness, but a difficult and hard way to a better math department.

Before ending, I would like to mention some other important ECMI-supporters, who passed away: Sten Ackermans (1936–1995), Marcello Anile (1948–2007), and Frank Hodnett (1939–2011) (Fig. 7). They helped to establish ECMI in their countries, in Eindhoven, Catania, and Limerick. They all gave ECMI a special flavor, worked enthusiastically for ECMI—we miss them all. Marcello was also a personal friend.

ECMI has an impressive past and I am sure, it has a good future. The "E," the "C" and the "M"—all are very strong. What we have to work for is to strengthen the "I." Without the "I," we loose our uniqueness. But as "whole" ECMI, we are unsubstitutable, we are unmistakable, we are—simply needed. Good luck, ECMI, for the next 25 years and beyond.

Kaiserslautern, Germany

Helmut Neunzert

Contents

Part I Circuits and Electromagnetic Devices

Part II Environment

Part III Fibers

Part IV Flow

Part V Medicine

Part VIII Methods

Part IX Education

Part I
Circuits and Electromagnetic Devices

Overview

The section on *Circuits and Electromagnetic Devices* contains five contributions. When going to nanoscales more details have to be modeled, for which additional unknowns and corresponding equations have to be introduced. The phenomena show dynamical behaviour and the equations involve nonlinear couplings. On the one hand the mathematical properties of the overall system have to be reviewed. For instance, do the new equations influence familiar statements, known at the macro scale, on well-posedness? Can we cast this into topological checks on the structure of the circuit network? How do we deal with multiscale effects, in time direction also known as multirate behaviour? How do we model transport in full details to allow for simulation of new materials? Can we identify local hotspots? How can we efficiently deal with temperature effects?

The first paper covers modeling of nanoscale circuit devices that exhibit memory. To model these within a circuit network additional unknowns have to be introduced that deal with the time history. This leads to extended DAEs (Differential-Algebraic Equations). The second paper considers index analysis for DAE systems, coming from branch-oriented modeling (in contrast to nodal analysis), or from hybrid circuit models, based on spanning tree concepts. The third paper concerns a multi-dimensional generalization in modeling transport through a heterojunction between materials in nanoscale organic photovoltaic devices. This is modeling on PDE-level. Here coupled multiscale features arise. The fourth paper concerns the simulation of nanoscale MOSFETs. Crystal heating is essential for a proper performance of the device. The last paper also includes coupling to heat, this time in electromagnetic heating. New techniques from dynamic iteration demonstrate that co-simulation can efficiently deal with the multirate time behaviour.

The paper by Ricardo Riaza: *Normal Hyperbolicity of Manifolds of Equilibria in Nonlinear Circuits with Mem-Devices*, deals with the network modeling of memory effects in resistors, capacitors and inductors. The new elements are called memristors, memcapacitances, and meminductors. For each device the effects are both nonlinear and dynamic, which can be modeled by an explicit linear differential equation and a nonlinear algebraic equation. The differential equation covers the time history of the memory effect. By this, the overall system of equations governing the circuit network are again a system of DAEs. For the well-posedness of the DC-problem (equilibrium point) conditions are formulated to guarantee that the manifold of such equilibria is normally hyperbolic and that it attracts all nearby trajectories. The paper has an extensive list of references.

Index analysis for DAEs of circuit networks usually is tuned to DAEs that come from Modified Nodal Analysis. In the paper by Ignacio García de la Vega and Ricardo Riaza: *Index Analysis of Branch-Oriented and Hybrid Models of Non-Passive Circuits*, index analysis is considered for DAEs that come from branch-oriented circuits or from hybrid formulations that are based on spanning tree concepts. For the first class of DAEs a complete characterization of index one and of index two is presented. For the second class of DAEs the cases of index zero and index one are treated.

In the paper by Matteo Porro, Carlo de Falco, Riccardo Sacco and Maurizio Verri: *Multiscale Modeling of Heterojunction Organic Photovoltaic Devices*, the modeling leads to a system of semilinear PDEs and ODEs. The dynamics of the excitation phenomena in the bulk leads to a parabolic problem that is coupled to an ODE that involves dissociation/recombination of excitations, electrons and holes into bonded pairs at the materials interface between the acceptor domain and the donor domain. In the acceptor domain, transport of photogenerated electrons is described by a second parabolic problem. In the donor domain there is a third parabolic problem for the holes transport. Time-domain simulations for planar device geometries, including a complex interface morphology, are shown.

The paper by Camiola V. Dario, Mascali Giovanni and Romano Vittorio: *Simulation of Nanoscale Double-Gate MOSFETs*, considers subband modeling based on the maximum entropy principle (MEP). Crystal heating is included, by which the electrical properties of the device can be affected. By hot electrons a phonon hot spot can be created, which increases the power density generated by the integrated circuits. The crystal heating is involved by the lattice temperature that enters the electron-phonon scattering and the production terms of the balance equations for the electron variables. The charge transport in the subbands involves non-parabolic effects through the Kane dispersion relation. Simulations are shown for a MOSFET configuration with an upper gate and a lower gate and with source and drain at the left and at the right, respectively. Time-domain simulation exploits ADI (Alternating Direction Implicit) techniques.

In the paper by Christof Kaufmann, Michael Günther, Daniel Klagges, Matthias Richwin, Sebastian Schöps and E. Jan W. ter Maten: *Coupled Heat-Electromagnetic Simulation of Inductive Charging Stations for Electric Vehicles*, efficient simulation of heat coupled to electromagnetic fields is considered. The interest by industry is clearly reflected in the author list. Co-simulation is a well-established technique to exploit multi-rate time integration. Here, within specific time windows, the solution of the electromagnetic field can well be approximated by solving it in the frequency domain. This solution can be improved by iteration techniques, similar to Dynamic Iteration. The current analysis involves only the dominant fourier mode of the solution of the electromagnetic field, but it can easily be extended to include more modes (like in Harmonic Balance). The authors also point out generalization to systems that can be formulated with two time scales in the time domain.

E. Jan W. ter Maten

Normal Hyperbolicity of Manifolds of Equilibria in Nonlinear Circuits with Mem-Devices

Ricardo Riaza

Abstract The memristor and other mem-devices are displaying a great impact on modern electronics. We examine in this communication certain dynamical features of circuits with memristors, memcapacitors and meminductors, related to the systematic presence of non-isolated equilibria in these nonlinear circuits.

1 Introduction

The *memory-resistor* or *memristor* is a nonlinear electronic device defined by a nonlinear charge-flux relation, whose existence was predicted by Chua in 1971 [4]. A device with a memristive characteristic was actually designed at the nanometer scale in 2008 [30]. The potential applications of this device in the design of non-volatile memories, pattern recognition, adaptive and learning systems, signal processing, etc., might make the memristor and related devices play a very significant role in electronics in the near future, specially at the nanometer scale. A lot of research is focused on this topic; cf. [2, 3, 6, 12–24, 26, 29, 31]. HP has announced that commercial memory chips based on the memristor will be released in 2013 [1]. The idea of a device with memory was extended to the reactive context in 2009 by introducing *memcapacitors* and *meminductors* [7].

In this communication we examine certain local dynamical features of circuits with such mem-devices. The form of the memristor constitutive relation is known to be responsible for the existence of a center manifold of equilibria, as detailed later; see e.g. [17]. These non-isolated equilibria will be displayed also by circuits with memcapacitors and meminductors. Along the lines of the pioneering work of Fiedler et al. [8–11], given an m-dimensional manifold of equilibria in any C^1,

R. Riaza (✉)
Depto. Matemática Aplicada TTI, ETSI Telecomunicación, Universidad Politécnica de Madrid, Ciudad Universitaria s/n, 28040 Madrid, Spain
e-mail: ricardo.riaza@upm.es

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_1,

continuous-time dynamical system, it is of interest to examine the *normal hyperbolicity* of such a manifold and the existence of *bifurcations without parameters* when the normal hyperbolicity fails. The idea is that at least m eigenvalues of the linearization about any of these equilibria do necessarily vanish: the manifold is said to be normally hyperbolic if the remaining eigenvalues are not in the imaginary axis.

The analysis proceeds in two steps. Section 2 addresses the aforementioned dynamical properties for circuits with memristors. Memcapacitors and meminductors are included in Sect. 3, and Sect. 4 briefly compiles some concluding remarks.

2 Circuits with Memristors

For the sake of simplicity we begin the analysis by considering circuits whose memdevices are only of memristive type. We will focus the attention on flux-controlled memristors, defined by a relation of the form $q = \xi(\varphi)$ [4]. By differentiating this relation we get the current-voltage characteristic $i = W(\varphi)v$, where $W(\varphi) = \xi'(\varphi)$ is the so-called *memductance* and depends on $\varphi = \int_{-\infty}^{t} v(\tau)d\tau$. The memory effect arises from the fact that the memductance (a generalization of the conductance of a nonlinear resistor) keeps track of the device history because of its dependence on an integral variable.

In terms of the loop and cutset matrices B, Q (see e.g. [5, 25]), such a circuit is modelled by the differential-algebraic system

$$C(v_c)v_c' = i_c \tag{1a}$$

$$L(i_l)i_l' = v_l \tag{1b}$$

$$\varphi_m' = v_m \tag{1c}$$

$$0 = i_m - W(\varphi_m)v_m \tag{1d}$$

$$0 = i_r - \gamma(v_r) \tag{1e}$$

$$0 = B_c v_c + B_l v_l + B_m v_m + B_r v_r + B_u V_s + B_j v_j \tag{1f}$$

$$0 = Q_c i_c + Q_l i_l + Q_m i_m + Q_r i_r + Q_u i_u + Q_j I_s. \tag{1g}$$

The loop matrix B is split as $(B_c\ B_l\ B_m\ B_r\ B_u\ B_j)$ in the statement of Kirchhoff laws within (1f) and (1g); note that B_c (resp. B_l, B_m, B_r, B_u, B_j) corresponds to the columns accommodating capacitors (resp. inductors, memristors, resistors, voltage sources, current sources). The same applies to the cutset matrix Q. Additionally, C and L stand for the capacitance and inductance matrices; resistors are assumed to be voltage-controlled by the characteristic $i_r = \gamma(v_r)$ and, for later use, the conductance matrix $\gamma'(v_r)$ will be denoted as G. Finally, V_s and I_s are the (DC) sources.

The main result to be reported in this communication is the one stated in Theorem 1 below. It extends to the memristive context the graph-theoretic analysis of qualitative properties carried out in [27, 28]. Note that the presence of a manifold of equilibria is an easy consequence of the fact that the variable φ_m is not actually involved in the equilibrium conditions which follow from forcing the right-hand side of (1) to vanish.

Theorem 1. *Assume that the capacitance, inductance, memductance and conductance matrices C, L, W, G in (1) are positive definite and that C, L are symmetric. Suppose that at least one of the following two sets of topological conditions holds:*

- *the circuit does not have VC-loops, VL-loops or ICL-cutsets; or*
- *it does not have IL-cutsets, IC-cutsets or VCL-loops.*

Then, locally around any equilibrium point the model (1) defines a local flow whose dimension is defined by the number of memristors and reactive elements. Additionally, there is a manifold of equilibria whose dimension is given by the number of memristors and which is normally hyperbolic and attracts all nearby trajectories.

The proof begins by showing that the number of eigenvalues in the matrix pencil describing the linearization of (1) equals the number of memristors and reactive elements, with a zero eigenvalue whose geometric and algebraic multiplicities are given by the number of memristors. The remainder of the proof essentially proceeds along the lines of [27, 28]: the absence of ICL-cutsets or VCL-loops rules out the presence of purely imaginary eigenvalues, and the positive definite assumption on the circuit matrices implies that all non-vanishing eigenvalues are actually located in the left-hand complex plane.

3 Memcapacitors and Meminductors

The result reported above can be extended to circuits with reactive mem-devices [7]. A *memcapacitor* is a nonlinear device governed by a relation of the form

$$q = C_m(\varphi)v, \tag{2}$$

where the memcapacitance C_m depends on $\varphi = \int v$. A *meminductor* is defined by

$$\varphi = L_m(q)i, \tag{3}$$

and the meminductance L_m now depends on $q = \int i$. These devices can be added to the previous model to yield

$$C(v_c)v_c' = i_c \tag{4a}$$

$$L(i_l)i_l' = v_l \tag{4b}$$

$$\varphi_{mc}' = v_{mc} \tag{4c}$$

$$q'_{mc} = i_{mc} \tag{4d}$$

$$\varphi'_{ml} = v_{ml} \tag{4e}$$

$$q'_{ml} = i_{ml} \tag{4f}$$

$$\varphi'_{m} = v_{m} \tag{4g}$$

$$0 = q_{mc} - C_m(\varphi_{mc})v_{mc} \tag{4h}$$

$$0 = \varphi_{ml} - L_m(q_{ml})i_{ml} \tag{4i}$$

$$0 = i_m - W(\varphi_m)v_m \tag{4j}$$

$$0 = i_r - \gamma(v_r) \tag{4k}$$

$$0 = B_c v_c + B_l v_l + B_{mc} v_{mc} + B_{ml} v_{ml} + B_m v_m + B_r v_r + B_u V_s + B_j v_j \tag{4l}$$

$$0 = Q_c i_c + Q_l i_l + Q_{mc} i_{mc} + Q_{ml} i_{ml} + Q_m i_m + Q_r i_r + Q_u i_u + Q_j I_s. \tag{4m}$$

Now the subscripts mc and ml correspond to memcapacitors and meminductors, respectively. As before, equilibrium points are defined by the vanishing of the right-hand side of (4). It is easy to check that at equilibrium all voltages and currents in memristors, memcapacitors and meminductors are null, and so they are q_{mc} and φ_{ml} because of (4h)–(4i). It then follows that the number of null eigenvalues equals the total number of mem-devices. As in the memristive case, the normal hyperbolicity and exponential stability of the manifold of equilibria can be guaranteed in the absence of the topological conditions arising in Theorem 1, provided that all claims about capacitive (resp. inductive) elements are now understood to stand for capacitors and memcapacitors (resp. inductors and meminductors).

4 Concluding Remarks

The results here reported provide a general framework which explains some of the qualitative properties displayed by specific memristive circuits in [12, 17]. A systematic analysis of Hopf bifurcations without parameters in this context, extending the results obtained in [17] for a specific example, is in the scope of future research.

Acknowledgements Research supported by Project MTM2010 -15102 of Ministerio de Ciencia e Innovación, Spain.

References

1. Adee, S.: Memristor inside. IEEE Spectrum, September 2010
2. Bao, B., Ma, Z., Xu, J., Liu, Z., Xu, Q.: A simple memristor chaotic circuit with complex dynamics. Int. J. Bifurcat. Chaos **21**, 2629–2645 (2011)
3. Biolek, D., Biolek, Z., Biolkova, V.: SPICE modeling of memristive, memcapacitive and meminductive systems. In: Proceedings of European Conference on Circuit Theory and Design, pp. 249–252 (2009)
4. Chua, L.O.: Memristor – the missing circuit element. IEEE Trans. Circuit Theory **18**, 507–519 (1971)
5. Chua, L.O., Desoer C.A., Kuh, E.S.: Linear and Nonlinear Circuits. McGraw-Hill, New York (1987)
6. Corinto, F., Ascoli, A., Gilli, M.: Analysis of current-voltage characteristics for memristive elements in pattern recognition systems. Int. J. Circuit Theory Appl. **40**(12), 1277–1320 (2012)
7. Di Ventra, M., Pershin, Y.V., Chua, L.O.: Circuit elements with memory: memristors, memcapacitors and meminductors. Proc. IEEE **97**, 1717–1724 (2009)
8. Fiedler, B., Liebscher, S.: Generic Hopf bifurcation from lines of equilibria without parameters: II. Systems of viscous hyperbolic balance laws. SIAM J. Math. Anal. **31**, 1396–1404 (2000)
9. Fiedler, B., Liebscher, S.: Takens-Bogdanov bifurcations without parameters, and oscillatory shock profiles. In: Broer, H., et al. (eds.) Global Analysis of Dynamical Systems, pp. 211–259. IOP, Bristol (2001)
10. Fiedler, B., Liebscher, S., Alexander, J.C.: Generic Hopf bifurcation from lines of equilibria without parameters: I. Theory. J. Differ. Equ. **167**, 16–35 (2000)
11. Fiedler, B., Liebscher, S., Alexander, J.C.: Generic Hopf bifurcation from lines of equilibria without parameters: III. Binary oscillations. Int. J. Bifurcat. Chaos **10**, 1613–1622 (2000)
12. Itoh, M., Chua, L.O.: Memristor oscillators. Int. J. Bifurcat. Chaos **18**, 3183–3206 (2008)
13. Itoh, M., Chua, L.O.: Memristor cellular automata and memristor discrete-time cellular neural networks. Int. J. Bifurcat. Chaos **19**, 3605–3656 (2009)
14. Itoh, M., Chua, L.O.: Memristor Hamiltonian circuits. Int. J. Bifurcat. Chaos **21**, 2395–2425 (2011)
15. Jeltsema D., van der Schaft, A.J.: Memristive port-Hamiltonian systems. Math. Comput. Model. Dyn. Syst. **16**, 75–93 (2010)
16. Kavehei, O., Iqbal, A., Kim, Y.S., Eshraghian, K., Al-Sarawi, S.F., Abbott, D.: The fourth element: characteristics, modelling and electromagnetic theory of the memristor. Proc. R. Soc. A **466**, 2175–2202 (2010)
17. Messias, M., Nespoli, C., Botta, V.A.: Hopf bifurcation from lines of equilibria without parameters in memristors oscillators. Int. J. Bifurcat. Chaos **20**, 437–450 (2010)
18. Muthuswamy, B.: Implementing memristor based chaotic circuits. Int. J. Bifurcat. Chaos **20**, 1335–1350 (2010)
19. Muthuswamy, B., Chua, L.O.: Simplest chaotic circuit. Int. J. Bifurcat. Chaos **20**, 1567–1580 (2010)
20. Muthuswamy, B., Kokate, P.P.: Memristor-based chaotic circuits. IETE Tech. Rev. **26**, 417–429 (2009)
21. Pershin, Y.V., Di Ventra, M.: Practical approach to programmable analog circuits with memristors. IEEE Trans. Circuits Syst. I **57**, 1857–1864 (2010)
22. Pershin, Y.V., Di Ventra, M.: Experimental demonstration of associative memory with memristive neural networks. Neural Netw. **23**, 881–886 (2010)
23. Pershin, Y.V., Di Ventra, M.: Memory effects in complex materials and nanoscale systems. Adv. Phys. **60**, 145–227 (2011)
24. Pershin, Y.V., Di Ventra, M.: Neuromorphic, digital and quantum computation with memory circuit elements. Proc. IEEE **100**(6), 2071–2080 (2012)
25. Riaza, R.: Differential-Algebraic Systems. Analytical Aspects and Circuit Applications. World Scientific, Singapore (2008)

26. Riaza, R.: Nondegeneracy conditions for active memristive circuits. IEEE Trans. Circuits Syst. II **57**, 223–227 (2010)
27. Riaza, R., Tischendorf, C.: Qualitative features of matrix pencils and DAEs arising in circuit dynamics. Dyn. Syst. **22**, 107–131 (2007)
28. Riaza, R., Tischendorf, C.: The hyperbolicity problem in electrical circuit theory. Math. Methods Appl. Sci. **33**, 2037–2049 (2010)
29. Riaza, R., Tischendorf, C.: Semistate models of electrical circuits including memristors. Int. J. Circuit Theory Appl. **39**, 607–627 (2011)
30. Strukov, D.B., Snider, G.S., Stewart, D.R., Williams, R.S.: The missing memristor found. Nature **453**, 80–83 (2008)
31. Yang, J.J., Pickett, M.D., Li, X., Ohlberg, D.A.A., Stewart, D.R., Williams, R.S.: Memristive switching mechanism for metal/oxide/metal nanodevices. Nat. Nanotechnol. **3**, 429–433 (2008)

Index Analysis of Branch-Oriented and Hybrid Models of Non-passive Circuits

Ignacio García de la Vega and Ricardo Riaza

Abstract We extend in this communication previous index analyses of branch-oriented and hybrid circuit models to a non-passive context. Specifically, in the absence of coupling effects, we present a complete characterization of index one and index two branch-oriented models, and index zero and index one hybrid models. The results are based on the structure of the forests of certain circuit minors.

1 Introduction

Differential-algebraic equations (DAEs) are nowadays systematically used in circuit simulation programs. This is a consequence of the fact that automatic methods to set up circuit models in a nonlinear context naturally generate them as a combination of both differential and algebraic equations. This is the case of nodal analysis techniques, such as MNA, used in SPICE and its commercial variants [3–5, 10, 12].

In this context, a major problem is the characterization of the DAE *index* of the circuit model. The index determines the numerical techniques that can be used in the simulation of the dynamics and characterizes several analytical properties of the circuit. Much research in this direction has been focused on the characterization of the index of nodal models [2, 3, 10, 12]. Under passivity assumptions, the index of nodal models is known to be not greater than two, according to the results in [3, 12].

Recent research has been directed to so-called *hybrid* models, whose origin can be traced back to [8] and which have been recently framed in a differential-algebraic formalism [6, 7, 11]. The hybrid equations arise as a reduction of *branch-oriented* models [9, 10], which avoid the use of node potentials in the formulation of the model.

I. García de la Vega • R. Riaza (✉)
Depto. Matemática Aplicada TTI, ETSI Telecomunicación, Universidad Politécnica de Madrid, Ciudad Universitaria s/n, 28040 Madrid, Spain
e-mail: ricardo.riaza@upm.es

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_2,

The index analysis carried out in [9, 10] for branch-oriented models and in [6, 7, 11] for hybrid systems applies to passive circuits, namely, to problems in which all circuit matrices (capacitance, inductance, conductance and, in the eventual presence of memristors, memductance) are positive definite. In this communication we report an extension of these results to non-passive circuits, by means of a non-trivial modification of the techniques introduced in [2]. In particular, as detailed in Sect. 2, we accommodate in the analysis of branch-oriented models both voltage- and current-controlled resistors and also topologically degenerate configurations (leading to index two models), in contrast to [2] which only accounts for voltage-controlled resistors and topologically nondegenerate configurations. Section 3 extends the results to hybrid models.

2 Branch-Oriented Models

By expressing Kirchhoff laws in terms of a reduced cutset matrix Q and a reduced loop matrix B (cf. e.g. [1,10]), the branch-oriented model of a nonlinear RLC circuit with independent sources can be written as

$$C(v_c)v_c' = i_c \tag{1a}$$

$$L(i_l)i_l' = v_l \tag{1b}$$

$$0 = Q_r i_r + Q_g i_g + Q_l i_l + Q_c i_c + Q_u i_u + Q_j i_s(t) \tag{1c}$$

$$0 = B_r v_r + B_g v_g + B_l v_l + B_c v_c + B_u v_s(t) + B_j v_j \tag{1d}$$

$$0 = v_r - f(i_r) \tag{1e}$$

$$0 = i_g - g(v_g), \tag{1f}$$

where we are using the subscripts r and g for current-controlled and voltage-controlled resistors, respectively, whereas l, c, u and j correspond to inductors, capacitors, voltage sources and current sources.

Topologically nondegenerate configurations, characterized by the absence of VC-loops (that is, loops composed of voltage sources and/or capacitors) and IL-cutsets (cutsets defined by current sources and/or inductors) are known to make (1) index one in a passive context [9]. The analysis when some of the resistors may become locally non-passive (that is, when some of the components of the characteristics f and g above may become negative at certain regions) is more intricate; as an extension of the results in [2], Theorem 1 below provides a full index one characterization in terms of *proper trees*, namely, spanning trees comprising all voltage sources and capacitors and neither current sources nor inductors.

Theorem 1. *Let the capacitance and inductance matrices C, L be non-singular. In the absence of resistive coupling effects, the model (1) is index one if and only if:*

(a) the circuit exhibits neither VC-loops nor IL-cutsets; and

(b) the sum of product of the incremental conductances of voltage-controlled twig resistors and the incremental resistances of current-controlled link resistors in proper trees does not vanish.

VC-loops and IL-cutsets lead to topologically degenerate configurations, and a full characterization of index two models for nodal analysis of non-passive circuits was not feasible along the lines of [2]; actually, this is still an open problem. For branch-oriented models such an index two characterization is possible, as detailed in Theorem 2 below. In the statement of this result, we use three reduced circuits, namely: the *resistive minor* obtained after short-circuiting voltage sources and capacitors and open-circuiting current sources and inductors; the *capacitive minor* defined by short-circuiting voltage sources and open-circuiting all other circuit elements except for capacitors; and the *inductive minor* obtained after open-circuiting current sources and short-circuiting all other devices except for inductors.

Theorem 2. *Consider a topologically degenerate, well-posed circuit in which the capacitance and inductance matrices are non-singular and which displays no coupling effects. As in item (b) of Theorem 1, assume that the sum of conductance-resistance products in the forests of the resistive minor defined above does not vanish. Then the model (1) is index two if and only if*

(i) neither the sum of capacitance products in the forests of the capacitive minor,

(ii) nor the sum of inductance products in the coforests of the inductive minor

do vanish.

3 Hybrid Circuit Models

Similar techniques can be used to characterize the index of so-called hybrid circuit models. These arise as a reduction of the model (1), when Kirchhoff laws (1c) and (1d) are based on a *normal reference tree*, that is, a spanning tree chosen to comprise all voltage sources and no current source, to have as many capacitors as possible, to include (among the ones satisfying the previous requirements) as many voltage-controlled resistors as possible, and to have (among the previous ones) as many current-controlled resistors as possible. As a byproduct, such a tree has as few inductors as possible. For the sake of simplicity, we will disregard voltage and current sources, so that the definition of a normal tree only involves the requirements stated above for capacitors, inductors and resistors.

This working setting makes it possible to express Kirchhoff laws as

$$\begin{pmatrix} v_{c_{\mathrm{co}}} \\ v_{g_{\mathrm{co}}} \\ v_{r_{\mathrm{co}}} \\ v_{l_{\mathrm{co}}} \end{pmatrix} = - \begin{pmatrix} K_{11} & 0 & 0 & 0 \\ K_{21} & K_{22} & 0 & 0 \\ K_{31} & K_{32} & K_{33} & 0 \\ K_{41} & K_{42} & K_{43} & K_{44} \end{pmatrix} \begin{pmatrix} v_{c_{\mathrm{tr}}} \\ v_{g_{\mathrm{tr}}} \\ v_{r_{\mathrm{tr}}} \\ v_{l_{\mathrm{tr}}} \end{pmatrix}$$

and

$$\begin{pmatrix} i_{c_{\text{tr}}} \\ i_{g_{\text{tr}}} \\ i_{r_{\text{tr}}} \\ i_{l_{\text{tr}}} \end{pmatrix} = \begin{pmatrix} K_{11}^T & K_{21}^T & K_{31}^T & K_{41}^T \\ 0 & K_{22}^T & K_{32}^T & K_{42}^T \\ 0 & 0 & K_{33}^T & K_{43}^T \\ 0 & 0 & 0 & K_{44}^T \end{pmatrix} \begin{pmatrix} i_{c_{\text{co}}} \\ i_{g_{\text{co}}} \\ i_{r_{\text{co}}} \\ i_{l_{\text{co}}} \end{pmatrix},$$

where the subscripts $_{\text{tr}}$ and $_{\text{co}}$ specify tree and cotree elements in a given normal tree. As detailed in [6, 7, 11], hybrid models eliminate all variables except for $v_{c_{\text{tr}}}$, $v_{g_{\text{tr}}}$, $i_{r_{\text{co}}}$ and $i_{l_{\text{co}}}$, to write the circuit equations in the form

$$\begin{aligned} (C_{\text{tr}}(v_{c_{\text{tr}}}) + K_{11}^T C_{\text{co}}(-K_{11} v_{c_{\text{tr}}}) K_{11}) v'_{c_{\text{tr}}} &= K_{21}^T h_{\text{co}}(-K_{21} v_{c_{\text{tr}}} - K_{22} v_{g_{\text{tr}}}) + K_{31}^T i_{r_{\text{co}}} + K_{41}^T i_{l_{\text{co}}} \\ (L_{\text{co}}(i_{l_{\text{co}}}) + K_{44} L_{\text{tr}}(K_{44}^T i_{l_{\text{co}}}) K_{44}^T) i'_{l_{\text{co}}} &= -K_{41} v_{c_{\text{tr}}} - K_{42} v_{g_{\text{tr}}} - K_{43} f_{\text{tr}}(K_{33}^T i_{r_{\text{co}}} + K_{43}^T i_{l_{\text{co}}}) \\ h_{\text{tr}}(v_{g_{\text{tr}}}) &= K_{22}^T h_{\text{co}}(-K_{21} v_{c_{\text{tr}}} - K_{22} v_{g_{\text{tr}}}) + K_{32}^T i_{r_{\text{co}}} + K_{42}^T i_{l_{\text{co}}} \\ f_{\text{co}}(i_{r_{\text{co}}}) &= -K_{31} v_{c_{\text{tr}}} - K_{32} v_{g_{\text{tr}}} - K_{33} f_{\text{tr}}(K_{33}^T i_{r_{\text{co}}} + K_{43}^T i_{l_{\text{co}}}). \end{aligned} \tag{2}$$

As shown in [6, 7, 11], the index of this model does not exceed one in a passive context. Again, this result can be extended to a non-passive setting in terms of the spanning forests of the resistive, capacitive and inductive minors introduced above. We make use of the so-called *resistor-acyclic condition* introduced in [6, 7, 11], which captures the configurations in which every voltage-controlled resistor defines a loop together with some capacitors, and every current-controlled resistor defines a cutset together with some inductors, so that the model (2) has no algebraic equations.

Theorem 3. *In the absence of capacitive and inductive coupling, the hybrid model (2) is index zero if and only if the resistor-acyclic condition is met, and the sums arising in items (i) and (ii) of Theorem 2 do not vanish.*

If the resistor-acyclic condition is not met, and the sums of capacitance and inductance products arising in items (i) and (ii) of Theorem 2 do not vanish, then the hybrid model (2) is index one if and only if the condition on the sum of conductance-resistance products depicted in item (b) of Theorem 1 is met.

It is worth emphasizing that the elimination of, say, index two variables in the branch-oriented model (1) provides a set of (hybrid) equations which retain the same non-degeneracy requirements (namely, the ones stated in items (b) of Theorem 1 and items (i) and (ii) of Theorem 2) in the index analysis, with the key difference that in this case these requirements yield a model whose index does not exceed one.

Acknowledgements Research supported by Project MTM2010-15102 of Ministerio de Ciencia e Innovación, Spain.

References

1. Chua, L.O., Desoer, C.A., Kuh, E.S.: Linear and Nonlinear Circuits. McGraw-Hill, New York (1987)
2. Encinas, A., Riaza, R.: Tree-based characterization of low index circuit configurations without passivity restrictions. Int. J. Circuit Theory Appl. **36**, 135–160 (2008)
3. Estévez-Schwarz, D., Tischendorf, C.: Structural analysis of electric circuits and consequences for MNA. Int. J. Circuit Theory Appl. **28**, 131–162 (2000)
4. Günther, M., Feldmann, U.: CAD-based electric-circuit modeling in industry. I: mathematical structure and index of network equations. Surv. Math. Ind. **8**, 97–129 (1999)
5. Günther, M., Feldmann, U.: CAD-based electric-circuit modeling in industry. II: impact of circuit configurations and parameters. Surv. Math. Ind. **8**, 131–157 (1999)
6. Iwata, S., Takamatsu, M.: Index minimization of differential-algebraic equations in hybrid analysis for circuit simulation. Math. Program. Ser. A **121**, 105–121 (2010)
7. Iwata, S., Takamatsu, M., Tischendorf, C.: Tractability index of hybrid equations for circuit simulation. Math. Comput. **81**, 923–939 (2012)
8. Kron, G.: Tensor Analysis of Networks. Wiley, London (1939)
9. Reiszig, G.: The index of the standard circuit equations of passive RLCTG-networks does not exceed 2. In: Proceedings of the 1998 IEEE International Symposium on Circuits and Systems (ISCAS'98), vol. 3, pp. 419–422 (1998)
10. Riaza, R.: Differential-Algebraic Systems. Analytical Aspects and Circuit Applications. World Scientific, Singapore (2008)
11. Takamatsu, M., Iwata, S.: Index characterization of differential-algebraic equations in hybrid analysis for circuit simulation. Int. J. Circuit Theory Appl. **38**, 419–440 (2010)
12. Tischendorf, C.: Topological index calculation of DAEs in circuit simulation. Surv. Math. Ind. **8**, 187–199 (1999)

Multiscale Modeling of Heterojunction Organic Photovoltaic Devices

Matteo Porro, Carlo de Falco, Riccardo Sacco, and Maurizio Verri

Abstract In this communication, we present a computational model for heterojunction Organic Photovoltaic (OPV) devices consisting of a system of semilinear PDEs and ODEs. The mathematical model is discussed, focusing on the transmission conditions at material interfaces, together with the numerical method used for its solution. Steady-state and transient simulations are performed on realistic devices with various interface morphologies.

1 Introduction and Motivation

An important class of OPVs is that of Organic Solar Cells (OSCs). In the design of efficient OSCs the impact of material interface morphology on performance is currently considered to be of paramount importance. For this reason, material scientists are putting much of their research effort into techniques for controlling

M. Porro (✉)
Dipartimento di Matematica "F. Brioschi", Politecnico di Milano, Piazza L. da Vinci 32, 20133 Milano, Italy

Center for Nano Science and Technology @PoliMi, Istituto Italiano di Tecnologia, Via Pascoli 70/3, 20133 Milano, Italy
e-mail: matteo1.porro@polimi.it

C. de Falco
Dipartimento di Matematica "F. Brioschi", Politecnico di Milano, Piazza L. da Vinci 32, 20133 Milano, Italy

CEN - Centro Europeo di Nanomedicina, Piazza L. da Vinci 32, 20133 Milano, Italy
e-mail: carlo.defalco@polimi.it

R. Sacco • M. Verri
Dipartimento di Matematica "F. Brioschi", Politecnico di Milano, Piazza L. da Vinci 32, 20133 Milano, Italy
e-mail: riccardo.sacco@polimi.it; maurizio.verri@polimi.it

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_3,

interfaces down to the nanoscale, for example by studying materials that have the ability to self-assemble into ordered nanostructures during the deposition process. For the same reason, computational models that allow to estimate device performance carefully accounting for the material interface geometry and the phenomena occurring on it are in high demand. Previous approaches in this direction can be found in [1] (for biplanar devices) and [8]. In this communication we present our work aimed at extending the model of [1] to treat arbitrary multidimensional morphologies.

2 Mathematical Model

Let Ω be an open subset of $\mathbb{R}^d$, $d = 1, 2, 3$, representing the geometrical model of an OSC and $\boldsymbol{\nu}$ be the unit outward normal vector over the boundary $\partial\Omega$. The device structure is divided into two open disjoint subregions, Ω_n (acceptor) and Ω_p (donor), separated by a regular surface Γ on which $\boldsymbol{\nu}_\Gamma$ is the unit normal vector oriented from Ω_p into Ω_n. The cell electrodes, cathode and anode, are denoted as Γ_C and Γ_A, respectively (see Fig. 1 for the 2D case).

Let X, n and p denote the volumetric densities of excitons, electrons and holes in the cell, respectively, P be the areal density of bonded pairs and φ be the electric potential. For any function $f : \Omega \to \mathbb{R}$, let $[[f]] := f_n - f_p$, f_n and f_p being the traces of f on Γ from Ω_n and Ω_p, respectively. Excitation phenomena occurring in the bulk are described by the parabolic problem:

$$\begin{cases} \dfrac{\partial X}{\partial t} - \nabla \cdot (D_X \nabla X) = G - \dfrac{X}{\tau_X} & \text{in } \Omega \setminus \Gamma \\ [[X]] = 0 & \text{on } \Gamma \\ [[-\boldsymbol{\nu}_\Gamma \cdot D_X \nabla X]] = \eta k_{\text{rec}} P - \dfrac{2H}{\tau_{\text{diss}}} X & \text{on } \Gamma \\ X = 0 & \text{on } \Gamma_{\text{C}} \cup \Gamma_{\text{A}} \\ X(\mathbf{x}, 0) = 0 & \forall \mathbf{x} \in \Omega. \end{cases} \tag{1a}$$

Dissociation/recombination of excitons, electrons and holes into bonded pairs at the material interface is described by the ODE:

$$\begin{cases} \dfrac{\partial P}{\partial t} = \dfrac{2H}{\tau_{\text{diss}}} X - (k_{\text{diss}} + k_{\text{rec}})\, P + 2H\gamma\, np & \text{on } \Gamma \\ P(\mathbf{x}, 0) = 0 & \forall \mathbf{x} \in \Gamma. \end{cases} \tag{1b}$$

Transport of photogenerated electrons in the acceptor domain Ω_n is described by the parabolic problem:

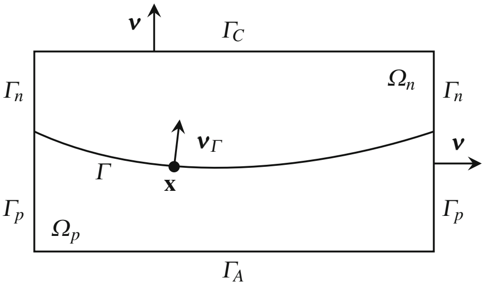


Fig. 1 Schematic representation of the mathematical domain

Table 1 Model parameters

Symbol	Parameter
μ_i, D_i	Mobility and diffusivity of species i, $i = X, n, p$
G	Exciton generation rate
τ_X, τ_{diss}	Exciton decay and dissociation times
k_{rec}, k_{diss}	Bonded pair recombination and dissociation rates
γ	Electron-hole recombination rate constant
η	Singlet exciton fraction
H	Active layer thickness

$$\begin{cases} \dfrac{\partial n}{\partial t} + \nabla \cdot \mathbf{J}_n = 0 & \text{in } \Omega_n \\ \mathbf{J}_n = -D_n \nabla n + \mu_n n \nabla \varphi & \text{in } \Omega_n \\ -\boldsymbol{\nu}_\Gamma \cdot \mathbf{J}_n = -k_{\text{diss}} P + 2H\gamma\, np & \text{on } \Gamma \\ -\kappa_n \boldsymbol{\nu} \cdot \mathbf{J}_n + \alpha_n n = \beta_n & \text{on } \Gamma_{\text{C}} \\ n(\mathbf{x}, 0) = 0 & \forall \mathbf{x} \in \Omega. \end{cases} \tag{1c}$$

A parabolic problem completely similar to (1c) describes hole transport in the donor domain Ω_p. The dependence of the electric potential and field on the space charge density in the cell is described by the Poisson equation:

$$\begin{cases} \nabla \cdot (-\varepsilon \nabla \varphi) = -q\, n & \text{in } \Omega_n \\ \nabla \cdot (-\varepsilon \nabla \varphi) = +q\, p & \text{in } \Omega_p \\ [[\varphi]] = [[-\boldsymbol{\nu}_\Gamma \cdot \varepsilon \nabla \varphi]] = 0 & \text{on } \Gamma \\ \varphi = 0 & \text{on } \Gamma_{\text{C}} \\ \varphi = V_{\text{appl}} + V_{\text{bi}} & \text{on } \Gamma_{\text{A}}. \end{cases} \tag{1d}$$

A list of the model parameters with their corresponding physical meaning is reported in Table 1. The PDE/ODE model (1) has been introduced in [3] and represents a

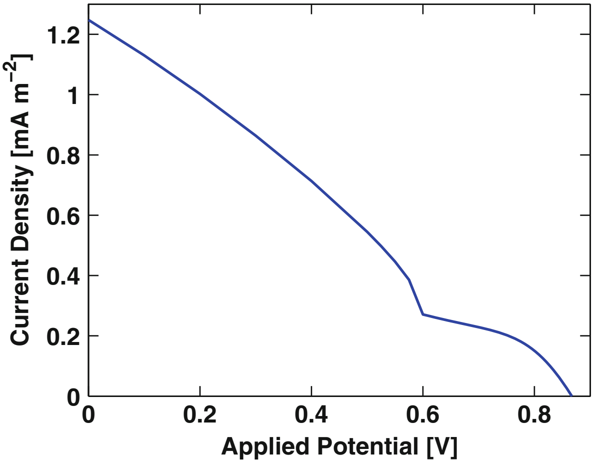


Fig. 2 $J - V$ characteristic for the finger-shaped heterostructure considered in [8]

multi-dimensional generalization of the 1D formulation proposed in [1]. System (1) is completed by periodic boundary conditions on $\Gamma_n \cup \Gamma_p$. We notice that the dissociation and recombination processes occurring at the donor-acceptor interface Γ are dealt with by the nonlinear transmission conditions (1a)$_3$ and (1c)$_2$, whose dependence on the local electric field magnitude and orientation is contained in the polaron dissociation rate constant k_{diss} [3].

3 Algorithms and Simulation Results

System linearization (by a quasi-Newton method) and approximation are carried out by adapting the approach used in [2]. Time advancing is treated using Rothe's method and adaptive BDF formulas, while the exponentially fitted Galerkin finite element method studied in [5] is used for spatial discretization. The interface conditions at the donor-acceptor interface are taken care of by means of the substructuring techniques described in [6].

Model (1) is here validated in both stationary and transient regimes. In a first set of simulations, we study the finger-shaped heterostructure considered in [8]. Figure 2 shows the output current-voltage characteristics predicted by our model, which is in excellent agreement with that computed in [8]. In a second set of simulations, we test the model in the time-dependent case. Figure 3 shows the cell current response under two different biasing conditions for a planar device geometry similar to that studied in [1]. In a third set of simulations, we test the ability of the model to describe the behaviour of a cell characterized by a complex

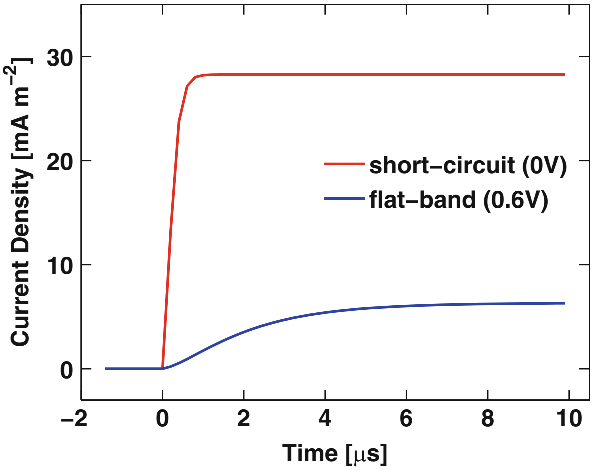


Fig. 3 Contact current density transient at two different voltage regimes

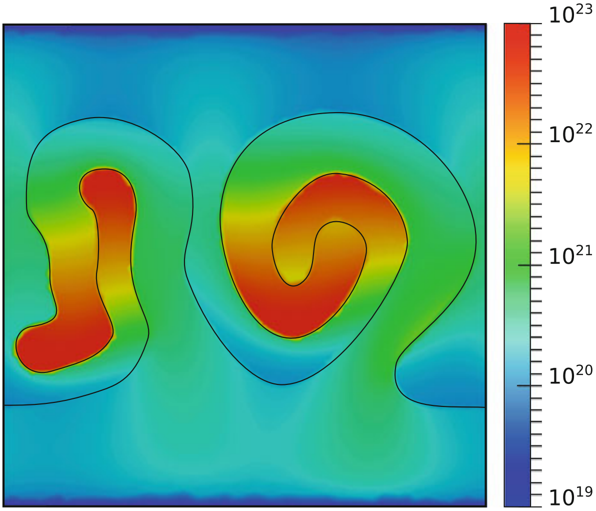


Fig. 4 Free carrier densities for a device with complex morphology

interface morphology. Figure 4 shows the free carrier densities computed for a "curly-shaped" geometry at short circuit working conditions. Ongoing activity is devoted to the investigation of the working principles of the light-harvesting device described in [4, 7].

References

1. Barker, J.A., Ramsdale, C.M., Greenham, N.C.: Modeling the current-voltage characteristics of bilayer polymer photovoltaic devices. Phys. Rev. B **67**, 075205 (2003)
2. de Falco, C., Sacco, R., Verri, M.: Analytical and numerical study of photocurrent transients in organic polymer solar cells. Comput. Methods Appl. Mech. Eng. **199**(25–28), 1722–1732 (2010)
3. de Falco, C., Porro, M., Sacco, R., Verri, M.: Multiscale modeling and simulation of organic solar cells. Comput. Methods Appl. Mech. Eng. **245–246**, 102–116 (2012)
4. Garbugli, M., Porro, M., Roiati, V., Rizzo, A., Gigli, G., Petrozza, A., Lanzani, G.: Light energy harvesting with nano-dipoles. Nanoscale **4**, 1728–1733 (2012)
5. Gatti, E., Micheletti, S., Sacco, R.: A new Galerkin framework for the drift-diffusion equation in semiconductors. East West J. Numer. Math. **6**, 101–136 (1998)
6. Hughes, T.J., Engel, G., Mazzei, L., Larson, M.G.: The continuous Galerkin method is locally conservative. J. Comput. Phys. **163**(2), 467–488 (2000)
7. Porro, M., de Falco, C., Verri, M., Lanzani, G., Sacco, R.: Multiscale simulation of organic heterojunction light harvesting devices. Int. J. Comput. Math. Electr. and Electron. Eng. **33** (2014, to appear)
8. Williams, J., Walker, A.B.: Two-dimensional simulations of bulk heterojunction solar cell characteristics. Nanotechnology **19**(42), 424011 (2008)

Simulation of Nanoscale Double-Gate MOSFETs

V. Dario Camiola, Giovanni Mascali, and Vittorio Romano

Abstract A nanoscale double-gate MOSFET is simulated by using a subband model based on the maximum entropy principle (MEP).

1 Mathematical Model

The main aim of the paper is to simulate the nanoscale silicon double gate MOSFET (hereafter DG-MOSFET) reported in Fig. 1, by including also the crystal heating which can influence the electrical properties of the device and pose severe restrictions on its performances. In fact phonons emitted by hot electrons create a phonon hot spot which increases the power density generated by the integrated circuits. This effect is becoming crucial by shrinking the dimension of the devices which is now below 100 nm, a length comparable with the wavelength of acoustic phonons [1, 2].

We consider a DG-MOSFET with length $L_x = 40\,\text{nm}$, width of the silicon layer $L_z = 8\,\text{nm}$ and oxide thickness $t_{ox} = 1\,\text{nm}$. The n^+ regions are 10 nm long. The doping in the n^+ regions is $N_D(x) = N_D^+ = 10^{20}\,\text{cm}^{-3}$ and in the n region is $N_D(x) = N_D^- = 10^{15}\,\text{cm}^{-3}$, with a regularization at the two junctions by a hyperbolic tangent profile.

Due to the symmetries and the dimensions of the device, the transport is, within a good approximation, one-dimensional and along the longitudinal direction

V.D. Camiola • V. Romano (✉)
Department of Mathematics and Computer Science, University of Catania,
viale A. Doric 6 95125, Catania, Italy
e-mail: camiola@dmi.unict.it; romano@dmi.unict.it

G. Mascali
Department of Mathematics and Computer Science, University of Calcaria,
and INFN-Gruppo c. Cosenza, 87036 Cosenza, Italy
e-mail: g.mascali@unical.it

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_4,

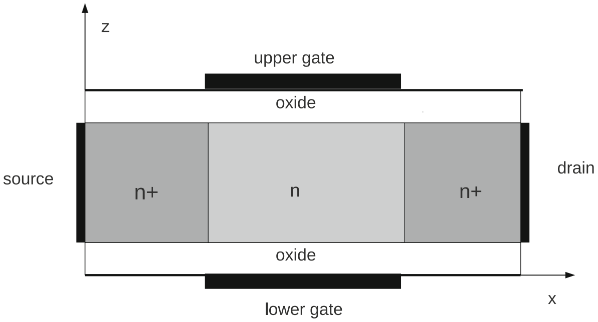


Fig. 1 Schematics representation of the simulated DG-MOSFET

with respect the two oxide layers, while electrons are quantized in the transversal direction. Six equivalent valleys are considered with a single effective mass $m^* = 0.32\, m_e$, m_e being the free electron mass.

Since the longitudinal length is of the order of a few tenths of nanometers, electrons as waves achieve equilibrium along the confining direction in a time which is much shorter than the typical transport time. Therefore we adopt a quasi-static description along the confining direction by using a coupled Schrödinger-Poisson system which leads to a subband decomposition, while transport along the longitudinal direction is described by a semiclassical Boltzmann equation for each subband.

Numerical integration of the Boltzmann-Schrödinger-Poisson system is very expensive from a computational point of view, for computer aided design (CAD) purposes (see references quoted in [3, 4]). In [3] we have formulated an energy transport model for the charge transport in the subbands by including the non parabolicity effects through the Kane dispersion relation. The model has been obtained, under a suitable diffusion scaling, from the Boltzmann equations by using the moment method and closing the moment equations with the Maximum Entropy Principle (MEP). Scatterings of electrons with acoustic and non polar optical phonons are taken into account. The parabolic subband case has been treated and simulated in [4].

A further issue is to include the crystal heating by adding an equation for the lattice temperature T_L in the same spirit as in [5, 6]

$$\rho c_V \frac{\partial T_L}{\partial t} - \mathrm{div}\,[K(T_L)\nabla T_L] = H, \tag{1}$$

with ρ and c_V silicon density and specific heat respectively. H is the phonon energy production given by

$$H = -n\, C_W + P_S\; \mathbf{J} \cdot \mathbf{E}, \tag{2}$$

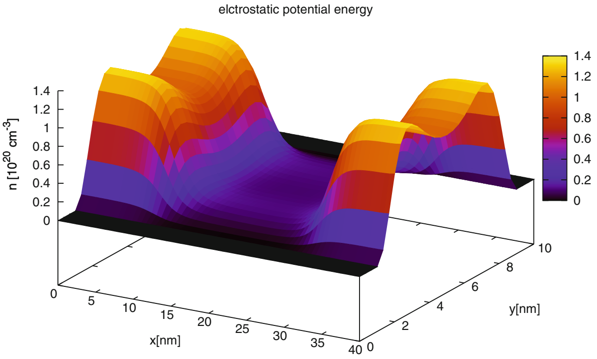


Fig. 2 Electron density when the applied potential between source and drain is $V_{SD} = 0.1$ V and source and gate are at the same potential

where P_S plays the role of a thermopower coefficient, nC_W is the electron energy production term with n electron density, and $\mathbf{J}$ is the current. The electron density is related to the surface density in each subband by the relation

$$n = \sum_{\nu} \rho_{\nu} |\phi_{\nu}|^2$$

where ϕ_{ν} are the envelope functions obtained solving the Schrödinger-Poisson system and the ρ_{ν}'s are the average surface densities in each subband ν. In [5] a more general model for H has been proposed.

We stress that the lattice temperature enters into the electron-phonon scattering and in turn in the production terms of the balance equations for the electron variables. It is crucial to address the importance of the crystal heating on the electric performance of the device.

2 Simulation Results

A suitable modification of the numerical scheme for the MEP energy transport-Schrödinger-Poisson system developed in [4] can be used by including also the discretization of the lattice temperature balance equation via an Alternating-Direction-Implicit (ADI) approach. Since the characteristic time of the crystal temperature is about one or two orders of magnitude longer than that of electrons, a multirate time step method as in [6] is a suitable choice.

In Figs. 2 and 3 we report some preliminary results. We note that there is a very high potential energy variation near the contacts. This could imply a noticeable raise of the crystal energy $k_B T_L$ around the drain and it is likely that the lattice temperature can approach the silicon melting temperature. The presence

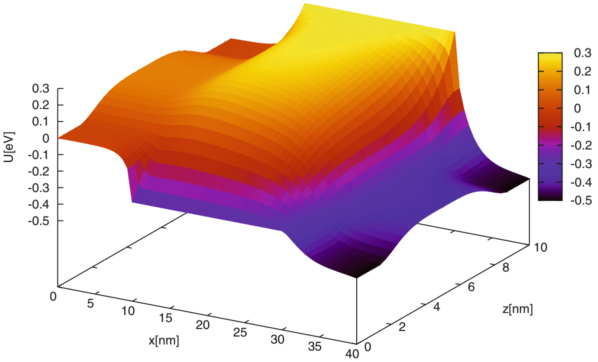


Fig. 3 Electrostatic energy when the applied potential between source and drain is $V_{SD} = 0.1$ V and source and gate are at the same potential

of strong electric fields could pose severe restrictions on the source/drain and source/gate voltages with stringent design constraints. These issues are currently under investigation by the authors.

Acknowledgements G.M. acknowledges the financial support by P.R.A., University of Calabria. V.R. acknowledges the financial support by P.R.A., University of Catania and the P.R.I.N. project 2010 "Kinetic and macroscopic models for particle transport in gases and semiconductors: analytical and computational aspects."

References

1. Sinha, S., Goodson, K.E.: Thermal conduction in sub-100 nm transistors. Microelectron. J. **37**, 1148–1157 (2006)
2. Rowlette, J.A., Goodson, K.E.: Fully coupled nonequilibrium electron-phonon transport in nanometer-scale silicon FETs. IEEE Trans. Electron Devices **55**, 220–232 (2008)
3. Mascali, G., Romano, V.: A non parabolic hydrodynamical subband model for semiconductors based on the maximum entropy principle. Math. Comput. Model. **55**, 1003–1020 (2012)
4. Camiola, V.D., Mascali, G., Romano, V.: Numerical simulation of a double-gate MOSFET with a subband model for semiconductors based on the maximum entropy principle. Continuum Mech. Thermodyn. **24**, 417–436 (2012)
5. Romano, V., Zwierz, M.: Electron-phonon hydrodynamical model for semiconductors. Z. Angew. Math. Phys. **61**, 1111–1131 (2010)
6. Romano, V., Rusakov, A.: 2d numerical simulations of an electron–phonon hydrodynamical model based on the maximum entropy principle. Comput. Methods Appl. Mech. Eng. **199**, 2741–2751 (2010)

Coupled Heat-Electromagnetic Simulation of Inductive Charging Stations for Electric Vehicles

Christof Kaufmann, Michael Günther, Daniel Klagges, Matthias Richwin, Sebastian Schöps, and E. Jan W. ter Maten

Abstract Coupled electromagnetic-heat problems have been studied for induction or inductive heating, for dielectric heating, for testing of corrosion, for detection of cracks, for hardening of steel, and more recently for inductive charging of electric vehicles. In nearly all cases a simple co-simulation is made where the electromagnetics problem is solved in the frequency domain (and which thus is assumed to be linear) and the heat equation in the time domain. One exchanges data after each time step (or after some change in the heat profile). However, the coupled problem is non-linear in the heat variable. In this paper we propose to split the time domain in windows in which we solve the electromagnetics problem in frequency

C. Kaufmann (✉)
Hochschule Bochum, Fachbereich Elektrotechnik und Informatik, Lennershofstraße 140, 44801 Bochum, Germany
e-mail: christof.kaufmann@hs-bochum.de

M. Günther
Bergische Universität Wuppertal, Fachbereich C, Lehrstuhl für Angewandte Mathematik/Numerische Analysis, Bendahler Straße 31, 42285 Wuppertal, Germany
e-mail: guenther@math.uni-wuppertal.de

D. Klagges • M. Richwin
Leopold Kostal GmbH & Co. KG, An der Bellmerei 10, 58513 Lüdenscheid, Germany
e-mail: d.klagges@kostal.de; m.richwin@kostal.de

S. Schöps
Technische Universität Darmstadt, Graduate School of Computational Engineering, Dolivostraße 15, 64293 Darmstadt, Germany
e-mail: schoeps@gsc.tu-darmstadt.de

E. Jan W. ter Maten
Bergische Universität Wuppertal, Fachbereich C, Lehrstuhl für Angewandte Mathematik/Numerische Analysis, Bendahler Straße 31, 42285 Wuppertal, Germany

Department of Mathematics & Computer Science, TU Eindhoven, CASA, PostBox 513, 5600 MB Eindhoven, The Netherlands
e-mail: termaten@math.uni-wuppertal.de; E.J.W.ter.Maten@tue.nl

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_5,

domain. We strengthen the coupling by iterations, for which we prove convergence. By this we obtain a higher accuracy, which will allow for larger time steps and also for higher order time integration. This fully exploits the multirate behavior of the coupled system. An industrial example illustrates the analysis.

1 Introduction

In todays development processes, simulation is becoming more and more important. One predicts physical behavior precisely—even for multiphysics systems, where many effects influence each other. In that sense the first prototypes can be replaced by simulation. This is called *virtual prototyping* and speeds up time-to-market considerably.

In this paper simulation of electromagnetic problems coupled with heat problems is considered. Well known applications are induction heating [11, 13], dielectric heating [8], e.g. used for microwave ovens, steel hardening of gears [10] and detection of cracks or corrosion in ships. We focus on the design process of an *inductive charging* station for electric vehicles.

In inductive charging the electromagnetic (EM) field is of main importance. It induces eddy currents in massive conductors. At the power levels used for charging of electric vehicles, these losses cause a significant amount of heat. The heat diffuses and changes temperature and properties of the materials, and thus also the EM field. These effects have to be considered in a two-way coupling: One way is the generation of heat via eddy current losses resulting from the EM field. The other is the influence of the temperature dependent material parameters on the EM field.

In contrast to [9] we focus here on the comparison of the different co-simulation methodologies.

2 Modeling

A simple time-domain model consists of the curl-curl equation (1) for the electromagnetic problem and the heat equation (2) to describe the heat diffusion. It can be stated as

$$\nabla \times (\mu^{-1} \nabla \times \mathbf{A}) + \varepsilon \frac{\partial^2 \mathbf{A}}{\partial t^2} + \sigma(T) \frac{\partial \mathbf{A}}{\partial t} = \mathbf{J}_{\text{src}} \tag{1}$$

$$\rho c \frac{\partial T}{\partial t} = \nabla \cdot (k \nabla T) + Q, \tag{2}$$

in which the heat source density Q comes from the power loss terms caused by the eddy current losses and the currents in the coil

$$Q(\mathbf{A},T) := \sigma(T)\,\frac{\partial \mathbf{A}}{\partial t}\cdot\frac{\partial \mathbf{A}}{\partial t} - \mathbf{J}_{\mathrm{src}}\cdot\frac{\partial \mathbf{A}}{\partial t}. \tag{3}$$

In (1) $\mathbf{A}$ is the magnetic vector potential, $\mathbf{J}_{\mathrm{src}}$ is the source current density. The electric conductivity σ is material and temperature dependent. The other material parameters (the magnetic permeability μ and the permittivity ε) are considered here as constant in time and do not dependent on the temperature T, but vary in space. For the heat equation, ρ and c are the mass density and the specific heat density, respectively; k is the thermal conductivity. All materials are assumed to be isotropic for simplicity of notation. Both equations must be equipped with appropriate boundary conditions (BC). The heat equation requires an initial value (IV) at start time.

3 Co-simulation

A simulation could be drawn out as one large system of equations, i.e. monolithic. Solving this system with classical time stepping methods would require to follow the demands of the fastest part of the system. In heat/electromagnetic problems this is usually the EM field. Here we assume a harmonic source current density, which determines the step size of the integrator. As first co-simulation scheme, we consider the one, that is closest to the monolithic approach: single rate co-simulation. Basically this is a Gauss-Seidel-type scheme, where each part uses the same time steps. The scheme is illustrated in Fig. 1. The single rate co-simulation approach is a simple and straight forward approach. Data of one part of the solution can be given directly into the next one. Alternatively, outer iterations can be used to increase the accuracy and stability. However, a lot of computational effort is spent in both subsystems due to the uniform time step, although the slow part does not need these steps. This observation is the base of the next scheme we consider.

The multirate co-simulation approach, illustrated in Fig. 2, makes use of the different time scales of the heat and EM phenomenons. Since the source current density (and thus magnetic vector potential) is faster changing than the temperature, there are more time steps needed for the curl-curl equation than for the heat equation. Clearly, here the advantage is the computational savings when solving the heat equation. On the other hand this approach is less straightforward than the single rate approach; one has to manage the more complex data transfer. A common way is the introduction of *synchronization time points* τ_i, as shown in Fig. 2. This requires to align the time stepping schemes of all subsystems. Another way to find the data at the desired time point is interpolation. Finally, iteration is still possible in this multirate approach, similar to dynamic iteration approaches, e.g. [1]. However, the main part of the computational costs is the integration of the curl-curl equation, which still has not changed in comparison to the single rate approach.

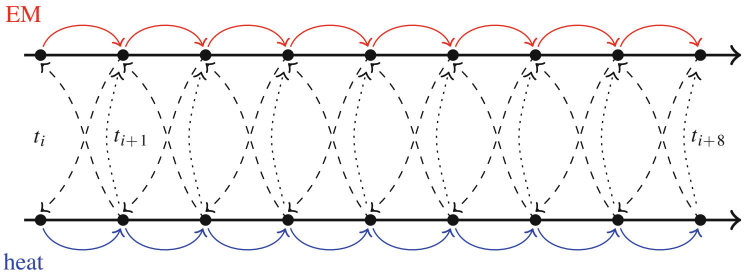


Fig. 1 Single rate co-simulation approach, see also [5]

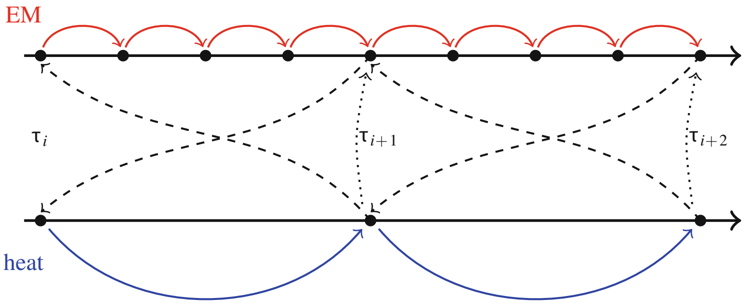


Fig. 2 Multirate co-simulation approach with synchronization time points τ_i

3.1 Frequency-Transient Model

More elaborated modeling significantly reduces the high computational costs for solving the curl-curl equation: for many applications it is accurate enough to average the power transferred within a time window $[\tau_i, \tau_{i+1}]$. Similarly, an averaged temperature, $\tilde{T}_i$, is used for the conductivity $\mathbf{M}_\sigma$ in the curl-curl equation. For metals $\sigma(T)$ is monotonically decreasing. The other material parameters (μ, ε) are assumed to be constant. That allows to solve the curl-curl equation in the frequency domain and to avoid the computation in time domain. We further assume that $\mathbf{J}_{\text{src}} = \hat{\mathbf{J}}_{\text{src}}\, e^{j\omega t}$. Then the coupled model for a time harmonic source current density can be stated as

$$(j\,\omega\,\mathbf{M}_\sigma(\tilde{T}_i) - \omega^2\,\varepsilon)\hat{\mathbf{A}}_c + \nabla\times(\mu^{-1}\,\nabla\times\hat{\mathbf{A}}_c) = \hat{\mathbf{J}}_{\text{src}} \tag{4}$$

$$\rho\,c\,\frac{\partial T}{\partial t} = \nabla\cdot(\mathbf{k}\nabla T) + \tilde{Q}_i(\tilde{T}_i), \tag{5}$$

where

$$\tilde{Q}_i(\tilde{T}_i) = \mathbf{M}_\sigma(\tilde{T}_i)\,\frac{\omega^2}{2}\left\|\hat{\mathbf{A}}_c(\tilde{T}_i)\right\|_c^2 - \frac{\omega}{2}\,\mathrm{Im}\left(\overline{\hat{\mathbf{A}}_c(\tilde{T}_i)}\cdot\hat{\mathbf{J}}_{\text{src}}\right). \tag{6}$$

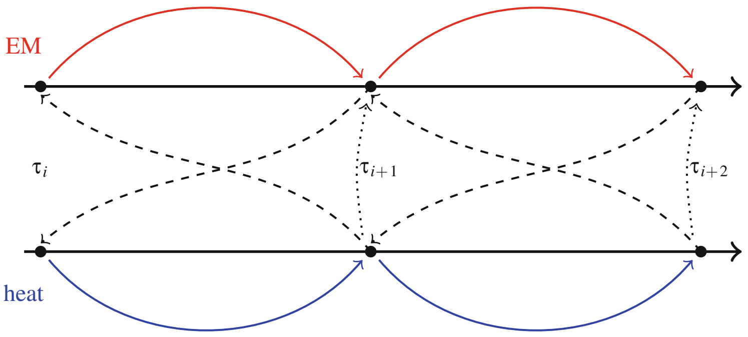


Fig. 3 Frequency-transient co-simulation approach, see also [5]

in which $\hat{\mathbf{A}}_c = \hat{\mathbf{A}}_c(\tilde{T}_i)$ and $\hat{\mathbf{J}}_{\mathrm{src}}$ are the first Fourier coefficients of the magnetic vector potential $\mathbf{A}$ and the source current density $\mathbf{J}_{\mathrm{src}}$, respectively. In (6) the norm is the complex norm; the overline indicates the complex conjugate. For further details of the derivation, see [9]. Please notice that there is still the parametric coupling from the heat equation via the conductivity $\mathbf{M}_\sigma$ to the curl-curl equation. Also the curl-curl equation is still coupled to the heat equation via the source term $\tilde{Q}_i$. Hence, the coupling is still two-way. However, the curl-curl equation (4) has become a purely algebraic equation whose solution depends on the temperature and, by this, implicitly on time. The scheme is illustrated in Fig. 3. By computing the curl-curl equation in frequency domain only one linear system is solved for one time step of the heat equation. Especially for high frequencies this saves a huge amount of computational time in comparison to the classical approaches. Also, the data transfer is straightforward. This setup has been the basis for several coupled EM-heat problems [3, 4, 8, 11], however typically without applying iterations.

When using an iterative scheme, convergence must be analyzed on beforehand. For the frequency-transient approach with an implicit Euler scheme for the time discretization of the heat equation, convergence can be proved [9]. This results in the following theorem:

Theorem 1. *We assume given BC and IV and for nonlinear materials, i.e., metals, a conductivity σ that is differentiable w.r.t. temperature T and $\partial\sigma/\partial T < 0$. Let the exact (monolithic) solution be denoted by $\mathbf{a}^*$ and $\mathbf{t}^*$, then the iteration is convergent for h small enough with*

$$\left\|\mathbf{t}^{(l+1)} - \mathbf{t}^*\right\| \leq c(\omega)\, h \left\|\mathbf{t}^{(l)} - \mathbf{t}^*\right\|,$$

where $c(\omega)$ is uniformly bounded for $\omega > \omega_0$ and $c(\omega) = \mathcal{O}\left(\frac{1}{\omega^2}\right)$ for sufficiently large ω.

For metals this implies that there are no additional step size restrictions for $\omega \to \infty$. Then, for higher frequencies the step size can be larger. In fact this is in good agreement with the high frequency applications found in literature.

Proof. Here we summarize the steps of the proof; for details see [9]. We assume the same space discretization for both subproblems to simplify notation. For the discretization in space we propose a lowest-order Finite Element Method (FEM) with Lobatto Quadrature or the Finite Integration Technique (FIT) [4, 12]. In the quadratic Q-term, the value of T only depends on the temperature in a local meshpoint. This simplifies the proof. For readability we drop the diacritic symbol ˆ.

Let $\mathbf{a}$ and $\mathbf{t}$ denote the discretized magnetic vector potential and temperature, respectively. At time τ_n we assume $\mathbf{a}, \mathbf{t}$ given. Then at $\tau_{n+1} = \tau_n + h$ we iteratively determine $\mathbf{t}^l \Rightarrow \mathbf{a}^{l+1} \Rightarrow \mathbf{t}^{l+1}$, etc. We assume that there is an exact solution without splitting error at τ_{n+1}: $\mathbf{a}^\star, \mathbf{t}^\star$. The *discretized curl-curl equation* (4) becomes in FIT-like notation [12]

$$[j\omega \mathbf{M}_\sigma(\mathbf{t}^l) - \omega^2 \mathbf{M}_\varepsilon + \mathbf{C}^\top \mathbf{M}_\nu \mathbf{C}]\mathbf{a}^{l+1} = \mathbf{j}_{\mathrm{src}},$$

with diagonal matrices for conductivity, permittivity and reluctivity, $\mathbf{M}_\sigma$, $\mathbf{M}_\varepsilon$, $\mathbf{M}_\nu$, discrete curl operators $\mathbf{C}$, $\mathbf{C}^\top$ and source current $\mathbf{j}_{\mathrm{src}}$. This gives an error equation

$$[\mathbf{X}_{\varepsilon\nu} + j\omega \mathbf{M}_\sigma^l](\mathbf{a}^{l+1} - \mathbf{a}^\star) = -j\omega[\mathbf{M}_\sigma^l - \mathbf{M}_\sigma^\star]\,\mathbf{a}^\star,$$

where we define for convenience

$$\mathbf{X}_{\varepsilon\nu} := -\omega^2 \mathbf{M}_\varepsilon + \mathbf{C}^\top \mathbf{M}_\nu \mathbf{C}, \qquad \mathbf{M}_\sigma^l := \mathbf{M}_\sigma(\mathbf{t}^l), \qquad \mathbf{M}_\sigma^\star := \mathbf{M}_\sigma(\mathbf{t}^\star).$$

Thus $\mathbf{a}^{l+1} = \mathbf{a}^\star + \mathbf{R}$, where

$$\mathbf{R} = -j\omega[\mathbf{X}_{\varepsilon\nu} + j\omega \mathbf{M}_\sigma^l]^{-1}[\mathbf{M}_\sigma^l - \mathbf{M}_\sigma^\star]\,\mathbf{a}^\star.$$

Hence

$$\|\mathbf{R}\| < \omega \|[\mathbf{X}_{\varepsilon\nu} + j\omega \mathbf{M}_\sigma^l]^{-1}\| \cdot \|\mathbf{M}_\sigma^l - \mathbf{M}_\sigma^\star\| \cdot \|\mathbf{a}^\star\|,$$

which asks for a uniform upper bound for the inverse operator and for Lipschitz continuity of $\mathbf{M}_\sigma$. Then the $\mathbf{a}^{l+1}$ are bounded.

The discretized version of the heat equation (5) is given in the following. For simplicity we assume time discretization by the implicit Euler scheme, we focus only on the quadratic term and disregard the other right-hand-side terms (rhs); $||.||$ is a vector of coordinate-wise norms

$$\begin{aligned}
[\mathbf{M}_{\rho,c} - h\tilde{\mathbf{S}}\mathbf{M}_k\tilde{\mathbf{S}}^\top](\mathbf{t}^{l+1} - \mathbf{t}^\star) &= \frac{1}{2}h\omega^2[\mathbf{M}_\sigma^{l+1}\|\mathbf{a}^{l+1}\|^2 - \mathbf{M}_\sigma^\star\|\mathbf{a}^\star\|^2] + \mathrm{rhs} \\
&= \frac{1}{2}h\omega^2[\mathbf{M}_\sigma^{l+1}\|\mathbf{a}^\star + \mathbf{R}\|^2 - \mathbf{M}_\sigma^\star\|\mathbf{a}^\star\|^2\,] + \mathrm{rhs} \\
&= \frac{1}{2}h\omega^2[\mathbf{M}_\sigma^{l+1} - \mathbf{M}_\sigma^\star]\,\|\mathbf{a}^\star + \mathbf{R}\|^2 + \mathscr{R} + \mathrm{rhs} \\
\text{with } \mathscr{R} &= \frac{1}{2}h\omega^2\mathbf{M}_\sigma^\star[< \mathbf{a}^\star, \mathbf{R} > + < \mathbf{R}, \mathbf{a}^\star > + < \mathbf{R}, \mathbf{R} >]
\end{aligned}$$

with material matrices $\mathbf{M}_{\rho,c}$ and $\mathbf{M}_k$ and the discrete divergence and gradient operators $\tilde{\mathbf{S}}$, $-\tilde{\mathbf{S}}^{\top}$, respectively [4]. It follows equivalently

$$[\mathbf{M}_{\rho,c} - h\tilde{\mathbf{S}}\mathbf{M}_k\tilde{\mathbf{S}}^{\top}](\mathbf{t}^{l+1} - \mathbf{t}^{\star}) - \frac{1}{2}h\omega^2[\mathbf{M}_{\sigma}^{l+1} - \mathbf{M}_{\sigma}^{\star}]\,\|\mathbf{a}^{\star} + \mathbf{R}\|^2 = \mathscr{R}.$$

We can rewrite

$$[\mathbf{M}_{\sigma}^{l+1} - \mathbf{M}_{\sigma}^{\star}]\,\|\mathbf{a}^{\star} + \mathbf{R}\|^2 = \mathrm{Diag}(\|\mathbf{a}^{\star} + \mathbf{R}\|^2)\,\mathrm{Vec}(\mathbf{M}_{\sigma}^{l+1} - \mathbf{M}_{\sigma}^{\star}).$$

For the last term we can apply the mean value theorem coordinate-wise and thus get

$$\mathrm{Vec}(\mathbf{M}_{\sigma}^{l+1} - \mathbf{M}_{\sigma}^{\star}) = \mathrm{Diag}(\mathbf{M}'_{\sigma,k})(\mathbf{t}^{l+1} - \mathbf{t}^{\star}).$$

Hence, for $\sigma'(T) < 0$ we have convergence for h small enough, but with good properties for varying ω. This completes the summary of the proof. □

4 Generalization

For an extended approach by Driesen and Hameyer [7], where also the complex phasor is allowed to vary slowly, the proof can be extended. In this case the curl-curl equation in frequency domain leads to a second order DAE after space discretization. Hence the error equation for the curl-curl equation needs to be integrated as well.

Another way of generalizing the frequency transient model is to include multifrequency excitation as needed, e.g., for non-smooth surfaces [10]. This is an easy way to allow approximations for other periodic waveforms of source currents. Other waveforms are important to approximate the current from power electronics, that control the primary coil. This gives way to a Harmonic Balance approach for the curl-curl equation. It also allows for including a nonlinear permeability μ. Otherwise the model can only be used for a working point of the magnetic material curve.

A more general fully multirate time domain model, that exploits different time scales, can be derived by using the MPDAE approach by Brachtendorf et al. [2]. By this, envelope simulation techniques from circuit simulation are applied to the coupled EM-heat problem.

5 Numerical Example

In this section results of a simulation for a model of an inductive charging system are shown. The simulation of a model with temperature independent conductivity, which results in a single way coupling, will be compared to the two-way coupling.

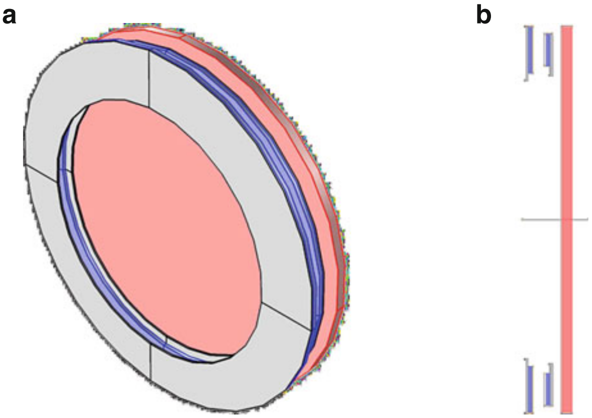


Fig. 4 Geometry of the simulated model. *From left to right*: ferrite (*gray*), primary coil (*blue*), air, secondary coil (*blue*), ferrite (*gray*), air, steel slice (*red*). The *left* coil represents the charging station and the *right* coil the coil behind the number plate in the car (Comsol). (**a**) 3d view on 2d-axisymmetric geometry. (**b**) Cut view

The frequency-transient model is applied to an inductive charging system for electric vehicles. The charging is done here through the number plate. The model consists of two copper coils with ferrite and air in between. The primary coil represents the charging station and the secondary coil the coil behind the number plate in the vehicle. To account for the challenges of a real world prototype, a steel bar with a constant permeability $\mu_r = 500$ is added behind the secondary coil. It models parts of the car body. The geometry is shown in Fig. 4.

The simulations have been run in Comsol [6] with appropriate settings to use the frequency-transient model as described in this paper. Simulation time is set to 20 min, the coils have 20 windings. The primary coil is excited by a current of 25 A at a (moderate) frequency of 10 kHz. The secondary has a zero current (no-load configuration). The conductivity σ in the independent case was chosen to be at room temperature (293.15 K), which is also the initial temperature.

The simulation with temperature independent conductivity shows a maximum temperature of 387.92 K, see Fig. 5a. The maximum temperature in case of an temperature dependent conductivity is 395.15 K, see Fig. 5b. The difference of about 7 K is due to the parameter coupling from the temperature to the curl-curl equation via the conductivity.

Remark that for this simulation only 17 time steps were needed to compute the results. When we compare that to simulating this problem with a monolithic model and assume 10 time steps per period of the source current, 120 million time steps would be necessary for simulation. This clearly shows the efficiency of simulation of the frequency-transient model compared to approaches, where the EM subsystem is solved in time domain.

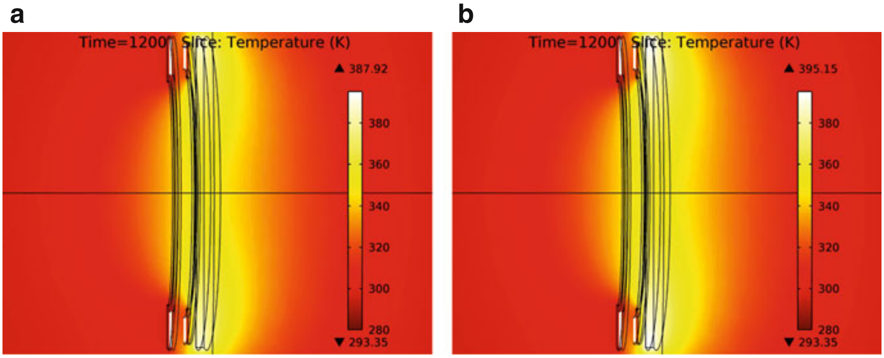


Fig. 5 Simulation of inductive charging system for e-cars after 20 min. The plot shows the temperature distribution (Comsol). (**a**) Temperature independent conductivity. (**b**) Temperature dependent conductivity

6 Conclusions

A frequency-transient model tailored for coupled heat-electromagnetic problems was described. An efficient multirate co-simulation approach was proposed for solving it. For this algorithm a convergence theorem for an iterative approach was proved for all frequencies. For metals and higher frequencies the speed of convergence increases. The numerical example confirms this result. The theorem also applies to approaches by Driesen and Hameyer [7] and implementations in Comsol [6]. In particular it applies to many cosimulation approaches for high frequency applications [3, 4, 8, 10, 11, 13]. From the analysis (see [9]) an optimal step size can be derived.

Acknowledgements This work is supported by the German BMBF in the context of the SOFA project (grant number 03MS648E). The fifth author is supported by the Excellence Initiative of the German Federal and State Governments and the Graduate School of Computational Engineering at Technische Universität Darmstadt.

References

1. Bartel, A., Brunk, M., Günther, M., Schöps, S.: Dynamic iteration for coupled problems of electric circuits and distributed devices. SIAM J. Sci. Comput. **35**(2), B315–B335 (2013)
2. Brachtendorf, H.G., Welsch, G., Laur, R., Bunse-Gerstner, A.: Numerical steady state analysis of electronic circuits driven by multi-tone signals. Electr. Eng. (Archiv fur Elektrotechnik) **79**, 103–112 (1996)
3. Chen, H., Tang, J., Liu, F.: Coupled simulation of an electromagnetic heating process using the finite difference time domain method. J. Microw. Power Electromagn. Energy **41**(3), 50–68 (2007)

4. Clemens, M., Gjonaj, E., Pinder, P., Weiland, T.: Numerical simulation of coupled transient thermal and electromagnetic fields with the finite integration method. IEEE Trans. Magn. **36**(4), 1448–1452 (2001)
5. Clemens, M., Schöps, S., Cimala, C., Gödel, N., Runke, S., Schmidthäusler, D.: Aspects of coupled problems in computational electromagnetics formulations. ICS Newslett. (International Compumag Society) **19**(2), 3–12 (2012)
6. COMSOL Multiphysics: Command Reference (2007). www.comsol.com
7. Driesen, J., Hameyer, K.: The simulation of magnetic problems with combined fast and slow dynamics using a transient time-harmonic method. Eur. Phys. J. Appl. Phys. **14**, 165–169 (2001)
8. Janssen, H.H.J.M., ter Maten, E.J.W., van Houwelingen, D.: Simulation of coupled electromagnetic and heat dissipation problems. IEEE Trans. Magn. **30**(5), 3331–3334 (1994)
9. Kaufmann, C., Günther, M., Klagges, D., Knorrenschild, M., Richwin, M., Schöps, S., ter Maten, E.J.W.: Efficient simulation of frequency-transient mixed co-simulation of coupled heat-electromagnetic problems. Math. Ind. **4** (2014, to appear).
10. Rudnev, V.: Induction Hardening of Gears and Critical Components. Gear Technology pp. 58–63 (Sept/Oct) and 47–53 (Nov/Dec) (2008). Part I and II
11. ter Maten, E.J.W., Melissen, J.B.M.: Simulation of inductive heating. IEEE Trans. Magn. **28**(2), 1287–1290 (1992)
12. Weiland, T.: Time domain electromagnetic field computation with finite difference methods. Int. J. Numer. Model. **9**(4), 295–319 (1996)
13. Will, J.: An optimal control problem in electromagnetic induction heating. Master's thesis, Chemnitz University of Technology, Department of Mathematics (2010)

Part II
Environment

Overview

The section on *Environment* contains seven contributions. Mathematical modelling and simulation in environmental science is a fast evolving field with the goal to understand key environmental issues, like water and air pollution, improving forecasts for environmental hazards, like earthquakes or volcanic eruptions. Next, the scientific results must be presented in a suitable information format to meet properly the needs of stakeholders and decision makers.

In this ECMI proceeding the first four papers are dealing with the problem of water pollution, asking for the optimal location measurement station along a river, modelling pollutant transport in groundwater flow and optimizing the shape design of wastewater canals. Related to the last topic is the paper of Neli Dimitrova on biodegradation of toxic substances in a wastewater treatment. The fifth paper considers the Unified Danish Eulerian Model (UNI-DEM), a large scale environmental model for long-range transport of air pollution. The last two papers finally show promising mathematical models and three dimensional numerical for volcano activities, either considering the hazard forecasting or the lava flow.

The first paper *Optimal Location of River Sampling Stations: A Case Study* by Lino J. Alvarez-Vazquez, Aurea Martínez, Miguel E. Vázquez-Méndez, A.W. Pollak and J. Jeffrey Peirce studies the optimal location of water pollution monitoring stations located along the length of the river by combining both numerical simulation of shallow water equations and optimization techniques. Finally, the resulting method is illustrated by a real case study of the Neuse River (North Caroline, USA).

The contribution by Neli Dimitrova: *Global Analysis of a Nonlinear Model for Biodegradation of Toxic Compounds in a Wastewater Treatment Process*, presents a rigorous mathematical stability analysis of a system of ordinary differential equations, describing the biodegradation of toxic substances in a wastewater treatment plant. Hereby properties like the equilibrium points of the considered model and their Lyapunov stability, the boundedness of solutions and their global asymptotic stability are investigated. This analysis could be useful to determine the parameter domain for a stable operation of the microbial process in a continuously stirred bioreactor. Finally, numerical simulations support the theoretical findings.

Pollutant transport in groundwater flow is a challenging topic due to the coupling effects between the different ground layers. In the paper by Amjad Ali, Winston L. Sweatman and Robert McKibbin: *Pollutant Transport and its Alleviation in Groundwater Aquifers*, the authors propose a simplified model for the transport of dissolved chemicals through groundwater aquifers. This simple model allows to simulate the typical natural stratification and changes in physical properties of the aquifer that occur between different geological layers. Finally the authors present an example where an instantaneous release of pollutant occurs and it is afterwards removed by the downstream release of a suitable pollutant removal agent.

In the fourth paper by Aurea Martínez, Lino J. Alvarez-Vázquez, Carmen Rodríguez, Miguel E. Vázquez-Méndez and Miguel A. Vilar: *Optimal Shape Design of Wastewater Canals in a Thermal Power Station*, the authors develop a strategy to determine an optimal geometry design of canals in a wastewater treatment plants of thermal power stations The underlying mathematical problem is stated as a control-constrained optimal control problem of partial differential equations and subsequently discretized via a characteristics/finite element method.

The paper by Zahari Zlatev, István Faragó and Agnes Havasi: *Mathematical Treatment of Environmental Models*, describes UNI-DEM, a large scale environmental model for long-range transport of air pollution which is a system of nonlinear partial differential equations. Next, this model is split into three sub-models that are transformed to ordinary differential equations by discretizing all spatial derivatives using a simple linear finite element method. Each subsystem is solved by an adequate ODE solver and finally a parallelization strategy is proposed.

In the paper by Gilda Currenti and Ciro Del Negro: *Model-Based Assessment of Geophysical Observations: From Numerical Simulations towards Volcano Hazard Forecasting* an integrated elastic 3-D model for magma migration and accumulation within the volcano edifice is considered and solved numerically by finite elements. The numerical model is validated using existing analytical solutions and was applied later for interpreting data from the Etna volcano during unrest periods. This approach calibrated with observable data might turn out useful in an accurate volcano hazard assessment and in scenario forecasting.

Finally the contribution by Marilena Filippucci, Andrea Tallarico and Michele Dragoni: *Thermal and Rheological Aspects in a Channeled Lava Flow*, deals with the three dimensional numerical simulation of the cooling of a lava flow. Hereby a 3D heat equation is solved numerically and the fraction of crust coverage is determined assuming that the lava rheology is pseudoplastic and dependent on temperature. The authors' findings are validated using data from the Mauna Loa (1984) lava flow indicating a strong link between the advective heat transport and the cooling rate of the lava.

Matthias Ehrhardt

Optimal Location of River Sampling Stations: A Case Study

Lino J. Alvarez-Vázquez, Aurea Martínez, Miguel E. Vázquez-Méndez, A.W. Pollak, and J. Jeffrey Peirce

Abstract Usual methods for monitoring and controlling river pollution include the establishment of water pollution monitoring stations located along the length of the river. The point where each station is located (known as sampling point) is of crucial importance if we want to obtain representative information about industrial and domestic pollution in the whole river, not only in the sampling points. In this work, the optimal location of sampling points is studied combining numerical simulation and optimization techniques. Based on a one dimensional system of partial differential equations, a mathematical formulation of the optimization problem is proposed, and it is solved for a real case on Neuse River (North Caroline, USA), where interesting conclusions are derived for the number of water quality sensors and their respective locations.

1 Introduction

Surface water quality directly impacts communities depending on these sources for potable use, recreation, agricultural supplies or commercial fishing. Water available for these purposes can be drastically impacted by contamination from municipal and

L.J. Alvarez-Vázquez (✉) • A. Martínez
Departamento de Matemática Aplicada II, E.I. Telecomunicación, Universidad de Vigo, 36310 Vigo, Spain
e-mail: lino@dma.uvigo.es; aurea@dma.uvigo.es

M.E. Vázquez-Méndez
Departamento de Matemática Aplicada, E.P.S., Universidad de Santiago de Compostela, 27002 Lugo, Spain
e-mail: miguelernesto.vazquez@usc.es

A.W. Pollak • J.J. Peirce
Department of Civil and Environmental Engineering, Duke University, Durham, NC 27708, USA
e-mail: peirce@duke.edu

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_6,

industrial discharges. One method for the effective monitoring and management of surface water quality is the establishment of real time, in situ monitoring systems. These systems provide the basis for future adaptive management schemes using data about the transport and fate of contaminants across environmental regions. Distributed monitoring systems are predicated on the development of new sensors capable of monitoring the contaminants of interest. From a practical viewpoint, a fundamental component of implementing this type of networks is the identification of the optimal locations to deploy environmental sensors or establish sampling sites [3, 13].

Numerous attempts at establishing standards for river sampling programs have been suggested. In 1971, Sharp [12] proposed the use of topographical optimization. This method, however, has been proven by Dixon et al. [4, 5] to not necessarily produce optimal sampling locations. Traditionally, proximity to affected populations or ease of access have ruled the installation of monitoring points, but the advent of deployable real time water quality sensors allows quality to be monitored at any point along the river. Ward [14] suggested placing sampling points near critical quality points, but Hren et al. [7] concluded that a focus on critical points could lead to biased assumptions about global water quality. A selection of good sampling sites also allows data from those sampling points to be extrapolated for understanding the distribution of contamination along the length of the whole river [1, 8]. Rather than focusing on critical points, the locations of water quality sampling points ought to be specific to the purposes of data collection about selected contaminants [11].

To illustrate the use of sampling program goals to direct the selection of optimal sensor locations, our focus is limited to a single sample contaminant, fecal coliform (FC) bacteria. Selecting this contaminant, the numerical model introduced by Alvarez-Vázquez et al. [1] in 2006 is used to determine the optimal locations for water quality sensors along a river for FC contamination sampling.

2 Setting of the Problem

The model developed by the authors in order to determine the optimal locations for FC sensing in rivers consists of three main parts:

1. The river section of total length L is divided into N segments $[a_{i-1}, a_i]$, for $i = 1, 2, \ldots, N$, with $a_0 = 0$ and $a_N = L$.
2. The average contamination in the transversal section $\rho(x, t)$ (at location x and at time t) is simulated for each $(x, t) \in [0, L] \times [0, T]$, where T represents the length of the time interval. Contamination sources are modelled as point inputs originating at known locations along the river reach.
3. The optimal sampling point in the i-th segment for $i = 1, 2, \ldots, N$ is located by identifying the point p_i which minimizes the difference between $\rho(p_i, t)$ and the average concentration along that segment. This optimization problem seeks to minimize the objective function $J(p)$ for $p = (p_1, p_2, \ldots, p_N)$:

$$J(p) = \sum_{i=1}^{N} \int_0^T (\rho(p_i, t) - c_i(t))^2 \, dt \tag{1}$$

where $c_i(t)$ denotes the averaged concentration for all points located in the i-th river segment $[a_{i-1}, a_i]$.

To compute the contamination at each location down river for each moment in the simulation, the model simultaneously solves two sets of equations, for the change in contamination due to FC loss and source input, as well as for river flow and the advection of contamination downstream. Neglecting molecular diffusion, the concentration of FC is obtained by solving the following initial-boundary value problem:

$$\left.\begin{array}{l}
\dfrac{\partial \rho}{\partial t} + u\dfrac{\partial \rho}{\partial x} + k\rho = \dfrac{1}{A}\displaystyle\sum_{j=1}^{V} m_j \delta(x - v_j) \quad \text{in } (0, L) \times (0, T), \\
\rho(0, t) = \rho_0(t) \quad \text{in } [0, T], \\
\rho(x, 0) = \rho^0(x) \quad \text{in } [0, L],
\end{array}\right\} \tag{2}$$

where $\delta(x - b)$ denotes de Dirac point representation of input contamination at particular location b; for each $j = 1, 2, \ldots, V$, $v_j \in (0, L)$ is the point where the j-th contamination source is located, and $m_j(t)$ is the mass flow rate of coliform concentration; k is the loss rate of coliform due to mortality, settling, etc.; and $A(x, t)$ and $u(x, t)$ denote the wetted area of the river cross section and the average water velocity, respectively. These parameters can be calculated by integrating the classical 1D shallow water equations for each point (x, t):

$$\left.\begin{array}{l}
\dfrac{\partial A}{\partial t} + \dfrac{\partial (Au)}{\partial x} = \displaystyle\sum_{j=1}^{V} q_j \delta(x - v_j) \quad \text{in } (0, L) \times (0, T), \\
\dfrac{\partial (Au)}{\partial t} + \dfrac{\partial (Au^2)}{\partial x} + gA\dfrac{\partial \eta}{\partial x} = \displaystyle\sum_{j=1}^{V} q_j V_j \cos(\beta_j) \delta(x - v_j) \\
\qquad + S_f \quad \text{in } (0, L) \times (0, T), \\
A(L, t) = A_L(t), \quad u(0, t) = u_0(t) \quad \text{in } [0, T], \\
A(x, 0) = A^0(x), \quad u(x, 0) = u^0(x) \quad \text{in } [0, L],
\end{array}\right\} \tag{3}$$

where $q_j(t)$ is the flow rate corresponding to the j-th contamination source, $V_j(t)$ is its velocity, and β_j is the angle between the j-th wastewater discharge and the main river; g is the gravity acceleration; S_f denotes the bottom friction stress (dependent on the Chézy coefficient, the gravity, and the area of the wet section); and $\eta(x, t)$ represents the height of the water surface with respect to a fixed reference level.

As originally proposed by Alvarez-Vázquez et al. [1], the model requires the number of segments N to be known. For this research, however, the value of N was varied through multiple simulations to determine if different values of number N

produced different optimal sensor locations along the river. Nevertheless, we must bear in mind that the knowledge of river system and distribution of contamination sources can also provide clues to a suitable range of N.

Thus, the problem of determining the optimal location of the sampling points can be subdivided into N one-dimensional uncoupled problems: for each $i = 1, 2, \ldots N$, we need to obtain the point $p_i \in [a_{i-1}, a_i]$, minimizing the corresponding part of the objective function:

$$J_i(p_i) = \int_0^T (\rho(p_i, t) - c_i(t))^2 \, dt \tag{4}$$

(that can be numerically computed, for instance, by the standard trapezoidal rule).

As it is well known, these N problems have solution, but no necessarily unique. An optimal solution can be obtained by any simple one-dimensional optimization method. In this work we have used the golden-section direct search method (see, for instance, monograph [2] for complete details on formulation and convergence of the optimization method).

3 Case Study and Final Results

The Neuse River was selected as the case study because of its vital public health and economic importance to the Piedmont and Coastal Plain of North Carolina. Noted as one of "America's Ten Most Endangered Rivers" in 1995, 1996, 1997, and 2007, the Neuse River supports a billion-dollar fishing industry after emptying into the nation's second largest estuary, the Pamlico Sound [6]. Meandering between open pit lagoons which store waste from the country's second largest swine industry, the river is polluted by swine waste contamination when storms flood the region [10]. Storms which cause the Neuse to flood can lead to the direct mixing of swine waste into river water, as experienced during the huge flooding caused by Hurricane Floyd in 1999. Inputs to the model concerning river flow were selected using river gauge data and historical records of flooding.

In this characterization of the case study, the results of optimal locations describe where to put water quality sensors to best understand the impact of swine waste on water quality. A short section of the river was selected to focus the analysis of the model results (see Fig. 1). The selected river section is 53.5 km in length and sits immediately before the Neuse's discharge near the city of New Bern, NC. This section was chosen because it contains a river gauge station and many swine waste lagoons in the river's floodplain. Using river gauge station data as inputs demonstrates the opportunity to implement this model as part of the design of a water quality monitoring system.

When applying our model to a river like the Neuse, inputs can be varied to study how the optimal sensor locations change as a result of different river conditions. The first model inputs that were varied between simulations included river flow and

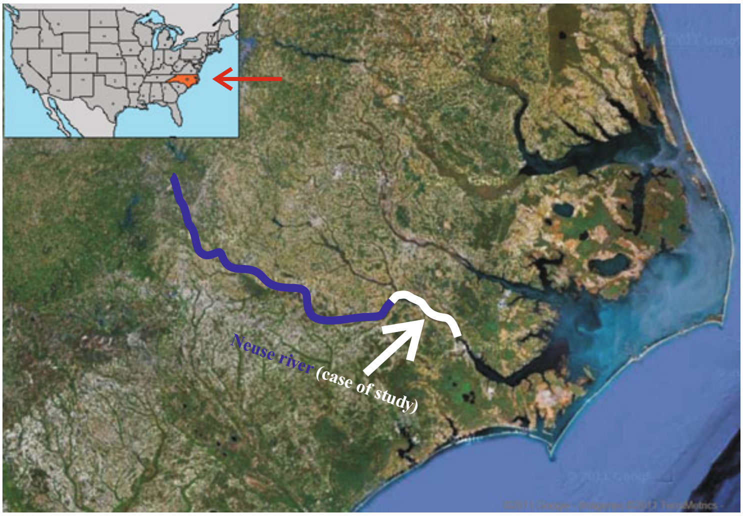


Fig. 1 Case study for our analysis: Neuse River (NC) satellite image. In *white*, the river section under study

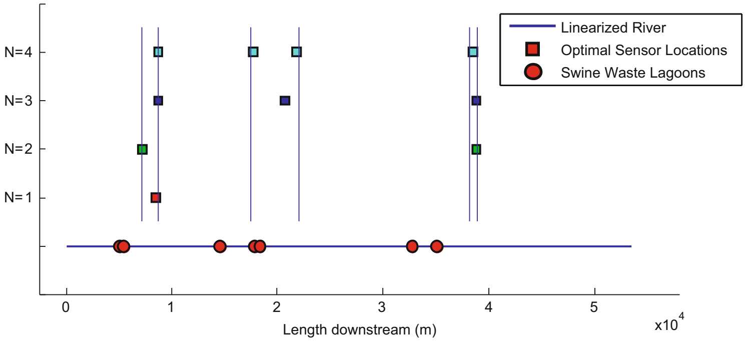


Fig. 2 Optimal water quality sensor locations from medium contamination scenario for all allowed values of $N = 1, \ldots, 4$

contamination levels. River conditions included small, medium, large, and extra large contamination levels and water flows. The fourth, most extreme case was included to analyze how complete mixing of swine waste lagoons would impact the location of optimal location of water quality sensors. The mixing of waste lagoons' entire contents into surface water is rare, but was well documented from the inundation caused by Hurricane Floyd [6].

The second set of variables used in this analysis of this siting model corresponds to the number of river segments N into which the selected 53.5 km section was divided in this case study. Seven modelled contaminant sources are geographically grouped into three distinct regions along the river reach (see Fig. 2), and simulations were run for each flow scenario for the values $N = 1, 2, 3$, and 4.

While Lettenmaier and Burges [9] concluded that the number of sampling points was generally more important than the location of each sampling point or the sampling frequency, varying N can allow an exploration into the importance of the number of sampling points to this case study.

Using four different flows and contamination scenarios combined with four different values of N produces an experimental design consisting of 16 model simulations. Each of the four contamination scenarios varied the contamination flow rates and water parameters such as water depth in the river and water velocity.

To illustrate our results, Fig. 2 shows that even as the number of modelled segments increases up to $N = 4$, the suggested sensor locations always group into three main regions. Comparing the results, for instance, for the case of medium contamination suggest that three sensors placed within the regions 7,200–8,700, 20,700–21,800, and 38,500–38,800 m provide optimal information about the contamination along the entire river. (These regions boundaries are illustrated by the pairs of vertical lines in Fig. 2). Achieved results are similar for the other contamination scenarios, showing that a fourth sensor looks unnecessary.

As a final conclusion, we can say that the application of specific mathematical models to optimize water quality sensor locations shows to be a powerful method which can be used for any river. The proposed strategy consisting of solving the optimization problem under multiple system scenarios provides the ability to study how optimal sensor locations are affected by anticipated contamination events and river flows.

Acknowledgements The financial support provided by the Projects MTM2012-30842 and MTM 2009-07749 of M.E.C. (Spain) and FEDER is gratefully acknowledged.

References

1. Alvarez-Vázquez, L.J., Martínez, A., Vázquez-Méndez, M.E., Vilar, M.A.: Optimal location of sampling points for river pollution control. Math. Comput. Simul. **71**, 149–160 (2006)
2. Bazaraa, M.S., Shetty, C.M.: Nonlinear Programming. Theory and Algorithms. Wiley, New York (1979)
3. Chang, N.B., Makkeasorn, A.: Optimal site selection of watershed hydrological monitoring stations using remote sensing and grey integer programming. Environ. Model. Assess. **15**, 469–486 (2010)
4. Dixon, W., Smyth, G.K., Chiswell, B.: Topologically optimum monitoring of rivers by approximation algorithms. In: Proceedings of International Symposium on Environmental Chemistry and Toxicology, Sydney (1996)
5. Dixon, W., Smyth, G.K., Chiswell, B.: Optimized selection of river sampling sites. Water Res. **33**, 971–978 (1999)
6. Environmental Management Commission, NC Division of Water Quality: Neuse River Basinwide Water Quality Plan. Chapter 217 (2009)
7. Hren, J., Childress, C.J.D., Norris, J.M., Chaney, T.H., Myers, D.N.: Regional water quality: evaluation of data for assessing conditions and trends. Environ. Sci. Technol. **24**, 1122–1127 (1990)

8. Karamouz, M., Kerachian, R., Akhbari, M., Hafez, B.: Design of river water quality monitoring networks: a case study. Environ. Model. Assess. **14**, 705–714 (2009)
9. Lettenmaier, D.P., Burges, S.J.: Design of trend monitoring networks. ASCE J. Environ. Eng. Div. **103**, 785–802 (1977)
10. National Resources Defense Council: America's Animal Factories: How States Fail to Prevent Pollution from Livestock Waste. Chapter 17 (1998)
11. Sanders, T.G., Ward, R.C., Loftis, J.C., Steele, T.D., Adrian, D.D., Yevjevich, V.: Design of Networks for Monitoring Water Quality. Water Resources Publications, Littleton (1983)
12. Sharp, W.E.: A topographical optimum water-sampling plan for rivers and streams. Water Resour. Res. **7**, 1641–1646 (1971)
13. Strobl, R.O., Robillard, P.D., Shannon, R.D., Day, R.L., McDonnell, A.J.: A water quality monitoring network design methodology for the selection of critical sampling points: Part I. Environ. Monit. Assess. **112**, 137–158 (2006)
14. Ward, P.R.B.: Prediction of mixing lengths for river flow gauging. ASCE J. Hydraul. Div. **99**, 1069–1081 (1973)

Global Analysis of a Nonlinear Model for Biodegradation of Toxic Compounds in a Wastewater Treatment Process

Neli Dimitrova

Abstract The paper presents rigorous mathematical stability analysis of a dynamic model, describing biodegradation of toxic substances in a wastewater treatment plant. Numerical simulations support the theoretical results.

1 Introduction

Toxicity of 1,2-dichloroethane (DCA), in particular for aquatic and atmospheric biotic systems, has been recently recognized as a serious ecological problem [4]. DCA is difficult to remove from aquatic media by physico-chemical methods due to its very low concentration. Therefore, biodegradation remains the only available alternative. A microbial strain, recently recommended as a "novelty" and capable to degrade DCA to its complete mineralization is *Klebsiella oxytoca VA 8391* [3, 4]. This strain was isolated from active sludge from a wastewater plant at the Luckoil Neftochim Rafinery in Burgas, Bulgaria. The identification was validated by the National Bank for Industrial Microorganisms and Cultures in Sofia, Bulgaria, and the strain was registered under the code number stated above.

We consider a continuous bioreactor model for DCA biodegradation by *Klebsiella oxytoca VA 8391* immobilized on granulated activated carbon. During the microbial process the immobilized cells can detach from the solid surface and live and grow in the liquid phase. The process is irreversible, i. e. free cells can not attach again the solid particles. The model is developed and validated in [4] by authors' own experiments.

N. Dimitrova (✉)
Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Acad. G. Bonchev Str., Bl. 8, 1113 Sofia, Bulgaria
e-mail: nelid@math.bas.bg

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_7,

2 Model Description

The continuous flow bioreactor model describing DCA biodegradation by *Klebsiella oxytoca VA 8391* immobilized on granulated activated carbon is presented by the following differential equations [4]

$$
\begin{aligned}
\dot{x}_1 &= (\mu_1(s) - D)x_1 + k_{im}x_{im} && (1)\\
\dot{x}_{im} &= (\mu_{im}(s) - k_{im})\, x_{im} && (2)\\
\dot{s} &= -\left(\frac{1}{\gamma}\mu_1(s) + \beta_1\right) x_1 - \left(\frac{1}{\gamma}\mu_{im}(s) + \beta_{im}\right) x_{im} && (3)\\
&\quad + D(s^{in} - s) - k_L a(1 - \mu_2(s))s\\
\dot{p} &= \left(\frac{1}{\gamma}\mu_1(s) + \beta_1\right) x_1 + \left(\frac{1}{\gamma}\mu_{im}(s) + \beta_{im}\right) x_{im} - Dp, && (4)
\end{aligned}
$$

where the dot over the phase variables means $\frac{d}{dt}$. The functions $\mu_1(s)$ and $\mu_{im}(s)$ are the specific growth rates of the free and the immobilized cells respectively, $\mu_2(s)$ is related to the adsorption capacity. The following functions are proposed in [4]:

$$\mu_1(s) = \frac{m_1 s}{k_s + s + s^2/k_i}, \quad \mu_{im}(s) = \frac{m_{im} s}{k_s + s + s^2/k_i}, \quad \mu_2(s) = \frac{m_2 s}{k + s}.$$

The growth rate functions $\mu_1(s)$ and $\mu_{im}(s)$ exhibit inhibition, i.e. they achieve their maximum at the point $s^m = \sqrt{k_s k_i}$. The function $\mu_2(s)$ is bounded and $\mu_2(s) < m_2$ is valid for all $s \geq 0$. The definition of the phase variables x_1, x_{im}, s and p as well as of the model parameters is given in Table 1.

In the bioreactor, the free cells are expected to consume easily the substrate necessarily for their growth, but they are more keen to be carried out by the flow. On the contrary, the immobilized cells have a more difficult access to the resources of the bulk fluid, but are more resistent to detachment induced by the hydrodynamical conditions. To predict this observation by the model, we assume that the following inequality holds true (see also the hypothesis (H5) below)

(H1) $m_{im} < m_1$

This inequality implies that $\mu_{im}(s) < \mu_1(s)$ for all $s > 0$.

3 Equilibrium Points of the Model and Their Lyapunov Stability

Denote by

$$\phi(s) = D(s^{in} - s) - k_L a(1 - \mu_2(s))s$$

Table 1 Definition of the model variables and parameters

	Definitions	Values
x_1	Concentration of free cells [kg m^{-3}]	–
x_{im}	Concentration immobilized cells [kg m^{-3}]	–
s	Substrate (DCA) concentration [kg m^{-3}]	–
p	Product (chloride) concentration [kg m^{-3}]	–
D	Dilution rate [h^{-1}]	5.9
k_{im}	Cell leakage factor [m h^{-1}]	0.01
s^{in}	Inlet substrate concentration s_2 [mmol/l]	0.05
β_1	Biodegradation rate constant due to free cells [h^{-1}]	0.001
β_{im}	Biodegradation rate constant due to immobilized cells [h^{-1}]	0.0015
γ	Yield coefficient for free biomass production [(kg cells)/(kg substr.)]	77.6
k	Parameter in the Langmuir isotherm	0.612
k_s	Saturation constant [kg m^{-3}]	0.26
k_i	Substrate inhibition constant [kg m^{-3}]	0.984
k_La	Volumetric mass transfer coefficient for DCA for adsorption [h^{-1}]	0.51
m_1	Maximum specific growth rate for free cells [h^{-1}]	0.972
m_2	Surface concentration limit of DCA in the Langmuir isotherm [g kg^{-1}]	0.63
m_{im}	Maximum specific growth rate for immobilized cells [h^{-1}]	0.18

the function included in the right-hand side of (3) and assume that the following inequality is satisfied:

(H2) $\max\{k_La,\ m_2\} < 1$.

It is straightforward to see, that $\frac{d}{ds}\phi(s) < 0$ for all $s \geq 0$; moreover, there exists a unique positive root ζ_0 of $\phi(s) = 0$ such that $\zeta_0 < s^{in}$ and further $\phi(s) \geq 0$ if $s \in [0, \zeta_0]$, and $\phi(s) < 0$ if $s > \zeta_0$.

The equilibrium points of the model are solutions of the form (x_1, x_{im}, s, p) of the nonlinear system, obtained from (1) to (4) by setting the right-hand sides equal to zero. We are looking for equilibrium points with nonnegative components due to physical evidence.

Proposition 1. *Under assumptions (H1) and (H2), the equilibrium points of the model are the following:*

(i) $E_0 = (0,\ 0,\ \zeta_0,\ 0)$*;*

(ii) $E_i = \left(\frac{\phi(\xi_i)}{\frac{1}{\gamma}D+\beta_1},\ 0, \xi_i,\ \frac{\phi(\xi_i)}{D}\right)$, $i = 1, 2$, *(with* $x_{im} = 0$*) where* ξ_i *are solutions of* $\mu_1(s) = D$*;* E_i *exist if and only if* $D \leq \max_{s>0}\ \mu_1(s) = \mu_1(s^m)$ *and* $\phi(\xi_i) > 0$.

(iii) $F_i = \left(x_1^{(i)}, x_{im}^{(i)}, \zeta_i, p^{(i)}\right)$, $i = 1, 2$, *where* ζ_i *are solutions of* $\mu_{im}(s) = k_{im}$*,* $x_1^{(i)} = \frac{k_{im}\phi(\zeta_i)}{\beta_{im}(D-\mu_1(\zeta_i))+k_{im}\left(\frac{1}{\gamma}D+\beta_1\right)}$, $x_{im}^{(i)} = \frac{D-\mu_1(\zeta_i)}{k_{im}}x_1^{(i)}$ *and* $p^{(i)} = \frac{x_1^{(i)}}{D}\left(\left(\frac{1}{\gamma}\mu_1(\zeta_i) + \beta_1\right) + \left(\frac{1}{\gamma}\mu_{im}(\zeta_i) + \beta_{im}\right)\frac{D-\mu_1(\zeta_i)}{k_{im}}\right)$*;* F_i *exist if and only if* $k_{im} \leq \max_{s>0}\mu_{im}(s) = \mu_{im}(s^m)$*,* $D > \mu_1(\zeta_i)$ *and* $\phi(\zeta_i) > 0$.

The point E_0 is called wash-out equilibrium. The existence of E_i corresponds to the case of free microbial culture without immobilized cells on the carrier. Practically the most important equilibria are the internal points F_i; the condition $k_{im} \leq \mu_{im}(s^m)$ describes the case of compensated immobilized cell leakage by growth within the particles.

Let $E \in \{E_0, E_1, E_2, F_1, F_2\}$ be any one of the equilibrium points, described above. Denote by $J(E)$ the Jacobian of (1)–(4) evaluated at E. The eigenvalues of $J(E)$ are the roots of the following characteristic equation (I denotes the (4×4)-unit matrix) $0 = |J(E) - \lambda I| = (-D - \lambda) \cdot (-\lambda^3 + a\lambda^2 - b\lambda + c)$, where the coefficients $a = a(E)$, $b = b(E)$ and $c = c(E)$ can be computed explicitly, using the well known invariants of the matrix $J(E)$. Obviously, $\lambda_4 = -D < 0$ is an eigenvalue of every equilibrium point $E \in \{E_0, E_1, E_2, F_1, F_2\}$. This means that there are no repelling steady states in the model. The other three eigenvalues are the roots of the cubic polynomial $g(\lambda) = -\lambda^3 + a\lambda^2 - b\lambda + c$. Using the Routh-Hurwitz criterion [5] for determining the signs of the real parts of the roots of $g(\lambda)$, we obtain the following

Proposition 2. *Let the hypotheses (H1) and (H2) be satisfied.*

(i) If $\mu_1(\zeta_0) < D$ and $\mu_{im}(\zeta_0) < k_{im}$ are fulfilled, the equilibrium point E_0 is locally asymptotically stable; otherwise E_0 is a saddle.
(ii) Let the assumptions of Proposition 1(ii) be satisfied. If $\mu_{im}(\xi_i) < k_{im}$, $i = 1, 2$, then E_1 is locally asymptotically stable and E_2 is a saddle equilibrium point. If $\mu_{im}(\xi_i) > k_{im}$, $i = 1, 2$, then E_1 and E_2 are saddle equilibrium points.
(iii) Let the assumptions of Proposition 1(iii) hold. Then F_1 is locally asymptotically stable and F_2 is a saddle equilibrium point.

4 Global Properties of the Solutions

The first three equations (1)–(3) do not depend on p. If we "compute" the solutions $x_1(t)$, $x_{im}(t)$, $s(t)$ and replace them in (4), we obtain a linear nonautonomous equation for p of the form $\dot{p} = -D\,p + \psi(t)$, which can be integrated directly. Therefore, we can omit the last equation (4) in the further considerations.

We impose additionally the following assumption on (1)–(3)

$$\text{(H3)} \quad \beta_1 < \beta_{im} < \frac{k_L a}{\gamma}, \qquad D > 1 - k_L a(1 - m_2)$$

Proposition 3. *Let the assumptions (H1)–(H3) be fulfilled. Then the set $\Omega = \big\{(x_1, x_{im}, s) : x_1 \geq 0, x_{im} \geq 0, s \geq 0, Ds^{in} \geq s + \beta_1 x_1 + \beta_{im} x_{im}\big\}$ is positively invariant for the model; all solutions are uniformly bounded for all $t \geq 0$ and thus exist for $t \in [0, +\infty)$.*

Experimental results show that the inlet substrate concentration s^{in} must be lower than the one corresponding to the maximum specific growth rate, i.e. s^{in} should be below the point s^m where substrate inhibition starts to be significant. Assume that the following inequalities are fulfilled:

(H4) $\quad s^{in} < s^m, \quad k_{im} < \mu_{im}(\zeta_0).$

It is not difficult to see that under assumptions (H1)–(H4), $s(t) < \zeta_0$ is valid for all sufficiently large $t > 0$. Moreover, since $\zeta_0 < s^{in}$ holds, assumption (H4) implies that the functions $\mu_1(s)$ and $\mu_{im}(s)$ are monotone increasing for $s \in [0, \zeta_0]$. Our last assumption is

(H4) $\quad D > \mu_1(s^{in}) + k_{im}$

The hypotheses (H1)–(H5) and Proposition 1 imply that there exist only two equilibrium points of (1)–(3) in Ω, namely E_0 and F_1; thereby F_1 is locally asymptotically stable, E_0 is a saddle equilibrium. We shall show that $F_1 = (x_1^{(1)}, x_{im}^{(1)}, \zeta_1)$ is globally asymptotically stable for the model.

Theorem 1. *Let the assumptions (H1)–(H5) be satisfied. Then the equilibrium point F_1 is globally asymptotically stable for (1)–(3) in the set Ω.*

Proof. It is enough to show that the stable manifold of E_0 lies exterior to the set Ω (cf. [6]). The negative eigenvalues of $E_0 = (0, 0, \zeta_0)$ are $\lambda_1 = \mu_1(\zeta_0) - D$ and $\lambda_2 = \frac{d}{ds}\phi(\zeta_0)$. Denote by $u = (u_1, u_2, u_3)$ and $v = (v_1, v_2, v_3)$ the corresponding eigenvectors. It is easy to see that $u_2 = 0$ and $qu_3 = -\left(\frac{1}{\gamma}\mu_1(\zeta_0) + \beta_1\right) u_1$ within $q = \mu_1(\zeta_0) - D - \frac{d}{ds}\phi(\zeta_0) > 0$. Therefore, u cannot be directed inside the positive octant. The same is valid for the eigenvector v, since the latter has the form $v = (0, 0, v_3)$ with $v_3 \neq 0$. Therefore, the stable manifold of E_0 does not intersect the interior of Ω, which implies that F_1 attracts all solutions with initial conditions in Ω, i.e. F_1 is a global attractor. This completes the proof.

5 Numerical Simulation

Consider the numerical coefficient values in Table 1 (last column). For these values, all the assumptions (H1)–(H5) are satisfied, and therefore Theorem 1 holds true.

Figure 1 visualizes results from computer experiments with an initial point $(x_1(0), x_{im}(0), s(0), p(0))$ from the set Ω, i.e. satisfying $Ds^{in} \geq s(0) + \beta_1 x_1(0) + \beta_{im} x_{im}(0)$. The solid circles correspond to experimental measurements, taken from [4].

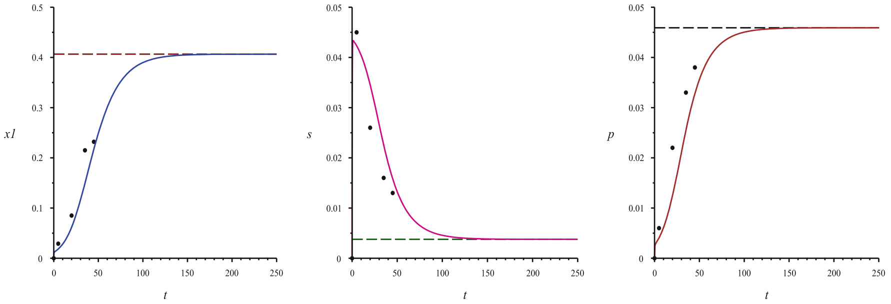


Fig. 1 Phase curves $x_1(t)$ (*left*), $s(t)$ (*middle*) and $p(t)$ (*right*); the *horizontal dashed lines* pass through the components of F_1. *Solid circles* denote experimental data

6 Conclusion

The paper presents global stability analysis of a practically validated ecological model for wastewater treatment. Most of the results are obtained and proved in [1,2]. The proof of the above Theorem 1 is new. Here, the computer simulations are compared with experimental measurements.

The present mathematical analysis of the model (1)–(4) could be useful to outline the parameter domain for stable operation of the microbial process in a continuously stirred bioreactor.

References

1. Borisov, M., Dimitrova, N.: Stability analysis in a model of 1,2-dichloroethane biodegradation by Klebsiella oxytoca va 8391 immobilized on granulated activated carbon. In: Todorov, M.D., Christov, C.I. (eds.) Proceedings of AMiTaNS'2011, vol. 1404, pp. 284–298. American Institute of Physics, Melville (2011)
2. Borisov, M., Dimitrova, N., Beschkov, V.: Stability analysis of a bioreactor model for biodegradation of xenobiotics. Int. J. Comput. Math. Appl. **64**(3), 361–373 (2012)
3. Mileva, A., Beshkov, V.: On the dechlorination capacity of the strain *Klebsiella oxytoca* on 1,2-dichloroethane. C. R. Acad. Bulg. Sci. **59**, 959–964 (2006)
4. Mileva, A., Sapundzhiev, Ts., Beshkov, V.: Modeling 1,2-dichloroethane biodegradation by *Klebsiella oxytoca va 8391* immobilized on granulated activated carbon. Bioprocess Biosyst. Eng. **31**, 75–85 (2008)
5. Sendov, Bl., Andreev, A., Kjurkchiev, N.: Numerical solution of polynomial equations. In: Ciarlet, P.G., et al. (eds.) Handbook of Numerical Analysis, vol. 3, pp. 625–778. North-Holland, Amsterdam (1999)
6. Smith, H.L., Waltman, P.: The theory of the chemostat. Dynamics of microbial competition. Cambridge University Press, Cambridge (1995)

Pollutant Transport and Its Alleviation in Groundwater Aquifers

Amjad Ali, Winston L. Sweatman, and Robert McKibbin

Abstract Dissolved chemicals are transported through groundwater aquifers by a mixture of advection with the underlying fluid flow and dispersion within that fluid. The aquifers can be modelled using a layered structure which simplifies the calculation of vertical transport. This simplified model still allows for the natural stratification which occurs in such systems and the changes in physical properties of the aquifer that occur between different geological layers. Equations are presented to calculate the subsequent concentration of the releases of chemicals into this system. A particular example is considered where an instantaneous release of pollutant occurs and it is subsequently remediated by the downstream release of a suitable pollutant removal agent.

1 Introduction

Pollutants released at or below ground level can be transported elsewhere by the flow of groundwater in subterranean aquifers. The pollution may occur as an instantaneous release such as at the location of an accidental spill or it may be a

A. Ali • W.L. Sweatman (✉) • R. McKibbin
Centre for Mathematics in Industry, Institute of Natural and Mathematical Sciences, Massey University, Auckland, New Zealand
e-mail: a.ali.1@massey.ac.nz; w.sweatman@massey.ac.nz; r.mckibbin@massey.ac.nz

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_8,

more gradual release such as the seepage of toxins from pre-existing rubbish dump sites. For some cases compensatory action may be taken to remediate the damaging effects by injecting suitable reagents downstream of the pollution source. Examples of groundwater contaminants include a variety of inorganic and organic chemicals or bacteriological compounds. The remediating pollution removal agents could be, for example, chemical oxidants such as oxygen, hydrogen peroxide, permanganate and persulfate. We develop a model to include such effects, and present a simple illustration of this new approach.

2 The Aquifer Dispersion Model

The model for the groundwater aquifer builds upon that developed in [1]. This assumes that the aquifer can be discretised into a number of distinct layers each of which has physical properties which do not depend upon height. The vertical discretisation is partially motivated by the natural layers that occur underground due to their geological formation. Such geological layers can have significant physical differences and their effects are included in our model. However, in general the thickness of an aquifer is very small compared to its lateral extent [2] and it is not unreasonable to assume that the mechanism of material transport will not vary greatly with height within thin homogeneous layers. It is possible to consider a gradual horizontal variation in the thickness profiles of layers and their physical properties within these models. The aquifer considered will be constrained at both its base and top. It is also possible to consider the case where the groundwater flow is unconstrained above with a phreatic surface.

Groundwater flows within the aquifer transporting dissolved or suspended substances. We introduce a particular case where there are two such substances present: a pollutant, perhaps from a leak or spill, and a pollutant removal agent, introduced to react with and alleviate the pollutant. Within the aquifer the transported substances, pollutant or removal agents, move in a similar way. They are transported both advectively with the groundwater flow and dispersively within the fluid. Horizontally, we consider the transport within a single layer. Vertically, the transport occurs between layers. Within a single layer the concentration varies with horizontal position but is independent of height within the layer. The concentration varies vertically between the different layers. As well as dispersive transfer of pollutant between neighbouring layers, fluid flux through the layer interfaces may carry (advect) dissolved pollutant with it across the layer interface.

Suppose, $\mathbf{q}(x, y) = (q_x, q_y)$ is the total volume flux vector of the fluid per unit width through the whole aquifer at the horizontal position (x, y) and $\mathbf{q_i}$ is that in the ith layer. Let $P_i(x, y, z)$ and $R_i(x, y, z)$ $[\mathrm{M\,L^{-3}}]$ be respectively the averaged concentrations of the pollutant and the removal agent over the layer thickness h_i in the ith layer of the aquifer. They satisfy

$$
\begin{aligned}
\phi_i h_i \dot{P}_i = & -\mathbf{q}_i \cdot \nabla P_i + \frac{\partial}{\partial x}\left(h_i \phi_i D_{H_i} \frac{\partial P_i}{\partial x}\right) + \frac{\partial}{\partial y}\left(h_i \phi_i D_{H_i} \frac{\partial P_i}{\partial y}\right) \\
& + \left(\tau_{i-1} + \frac{1 + \operatorname{sgn} r_{i-1}}{2} r_{i-1}\right) P_{i-1} \\
& + \left(-\tau_{i-1} + \frac{1 - \operatorname{sgn}(r_{i-1})}{2} r_{i-1}\right) P_i \\
& + \left(-\tau_i - \frac{1 + \operatorname{sgn}(r_i)}{2} r_i\right) P_i \\
& + \left(\tau_i - \frac{1 - \operatorname{sgn}(r_i)}{2} r_i\right) P_{i+1} + \phi_i h_i f_{P_i} - \phi_i h_i k_P P_i R_i \qquad (1)
\end{aligned}
$$

$$
\begin{aligned}
\phi_i h_i \dot{R}_i = & -\mathbf{q}_i \cdot \nabla R_i + \frac{\partial}{\partial x}\left(h_i \phi_i D_{H_i} \frac{\partial R_i}{\partial x}\right) + \frac{\partial}{\partial y}\left(h_i \phi_i D_{H_i} \frac{\partial R_i}{\partial y}\right) \\
& + \left(\tau_{i-1} + \frac{1 + \operatorname{sgn} r_{i-1}}{2} r_{i-1}\right) R_{i-1} \\
& + \left(-\tau_{i-1} + \frac{1 - \operatorname{sgn}(r_{i-1})}{2} r_{i-1}\right) R_i \\
& + \left(-\tau_i - \frac{1 + \operatorname{sgn}(r_i)}{2} r_i\right) R_i \\
& + \left(\tau_i - \frac{1 - \operatorname{sgn}(r_i)}{2} r_i\right) R_{i+1} + \phi_i h_i f_{R_i} - \phi_i h_i k_R P_i R_i \qquad (2)
\end{aligned}
$$

where ϕ_i [-] is the porosity of the solid matrix of the aquifer in the ith layer, D_{H_i} and D_{V_i} [$\mathrm{L}^2\,\mathrm{T}^{-1}$] are respectively the coefficients of horizontal and vertical dispersion, and $r_i(x, y)$ is the directed interlayer fluid transfer from the ith layer to the $i + 1$th layer normal to the layer interface. The interlayer dispersive transfer coefficient between the ith and the $i + 1$th layer $\tau_i(x, y)$ is estimated using $1/\tau_i = h_i/(2\phi_i D_{V_i}) + h_{i+1}/(2\phi_{i+1} D_{V_{i+1}})$ for internal layer boundaries and $\tau_i = 0$ at the base and top of the aquifer. The functions $f_{P_i}(x, y, t)$ and $f_{R_i}(x, y, t)$ are respectively source terms for pollutant and removing agent at the point (x, y) of the ith layer averaged over the layer thickness, and k_P and k_R [$(\mathrm{M}\,\mathrm{L}^{-3})^{-1}\,\mathrm{T}^{-1}$] are respectively the rates of decay of pollutant and pollutant removing agent as a result of their chemical reaction. There is further discussion of the transport of a single substance alone in [3]. We note, that as the pollutant and pollutant removal agent are transported in a similar way, the equations for the two substances only differ in their last two terms: the source term and interaction term.

Table 1 Parameters used in the illustration (Fig. 1)

Parameter names	Values	Units
Total volumetric flux q_x	10	$\mathrm{m}^2\,\mathrm{day}^{-1}$
Porosity ϕ	0.1	–
Horizontal dispersion coefficient D_{H_i}	0.2	$\mathrm{m}^2\,\mathrm{day}^{-1}$
Vertical dispersion coefficient D_{V_i}	0.06	$\mathrm{m}^2\,\mathrm{day}^{-1}$
Degradation rate coefficient for pollutant k_P	3	$\left(\mathrm{kg\,m}^{-3}\right)^{-1}\,\mathrm{day}^{-1}$
Degradation rate coefficient for pollutant removing agent k_R	1	$\left(\mathrm{kg\,m}^{-3}\right)^{-1}\,\mathrm{day}^{-1}$

3 An Illustration of the Model

We present an illustration of the model. For simplicity, each horizontal layer is taken to be homogeneous and uniform in thickness. Five layers are used to represent each of three geological strata. The middle of the aquifer is composed of sand and gravel with permeability $10^{-9}\,\mathrm{m}^2$ and the upper and lower layers are composed of clean sand with permeability $10^{-10}\,\mathrm{m}^2$ [4]. The values of other parameters used are shown in Table 1. There is no interlayer fluid transfer (r_i), but there will be dispersive transfer of the species due to concentration gradients across the layer interfaces. The ratios of the vertical transverse dispersion constants to the horizontal ones are small [5]. Variation in the y-direction has been suppressed and we consider the concentration dependence solely in the x and z directions.

We consider an instantaneous release of 2 units of a pollutant, such as might happen as the result of an industrial accidental spill. Figure 1 shows concentration profiles for the pollutant at 3 and 6 days subsequent to its release. The point of pollutant release is marked as a red rectangle. There are two simulations, with and without pollutant removal agent. In the latter, after a 2-day delay, a pollutant removal agent is released 10 m downstream of the site of pollutant release (marked as a green rectangle in Fig. 1). This release is taken to be continuous with a constant rate of 2 units per day. In their interaction, each unit of pollutant removal agent can remove 3 units of pollutant. For this example, both the releases occur within the fourth layer (about 2 m) below the top of the aquifer.

As the central layers of sand and gravel are more permeable than the clean sand layers above and below, the underlying horizontal groundwater flow is more rapid there. The corresponding advected pollutant and removal agent concentrations can be seen in Fig. 1.

4 Conclusions

In this short paper, a simple model has been presented for calculating the concentrations of a pollutant and a pollutant removal agent transported by groundwater flow within an aquifer. An example illustrates the process, showing the effect of remedial action.

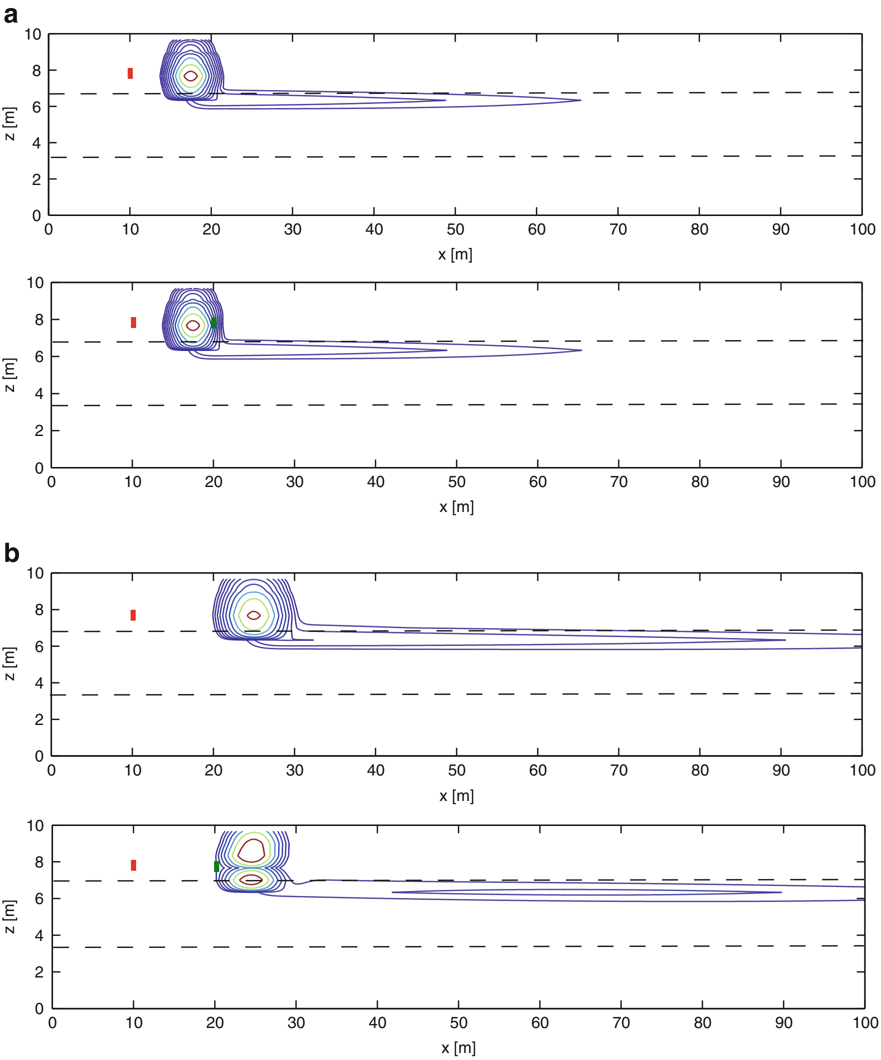


Fig. 1 Contour plots of the concentration of the pollutant subsequent to release. The *top pair* of graphs are for 3 days after the pollutant release and the *bottom pair* are for 6 days after. Within each set, the *top figure* is a contour plot of what the pollutant concentration would be without any pollutant removal agent and the *bottom figure* is the pollutant concentration having included the effect of the pollutant removal agent. The release positions of the pollutant and pollutant removal agent are marked as *red* and *green rectangles*, respectively

References

1. McKibbin, R.: Groundwater pollutant transport: transforming layered models to dynamical systems. An. Şt. Univ. Ovidius Constanţa. Ser. Mat. **17**(3), 183–196 (2009)
2. Bear, J., Bachmat, Y.: Introduction to Modeling of Transport Phenomena in Porous Media. Kluwer, Dordrecht (1991)
3. McKibbin, R.: Some aspects of modelling pollution transport in groundwater aquifers. In: Weiranto, L.H., Pudjaprasetya, S. (eds.) Proceedings of CIAM 2010, Conference on Industrial and Applied Mathematics, 6–8 July 2010, pp. 9–16. Institut Teknologi Bandung, Indonesia (2010)
4. Bear, J., Verruijt, A.: Modeling Groundwater Flow and Pollution. D. Reidel, Dordrecht (1978)
5. Gelhar, L.W., Welty, C., Rehfeldt, K.R.: A critical review of data on field-scale dispersion in aquifers. Water Resour. Res. **28**(7), 1955–1974 (1992)

Optimal Shape Design of Wastewater Canals in a Thermal Power Station

Aurea Martínez, Lino J. Alvarez-Vázquez, Carmen Rodríguez, Miguel E. Vázquez-Méndez, and Miguel A. Vilar

Abstract Inside the canals of wastewater treatment plants of thermal power stations usually produces in a natural way a deposition of particles in suspension, which causes a change in geometry of the bottom of channel, with the consequent appearance of accumulated sludge and growth of algae and vegetation. This fact may lead to a misfunction of the purification process in the plant. Our main aim focuses on the optimal design of the geometry of such canals to avoid the difficulties derived from these processes. The problem can be formulated as a control-constrained optimal control problem of partial differential equations, and discretized via a characteristics/finite element method. For a simplified case study (canals of rectangular section), theoretical and applicable results are presented.

1 Introduction

A thermal power station is a facility used for generating electrical energy from the energy released as heat, usually by combustion of fossil fuels like oil, natural gas or coal. This heat is used by a thermodynamic cycle to move a conventional alternator and produce energy.

A. Martínez (✉) • L.J. Alvarez-Vázquez
Departamento de Matemática Aplicada II, E.I. Telecomunicación, Universidad de Vigo,
36310 Vigo, Spain
e-mail: aurea@dma.uvigo.es; lino@dma.uvigo.es

C. Rodríguez
Departamento de Matemática Aplicada, Fac. Matemáticas, Universidad de Santiago
de Compostela, 15782 Santiago, Spain
e-mail: carmen.rodriguez@usc.es

M.E. Vázquez-Méndez • M.A. Vilar
Departamento de Matemática Aplicada, E.P.S., Universidad de Santiago de Compostela,
27002 Lugo, Spain
e-mail: miguelernesto.vazquez@usc.es; miguel.vilar@usc.es

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_9,

One problem faced by these plants is their need for cooling, as they need to evacuate about half of the total thermal power. Conventional techniques need to employ large amounts of water which is returned to environment after suffering a significant temperature drop.

In these traditional systems, water in circulation that cools the condenser expels the heat extracted to the atmosphere through cooling towers (large hyperboloid shaped structures which identify these plants). We must take into account the need to purge part of the salts contained in the evaporated water degrading its quality (in the towers cooling, due to evaporation, increases saline concentration). So, to avoid problems in the system, purges are made in the towers, and this removed liquid effluent must also be treated. Because of these facts it is necessary to design and build for the thermal power station a wastewater treatment plant in order to give a specialized treatment of water that is generated.

A wastewater treatment plant of these characteristics is intended to obtain from wastewater, using different physical, chemical or biological techniques, a effluent water of improved quality characteristics, using as reference certain standard parameters. Inside this treatment plant water transfers occur between different containers through canals. In these canals usually produces in a natural way a deposition of particles in suspension, which causes a change in geometry of the bottom of channel, with the consequent appearance of accumulated sludge and growth of algae and vegetation. This fact may lead to a bad operation of the purification process in the plant. Our objective will then focus on the optimal design of the geometry of such canals to avoid the problems outlined above.

2 Modelling and Resolution

In order to avoid problems arising from the accumulation of vegetation and sludge in the canals of a wastewater treatment plant in a power station, we will try to optimize the design of the section of these channels so as to minimize these negative effects.

When formulating the problem mathematically we need to deal with the hydrodynamics (the shallow water equations in the domain that forms the channel), the transport of sediments (a convection-reaction-diffusion equation), and the deposition of sediment in the canal bottom [4–10]. Thus, for a canal of length L and for a time interval of length T, the state system modelling the sedimentation process is given by the following set of coupled, nonlinear partial differential equations:

$$\begin{cases} \dfrac{\partial A}{\partial t} + \dfrac{\partial Q}{\partial x} = 0 \qquad \text{in } (0,L) \times (0,T) \\ \dfrac{\partial Q}{\partial t} + \dfrac{\partial}{\partial x}\left(\dfrac{Q^2}{A}\right) + gA\dfrac{\partial}{\partial x}(b + z + H) = 0 \qquad \text{in } (0,L) \times (0,T) \\ \dfrac{\partial}{\partial t}(Ac) + \dfrac{\partial}{\partial x}(Qc) - \dfrac{\partial}{\partial x}\left(kA\dfrac{\partial c}{\partial x}\right) - \phi = 0 \qquad \text{in } (0,L) \times (0,T) \\ \rho_s(1-\eta)\dfrac{\partial A_s}{\partial t} + \phi = 0 \qquad \text{in } (0,L) \times (0,T) \end{cases} \tag{1}$$

with boundary conditions:

$$\begin{cases} A(L,t) = A_L(t) & \text{in } (0,T) \\ Q(0,t) = Q_0(t) & \text{in } (0,T) \\ c(0,t) = c_0(t) & \text{in } (0,T) \\ k\,\dfrac{\partial c}{\partial x}(L,t) = c_L(t) & \text{in } (0,T) \\ A_s(0,t) = A_{s0}(t) & \text{in } (0,T) \end{cases} \tag{2}$$

and initial conditions:

$$\begin{cases} A(x,0) = A^0(x) & \text{in } (0,L) \\ Q(x,0) = Q^0(x) & \text{in } (0,L) \\ c(x,0) = c^0(x) & \text{in } (0,L) \\ A_s(x,0) = A_s^0(x) & \text{in } (0,L) \end{cases} \tag{3}$$

where $A(x,t)$ is the wet section (area of the canal section occupied by water), $A_s(x,t)$ is the sedimented section, $Q(x,t)$ is the water flow rate across the section (that is, $Q = Au$ with $u(x,t)$ the averaged velocity of water), g is the gravity acceleration, $b(x)$ gives canal bottom geometry, $z(x,t)$ is the height of settled sediment (related to sedimented section by a bijective function $A_s = S(z)$ depending on section shape for the cases of known shapes: rectangular, circular, trapezoidal, etc.), $H(x,t)$ is the height of water (in a similar way to previous case, $A = S(z+H) - S(z)$), $c(x,t)$ is the concentration of sediment in suspension, k is the diffusion coefficient, ϕ is the sediment exchange rate with the bed (in this case taken as $\phi = \frac{v_f}{\gamma} A(c^* - c)$ with v_f the settling velocity of sediment, γ an adaptation length and c^* the sediment transport capacity), ρ_s is the density of sediment, and $\eta \in [0,1]$ represents the bed porosity.

We assume that, originally, the canal presents a rectangular shape with a width D and a depth E. In order to minimize the negative effects of sedimentation, we try to obtain a new optimized trapezoidal section. At this point it is especially important the choice of design variables. From the point of view merely geometric, there exist two straightforward design variables (see Fig. 1): the width of the modified canal bottom w, and the angle of the lateral wall α.

In this case, the function relating trapezoidal section a to height h is given by function:

$$a = S(h) = wh + \frac{\tan(\alpha)}{2} h^2 \tag{4}$$

whose inverse can be also easily computed:

$$h = S^{-1}(a) = \frac{\sqrt{w^2 + 2\tan(\alpha)a} - w}{\tan(\alpha)} \tag{5}$$

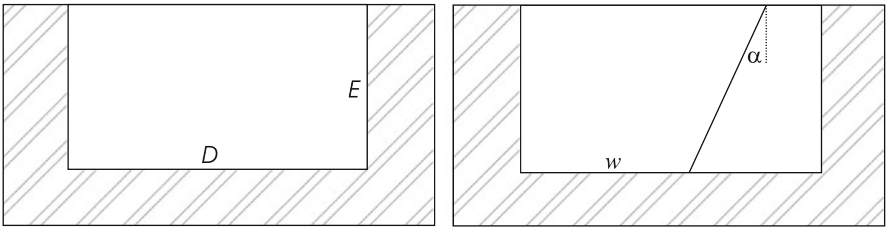


Fig. 1 Original rectangular canal section (*left*). Optimized trapezoidal canal section (*right*)

In order to assure the operation and effectiveness of the canal and its structural stability we need to impose several constraints on the design variables, for instance, bound constraints on the control of the type:

$$\underline{w} \leq w \leq \overline{w} \tag{6}$$

$$\underline{\alpha} \leq \alpha \leq \overline{\alpha} \tag{7}$$

By geometrical reasons, in our study we propose the values $\underline{w} = \frac{D}{4}$, $\overline{w} = D - E$, $\underline{\alpha} = 0$, and $\overline{\alpha} = \frac{\pi}{4}$.

Finally, since we are interested in reducing the accumulation of vegetation and sludge in the canals of the wastewater treatment plant in a power station, we will try to minimize the area of the sedimented section A_s. With this purpose in mind, we define the objective function $J(w, \alpha)$ to be minimized as:

$$J = \frac{1}{2} \int_0^T \int_0^L A_s^2 \, dx \, dt \tag{8}$$

Thus, a mathematical formulation of our optimization problem (P) can read as: Finding the optimal design variables (w, α) such that, satisfying the state system (1)–(3) and the control constrains (6)–(7), minimize the cost function J given by (8).

3 A Case Study

As a first step in our study we present a simplified case: We assume both lateral walls to be vertical (that is, we fix angle α to 0), then the only design variable is the width w of the canal (which presents now a rectangular section).

In this simplified case, function S (and its inverse) takes a much simpler form:

$$a = S(h) = wh \tag{9}$$

$$h = S^{-1}(a) = \frac{a}{w} \tag{10}$$

Under these conditions, and introducing in a classical way an adjoint state (r, p, s, v), solution of the linear adjoint system [3]:

$$\begin{cases} -\dfrac{\partial r}{\partial t} + \dfrac{Q^2}{A^2}\dfrac{\partial p}{\partial x} - g\dfrac{1}{w}A\dfrac{\partial p}{\partial x} + g\left(b' + \dfrac{1}{w}\dfrac{\partial A_s}{\partial x}\right)p + \dfrac{Qc}{A^2}\dfrac{\partial s}{\partial x} \\ \qquad + \dfrac{1}{\rho_s(1-\eta)}\dfrac{\nu_f}{\gamma}(c^* - c)v = 0 \qquad \text{in } (0, L) \times (0, T) \\ -\dfrac{\partial p}{\partial t} - \dfrac{\partial r}{\partial x} - 2\dfrac{Q}{A}\dfrac{\partial p}{\partial x} - \dfrac{c}{A}\dfrac{\partial s}{\partial x} = 0 \qquad \text{in } (0, L) \times (0, T) \\ -\dfrac{\partial s}{\partial t} - \dfrac{Q}{A}\dfrac{\partial s}{\partial x} - \dfrac{\partial}{\partial x}\left(k\dfrac{\partial s}{\partial x}\right) + \dfrac{\nu_f}{\gamma}s - \dfrac{1}{\rho_s(1-\eta)}\dfrac{\nu_f}{\gamma}Av \\ \qquad = 0 \qquad \text{in } (0, L) \times (0, T) \\ -\dfrac{\partial v}{\partial t} - g\dfrac{1}{w}\dfrac{\partial}{\partial x}(Ap) = A_s \qquad \text{in } (0, L) \times (0, T) \end{cases} \tag{11}$$

with boundary conditions:

$$\begin{cases} \left\{\dfrac{Q^2}{A^2}p - g\dfrac{1}{w}Ap + \dfrac{Qc}{A^2}s\right\}(0, t) = 0 \qquad \text{in } (0, T) \\ \left\{r + 2\dfrac{Q}{A}p + \dfrac{c}{A}s\right\}(L, t) = 0 \qquad \text{in } (0, T) \\ \left\{\dfrac{Q}{A}s + k\dfrac{\partial s}{\partial x}\right\}(L, t) = 0 \qquad \text{in } (0, T) \\ s(0, t) = 0 \qquad \text{in } (0, T) \\ A(L, t)\,p(L, t) = 0 \qquad \text{in } (0, T) \end{cases} \tag{12}$$

and final conditions:

$$\begin{cases} r(x, T) = 0 \qquad \text{in } (0, L) \\ p(x, T) = 0 \qquad \text{in } (0, L) \\ s(x, T) = 0 \qquad \text{in } (0, L) \\ v(x, T) = 0 \qquad \text{in } (0, L) \end{cases} \tag{13}$$

we can derive an optimality condition in order to characterize the optimal solution of problem (P):

Theorem 1. *Let $w \in [\underline{w}, \overline{w}]$ be an optimal solution of the control problem (P), minimizing the objective function $J(w)$ given by (8) in the admissible interval $[\underline{w}, \overline{w}]$. Then, there exist (A, Q, c, A_s), solution of the state system (1)–(3), and (r, p, s, v), solution of the adjoint system (11)–(13), such that:*

$$(\tilde{w} - w)\int_0^T\int_0^L \frac{\partial}{\partial x}(A + A_s)Ap\,dx\,dt \geq 0, \quad \forall \tilde{w} \in [\underline{w}, \overline{w}]. \tag{14}$$

It is worthwhile remarking here that above optimality condition (14) gives, in fact, an expression for the derivative of the cost function J, that can be used in the numerical computation of the optimal solution [1, 2]. Some preliminary numerical results for this simplified study case are currently being developed by the authors, where the resolution of the state and the adjoint systems is performed by means of a combination of the method of characteristics for upwinding the time derivative, and the Lagrange finite element method for dealing with the space discretization [3]. These numerical results will be the subject of a forthcoming publication.

Acknowledgements The financial support provided by the Projects MTM2012-30842 and MTM 2009-07749 of Ministerio de Economía y Competitividad (Spain) and FEDER is gratefully acknowledged.

References

1. Alvarez-Vázquez, L.J., Martínez, A., Rodríguez, C., Vázquez-Méndez, M.E., Vilar, M.A.: Optimal shape design for fishways in rivers. Math. Comput. Simul. **76**, 218–222 (2007)
2. Alvarez-Vázquez, L.J., Martínez, A., Vázquez-Méndez, M.E., Vilar, M.A.: An optimal shape problem related to the realistic design of river fishways. Ecol. Eng. **32**, 293–300 (2008)
3. Alvarez-Vázquez, L.J., Martínez, A., Vázquez-Méndez, M.E., Vilar, M.A.: An application of optimal control theory to river pollution remediation. Appl. Numer. Math. **59**, 845–858 (2009)
4. Burger, R., Liu, C., Wendland, W.L.: Existence and stability for mathematical models of sedimentation-consolidation processes in several space dimensions. J. Math. Anal. Appl. **264**, 288–310 (2001)
5. Jiang, J., Mehta, A.J.: Fine-grained sedimentation in a shallow harbor. J. Coast. Res. **17**, 389–393 (2001)
6. Mehta, A.J.: On estuarine cohesive sediment behaviour. J. Geophys. Res. **94**, 14303–14314 (1989)
7. van der Ham, R., Winterwerp, J.C.: Turbulent exchange of fine sediments in a tidal channel in the Ems/Dollard estuary. Part II: analysis with a 1D numerical model. Cont. Shelf Res. **21**, 1629–1647 (2001)
8. van Rijn, L.C.: Principles of sediment transport in rivers, estuaries and coastal seas. MIT Press, Cambridge (1993)
9. Winterwerp, J.C.: A simple model for turbulence induced flocculation of cohesive sediment. J. Hydraul. Res. **36**, 309–326 (1998)
10. Wu, Y., Falconer, R.A., Uncles, R.J.: Modelling of water flows and cohesive sediment fluxes in the Humber Estuary, UK. Mar. Pollut. Bull. **37**, 182–189 (1998)

Mathematical Treatment of Environmental Models

Zahari Zlatev, István Faragó, and Ágnes Havasi

Abstract Large-scale environmental models can successfully be used in different important for the modern society studies as, for example, in the investigation of the influence of the future climatic changes on pollution levels in different countries. Such models are normally described mathematically by non-linear systems of partial differential equations, which are defined on very large spatial domains and have to be solved numerically on very long time intervals. Moreover, very often many different scenarios have also to be developed and used in the investigations. Therefore, both the storage requirements and the computational work are enormous. The great difficulties can be overcome only if the following four tasks are successfully resolved: (a) fast and sufficiently accurate numerical methods are to be selected, (b) reliable and efficient splitting procedures are to be applied, (c) the cache memories of the available computers are to be efficiently exploited and (d) the codes are to be parallelized.

1 Description of a Large Scale Environmental Model

For the sake of simplicity we shall restrict ourselves on the area of long-range transport of air pollution and to a particular model (UNI-DEM, the Unified Danish Eulerian Model, [9]), but most of the results can easily be extended to other environmental models. UNI-DEM is described mathematically by the following system of partial differential equations (PDEs):

Z. Zlatev
Department of Environmental Science, Aarhus University, Roskilde, Denmark
e-mail: zz@dmu.dk

I. Faragó • Á. Havasi (✉)
Department of Applied Analysis and Computational Mathematics and HAS, ELTE Research Group "Numerical Analysis and Large Networks", Eötvös Loránd University, Budapest, Hungary
e-mail: faragois@cs.elte.hu; havasia@cs.elte.hu

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_10,

$$
\begin{aligned}
\frac{\partial c_i}{\partial t} &= -\,u\frac{\partial c_i}{\partial x} - v\frac{\partial c_i}{\partial y} && \text{- horizontal transport}\\
&+ \frac{\partial}{\partial x}\left(K_x\frac{\partial c_i}{\partial x}\right) + \frac{\partial}{\partial y}\left(K_y\frac{\partial c_i}{\partial y}\right) && \text{- horizontal diffusion}\\
&+ Q_i(t,x,y,z,c_1,c_2,\ldots,c_q) + E_i(t,x,y,z) && \text{- chemical reactions + emission}\\
&+ (k_{1i} + k_{2i})c_i && \text{- dry and wet depositions}\\
&- w\frac{\partial c_i}{\partial z} + \frac{\partial}{\partial z}\left(K_z\frac{\partial c_i}{\partial z}\right), && \text{- vertical transport}\\
& i = 1,2,\ldots,q && \text{- number of equations}\\
& && \text{(chemical species)}
\end{aligned}
\tag{1}
$$

The different quantities involved in (1) are briefly described below:

- $c_i = c_i(t,x,y,z)$ is the concentration of the chemical species i at point (x,y,z) of the space domain and at time t of the time-interval,
- $u = u(t,x,y,z)$, $v(t,x,y,z)$ and $w = w(t,x,y,z)$ are wind velocities (along the Ox, Oy and Oz directions, respectively) at the spatial point (x,y,z) and time t,
- $K_x = K_x(t,x,y,z)$, $K_y = K_y(t,x,y,z)$ and $K_z = K_z(t,x,y,z)$ are diffusivity coefficients at the spatial point (x,y,z) and time t (it is often assumed that K_x and K_y are non-negative constants, while the calculation of K_z is normally rather complicated),
- $k_{1i} = k_{1i}(t,x,y,z)$ and $k_{2i} = k_{2i}(t,x,y,z)$ are deposition coefficients (dry and wet deposition respectively) of chemical species i at the spatial point (x,y,z) and time t of the time-interval. It should be mentioned here that for some of the species these coefficients are non-negative constants. The wet deposition coefficients k_{2i} are equal to zero when it is not raining.
- $E_i(t,x,y,z)$ is emission source for chemical species i at the spatial point (x,y,z) and time t of the time-interval.

2 Splitting the Model

The mathematical model defined by (1) is normally split (see [9]) into the following three sub-models:

$$
\frac{\partial c_i^{(1)}}{\partial t} = -w\frac{\partial c_i^{(1)}}{\partial z} + \frac{\partial}{\partial z}\left(K_z\frac{\partial c_i^{(1)}}{\partial z}\right), \tag{2}
$$

$$
\frac{\partial c_i^{(2)}}{\partial t} = -u\frac{\partial c_i^{(2)}}{\partial x} - v\frac{\partial c_i^{(2)}}{\partial y} + \frac{\partial}{\partial x}\left(K_x\frac{\partial c_i^{(2)}}{\partial x}\right) + \frac{\partial}{\partial y}\left(K_y\frac{\partial c_i^{(2)}}{\partial y}\right), \tag{3}
$$

$$\frac{\partial c_i^{(3)}}{\partial t} = -Q_i(t,x,y,z,c_1^{(3)},c_2^{(3)},\ldots,c_q^{(3)}) + E_i(t,x,y,z) + (k_{1i}+k_{2i})c_i^{(3)}, \quad (4)$$

The first of these three sub-models describes the vertical exchange. The second sub-model describes the combination of the horizontal transport (the advection) and the horizontal diffusion. The last sub-model describes the chemical reactions together with the emission sources and the deposition terms.

Note that the three sub-models are fully defined by (2)–(4), but the splitting procedure is not. It will be completely determined only when it is explained how these sub-models are combined. The simple sequential splitting procedure is applied in UNI-DEM. It is obtained in the following way. Assume that the space domain is discretized by using a grid with $N_x \times N_y \times N_z$ grid-points, where N_x, N_y and N_z are the numbers of the grid-points along the grid-lines parallel to the Ox, Oy and Oz axes. Assume further that the number of chemical species involved in the model is $N_s = q$. Finally, assume that approximate values of the concentrations (for all species and at all spatial grid-points) have been found for some $t = t_n$. These values can be considered as components of a vector-function $c(t_n, x_i, y_j, z_k) \in \mathbb{R}^{N_x \times N_y \times N_z \times N_s}$. The next time-step, time-step $n+1$ (at which approximations of the concentrations are found at $t_{n+1} = t_n + \Delta t$ where Δt is some increment), can be performed by solving successively the three sub-models. The values of $c(t_n, x_i, y_j, z_k)$ are used as an initial condition in the solution of (2). The solution of (2) is used as an initial condition of (3). Finally, the solution of (3) is used as an initial condition of (4). The solution of (4) is accepted as an approximation to $c(t_{n+1}, x_i, y_j, z_k)$. In this way, everything is prepared to start the calculations in the next time-step, step $n+2$.

The major advantage of any splitting procedure based on the above three sub-models is due to the fact that no extra boundary conditions are needed when (2)–(4) are used. This is true not only for the sequential splitting procedure sketched above, but also for any other splitting procedure based on the sub-models defined by (2)–(4).

3 Choice of Numerical Methods

Assume that the spatial derivatives are discretized by some numerical algorithm (it must be mentioned here that different numerical algorithms can be applied in the different sub-models and this is one of the big advantages of using splitting techniques: for each sub-model one can select the most suitable algorithm). Then the three systems of PDEs represented by (2)–(4) will be transformed into three systems of ODEs (ordinary differential equations):

$$\frac{dg^{(1)}}{dt} = f^{(1)}(t, g^{(1)}), \quad \frac{dg^{(2)}}{dt} = f^{(2)}(t, g^{(2)}), \quad \frac{dg^{(3)}}{dt} = f^{(3)}(t, g^{(3)}). \quad (5)$$

The components of functions $g^{(m)}(t) \in \mathbb{R}^{N_x \times N_y \times N_z \times N_s}$, $m = 1, 2, 3$ are approximations at time t of the concentrations at all spatial grid-points and for all species. The components of functions $f^{(m)}(t) \in \mathbb{R}^{N_x \times N_y \times N_z \times N_s}$, $m = 1, 2, 3$ depend both on quantities involved in the right-hand-side of (1) and on the particular numerical algorithms that are used in the discretization of the spatial derivatives.

A simple linear finite element method is used to discretize the spatial derivatives in (2) and (3). The spatial derivatives can also be discretized by using other numerical methods as, for example, a pseudo-spectral discretization, a semi-Lagrangian discretization (which can be used only to discretize the first-order derivatives, i.e., the advection part should not be combined with the diffusion part when this method is to be applied) and methods producing non-negative values of the concentrations.

The first system of ODEs in (5) can be solved by using many classical time-integration methods. The well-known θ-method is currently used in UNI-DEM.

Predictor-corrector (PC) methods with several different correctors, which are fully discussed in [6], are used in the solution of the second ODE system in (5). The correctors are carefully chosen so that the stability properties of the method can be enhanced. If the code judges the time-stepsize to be too large for the currently used PC method (and may lead to unstable computations), then it switches to a more stable (but also more expensive, because more corrector formulae are used in order to obtain better stability) PC scheme. On the other hand, if the code judges that the stepsize is too small for the currently used PC method, then it switches to a not so stable but more accurate PC scheme (which is using less corrector formulae and, therefore, is less expensive). In this way the code is trying both to keep the same stepsize and to optimize the performance. More details about this strategy can be found in [6].

The solution of the third system in (5) is much more complicated, because this system is both time-consuming and very stiff. Often the QSSA (Quasi-Steady-State-Approximation) method is used in this part of the model. It is simple and relatively stable but not very accurate (therefore it has to be run with a small time-stepsize). An improved QSSA method was implemented in UNI-DEM. The classical numerical methods for stiff ODE systems (such as the Backward Euler Method, the Trapezoidal Rule and Runge-Kutta algorithms) lead to the solution of non-linear systems of algebraic equations and, therefore, they are normally more expensive. On the other hand, these methods can be incorporated with an error control and perhaps with larger time-steps. Partitioning can also be used. Some convergence problems related to the implementation of partitioning have been studied in [7]. More details about the numerical algorithms can be found in [9].

4 Applying Parallelization

Another great advantage of using splitting is the appearance of many natural parallel tasks. It is easy to see that (a) the first system in (5) contains $N_x \times N_y \times N_s$ independent tasks (for each chemical compound, each system along a vertical grid-line can be treated independently), (b) the second system in (5) contains $N_x \times N_y \times N_z$ independent tasks (the chemical compounds at each grid-point can be treated independently of the chemical compounds at the other grid-points) and (c) the third system in (5) contains $N_z \times N_s$ independent tasks (for each chemical compound the system along a horizontal grid-plane can be treated independently). These parallel tasks, which appear in a natural way when any splitting based on (2)–(4) is applied, were efficiently exploited during the parallelization process. Furthermore, standard parallel tools, OpenMP and MPI, have been extensively used. Much more details can be found in [1,9].

5 Applications

UNI-DEM has been used in many different studies (many of them are reported in [9]). Investigations of the influence of the climate changes on pollution levels in Europe [8] and Hungary with its surroundings [10] have recently been carried out.

6 Conclusions

Assume that $N_x = N_y = 480, N_z = 10, N_s = 35$ are used (this was the case in [8, 10]). Then the number of equations is 80,640,000 and 213,120 time-steps are needed to perform calculations with meteorological and emission data covering a whole year. Moreover, calculations over a long time-period (16 years) were needed in [8, 10]. It is clear that it was possible to resolve the enormous tasks only if (a) efficient splitting procedures are used, (b) suitable numerical methods are selected for each sub-model and (c) parallel computations are applied. It should nevertheless be pointed out that further improvements in connection with the tasks related to (a)–(c) are highly desirable.

Much more details about the mathematical treatment of large environmental models can be found in [2]. More precisely the splitting techniques are treated in [3,4], the organization of parallel computations described in [11] and the handling of the most difficult part, the sub-model containing the chemical reactions is discussed in [5]. Different applications of environmental models are also reported in [2].

References

1. Alexandrov, V., Owczarz, W., Thomsen, P.G., Zlatev, Z.: Parallel runs of large air pollution models on a grid of SUN computers. Math. Comput. Simul. **65**, 557–577 (2004)
2. Faragó, I., Havasi, Á., Zlatev, Z. (eds.): Advanced Numerical Methods for Complex Environmental Models: Needs and Availability. Bentham, Oak Park (2013)

3. Faragó, I., Havasi, Á., Zlatev, Z.: Implementation of splitting procedures. In: Faragó, I., Havasi, Á., Zlatev, Z. (eds.) Advanced Numerical Methods for Complex Environmental Models: Needs and Availability, pp. 79–125. Bentham, Oak Park (2013)
4. Faragó, I., Havasi, Á., Zlatev, Z.: Application of splitting in an air pollution model. In: Faragó, I., Havasi, Á., Zlatev, Z. (eds.) Advanced Numerical Methods for Complex Environmental Models: Needs and Availability, pp. 126–165. Bentham, Oak Park (2013)
5. Faragó, I., Havasi, Á., Zlatev, Z.: Treatment of the chemical reactions in an air pollution model. In: Faragó, I., Havasi, Á., Zlatev, Z. (eds.) Advanced Numerical Methods for Complex Environmental Models: Needs and Availability, pp. 54–78. Bentham, Oak Park (2013)
6. Zlatev, Z.: Application of predictor-corrector schemes with several correctors in solving air pollution problems. BIT **24**, 700–715 (1984)
7. Zlatev, Z.: Partitioning ODE systems with an application to air pollution models. Comput. Math. Appl. **42**, 817–832 (2001)
8. Zlatev, Z.: Impact of future climate changes on high ozone levels in European suburban areas. Clim. Change **101**, 447–483 (2010)
9. Zlatev, Z., Dimov, I.: Computational and Numerical Challenges in Environmental Modelling. Elsevier, Amsterdam (2006)
10. Zlatev, Z., Havasi, Á., Faragó, I.: Influence of climatic changes on pollution levels in Hungary and its surrounding countries. Atmosphere **2**, 201–221 (2011)
11. Zlatev, Z., Georgiev, K., Dimov, I.: Parallel computations in a large-scale air pollution model. In: Faragó, I., Havasi, Á., Zlatev, Z. (eds.) Advanced Numerical Methods for Complex Environmental Models: Needs and Availability, pp. 166–201. Bentham, Oak Park (2013)

Model-Based Assessment of Geophysical Observations: From Numerical Simulations Towards Volcano Hazard Forecasting

Gilda Currenti and Ciro Del Negro

Abstract Geodetic, gravity and magnetic field changes, produced by mass and stress redistributions accompanying magma migration and accumulation within the volcano edifice, are numerically computed by an integrated elastic 3-D model based on Finite Element Method (FEM). Firstly, comparisons are made between analytical and numerical solutions to validate the numerical model and to estimate the perturbations caused by medium heterogeneity and topographic features. Successively, the integrated numerical procedure was applied to interpret geophysical observations collected at Etna volcano during unrest periods. The obtained results highlight that heterogeneity and topography engender deviations from analytical results in the geophysical changes and, hence, the disregard of these complexities could lead to an inaccurate estimate of source parameters in inversion procedure. The FEM approach allows for considering a picture of a fully 3D model of Etna volcano, which advance the reliability of model-based assessments of geophysical observations. This approach, based on observable data and complemented by physical modeling techniques, makes the step ahead in the volcano hazard assessment and in the understanding of the underlying physics and poses the basis for future developments of scenario forecasting.

1 Introduction

Volcano unrest generates a wide variety of geophysical signals, which can be observed before and during eruptive processes. In particular, ground deformation, gravity and magnetic changes in volcanic areas are generally recognized as reliable indicators of unrest, resulting from the magma accumulation and migration

G. Currenti (✉) • C.D. Negro
Istituto Nazionale di Geofisica e Vulcanologia, Piazza Roma 2, Catania, Italy
e-mail: gilda.currenti@ct.ingv.it; ciro.delnegro@ct.ingv.it

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_11,

from depth. Continuous measurements of these geophysical signals are useful for detecting magma recharging phases and imaging spatio-temporal evolution of propagating dikes. These geophysical signals are generally interpreted separately from each other and the consistency of interpretations from these different methods is qualitatively checked only a posteriori. An integrated approach based on different geophysical data should prove a more efficient and accurate procedure for inferring magmatic sources and minimizing interpretation ambiguities.

Over the last decade at Etna volcano, where volcanological tradition is consolidated, and scientific and technological standards are highly advanced, geodetic, gravity and magnetic investigations have been playing an increasingly important role in studying the eruptive processes [2, 3, 5, 8, 10]. A series of analytical solutions, based on a homogeneous elastic half-space model, have been devised and widely used in literature [11, 12, 14] for modelling ground deformation, gravity and magnetic variations due to volcanic sources. It is worth noting that Etna volcano is elastically inhomogeneous, as indicated by geological evidences and seismic tomography [4, 13], and that rigidity layering and heterogeneities are likely to affect the magnitude and pattern of observed signals. To overcome these intrinsic limitations and provide more realistic models, which allows considering topographic effects as well as complicated distribution of medium properties, we exploited the Finite Element Method (FEM). This procedure allows joint evaluation of geophysical changes caused by dislocation and overpressure sources in a 3D formulation.

The 2008 Etna eruption offers an exemplary case study to validate the capability of the proposed integrated approach for imaging the intrusive process occurring in the northern flank of the volcano [7, 10]. The main objective is to solve the scientific challenge of developing numerical models of the involved magmatic process, in which the output from numerical predictions are compared with available geophysical observations to provide a quantitative estimate of the volcano internal state and to constrain the active magmatic source. Numerical solutions for deformation, gravity and magnetic fields are obtained by modelling the source intrusion as an extension fracture driven by a magmatic overpressure, which is a realistic representation of an intrusive dike. Combined geophysical investigations provide a quantitative estimate of the source model parameters and the involved mechanisms helpful in the assessment of volcano hazard.

2 Numerical Model

The deformation and stress field produced by magmatic sources usually occur very slowly, so the rock is in static equilibrium and the displacement can be found by solving the equations of equilibrium. In the case where rock behaves elastically, the equations of equilibrium are coupled with constitutive Hooke's law giving the following set of equations [9]:

$$
\begin{aligned}
\nabla \cdot \boldsymbol{\sigma} &= 0 \\
\boldsymbol{\sigma} &= \lambda\, tr(\boldsymbol{\epsilon})\mathbf{I} + 2\mu\boldsymbol{\epsilon} \\
\boldsymbol{\epsilon} &= \frac{1}{2}\left(\nabla\mathbf{u} + (\nabla\mathbf{u})^T\right)
\end{aligned}
\tag{1}
$$

where $\boldsymbol{\sigma}$ and $\boldsymbol{\epsilon}$ are the stress and strain tensors, respectively, $\mathbf{u}$ the deformation vector and λ and μ are the Lame's elastic medium parameters. When the medium is homogeneous Eq. (1) result in the Cauchy-Navier equation:

$$(\lambda + \mu)\nabla(\nabla \cdot \mathbf{u}) + \mu\nabla^2\mathbf{u} = 0 \tag{2}$$

Subsurface stress and deformation fields caused by dislocation and pressure sources necessarily alter the density distribution and the magnetization of the surrounding rocks that, in turn, affects the gravity and the magnetic fields, respectively. The gravity change Δg can be calculated by solving the following boundary value problem for the gravitational potential ϕ_g [1]:

$$
\begin{aligned}
\nabla^2\phi_g &= -4\pi G\Delta\rho \\
\Delta g &= -\frac{\partial\phi_g}{\partial z}
\end{aligned}
\tag{3}
$$

where G is the gravitational constant and $\Delta\rho$ is the density distribution change given by:

$$\Delta\rho = \Delta\rho_1 - \rho_1\nabla \cdot \mathbf{u} - \mathbf{u} \cdot \nabla\rho_0 \tag{4}$$

On the right side of Eq. (4), the first term is the density change related to the arrival of the new mass from depth, the remaining two from the linearized version of the continuity equation for the material already present in the elastic medium [1]. Particularly, the second term results from the volume change arising from the compressibility of the medium and the third term originates from the displacement of density boundaries in heterogeneous media. As for the magnetic field, the piezomagnetic change can be described by the scalar potential formulation [12]:

$$
\begin{aligned}
\nabla^2\phi_m &= 4\pi\nabla \cdot \mathbf{J} \\
\mathbf{J} &= \frac{3}{2}\beta\boldsymbol{\sigma'} \cdot \mathbf{J_0}
\end{aligned}
\tag{5}
$$

where ϕ_m is the piezomagnetic scalar potential, $\mathbf{J}$ the magnetization change, $\mathbf{J_0}$ the initial magnetization, β the stress sensitivity and $\boldsymbol{\sigma'}$ the deviatoric stress tensor. Equations (1), (3) and (5) show that magnetic and gravity field changes are related to the deformation and stress fields of the elastic medium. Therefore, the deformation field and the changes in potential fields produced by volcanic sources need to be jointly modelled. Starting from the numerical solution of elastic deformation and its derivatives (1), the gravity and piezo-magnetic changes are computed using Eqs. (3) and (5) by FEM technique. FEM solutions strongly depend on numerical parameters not known a priori, such as the domain extension, the mesh resolution and the boundary conditions, and, hence, it is necessary to calibrate the model. Preliminarily, some benchmark tests were carried out to compare the analytical results with numerical ones assuming an homogeneous half-space medium [5–7].

3 3D Model of the 2008 Magmatic Intrusion at Etna

The numerical procedure was applied to model the magmatic intrusion occurring along the north flank of Mt Etna on 13 May 2008. A fully three-dimensional elastic Finite Element model of Mt Etna was designed to evaluate the ground deformation, magnetic and gravity changes. A computational domain of 100 × 100 × 50 km is considered for the deformation field calculations. The 3D topography of Mt Etna, which is rather asymmetric with a prominent mass deficit in correspondence of Valle del Bove, was taken into account using a Digital Elevation Model from the 90 m Shuttle Radar Topography Mission (SRTM) data and a bathymetry model from the GEBCO database (http://www.gebco.net/). The computational domain was meshed into 215,009 isoparametric, and arbitrarily distorted tetrahedral elements connected by 38,007 nodes. Lagrange cubic shape functions are used in the computations, since the use of lower order elements worsens the accuracy of stress field solutions. Zero displacements are assigned at the bottom and the lateral boundaries of the domain, while the upper boundary representing the ground surface is stress free. The intrusion source is simulated as a dislocation surface by introducing the mesh nodes in pairs along the surface rupture and assigning a tensile opening between pair nodes. In order to solve the Poisson's Equations (3) and (5), the potential or its normal derivatives are to be assigned at the boundaries of the domain, which is extended along the z direction to 50 km to finally obtain a 100 × 100 × 100 km computational domain for ensuring the continuity of the gravity and magnetic potential on the ground surface. Along the external boundaries, zero gravity potential is specified using Dirichlet boundary conditions, while the magnetic field is assumed to be tangential by assigning a Neumann condition on the magnetic potential. The magnetic problem is made unique by setting the potential to zero at an arbitrary point on the external boundary. Heterogeneous distribution of magneto-elastic properties is included in the model by considering seismic tomography investigations [4] and geological models [13].

Numerical results are compared with geophysical observations from ground-based stations (GPS, magnetic, and gravity data) and satellite platform (DInSAR data from ENVISAT satellite) to constrain the source parameters of the magmatic intrusion (Fig. 1). To improve the fit to the data and make the model more realistic, we solved for a distributed opening model over the dislocation surface. The resolved

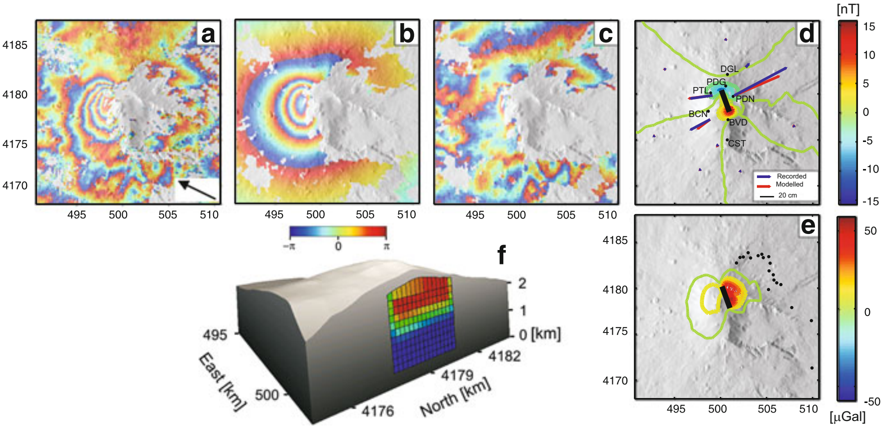


Fig. 1 Integrated numerical model to interpret DInSAR, GPS, magnetic and gravity data acquired during the onset of the 2008 Etna eruption. Observed interferogram for the descending scene pair ENVISAT 080507-080716 (**a**). Computed (**b**) and residual interferograms (**c**). Observed (*blue arrows*) and computed (*red arrows*) displacements at the summit permanent GPS stations (**d**). Computed magnetic (**d**, contour lines at 2 nT) and gravity (**e**, contour lines at 10 μGal) changes. Opening distributions obtained from the inversion of geophysical data (**f**)

opening distribution shows that the intrusion is quite shallow with a mean opening less than 2 m, likely to represent the zone of magma filled fracture in the northern part of the volcano (Fig. 1). The model well fits the deformation pattern derived from DInSAR and GPS data. The rewrapped modelled displacements from the distributed opening model enhances the fringes gradient and resemble quite well the overall feature of the DInSAR observations. The match between the observed and the computed magnetic changes is quite good at most stations. The total gravity change reaches a maximum amplitude of about 60 μGal in proximity of the intrusion where unfortunately no data are available. Indeed, the computed gravity field vanishes within 3–4 km from the magma intrusion and does not show significant changes at the gravity benchmarks, where in agreement with the model, no gravity variations were recorded.

4 Conclusions

A coupled numerical problem was set up to estimate ground deformation, gravity and magnetic changes produced by stress redistribution accompanying magma migration within the volcano edifice. The integrated numerical procedure was applied to image the magmatic intrusion occurring in the northern flank of Etna during the onset of the 2008 eruption. By giving a fairly complete picture of the magmatic intrusion, geophysical data combined with the numerical modelling procedure have proven to be useful for interpreting the observations and constraining the magmatic source parameters. The FEM-based approach improves the reliability of model-based inference of geophysical observations gathered during monitoring of volcanic unrest contributing to a more accurate evaluation of the hazard assessment.

Acknowledgements Satellite data were provided by ESA, through CAT1- 7926 project. We are grateful to Carl W Gable for giving hints on meshing procedures and to C A Williams for suggestions on FEM developments.

References

1. Bonafede, M., Mazzanti, M.: Modelling gravity variations consistent with ground deformation in the campi flegrei caldera (Italy). J. Volcanol. Geoth. Res. **81**, 137–157 (1998)
2. Bonforte, A., Bonaccorso, A., Guglielmino, F., Palano, M., Puglisi, G.: Feeding system and magma storage beneath Mt. Etna as revealed by recent inflation/deflation cycles. J. Geophys. Res. **113** (2008). doi:10.1029/2007JB005334
3. Carbone, D., Currenti, G., Del Negro, C.: Elastic model for the gravity and elevation changes prior to the 2001 eruption of etna volcano. Bull. Volcanol. **69**, 553–562 (2007). doi:10.1007/s00445-006-0090-5
4. Chiarabba, C., Amato, A., Boschi, E., Barberi, F.: Recent seismicity and tomographic modeling of the mount etna plumbing system. J. Geophys. Res. **105**, 923–10938 (2000)
5. Currenti, G.: Numerical evidences enabling to reconcile gravity and height changes in volcanic areas. Geophys. J. Int. (2013). doi:10.1093/gji/ggt507
6. Currenti, G., Del Negro, C., Di Stefano, A., Napoli, R.: Numerical simulation of stress induced piezomagnetic fields at etna volcano. Geophys. J. Int. **179**, 1469–1476 (2009). doi:10.1111/j.1365-246X.2009.04381.x
7. Currenti, G., Napoli, R., Di Stefano, A., Greco, F., Del Negro, C.: 3D integrated geophysical modeling for the 2008 magma intrusion at etna: constraints on rheology and dike overpressure. Phys. Earth Planet. Inter. (2011). doi:10.1016/j.pepi.2011.01.002
8. Del Negro, C., Currenti, G., Solaro, G., Greco, F., Pepe, A., Napoli, R., Pepe, S., Casu, F., Sansosti, E.: Capturing the fingerprint of etna volcano activity in gravity and satellite radar data. Sci. Rep. **3** (2013). doi:10.1038/srep03089
9. Fung, Y.: Foundations of Solid Mechanics. Prentice-Hall, Englewood Cliffs (1965)
10. Napoli, R., Currenti, G., Del Negro, C., Greco, F., Scandura, D.: Volcanomagnetic evidence of the magmatic intrusion on 13th May 2008 etna eruption. Geophys. Res. Lett. **35** (2008). doi:10.1111/j.1365-246X.2010.04769.x
11. Okubo, S.: Gravity and potential changes due to shear and tensile faults in a half-space. J. Geophys. Res. **97**, 7137–7144 (1992)
12. Sasai, Y.: Tectonomagnetic modeling on the basic of the linear piezomagnetic effect. Bull. Earthquake Res. Inst. **66**, 585–722 (1991)
13. Tibaldi, A., Groppelli, G.: Volcano-tectonic activity along structures of the unstable ne flank of Mt. Etna (Italy) and their possible origin. J. Volcanol. Geoth. Res. **115**, 277–302 (2002)
14. Utsugi, M., Nishida, Y., Sasai, Y.: Piezomagnetic potentials due to an inclinated rectangular fault in a semi-infinite medium. Geophys. J. Int. **140**, 479–492 (2000)

Thermal and Rheological Aspects in a Channeled Lava Flow

Marilena Filippucci, Andrea Tallarico, and Michele Dragoni

Abstract We investigated the cooling of a lava flow in the steady state considering that lava rheology is pseudoplastic and dependent on temperature. We consider that cooling of the lava is caused by thermal radiation at the surface into the atmosphere and thermal conduction at the channel walls and at the ground. The heat equation is solved numerically in a 3D computational domain. The fraction of crust coverage is calculated under the assumption that the solid lava is a plastic body with temperature dependent yield strength. We applied the results to the Mauna Loa (1984) lava flow. Results indicate that the advective heat transport significantly modifies the cooling rate of lava slowing down the cooling process also for gentle slope.

1 Introduction

In lava flows the mechanism of cooling and solidification plays a very important role in controlling the flow dynamics. Heat convection (free and forced) is a heat loss mechanism that acts for all the life time of sub-aerial lava flows. Heat radiation, due to the proportionality with T^4, is the dominant mechanism of cooling at high temperatures. Many authors agree that the transition between radiation-dominated and convection-dominated cooling takes place when lava temperature reaches about 400–600 °C [2, 8, 9, 11, 12, 14, 15]. Heat advection as source of heat has been neglected under particular conditions [2, 11, 14]. Keszthelyi and Denlinger [11], studying the initial cooling of a pahoehoe lava flow neglected the

M. Filippucci (✉) • A. Tallarico
Dipartimento di Scienze della Terra e Geoambientali, University of Bari, Bari, Italy
e-mail: marilena.Filippucci@gmail.com; andrea.tallarico@uniba.it

M. Dragoni
Dipartimento di Fisica e Astronomia, University of Bologna, Bologna, Italy
e-mail: michele.dragoni@unibo.it

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_12,

effect of advection and explained this choice observing that advective heat, being the product of velocity and temperature gradient in the flow direction, should be zero everywhere. In fact, the molten lava is mostly isothermal and the temperature gradient is negligible. Neri [14] neglected the effect of advection with respect to the heat production due to crystallization in the solidification process of a cooling lava flow. On the basis of surface temperature measurements of active pahoehoe flows, Ball et al. [2] stated that the advective heat transport is unimportant as long as the lava surface moves at the same velocity as the underlying layers but it becomes not negligible once the velocity of the crust is smaller than the velocity of the underlying lava. The importance of heat advection has been studied for volcanic conduits and for lava flows by [6, 13] and numerically by [4]. Filippucci et al. [4] assumed that lava rheology is a function of temperature and strain rate as retrieved by [10] and assumed that lava cooling is caused by two different mechanisms: heat radiation into the atmosphere and heat conduction through the channel levees and the ground. Using two different effusion temperatures, the authors observed that, as an effect of the heat advection, the hotter lava, although it is subjected to higher heat radiation into the atmosphere, cools slower than the colder one because it flows faster. So the advective heat transport strongly influences the cooling dynamics of the lava flow. Filippucci et al. [4] used the geometrical and physical parameters of the Mt Etna lava channel as described by [1]. We adopt the numerical code developed by [4] and apply it to the Mauna Loa, 1984, basaltic 'a'a lava channel flow whose geometrical, physical and reological parameters are collected by [7]. Following [7], within the upper reaches of the flow, all movement became concentrated in a central channel of stable geometry over underlying slopes ranging between 1° and 9°. The aim is to study the cooling of a lava flow with higher effusion temperature with respect to the Mt Etna case study of [4] and flowing down on a gentle slope.

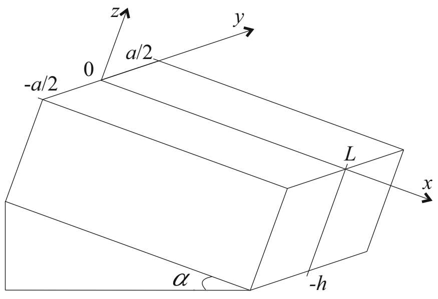
2 Dynamical, Rheological and Thermal Model

We consider a viscous fluid flowing in the x direction in an inclined rectangular channel, with the cross section parallel to the yz plane. The width of the channel is a and the thickness is h; the slope of the inclined plane is α. The channel and the coordinate system are shown in Fig. 1. The flow is assumed laminar and subjected to the gravity force. The fluid is assumed isotropic and incompressible, with constant density ρ. The equation of motion in the steady state is:

$$\rho g \sin\alpha + \frac{\partial}{\partial y}\left(\eta_a \frac{\partial v_x}{\partial y}\right) + \frac{\partial}{\partial z}\left(\eta_a \frac{\partial v_x}{\partial z}\right) = 0 \tag{1}$$

where v_x is the x component of velocity, g is the acceleration of gravity and η_a is the apparent viscosity which depends on temperature T. The apparent viscosity of a power-law fluid is:

Fig. 1 Flow segment



$$\eta_a(x, y, z, T) = k(T)\left[\left(\frac{\partial v_x}{\partial y}\right)^2 + \left(\frac{\partial v_x}{\partial z}\right)^2\right]^{\frac{n(T)-1}{2}} \tag{2}$$

where both k and n depend on T. The dynamic boundary conditions are:

$$v_x\left(\pm\frac{a}{2}, z\right) = 0\,;\ \ v_x(y, -h) = 0 \tag{3}$$

$$\frac{\partial v_x}{\partial y}(0, z) = 0\,;\ \ \frac{\partial v_x}{\partial z}(y, 0) = 0 \tag{4}$$

Lava viscosity depends on strain rate through Eq. (2). In particular, fluid consistency k depends on T through an exponential function and the power-law exponent n depends on T through a linear function as found by [10] and used by [4]. The empirical functions for rheological parameters have been retrieved by [10] for the basaltic melt of Sommata (Vulcano island, Italy) as representative of the pseudoplastic rheology:

$$k(T) = k_0 \mathrm{e}^{p_1 + \frac{p_2}{T}} \tag{5}$$

$$n(T) = 1 + p_3 + p_4 T \tag{6}$$

Values of the parameters p_1, p_2, p_3 and p_4 are in Table 1.

The numerical solution of (1), using the finite volume method, is given by [3].

We assume that lava starts cooling as it exits from the vent with an effusion temperature T_0. The high lava emissivity ε and the high lava temperature T_0 imply that heat exchange at the lava surface occurs mainly by radiation into the atmosphere [14]. We assume a radiative heat flux from the upper lava surface

$$q_r = \sigma \varepsilon T_u^4 \tag{7}$$

where σ is the Stefan-Boltzmann constant, ε is the surface emissivity of lava and T_u is the temperature of the upper surface $z = 0$. We assume that the atmospheric

Table 1 Values of the model parameters

Parameter	Description	Value unit
a	Channel width	5 m
c_p	Specific heat capacity	837 J kg^{-1} K^{-1}
g	Acceleration of gravity	9.8 m s^{-2}
h	Channel thickness	5 m
q_c	Lateral and basal heat flux	1,000 W m^{-2}
k_0	Rheological parameter	1 Pa s^n
p_1	Rheological parameter	-18.71
p_2	Rheological parameter	33.4 10^3 K
p_3	Rheological parameter	-1.35
p_4	Rheological parameter	0.85 10^{-3} K^{-1}
K	Thermal conductivity	3 W K^{-1} m^{-1}
L	Channel length	100 m
T_s	Solidus temperature	1,253 K
T_e	Effusion temperature	1,140 K
α_1	Channel slope	5°
α_2	Channel slope	9°
ε_c	Thermal emissivity	1
ρ	Density	2,800 kg m^{-3}
σ	Stefan constant	5.668108 W m^{2}K^4
χ	Thermal diffusivity	1.28 10^{-6} m^2 s^{-1}

temperature is negligible with respect to T_u. We assume that the conductive heat loss through the levees and the ground can be represented by a constant heat flux q_c (1) as used by [4].

We neglect viscous dissipation and the heat of crystallization and therefore we do not consider any internal heat source. The heat equation at the steady state is then:

$$v_x \frac{\partial T}{\partial x} = \chi \left(\frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2} \right) \tag{8}$$

where χ is the thermal diffusivity given by:

$$\chi = \frac{K}{\rho c_p} \tag{9}$$

where K is the thermal conductivity and c_p is the specific heat capacity. We assume that K does not depend on temperature. The thermal boundary conditions are the radiative heat flux q_r at the upper surface, the constant heat flux q_c at the levees and at the ground and the symmetry of the problem with respect to the xz plane:

$$T(x = 0) = T_0 \tag{10}$$

$$\frac{\partial T}{\partial y}(y = 0) = 0 \tag{11}$$

$$\frac{\partial T}{\partial y}\left(y = \pm\frac{a}{2}\right) = -\frac{q_c}{K} \tag{12}$$

$$\frac{\partial T}{\partial z}(z = 0) = -\frac{q_r}{K} \tag{13}$$

$$\frac{\partial T}{\partial z}(z = -h) = -\frac{q_c}{K} \tag{14}$$

We consider a flow segment with length L (Fig. 1).

The numerical solution and tests of (8), using the finite volume method are given by [4].

The advective term in the heat equation makes the temperature dependent on the flow velocity, which in turn depends on temperature through the apparent viscosity (2). So, the dynamical and the thermal problem are mutually dependent. Since we consider only a short segment of a lava flow, the thickness can be assumed as constant. This allows a reduction of the computational time cost required to solve the 3D problem and avoids remeshing of the computational domain. The iterative procedure to solve Eqs. (1) and (8) is made of the following steps:

1. set an initial value of temperature $T(x, y, z)$ in the domain;
2. compute k and n given by (5) and (6) in the domain;
3. solve the equation of motion (1) to obtain $v_x(x, y, z)$ in the domain;
4. solve the heat equation (8) in the domain and update $T(x, y, z)$;
5. back to point (2).

Defining R as the sum of N_{CV} residuals in L_1 norm between the temperature T at the present iteration and that at the preceding one T^*:

$$R = \sum_{k=1}^{N_{CV}} |T - T^*|_k, \tag{15}$$

the procedure stops if R falls below 10^{-4} K.

3 Results

We evaluated temperature and velocity fields for the steady state case assuming two different channel slopes α_1 and α_2 (Table 1). First, we consider the case $\alpha = \alpha_1$. Figure 2a shows the temperature contour map in the plane $z = 0$. Vertical profiles

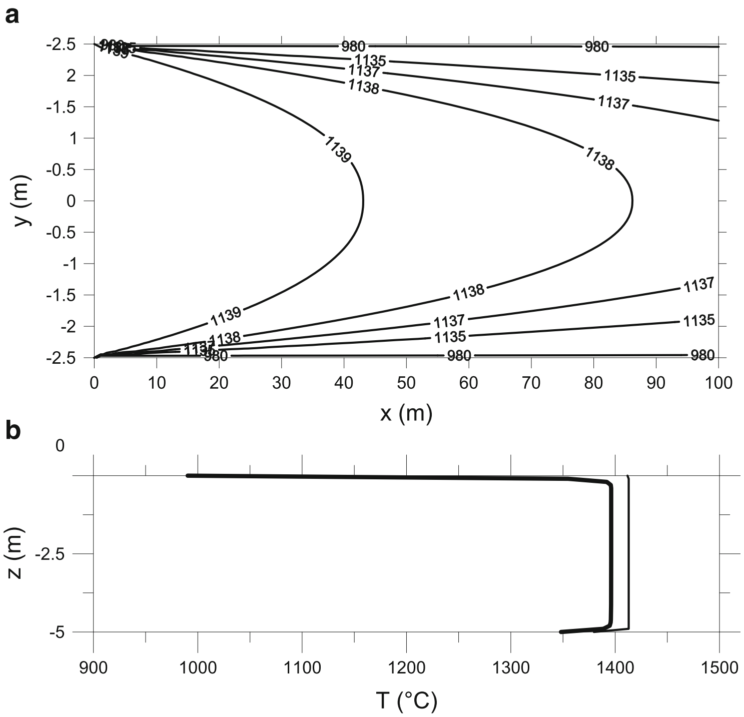


Fig. 2 Temperature T at the steady state for $\alpha = \alpha_1$. (**a**) Contour maps of T on the plane $z = 0$; (**b**) *thin line*: vertical profile of T at the channel center ($x = L$, $y = 0$); *thick line*: vertical profile of T at the channel levee ($x = L$, $y = \pm a/2$)

of temperature at $y = 0$ and $y = \pm a/2$ are in Fig. 2b. In Fig. 4 the velocity contour maps at $x = 0$ (Fig. 4a) and at $x = L$ (Fig. 4c), at $z = 0$ (Fig. 4b) and at $y = 0$ (Fig. 4d) are shown. As it can be observed the advective heat transport modifies the temperature field so that T is maximum at the center of the channel surface, where the velocity is maximum. It can be also observed that T is minimum at the channel levees where the velocity is null. When the channel slope is α_1, at $x = L$ the steady state lava temperature at the levees can reach $T = 692\,°\mathrm{C}$ while at the center is equal to the effusion temperature, that is $T = 1{,}140\,°\mathrm{C}$. The lava velocity is approximately the same at the vent and at the channel outflow section. Secondly, we consider the $\alpha = \alpha_2$ case. As before, contour maps and profiles of temperature are in Fig. 3 while contour maps of velocity are in Fig. 5. When the channel slope is α_2, at $x = L$ the levees are at $T = 717\,°\mathrm{C}$ while the center is still at the effusion temperature. As before, lava velocity does not decrease from the vent to the channel outflow section but it is much lower than that pf the previous case. In both cases, the maximum velocity is at the channel surface, that is at $z = 0$ indicating that lava flow preserves the structure of the channel flow.

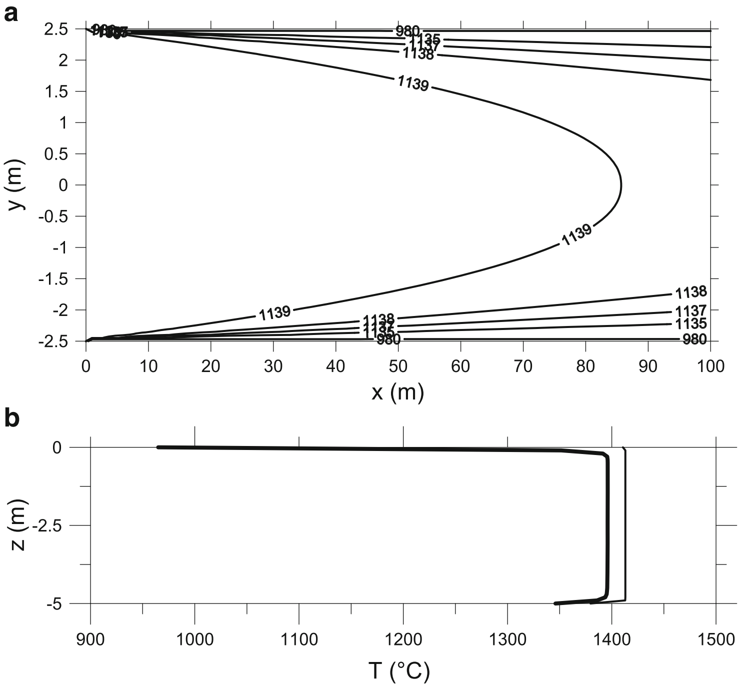


Fig. 3 Temperature T at the steady state for $\alpha = \alpha_2$. (**a**) Contour maps of T on the plane $z = 0$; (**b**) *thin line*: vertical profile of T at the channel center ($x = L$, $y = 0$); *thick line*: vertical profile of T at the channel levee ($x = L$, $y = \pm a/2$)

4 Discussion and Conclusion

We consider a segment of the channel close to the vent and far from the flow front, so that the flow dynamics is not influenced by the flow front. In the present model, we assume that lava rheology is temperature and strain rate dependent, that lava flows in an inclined rectangular channel under the gravity force and that lava cooling is caused by two different mechanisms: heat radiation into the atmosphere and heat conduction through the channel levees and the ground.

We neglect thermal convection since cooling of sub-aerial basaltic lava flows is initially dominated by radiative cooling. Only on time scales of tens of minutes and longer, heat loss via forced atmospheric convection (i.e. cooling by the wind) is predicted to dominate and the transition from radiation-dominated to convection-dominated cooling is found to be at about 400–600 °C [8, 11, 14]. In our simulation, we lower this threshold in a very thin part of the channel surface near the channel levees for the steady state case, so the error due to this approximation is supposed to be minimal.

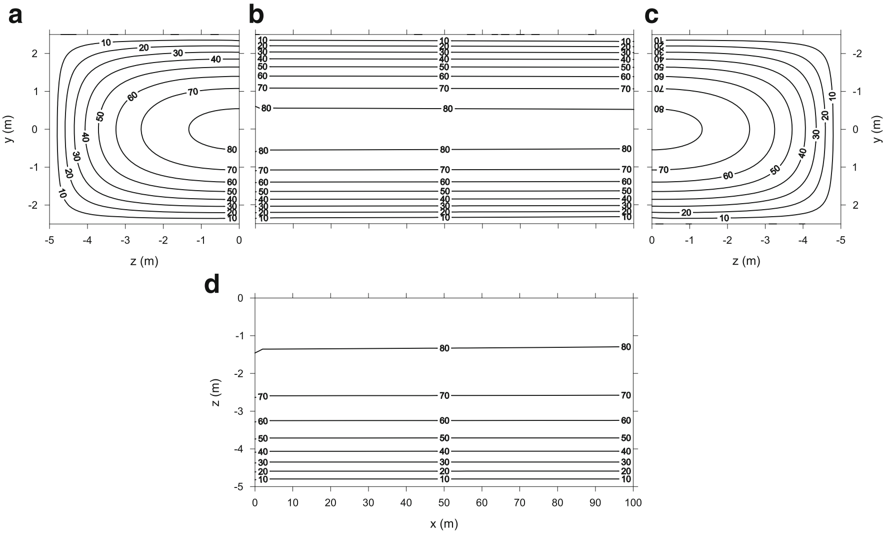


Fig. 4 Contour maps of velocity v_x at the steady state for $\alpha = \alpha_1$. (**a**) Contour map on the plane $x = 0$; (**b**) contour map on the plane $z = 0$; (**c**) contour map on the plane $x = L$; (**d**) contour map on the plane $y = 0$

We study the cooling process in the steady state two different channel slopes α_1 and α_2.

The resulting temperature fields show that the two static zones close to the channel levees that can be observed in [4] do not appear in this case study for both the values of the channel slope. The reason is that the effusion temperature of Mauna Loa lava flow is higher than the Etna lava flow. The temperature is lower than the solidus temperature T_s for both the values of α very close to che channel levees in a very narrow zone, so it is expected that the upper crust forms at the channel levees very far from the vent. This is confirmed by direct observation reported by [7] who showed that the stabilized channel is 10 km long and only in the final part, in a zone of the channel that is called transitional channel, a pahoehoe crust starts forms near the channel levees, while in the central part of the channel lava flows with poorly developed crust.

The temperature and strain dependent reological model of [10] gives very high values of viscosity which translate in very high values of velocity, as it can be observed in Figs. 3 and 5, even if it is evident the sensitivity of the solution to the value of the channel slope α.

The two different values of the effusion temperature underline the role of viscosity and of heat advection in the cooling process, since the steeper channel flows much faster than the less inclined, although the differences among the two values of α are not so high. So the temperature dependence of the viscosity function strongly affects the dynamics of the lava flow. Regarding the differences in the thermal aspects of the two case studies, we cannot observe great differences because

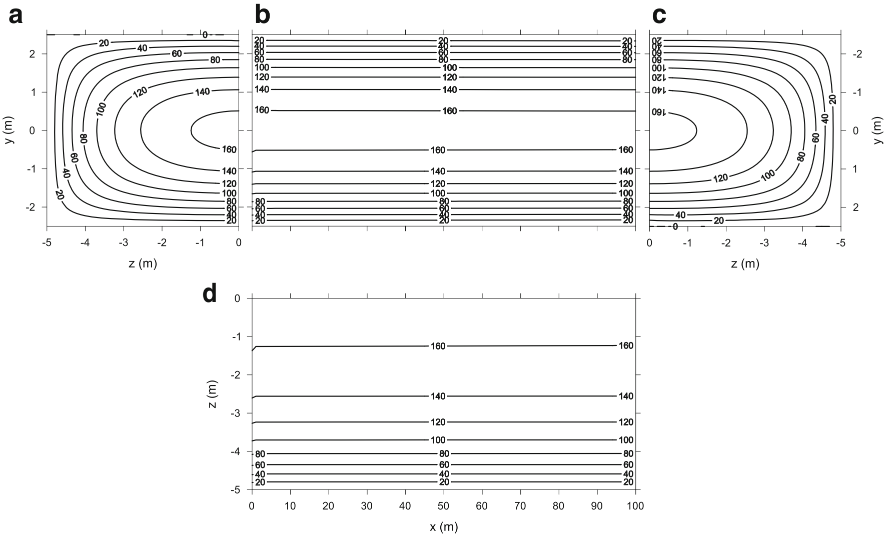


Fig. 5 Contour maps of velocity v_x at the steady state for $\alpha = \alpha_2$. (**a**) Contour map on the plane $x = 0$; (**b**) contour map on the plane $z = 0$; (**c**) contour map on the plane $x = L$; (**d**) contour map on the plane $y = 0$

of the proximity to the vent. The channel is approximately isothermal except for the levees and ground where the cooling can be observed also in the first 100 m, as it can be seen in Figs. 2b and 3b.

As observed by [16] on active lava flows on Kilauea volcano, Hawaii, the high temperatures at the center of the channel, where the crust is very thin, and the low temperatures at the levees of the channel, where the crust is thicker and stable, can explain how the gradual inwards growth of lateral crust may cause the tube formation in the classic zipper fashion [5].

References

1. Bailey, J.E., Harris, A.J.L., Dehn, J., Calvari, S., Rowland, S.K.: The changing morphology of an open lava channel on Mt. Etna. Bull. Volcanol. **68**, 497–515 (2006)
2. Ball, M., Pinkerton, H., Harris, A.J.L.: Surface cooling, advection and the development of different surface textures on active lavas on Kilauea, Hawai'i. J. Volcanol. Geoth. Res. **173**(1–2), 148–156 (2008)
3. Filippucci, M., Tallarico, A., Dragoni, M.: A three-dimensional dynamical model for channeled lava flow with nonlinear rheology. J. Geophys. Res. **115**, B05202 (2010)
4. Filippucci, M., Tallarico, A., Dragoni, M.: Role of heat advection in a channeled lava flow with power law, temperature-dependent rheology. J. Geophys. Res. **118**(6), 2764–2776 (2013)
5. Greeley, R.: Observations of actively forming lava tubes and associated structures Hawaii. Mod. Geol. **2**, 207–223 (1971)

6. Gregg, T.K.P., Fink, J.H.: A laboratory investigation into the effect of slope on lava flow morphology. J. Volcanol. Geoth. Res. **96**, 145–159 (2000)
7. Harris, A.J.L., Rowland, S.K.: FLOWGO: a kinematic thermo-rheological model for lava flowing in a channel. Bull. Volcanol. **63**, 20–44 (2001)
8. Harris, A.J.L., Flynn, L.P., Keszthelyi, L., Mouginis-Mark, P.J., Rowland, S.K., Resing, J.A.: Calculation of lava effusion rates from Landsat TM data. Bull. Volcanol. **60**, 152–171 (1998)
9. Head, J.W., Wilson, L.: Volcanic processes and landforms on Venus: theory, predictions, and observations. J. Geophys. Res. **91**, 9407–9446 (1986)
10. Hobiger, M., Sonder, I., Büttner, R., Zimanowski, B.: Viscosity characteristics of selected volcanic rock melts. J. Volcanol. Geoth. Res. **200**(1–2), 27–34 (2011)
11. Keszthelyi, L., Denlinger, R.: The initial cooling of pahoehoe flow lobes. Bull. Volcanol. **58**, 5–18 (1996)
12. Keszthelyi, L., Harris, A.J.L., Dehn, J.: Observations of the effect of wind on the cooling of active lava flows. Geophys. Res. Lett. **30**, 1989 (2003)
13. Merle, O.: Internal strain within lava flows from analogue modelling. Volcanol. Geoth. Res. **81**(3–4), 189–206 (1998)
14. Neri, A.: A local heat transfer analysis of lava cooling in the atmosphere: application to thermal diffusion-dominated lava flows. J. Volcanol. Geoth. Res. **81**, 215–243 (1998)
15. Patrick, M., Dehn, J., Dean, K.: Numerical modelling of lava flow cooling applied to the 1997 Okmok eruption: comparison with advanced very high resolution radiometer thermal imagery. J. Geophys. Res. **110**, B02210 (2005)
16. Pinkerton, H., James, M.R., Jones, A.: Surface temperature measurements of active lava flows on Kilauea volcano, Hawai'i. J. Volcanol. Geoth. Res. **113**(1–2), 159–176 (2002)

Part III
Fibers

Overview

Fiber spinning, fiber suspension flows, fiber micro-structures are the topics of this section *Fibers*. The dynamics and behavior of fibers play an important role in nonwoven production, glass-wool manufacturing and paper forming. In these processes slender objects, such as oriented particles, elastic threads or viscous/viscoelastic jets, move due to mechanical or aerodynamic forces and interact with each other, outer boundaries and/or surrounding flows. The quality of fiber fabrics crucially depends on the properties of the micro-structure.

The following nine contributions present new models and methods for different aspects, bringing together asymptotics, stochastics and numerics.

The first two papers address the simulation of fiber spinning, using asymptotic Cosserat formulations for the jet dynamics. In *On Viscoelastic Fiber Spinning: Die Swell Effect in the 1D Uniaxial UCM Model* Maike Lorenz et al. investigate the phenomenon of die swell occurring in drawing processes of viscoelastic jets. The asymptotically derived upper convected Maxwell (UCM) model has an hyperbolic character such that the existence regime of solutions depend crucially on the physical parameters and the boundary conditions. These restrictions hold also true in the viscous limit when the Weissenberg number vanishes (viscous string model). The viscous Cosserat rod model that covers additional angular momentum effects overcomes the restrictions of the string model and is valid for all possible parameters. As it converges to the string model in the slenderness limit, it can be understood as regularized string model. The partial and ordinary differential system is highly stiff since it contains the slenderness parameter. In *Numerical Treatment of Non-Stationary Viscous Cosserat Rod in a Two-Dimensional Eulerian Framework* Walter Arne et al. propose a numerical scheme based on a semidiscretization with finite volumes in space. The time integration is performed with stiffly accurate Radau methods. Numerical results are shown for rotational spinning.

The third and fourth papers deal with fiber-fluid interactions. Due to the slender geometry the effect of a single fiber on a surrounding flow field is small and neglected in the majority of cases. This yields a one-way coupling. However, when considering fiber curtains or fiber bundles in a flow, there is the need of a two-way coupling. In *Asymptotic Modeling Framework for Fiber-Flow Interactions in a Two-Way Coupling* Thomas M. Cibis et al. present an asymptotic modeling framework for such a two-way coupling between fibers and flow. It is based on slender-body theory and the modeling of exchange functions in terms of drag forces and heat sources. The exchange functions are incorporated in the conservation equations for linear momentum and energy with respect to flow and fibers and satisfy a generalized action-reaction principle. The concept is applied to a rotational spinning process for glass wool production. In *Efficient Simulation of Random Fields for Fiber-Fluid Interactions in Isotropic Turbulence* Florian Hübsch et al. consider a fiber dynamics in a turbulent flow. The flow fluctuations are modeled as random field in $\mathbb{R}^4$ on top of a statistic k-ϵ turbulence formulation and the interactions are described as a one-way coupling with a stochastic aerodynamic drag force on the fiber. The focus lies on the construction and efficient simulation/sampling of the fluctuations, therefore the special covariance structure of the random field (isotropy, homogeneity and decoupling of space and time) is exploited.

In *On Stability of a Concentrated Fiber Suspension Flow* by Uldis Strautins a linear stability analysis of a fiber suspension flow in a channel domain is performed using a modified Folgar-Tucker equation. Two kinds of potential instability are identified: whereas one is associated with overcritical Reynolds number, the other one is present for any Reynolds numbers since it is associated with certain perturbations in fiber orientation field. The second type of instability leads to initially growing transient perturbations in the micro-structure. It is shown that both types of instability lead to instability of the bulk velocity field. The presence of fibers increases the stability region.

The last three papers address the simulation and/or investigation of fiber micro-structures. In *Microstructure Simulation of Paper Forming* Erik Svenning et al. present a numerical framework designed to simulate a paper forming process. This process includes strong fluid-structure interaction and complex geometries. The fluid flow solver employs immersed boundary methods to compute the flow around the fibers without the necessity of a boundary conforming grid. The fibers are described by a Euler-Bernoulli beam equation. The contact is realized by a penalty-based model. For production processes of nonwoven materials, a monolithic numerical treatment is not possible for computational reasons due to the fiber concentration and geometry. Hence, in *Three-Dimensional Fiber Lay-Down in an Industrial Application* Johannes Maringer et al. propose surrogate fiber lay-down models that describe the form of deposited endless fibers with help of stochastic differential equations (degenerated diffusion processes). The model parameters are estimated from the models of first principles (Cosserat theory for single fibers) in combination with measurements of the resulting nonwoven. In the paper the adaptation of a three-dimensional model to a typical industrial process is discussed. Apart from such lay-down models there exist several other stochastic models for fiber Microstructure s (system): systems of straight non-overlapping fibers, systems of overlapping bending fibers, or fiber systems created by sedimentation. In *3d*

Modeling of Dense Packings of Bended Fibers Hellen Altendorf and Dominique Jeulin present a stochastic model that generalizes the force-biased packing approach to fibers represented as chains of balls. The starting configuration is a Boolean system of fibers modeled by random walks, where two parameters in the multivariate von Mises-Fisher orientation distribution control the bending. The points of the random walk are associated with a radius and the current orientation. The resulting chains of balls are interpreted as fibers. The final fiber configuration is obtained as an equilibrium between repulsion forces avoiding crossing fibers and recover forces ensuring the fiber structure. This approach can provide high volume fractions. Alternatively, a intelligent placing strategy is exploited that turns out to be very efficient for intermediate volume fractions.

Nicole Marheineke

On Viscoelastic Fiber Spinning: Die Swell Effect in the 1D Uniaxial UCM Model

Maike Lorenz, Nicole Marheineke, and Raimund Wegener

Abstract This work deals with a stationary viscoelastic jet under gravitational forces described by an upper convected Maxwell (UCM) model. For spinning processes we demonstrate that a die swell-like behavior of the solution is in general possible for the asymptotically derived one-dimensional model equations. Nevertheless, to use the model for the prediction of a die swell appropriate boundary conditions or the inclusion of further effects such as surface tension have to be considered. Moreover, the regime of existence of solutions for drawing processes is determined numerically.

1 Introduction

In the production of fibers a molten polymer is pressed out of a nozzle forming a curved jet in a surrounding air flow. The spinning of such a viscoelastic jet can be described as a three-dimensional free boundary value problem using an upper convected Maxwell model. For simplicity, we neglect surface tension, temperature and aerodynamic effects. Applying slender-body theory similar to [9], an asymptotic one-dimensional model based on the transient, arc-length parameterized jet's centerline is derived in [8]. For time $t \in \mathbb{R}_0^+$ and arc-length parameter $s \in [0, L(t)]$, this dimensionless model is given by the balance of mass and momentum

N. Marheineke (✉)
Department Mathematik, Friedrich-Alexander-Universität Nürnberg-Erlangen, Cauerstr. 11, 91058 Erlangen, Germany
e-mail: marheineke@math.fau.de

R. Wegener
Fraunhofer Institut für Techno- und Wirtschaftsmathematik, Fraunhofer Platz 1, 67663 Kaiserslautern, Germany
e-mail: raimund.wegener@itwm.fraunhofer.de

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*, Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_13,

$$\begin{aligned}\partial_t A + \partial_s(uA) &= 0,\\ \mathrm{Re}(\partial_t(A\mathbf{v}) + \partial_s(uA\mathbf{v})) &= \partial_s(A\sigma\partial_s\boldsymbol{\gamma}) + A\mathbf{f},\end{aligned} \tag{1}$$

the constitutive equations

$$\begin{aligned}\mathrm{We}(\partial_t p + u\partial_s p + p\partial_s u) + p &= -\partial_s u,\\ \mathrm{We}(\partial_t\sigma + u\partial_s\sigma - (3p + 2\sigma)\partial_s u) + \sigma &= 3\partial_s u\end{aligned} \tag{2}$$

and the dynamics of the center-line

$$\partial_t\boldsymbol{\gamma} + u\partial_s\boldsymbol{\gamma} = \mathbf{v}, \quad \|\partial_s\boldsymbol{\gamma}\| \quad = 1 \tag{3}$$

with the cross-sectional area A, the momentum-associated velocity $\mathbf{v}$, the intrinsic velocity (speed) u, the pressure p, the stress component σ and the center-line $\boldsymbol{\gamma}$. The outer forces $\mathbf{f}$ are considered to be given. The evolution of the jet length is described by $\mathrm{d}L(t)/\mathrm{d}t = u(L(t), t)$ with $L(0) = 0$. The boundary conditions at the nozzle are

$$\begin{aligned}&A(0,t) = 1 &\qquad& u(0,t) = 1\\ &\sigma(0,t) = \sigma_0 && \\ &\boldsymbol{\gamma}(0,t) = \boldsymbol{\gamma}_0 && \partial_s\boldsymbol{\gamma}(0,t) = \boldsymbol{\tau}_0,\end{aligned}$$

and at the free end of the jet we have $\sigma(L(t), t) = 0$. The dimensionless parameters are the Reynolds number Re and the Weissenberg number We denoting the ratio of inertial forces and viscous forces and the ratio of relaxation time and process time, respectively.

In this work we investigate the applicability and properties of the one-dimensional model, in particular whether we can simulate a die swell. A die swell, also called extrudate swell, is an effect observed in many extrusion processes with viscoelastic fluids. Here, a fluid exits from a capillary into the air such that a jet forms with a diameter significantly larger than the diameter of the nozzle [3], for photos see [7]. A sketch is given in Fig. 1a. In fiber spinning processes the forming of a die swell is undesirable since it changes the flow properties of the non-Newtonian fluid and consequently the quality of the resulting fabric. Hence, the understanding and prediction of this phenomenon is of interest to industry. So far, simulations for the UCM fluid showing this effect are based on the full two- or three-dimensional models, see e.g. [4].

The paper is organized as follows. First we derive the stationary model equations from (1) to (3) for an arbitrary curved jet. They can be used for describing a spun jet exposed to gravity. Then we investigate for the special uniaxial case (straight jet)—where the jet is pointing in the direction of the gravitational force—, whether the model allows for a die swell. In addition, we determine numerically the regime of existence of solutions in the space of the dimensionless parameters.

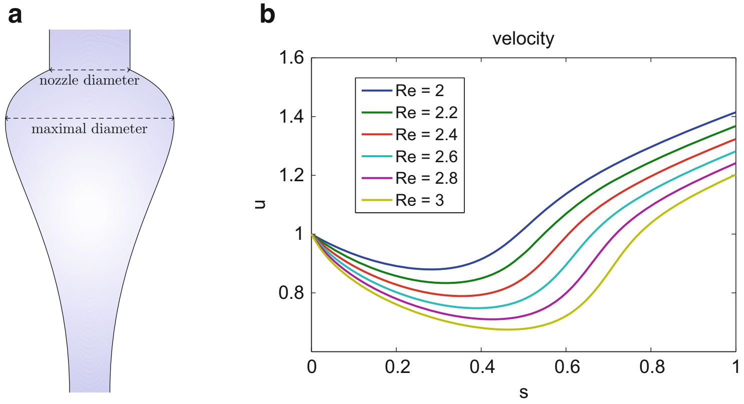


Fig. 1 Die swell in an extrusion process. Simulation results of the stationary uniaxial UCM model (6) with $b(1) = -0.3$, $a(0) = -1$, $\mathrm{Fr} = 1, \mathrm{We} = 2$, and $\mathrm{Re} \in [2, 3]$. The plotted velocity is related to the cross-sections by $u = 1/A$. (**a**) Sketch. (**b**) Simulations

2 Stationary One-Dimensional UCM Model for Straight Jets

Let us consider a fiber spinning process which lasts long enough such that the jet up to a length L can be regarded as stationary. Without loss of generality let be $L = 1$. Then, the corresponding stationary UCM model for $s \in [0, 1]$ is given by

$$\begin{aligned}
\partial_s \boldsymbol{\gamma} &= \boldsymbol{\chi} \\
q_1 \, \partial_s \boldsymbol{\chi} &= \frac{1}{u}\mathbf{f} - \frac{1}{u}(\mathbf{f} \cdot \boldsymbol{\chi})\boldsymbol{\chi} \\
q_2 \, \partial_s u &= \sigma - \mathrm{We}\, u \, \mathbf{f} \cdot \boldsymbol{\chi} \\
\partial_s \sigma &= \partial_s u \left(\mathrm{Re}\, u + \frac{\sigma}{u}\right) - \mathbf{f} \cdot \boldsymbol{\chi} \\
\mathrm{We}\, \partial_s p &= -\frac{1}{u}\left(\partial_s u (1 + \mathrm{We}\, p) + p\right) \\
\|\boldsymbol{\chi}\| &= 1
\end{aligned} \tag{4}$$

$$q_1 = \mathrm{Re}\, u - \frac{\sigma}{u}, \qquad q_2 = 3 + \mathrm{We}\,(\sigma + 3p) - \mathrm{Re}\mathrm{We}\, u^2,$$

supplemented with appropriate boundary conditions. In order to obtain the model we neglect the time dependence in (1)–(3) and convert it to a system of first order ordinary differential equations (ODE). Therefore we introduce the tangent of the jet $\boldsymbol{\chi}$. Since the mass flux is constant, the cross-sectional area is related to the velocity and drops out of the equations, i.e. $A \equiv 1/u > 0$ holds.

The solvability of this model is restricted and depends critically on the parameter regime and the boundary conditions. The viscoelastic UCM model includes the viscous case when $\mathrm{We} = 0$. From the viscous case it is known that the term q_1, which is a monotonically increasing function in s, crucially determines the applicability of the model equations, cf. [1, 2]. For $q_1(s) = 0$ a singularity occurs that is only removable by help of an appropriate choice of closure conditions [6]. For the viscoelastic model the term q_2 leads additionally to limitations which are even more difficult to predict since q_2 is not monotone. In the uniaxial case of interest the term q_1 is not present and we can exclusively focus on q_2.

A stationary uniaxial (straight) jet is described by

$$\begin{aligned} q_2 \partial_s u &= \sigma - \mathrm{WeB}u \\ \partial_s \sigma &= \partial_s u \left(\mathrm{Re}u + \frac{\sigma}{u} \right) - \mathrm{B} \\ \mathrm{We}\partial_s p &= -\frac{1}{u} \left(\partial_s u \left(1 + \mathrm{We}\, p\right) + p \right) \\ q_2 &= 3 + \mathrm{We} \left(\sigma + 3p \right) - \mathrm{ReWe}u^2. \end{aligned} \tag{5}$$

We deduce this system of equations from (4) by setting $\boldsymbol{\gamma} = (0, 0, -s)$ in a gravitational configuration where the outer forces are $\mathbf{f} = \mathrm{B}(0, 0, -1)$ and $\mathrm{B} = \mathrm{Re}/\mathrm{Fr}^2$. Here, Fr is the Froude number which denotes the ratio of inertial and gravitational forces. See also [5, 10] for an asymptotic derivation and for existence and uniqueness results to $\mathrm{Re} = 0$.

Depending on the boundary conditions the term q_2 in (5) may limit the regime of existence if it contains a root. Hence, we cannot expect solutions for the whole parameter space $(\mathrm{Re}, \mathrm{Fr}, \mathrm{We})$, whereas the viscous model ($\mathrm{We} = 0$) has solutions for all $(\mathrm{Re}, \mathrm{Fr})$. Another difference to the viscous model is that the viscoelastic model has solutions with a minimum in the velocity u for $s \in (0, 1)$. For the viscous model one can easily show that any extremal value of u in $(0, 1)$ is a maximum. Since in the stationary case we have the relation $A = 1/u$ for the cross-sectional area A we investigate whether this behavior enables us to reproduce a die swell which is observed in experiments with viscoelastic fluids.

For the numerical solution of the arising boundary value problems we use a Runge–Kutta collocation method. The resulting systems of non-linear equations are solved via Newton's method [11].

2.1 Drawing Processes and Die Swell

This section addresses the question whether the uniaxial UCM model (5) allows for solutions with a die swell. To simplify the investigations we restate (5) in terms of $b = \sigma - \mathrm{WeB}\, u$ and $a = q_2$:

$$
\begin{aligned}
\partial_s u &= \frac{b}{a} \\
\partial_s b &= \frac{b}{a}\left(\frac{b + \mathrm{Re}u^2}{u}\right) - \mathrm{B} \\
\partial_s a &= \frac{b}{a} 2\mathrm{We}\left(\mathrm{BWe} + \frac{b}{u} - \mathrm{Re}u\right) - \mathrm{Re}u + \frac{1}{\mathrm{We}u}(3 - a).
\end{aligned} \tag{6}
$$

For the boundary conditions we impose

$$
u(0) = 1, \quad b(1) = D_1, \quad a(0) = D_2 \tag{7}
$$

for some $D_1 \in \mathbb{R}$ and $D_2 \in \mathbb{R} \setminus \{0\}$. In particular, we set $D_1 = 0$ to achieve a *constant velocity end* with $\partial_s u(1) = 0$ or alternatively $D_1 = -\mathrm{WeB}\, u(1)$ for a *stress free end* corresponding to $\sigma(1) = 0$.

Definition 1. Let (u, b, a) be a continuously differentiable solution of (6) for arbitrary but fixed boundary conditions. We call $s^* \in (0, 1)$ a point where a die swell occurs if $\partial_s u(s^*) = 0$ holds true and $u(s^*)$ is a local minimum.

Among all possible solutions we are only interested in drawing processes as defined below which we consider to be the physically relevant solutions in this scenario.

Definition 2 (Drawing Process). We call the solution of the ODE system (6) a *drawing process*

- *without die swell* if $\partial_s u(s) > 0$ for all $s \in [0, 1)$,
- *with die swell* if there exists exactly one $s^* \in (0, 1)$ with $\partial_s u(s^*) = 0$, $\partial_s u(s) < 0$ for all $0 \leq s < s^*$ and $\partial_s u(s) > 0$ for all $s^* < s < 1$.

With the following Lemmata we can exclude the occurrence of a die swell for the boundary condition $b(1) = 0$ (constant velocity end).

Lemma 1. *Let* $(\mathrm{Re}, \mathrm{Fr}, \mathrm{We})$ *be given with* $\mathrm{B} \neq 0$ *and suppose that* (u, b, a) *are continuously differentiable solutions of (6) for arbitrary but fixed boundary conditions. Suppose that* $a \neq 0$ *for all* $s \in [0, 1]$. *Then* b *can have at most one root on* $[0, 1]$.

Proof. For any root s^* in b we find that $\partial_s b(s^*) = -\mathrm{B}$. By continuity of b only one root can occur.

Lemma 2. *Let* $(\mathrm{Re}, \mathrm{Fr}, \mathrm{We})$ *and some* $D \in \mathbb{R} \setminus \{0\}$ *be given such that* $a(0) = D$ *or* $a(1) = D$. *Suppose that a continuously differentiable solution of (6) exists with* $u(0) = 1$, $b(1) = 0$ *and* $a \neq 0$ *for all* $s \in [0, 1]$. *Then this cannot be a drawing process with a die swell. Furthermore, this is only a drawing process if* $D > 0$.

Proof. For a drawing process with a die swell a root of $\partial_s u$ is required on $(0, 1)$. This corresponds to a root in b at some $s^* \in (0, 1)$. Due to the boundary condition and Lemma 1 this is not possible.

If $D < 0$ also $a < 0$ holds true for all $s \in [0, 1]$ and hence $b(1) = 0$ enforces a minimum in u such that the velocity is monotonically decreasing on [0, 1]. This is not a drawing process.

By the previous considerations we know that a die swell cannot occur for $b(1) = 0$. Hence we impose the boundary condition $\sigma(1) = 0$ ensuring a stress free end, i.e. in terms of b, a we set $b(1) = -\mathrm{WeB}\, u(1) < 0$. To ensure a drawing process a needs to be negative and we choose $a(0) < 0$. For given (Re, Fr, We) we can decide by the sign of $b(0)$ whether a numerically determined solution is a drawing process with a die swell ($b(0) > 0$) or without a die swell ($b(0) \leq 0$). For all tested parameter triples (Re, Fr, We) the solutions belong to drawing processes without a die swell. Nevertheless, we can impose $b(1) < 0$, $a(0) < 0$ and find parameters for which a die swell occurs. One example is given in Fig. 1b for $\mathrm{Fr} = 1, \mathrm{We} = 2$, $b(1) = -0.3$, $a(0) = -1$ and different Re. The velocity develops a clear minimum and finally increases to values larger than the inflow velocity. This corresponds to a jet with a diameter that first increases to values larger than the diameter of the nozzle (cf. Fig. 1a).

2.2 Regime of Existence

The parameter regime (Re, Fr, We) where solutions exist depends essentially on the run of the non-monotone function q_2 for given boundary conditions. For its determination we have to detect the roots of q_2. The stress free boundary condition $\sigma(1) = 0$ is not suited for a systematic search since q_2 develops a root inside the considered interval [0, 1]. On the contrary, for the constant velocity boundary condition $b(1) = 0$, i.e. (7) with $D_1 = 0$ and $D_2 > 0$, we find that the minimum of $q_2 = a$ occurs at $s = 1$. Hence, for this case which excludes the occurrence of a die swell, we can carry out a systematic numerical search in the three-dimensional parameter space using the following approach: the aim is to find those parameters (Re, Fr, We) for which a solution of (7) exists with $a(1) = 0$. For the numerical treatment we consider We variable and impose the additional boundary condition $a(1) = \delta$ for $0 < \delta \ll 1$.

For $\delta = 0.1$, $D_2 = 1$ and $\mathrm{Fr} = 2$ the numerical result is shown in Fig. 2a. The blue curves visualize the limiting curves found for varying Re, the green circles mark the parameters $(\mathrm{Re}, \mathrm{We}) \in [10^{-2}, 10^{2}] \times [10^{-7}, 6.2]$ for which numerical solutions exist. One observes that the area enclosed by both limiting curves does not allow for solutions. We also notice that we have to consider either We or Re variable and combine both results to capture all regions of the limiting curve depending on the gradient.

With a combined search we can additionally vary Fr and obtain the corresponding limiting surface shown in Fig. 2b. In consistence with Fig. 2a no solutions can be found for (Re, Fr, We) inside the volume enclosed by the limiting surface. Close to the viscous case ($\mathrm{We} \ll 1$) no limitations occur. For moderate values of We only

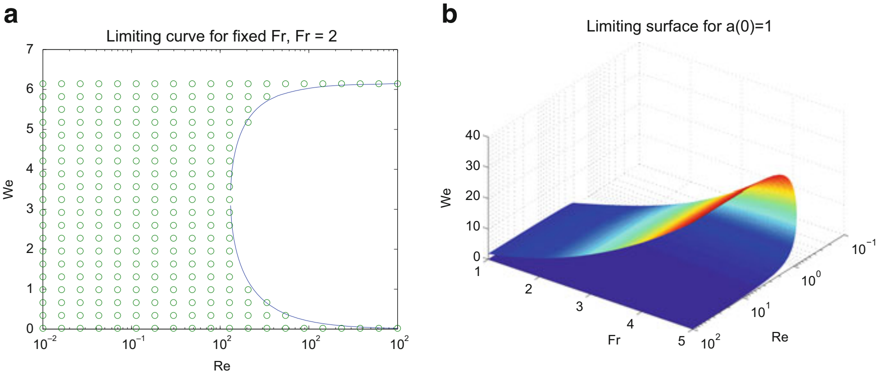


Fig. 2 Parameter regime of existence of solutions for $u(0) = 1, b(1) = 0, a(0) = 1$. (**a**) Limiting curves. (**b**) Limiting surface

small Reynolds numbers imply solutions whereas larger Reynolds numbers require We to lie above the upper part of the limiting surface. This part grows more than linearly in We for increasing Froude numbers. Increasing D_2 leads to comparably shaped limiting surfaces with a less restrictive regime of existence.

3 Conclusion and Outlook

The asymptotic upper convected Maxwell model allows for solutions forming a die swell. For the stationary uniaxial gravitational spinning set-up the occurrence of a die swell can be analytically excluded for the boundary condition of a constant velocity end $\partial_s u(1) = 0$ and no numerical examples are found for the stress free end $\sigma(1) = 0$. Nevertheless, for artificial boundary conditions the desired shape can be obtained. It is still an open question whether reasonable boundary conditions can be physically motivated to simulate the die swell or whether further effects such as surface tension need to be included to obtain results that are comparable with experimental data and meaningful for the prediction of a die swell.

In contrast to the uniaxial viscous model, the applicability of the viscoelastic one is limited to certain parameter ranges due to the occurring roots of the non-monotone function q_2. This makes the simulation of some industrial spinning processes not only difficult but impossible. The investigation of alternative asymptotic models that overcome this problem is left to future work.

Acknowledgements This work has been supported by the Center for Mathematical and Computational Modeling $(CM)^2$, Kaiserslautern (Rhineland-Palatinate research initiative)

References

1. Arne, W., Marheineke, N., Meister, A., Wegener, R.: Numerical analysis of Cosserat rod and string models for viscous jets in rotational spinning processes. Math. Models Methods Appl. Sci. **20**(10), 1941–1965 (2010)
2. Arne, W., Marheineke, N., Wegener, R.: Asymptotic transition from Cosserat rod to string models for curved viscous inertial jets. Math. Models Methods Appl. Sci. **21**(10), 1987–2018 (2011)
3. Bird, R.B., Armstrong, R.C., Hassager, O.: Dynamics of Polymeric Liquids. Wiley, New York (1987)
4. Crochet, M.J., Keunings, R.: Die swell of a Maxwell-fluid: numerical prediction. J. Nonnewton. Fluid Mech. **7**, 199–212 (1980)
5. Hagen, T.C.: On viscoelastic fluids in elongation. Adv. Math. Res. **1**, 187–205 (2002)
6. Hlod, A., Aarts, A.C.T., van de Ven, A.A.F., Peletier, M.A.: Mathematical model of falling of a viscous jet onto a moving surface. Eur. J. Appl. Math. **18**(6), 659–677 (2007)
7. Joseph, D.D.: Fluid Dynamics of Viscoelastic Liquids. Springer, New York (1990)
8. Lorenz, M., Marheineke, N., Wegener, R.: On an asymptotic upper convected Maxwell model for a viscoelastic jets. Proc. Appl. Math. Mech. **12**, 601–602 (2012)
9. Marheineke, N., Wegener, R.: Asymptotic model for the dynamics of curved viscous fibres with surface tension. J. Fluid Mech. **622**, 345–369 (2009)
10. Schultz, W.: Slender viscoelastic fiber flow. J. Rheol. **31**(8), 733–750 (1987)
11. Shampine, L.F., Gladwell, I., Thompson, S.: Solving ODEs with MATLAB. Cambridge University Press, Cambridge (2003)

Numerical Treatment of Non-stationary Viscous Cosserat Rod in a Two-Dimensional Eulerian Framework

Walter Arne, Nicole Marheineke, Andreas Meister, and Raimund Wegener

Abstract This work deals with the modeling and simulation of the dynamics of a slender viscous jet as it arises in spinning processes. There exist two classes of asymptotic one-dimensional models for such a jet, string and more complex rod models, that are given by systems of partial and ordinary differential equations. In this paper, we present non-stationary simulations of a rod in an Eulerian framework for arbitrary parameter ranges of 2d spinning where the string models failed so far. The numerical treatment is based on a finite volume approach with mixed central, up- and downwinded differences, the time integration is performed by a Radau method.

1 Introduction

Considering the spinning of highly viscous fluids, the unrestricted motion of a non-stationary jet's center-line plays an important role [1]. Typical industrial applications are e.g. drawing, tapering and spinning of glass/polymer fibers [2, 3] or pellet manufacturing [4,5]. For the numerical simulation there exist two classes of asymptotic one-dimensional models: string and rod models. Whereas the string models

W. Arne (✉) • R. Wegener
Fraunhofer Institut für Techno- und Wirtschaftsmathematik (ITWM), Fraunhofer Platz 1,
67663 Kaiserslautern, Germany
e-mail: walter.arne@itwm.fraunhofer.de; raimund.wegener@itwm.fraunhofer.de

N. Marheineke
Lehrstuhl Angewandte Mathematik 1, FAU Erlangen-Nürnberg, Cauerstr. 11,
91058 Erlangen, Germany
e-mail: marheineke@math.fau.de

A. Meister
Fachbereich Mathematik und Naturwissenschaften, Universität Kassel, Heinrich Plett Str 40,
34132 Kassel, Germany
e-mail: meister@mathematik.uni-kassel.de

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_14,

consist of balance equations for mass and linear momentum, the more complex rod models also contain an angular momentum balance [6, 7]. The applicability of the string model with asymptotically derived boundary conditions [8] turned out to be restricted to certain parameter ranges. Already for jets in a stationary, rotational 2d scenario no solutions exist for $\mathrm{ReRb}^2 < 1$ with Rossby number $\mathrm{Rb} \ll 1$ [9]; the numerical evidence of this inviscid bound was specified analytically in [10]. The limitation can be partly overcome by a modification of the closure conditions motivated by Hlod et al. [11, 12]. However, there is still a parameter range for which an existence gap of string solutions is observed [10]. The viscous Cosserat rod theory raises hope to open the parameter ranges of practical interest and time-dependencies to simulation. Based on the work by Ribe [13] we developed a modified incompressible Cosserat rod model [14] that reduces asymptotically to the string equations for a vanishing slenderness parameter. So far, stationary rod simulations have been applied successfully in the study of a fluid-mechanical sewing machine [15] and the design of a glass wool production process [16]. In this paper, we present non-stationary rod simulations for a 2d inflow-outflow problem in an Eulerian framework. In long-time behavior they converge to the stationary results. The applicability is unrestricted such that they allow the study of practically relevant parameter ranges where the string models failed [8]. The proposed numerical scheme can be generalized to arbitrary 3d flow situations, including inflow problems with increasing jet length and free end [17].

The paper is structured as follows. Focusing on the rotational 2d spinning scenario of [8, 10] we first introduce the Cosserat rod model. Then, we present the numerical approach and finally discuss the simulation results.

2 Viscous Rod Model

In rotational spinning processes viscous liquid jets leave small nozzles located on the curved face of a circular cylindrical drum rotating about its symmetry axis, Fig. 1. They move and grow due to gravity and aerodynamic forces. At the nozzle, the jet's velocity, cross-sectional area, direction and curvature are prescribed. The jet end is characterized by stress-free conditions for inner contact forces and couples.

This work aims at the numerical handling of the time-dependencies. Therefore we focus on the rotational 2d scenario of [8, 10, 14] neglecting gravity, aerodynamic forces, surface tension and temperature effects. Note that once the numerical concept is established, these effects can be easily added as it is done for stationary considerations [16]. We consider a jet of certain length, i.e. inflow-outflow set-up with time-independent flow domain, and study the effects of viscosity and rotation on the non-stationary jet's center-line. Due to the slender geometry the jet dynamics can be reduced to a one-dimensional description by averaging the underlying balance laws over its cross-sections. The special Cosserat rod theory consists hereby of two constitutive elements, a curve specifying the position (center-line) and an orthonormal director triad characterizing the orientation of the cross-sections, for

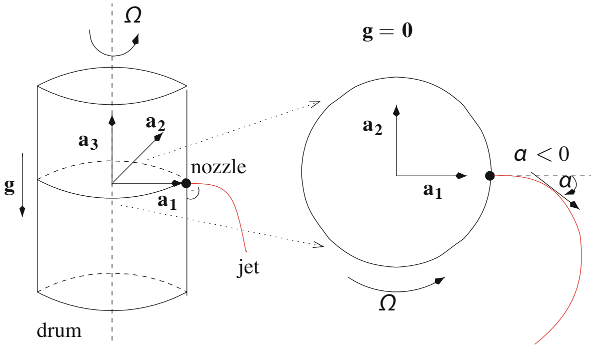


Fig. 1 Rotational 3d spinning process and simplified 2d set-up for gravity $\mathbf{g} = \mathbf{0}$

details see [18]. Formulated in the rotating framework (Fig. 1) the corresponding 2d rod equations in Eulerian description are given by [14],

$$\begin{aligned} \mathsf{R}(\alpha) \cdot \partial_t \mathsf{r} &= \mathsf{v} - u\mathsf{e}_2 \\ \partial_t \alpha &= \omega - u\kappa \\ \partial_s (u\mathsf{e}_2) &= \partial_s \mathsf{v} + \kappa \mathsf{v}^\perp + \omega \mathsf{e}_1 \\ \partial_t \kappa + \partial_s (u\kappa) &= \partial_s \omega \\ \partial_t A + \partial_s (uA) &= 0 \\ \rho \partial_t (A\mathsf{v}) + \rho \partial_s (uA\mathsf{v}) &= \partial_s \mathsf{n} + \kappa \mathsf{n}^\perp - \rho A \omega \mathsf{v}^\perp + \mathsf{k} \\ \rho \partial_t (I\omega) + \rho \partial_s (uI\omega) &= \partial_s m - n_1 + l \end{aligned} \tag{1}$$

with

$$\mathsf{R}(\alpha) = \begin{pmatrix} \sin\alpha & -\cos\alpha \\ \cos\alpha & \sin\alpha \end{pmatrix}$$

$$\mathsf{k} = -2\rho A \Omega \mathsf{v}^\perp + \rho A \Omega^2 \mathsf{R}(\alpha) \cdot \mathsf{r}, \qquad l = \rho I \Omega \, \partial_s u$$

and material laws

$$n_2 = 3\mu A \partial_s u, \qquad m = 3\mu I \, \partial_s \omega, \qquad I = \frac{A^2}{4\pi}.$$

Here e_i, $i = 1, 2$ denote the canonical basis vectors in $\mathbb{R}^2$, moreover $\mathsf{x}^\perp = (x_1, x_2)^\perp = (-x_2, x_1)$ for all $\mathsf{x} \in \mathbb{R}^2$. The four kinematic and three dynamic (balance) equations describe the variables of jet's center-line $\mathsf{r} = (r_1, r_2)$, angle α determining the tangent (i.e. director triad in 2d), curvature κ, cross-section A, velocity $\mathsf{v} = (v_1, v_2)$, angular speed ω, convective speed u and inner shear force n_1. The inner traction n_2 and couple m are specified by the material laws. Further,

k present the centrifugal and Coriolis forces and l the corresponding outer couples. The closed system contains seven physical parameters: jet density ρ, viscosity μ, jet length L, diameter D and velocity U at the nozzle as well as drum radius R and drum angular speed Ω. These induce four dimensionless numbers characterizing the spinning: Reynolds $\mathrm{Re} = \rho U R/\mu$ and Rossby numbers $\mathrm{Rb} = U/(\Omega R)$ as well as $\ell = L/R$ and $\epsilon = D/R$ as length ratios between jet length, nozzle diameter respectively and drum radius. For the subsequent numerical treatment, we make (1) dimensionless by scaling the quantities with the following reference values:

$$\begin{aligned}
&s_0 = r_0 = R, \quad t_0 = v_0/r_0, \quad \kappa_0 = 1/r_0,\\
&v_0 = u_0 = U, \quad \omega_0 = r_0/v_0, \quad A_0 = \pi D^2/4,\\
&n_0 = \mu A_0 v_0/r_0 = \pi\rho v_0^2 r_0^2 \epsilon^2/(4\mathrm{Re}),\\
&m_0 = \mu A_0^2 v_0/(\pi r_0^2) = \pi\rho v_0^2 r_0^3 \epsilon^4/(16\mathrm{Re})
\end{aligned}$$

The dimensionless rod model reads

$$\begin{aligned}
\mathsf{R}(\alpha)\cdot\partial_t \mathsf{r} &= \mathsf{v} - u\mathsf{e}_2\\
\partial_t \alpha &= \omega - u\kappa\\
\partial_s(u\mathsf{e}_2) &= \partial_s \mathsf{v} + \kappa \mathsf{v}^\perp + \omega \mathsf{e}_1\\
\partial_t \kappa + \partial_s(u\kappa) &= \partial_s \omega\\
\partial_t A + \partial_s(uA) &= 0\\
\partial_t(A\mathsf{v}) + \partial_s(uA\mathsf{v}) &= \frac{1}{\mathrm{Re}}(\partial_s \mathsf{n} + \kappa \mathsf{n}^\perp) - A\omega \mathsf{v}^\perp + \mathsf{k}\\
\partial_t(A^2\omega) + \partial_s(uA^2\omega) &= \frac{4}{\mathrm{Re}}\partial_s m - \frac{16}{\epsilon^2 \mathrm{Re}} n_1 + l
\end{aligned} \tag{2}$$

with outer forces, couples and material laws

$$\begin{aligned}
\mathsf{k} = -\frac{2}{\mathrm{Rb}} A\mathsf{v}^\perp + \frac{1}{\mathrm{Rb}^2} A\mathsf{R}(\alpha)\cdot\mathsf{r}, &\qquad l = \frac{1}{\mathrm{Rb}} A^2 \partial_s u,\\
n_2 = 3A\partial_s u, &\qquad m = \frac{3}{4} A^2 \partial_s \omega.
\end{aligned}$$

Boundary conditions for rotational spinning are

$$\begin{aligned}
&\mathsf{r}(0,t) = \mathsf{e}_1, \quad \alpha(0,t) = 0, \quad \kappa(0,t) = 0,\\
&A(0,t) = 1, \quad \mathsf{v}(0,t) = \mathsf{e}_2, \quad \omega(0,t) = 0,\\
&u(0,t) = 1, \quad \mathsf{n}(\ell,t) = 0, \quad m(\ell,t) = 0.
\end{aligned}$$

Appropriate initial conditions are presented later on.

3 Numerical Treatment

Finite volume schemes are well-established for the numerical solution of time-dependent partial differential equations for various applications [19]. To set up the concept for our problem we rewrite (2) in a more convenient formulation, whereby we define $\mathsf{0}_k$ as the zero vector in $\mathbb{R}^k$. We introduce the vector of unknowns ϕ. To take account of the differential-algebraic structure of the model, we additionally consider the mapping

$$\phi = (n_1, u, \mathsf{r}, \alpha, \kappa, A\mathsf{v}, A^2\omega), \qquad \mathsf{z}(\phi) = (0_2, \mathsf{r}, \alpha, \kappa, A\mathsf{v}, A^2\omega)$$

that consists of the variables having an evolution equation in (2). Finite volume schemes are based on the integral form of the governing equations that are expressed in terms of flux functions and source terms. Therefore, we summarize the constituents with respect to their physical meaning and later used numerical approximation (the upper index u, d, c indicates thereby the respective fluxes considered for up-, downwinded and central differences)

$$\begin{aligned}
\mathsf{f}^u(\phi) &= (Av_1/A, Av_2/A - u, \mathsf{0}_3, A^2\omega/A^2 - u\kappa, -uA, -uAv_1,\\
&\quad -uAv_2, -uA^2\omega)\\
\mathsf{f}^d(\phi) &= (\mathsf{0}_7, \frac{1}{\mathrm{Re}} n_1, \mathsf{0}_2)\\
\mathsf{f}^c(\phi, \partial_s \mathsf{h}(\phi)) &= (\mathsf{0}_9, \frac{3}{\mathrm{Re}}(A^2 \partial_s (A^2\omega/A^2)))\\
\mathsf{q}(\phi, \partial_s \phi) &= (\mathsf{0}_7, -\frac{3}{\mathrm{Re}} A\kappa \partial_s u, 0, \frac{1}{\mathrm{Rb}} A^2 \partial_s u)
\end{aligned}$$

with $\mathsf{h}(\phi) = (\mathsf{0}_9, A^2\omega/A^2)$. The remaining source terms are collected in $\mathsf{g}(\phi)$. Due to this dispartment where the closure relations are incorporated, the system (2) becomes

$$\partial_t \mathsf{z}(\phi) = \partial_s \mathsf{f}^c(\phi, \partial_s \mathsf{h}(\phi)) + \partial_s \mathsf{f}^u(\phi) + \partial_s \mathsf{f}^d(\phi) + \mathsf{q}(\phi, \partial_s \phi) + \mathsf{g}(\phi) \ . \tag{3}$$

Concerning the fixed jet length ℓ we introduce an equidistant space discretization of the interval $[0, \ell]$ in the form

$$\triangle s = \frac{\ell}{N}, \quad s_{(j+1)/2} = j\,\frac{\ell}{2N}, \quad j = 0, \dots, 2N.$$

The idea is now to integrate (3) over the control volumes $[s_{i-1/2}, s_{i+1/2}]$, $i = 1, \dots, N$ and to set up a differential algebraic system (DAE) in time for the cell averages ϕ_i of the unknown quantities,

$$\phi_i(t) := \frac{1}{\triangle s} \int_{s_{i-1/2}}^{s_{i+1/2}} \phi(s,t)\, ds,\ i = 1, \dots, N. \tag{4}$$

For this procedure we have to approximate all constituents in terms of $\phi_i(t)$, in particular the fluxes f^u, f^d, f^c at the points $(s_{i+1/2}, t)$, using the respective boundary conditions. We obtain

$$\begin{aligned}
&\mathsf{f}^u(\phi(s_{i+1/2},t)) \approx \mathsf{f}^u(\phi_i(t),t), && i = 1, \dots, N,\\
&\mathsf{f}^d(\phi(s_{i+1/2},t)) \approx \mathsf{f}^d(\phi_{i+1}(t),t), && i = 0, \dots, N-1,\\
&\mathsf{f}^c(\phi(s_{i+1/2},t), \partial_s \mathsf{h}(\phi(s_{i+1/2},t)) \approx \mathsf{f}^c\left(\frac{\phi_i(t)+\phi_{i+1}(t)}{2}, \frac{\mathsf{h}(\phi_{i+1}(t)) - \mathsf{h}(\phi_i(t))}{\triangle s}\right),\\
& && i = 0, \dots, N.
\end{aligned}$$

The volume integrals for $\mathsf{q}(\phi, \partial_s \phi)$ are discretized by means of

$$\frac{1}{\triangle s} \int_{s_{i-1/2}}^{s_{i+1/2}} \mathsf{q}(\phi, \partial_s \phi)\, ds \approx \mathsf{q}\left(\phi_i(t), \frac{\phi_i(t) - \phi_{i-1}(t)}{\triangle s}\right)$$

for $i = 2, \dots, N$. For $\mathsf{z}(\phi(s,t))$ and $\mathsf{g}(\phi(s,t))$ we use the cell average approximation (4). Considering the boundaries at the nozzle ($s = 0$) and at the jet end ($s = \ell$), the proposed discretizations make use of the respective boundary conditions (2) in a natural way. At the nozzle, for which we abbreviate the posed conditions by $\phi(0,t)$, we approximate

$$\mathsf{f}^c(\phi(s_{1/2},t), \partial_s \mathsf{h}(\phi(s_{1/2},t))) \approx \mathsf{f}^c\left(\phi(0,t), \frac{\mathsf{h}(\phi_1(t)) - \mathsf{h}(\phi(0,t))}{\triangle s/2}\right),$$

and at the stress (friction) free jet end we take $\mathsf{f}^c(\phi(s_{N+1/2},t), \partial_s \mathsf{h}(\phi(s_{N+1/2},t))) \approx \mathsf{0}$. Moreover we set $\mathsf{f}^u(\phi(s_{1/2},t)) \approx \mathsf{f}^u(\phi(0,t))$ as well as $\mathsf{f}^d(\phi(s_{N+1/2},t)) \approx 0$. Finally, we determine the approximation of the source term q according to the first control volume. In accordance to the general procedure we have

$$\frac{1}{\triangle s} \int_{s_{1/2}}^{s_{3/2}} \mathsf{q}(\phi(s,t))\, ds \approx \mathsf{q}\left(\phi_1(t), \frac{\phi_1(t) - \phi(0,t)}{\triangle s}\right).$$

The time integration of the DAE of index 2 for the ϕ_i is performed with a standard Runge-Kutta method for stiff problems (Radau II) [20], and the resulting nonlinear system of equations is solved with a Newton method.

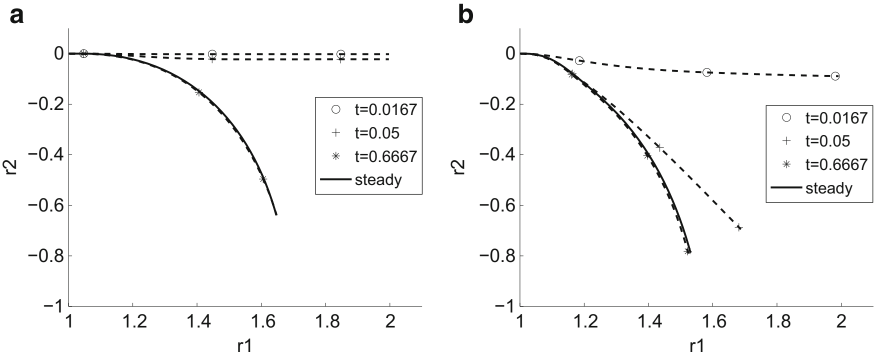


Fig. 2 Jet dynamics over time for Re = 1, $\epsilon = 0.1$, $\ell = 1$ and varying Rb. (**a**) Rb = 1. (**b**) Rb = 0.1

4 Simulation Results

In this section we present instationary rod simulations for the rotational 2d spinning set-up (Fig. 1) and compare their longtime behavior with the well-established stationary results of [10, 14]. Thereby, the length ratios are exemplary chosen as $l = 1$ and $\epsilon = 0.1$. For the non-stationary case we use the following initialization (straight jet):

$$r_1(s,0) = s + 1, \qquad r_2(s,0) = 0, \qquad \alpha(s,0) = 0,$$
$$\kappa(s,0) = 0, \qquad u(s,0) = 1, \qquad n_1(s,0) = 0,$$
$$A(s,0) = 1, \qquad A\mathsf{v}(s,0) = (0,1), \qquad A^2\omega(s,0) = 0.$$

Figure 2a, b show the evolution of the jet's centerline over time for different parameters; in particular three time points are depicted. We observe a clear convergence of the instationary solutions to the stationary ones as time increases ($t \to \infty$). The case Re = Rb = 1 of Fig. 2a lies in the parameter regime where also the string model [8] is applicable yielding similar results as $\epsilon \to 0$ in consistency to the theoretical studies. New and very promising for the future investigations is the case Re = 1, Rb = 0.1 of Fig. 2b. So far, no instationary simulations exist for this regime which is very interesting for practical applications. Industrial processes run with very fast rotations (Rb $\ll$ 1) causing strong bending and distinct non-steady effects.

5 Conclusion and Outlook

For the first time non-stationary simulations of the viscous Cosserat rod model are shown. They open the parameter range of industrial relevance—where the simpler string models fail—to systematic numerical investigations. The proposed numerical scheme allows the straightforward extension to free 3d spinning, i.e. inflow set-up with enlarging domain [17]. For future work we plan to incorporate aerodynamic forces and temperature effects. This is essential for the study and design/optimization of industrial processes.

Acknowledgements This work has been supported by the German Bundesministerium für Bildung und Forschung (BMBF), Project 03MS606.

References

1. Wong, D.C.Y., Simmons, M.J.H., Decent, S.P., Parau, E.I., King, A.C.: Break up dynamics and drop size distributions created from curved liquid jets. Int. J. Multiphase Flow **30**, 499–520 (2004)
2. Pearson, J.R.: Mechanics of Polymer Processing. Elsevier, Amsterdam (1985)
3. Klar, A., Marheineke, N., Wegener, R.: Hierarchy of mathematical models for production processes of technical textiles. Z. Ang. Math. Mech. **89**(12), 345–369 (2009)
4. Decent, S.P., King, A.C., Simmons, M.J.H., Parau, E., Wallwork, I.M., Gurney, C.J., Uddin, J.: Free jets spun from a prilling tower. J. Eng. Math. **42**, 265–282 (2009)
5. Parau, E.I., Decent, S.P., Simmons, M.J.H., Wong, D.C.Y., King, A.C.: Nonlinear viscous liquid jets from a rotating orifice. J. Eng. Math. **57**(2), 159–179 (2006)
6. Entov, V.M., Yarin, A.L.: The dynamics of thin liquid jets in air. J. Fluid Mech. **140**, 91–111 (1984)
7. Yarin, A.L.: Free Liquid Jets and Films: Hydrodynamics and Rheology. Longman, New York (1993)
8. Panda, S., Marheineke, N., Wegener, R.: Systematic derivation of an asymptotic model for the dynamics of curved viscous fibers. Math. Methods Appl. Sci. **31**(10), 1153–1173 (2008)
9. Götz, T., Klar, A., Unterreiter, A., Wegener, R.: Numerical evidence for the non-existing of stationary solutions of the equations describing rotational fiber spinning. Math. Models Methods Appl. Sci. **18**(10), 1829–1844 (2008)
10. Arne, W., Marheineke, N., Wegener, R.: Asymptotic transition from Cosserat rod to string models for curved viscous inertial jets. Math. Models Methods Appl. Sci. **21**(10), 1987–2018 (2011)
11. Hlod, A.: Curved jets of viscous fluid: interactions with a moving wall. Ph.D. thesis, Mathematics Department, TU Eindhoven (2009)
12. Hlod, A., Aarts, A.C.T., van de Ven, A.A.F., Peletier, M.A.: Three flow regimes of viscous jet falling onto a moving surface. IMA J. Appl. Math. **77**, 659–677 (2012)
13. Ribe, N.: Coiling of viscous jets. Proc. R. Soc. Lond. A **2051**, 3223–3239 (2004)
14. Arne, W., Marheineke, N., Meister, A., Wegener, R.: Numerical analysis of Cosserat rod and string models for viscous jets in rotational spinning processes. Math. Models Methods Appl. Sci. **20**(11), 1941–1965 (2010)
15. Ribe, N.M., Lister, J.R., Chiu-Webster, S.: Stability of a dragged viscous thread: onset of ‘stitching’ in a fluid-mechanical ‘sewing machine’. Phys. Fluids **18**, 124105 (2006)

16. Arne, W., Marheineke, N., Schnebele, J., Wegener, R.: Fluid-fiber-interaction in rotational spinning process of glass wool production. J. Math. Ind. **1**, 2 (2011)
17. Arne, W., Marheineke, N., Meister, A., Wegener, R.: Finite volume approach for the instationary Cosserat rod model describing spinning of viscous jets. arXiv:1207.0731 (2012)
18. Antman, S.S.: Nonlinear Problems of Elasticity. Springer, New York (2006)
19. Versteeg, H., Malalasekera, W.: An Introduction to Computational Fluid Dynamics: The Finite Volume Method. Pearson, Harlow (2007)
20. Hairer, E., Wanner, G.: Solving Ordinary Differential Equations II. Springer, Berlin (2004)

Asymptotic Modeling Framework for Fiber-Flow Interactions in a Two-Way Coupling

Thomas Martin Cibis, Nicole Marheineke, and Raimund Wegener

Abstract In this work we describe fiber-flow interactions by help of a two-way coupling approach that is based on slender-body theory and the modeling of exchange functions in terms of drag forces and heat sources. The exchange functions are incorporated in the conservation equations for linear momentum and energy with respect to flow and fibers and satisfy a generalized action-reaction principle.

1 Introduction

In the production of nonwoven fabrics thousands of long slender fibers are spun and entangled by turbulent air flows before they lay down onto a conveyor belt. There they form a nonwoven material. The quality of the fabric in key parameters such as homogeneity and thickness is largely determined by the fiber-flow interactions [1]. Therefore, the understanding of the behavior of the fibers in the flow is of great importance. Our goal is the fast computation of thousands of fibers with high stretching in air flows. The monolithic direct numerical simulation and approximations such as the immersed boundary method [2], are not applicable, since the required fine resolution is computationally too complex and expensive (too memory-demanding and time-consuming). Recently developed asymptotic modeling approaches [3–5], in contrast, provide promising results. They are based on slender-body theory and describe the interactions by external source terms

T.M. Cibis (✉) • N. Marheineke
Department Mathematik, Friedrich-Alexander-Universität Nürnberg-Erlangen, Cauerstr. 11, 91058 Erlangen, Germany
e-mail: cibis@math.fau.de; marheineke@math.fau.de

R. Wegener
Fraunhofer Institut für Techno- und Wirtschaftsmathematik, Fraunhofer Platz 1, 67663 Kaiserslautern, Germany
e-mail: wegener@itwm.fraunhofer.de

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*, Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_15,

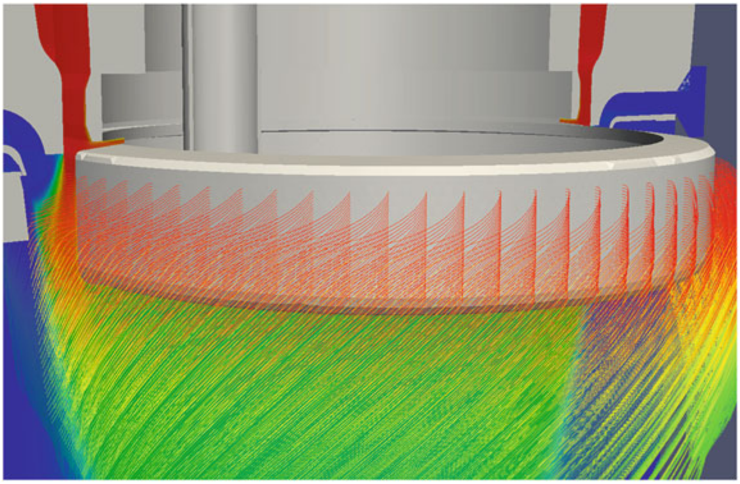


Fig. 1 Simulation of a rotational spinning process with about 30,000 fibers [5, 11]

(drag and heat source) in the conservation equations for linear momentum and energy. Observations in experiments show that the effect of the flow on a single fiber can be enormous. The "reaction" of a single thin fiber, however, is hardly perceptible. Therefore, earlier publications postulated, that its influence can (generally) be neglected, e.g. [3, 4]. This leads to the concept of one-way coupling, which takes into account the effect of the flow on the fibers and neglects the effect of the fibers on the flow. In fiber bundles or curtains, however, a "reaction" of the fibers on the flow and thus on other fibers is clearly observable. Here, it is not appropriate to neglect the "reaction" any more. But shall effect and its "reaction" be gathered, it is desirable to fulfill the action-reaction principle.

In this paper we present an asymptotic modeling framework and discuss on top of the already established one-way coupling the extension to the two-way coupling focusing on the drag force. The approach is applied to the simulation of a rotational spinning process (Fig. 1).

2 Modeling of Fibers and Air Flow

In the following we introduce the underlying models for the fibers and the flow. Thereby we choose the most general form of presentation focusing on the interaction. Specific boundary conditions, outer forces and material laws, which are required to close the system, depend on the considered application and can be straightforward included, so we neglect them here and refer for instance for elastic fibers to [1] and for viscous jets to [6].

We describe the fibers in the sense of the special Cosserat rod theory as one-dimensional objects, i.e. as curves, in the Euclidean space $\mathbb{E}^3$ [7]. In particular, a single long thin fiber is represented by a curve $\boldsymbol{r}: \mathscr{I}_{\mathscr{T}} \to \Omega \subseteq \mathbb{E}^3$, along which the orthonormal triad $(\boldsymbol{d_1}, \boldsymbol{d_2}, \boldsymbol{d_3})$ with $\boldsymbol{d_i}: \mathscr{I}_{\mathscr{T}} \to \mathbb{E}^3$ for $\boldsymbol{i} \in \{\boldsymbol{1}, \boldsymbol{2}, \boldsymbol{3}\}$ characterizes the orientation of the cross sections. The domain of definition is $\mathscr{I}_{\mathscr{T}} = \{(s,t) \in \mathbb{R}^2 : s \in [0, l(t)], t > 0\}$ with the curve-parameterization s and the time t. With the function l, it is possible to vary the length of the fiber as a function of time. For further description of the fiber the tangent $\boldsymbol{\tau}$, the linear velocity $\boldsymbol{v}$, the convection speed along the fiber u, the generalized curvature $\boldsymbol{\kappa}$ and the angular velocity $\boldsymbol{\omega}$, the line density (ρA), the inner tension forces $\boldsymbol{n}$, the external non-aerodynamic forces $\boldsymbol{f}$ (such as the gravitational force), the aerodynamic forces $\boldsymbol{f_{air}}$, the angular momentum line density $\boldsymbol{h}$, the inner torque $\boldsymbol{m}$, the external torque $\boldsymbol{l}$, the specific heat capacity c_p, the temperature T, the external non-aerodynamic heat sources q and the aerodynamic heat sources q_{air} are used. These satisfy the kinematic equations

$$\begin{aligned} \partial_t \boldsymbol{r} &= \boldsymbol{v} - u\boldsymbol{\tau}, & \partial_t \boldsymbol{d_\alpha} &= (\boldsymbol{\omega} - u\boldsymbol{\kappa}) \times \boldsymbol{d_\alpha}, \boldsymbol{\alpha} \in \{\boldsymbol{1}, \boldsymbol{2}\}, \\ \partial_s \boldsymbol{r} &= \boldsymbol{\tau}, & \partial_s \boldsymbol{d_\alpha} &= \boldsymbol{\kappa} \times \boldsymbol{d_\alpha}, \boldsymbol{\alpha} \in \{\boldsymbol{1}, \boldsymbol{2}\} \end{aligned}$$

and the dynamic equations, namely the conservation equations for mass, linear momentum, angular momentum and energy,

$$\begin{aligned} \partial_t(\rho A) + \partial_s(u(\rho A)) &= 0, \\ \partial_t((\rho A)\boldsymbol{v}) + \partial_s(u(\rho A)\boldsymbol{v}) &= \partial_s \boldsymbol{n} + \boldsymbol{f} + \boldsymbol{f}_{air}, \\ \partial_t \boldsymbol{h} + \partial_s(u\boldsymbol{h}) &= \partial_s \boldsymbol{m} + \boldsymbol{\tau} \times \boldsymbol{n} + \boldsymbol{l}, \\ c_p(\partial_t((\rho A)T) + \partial_s(u(\rho A)T)) &= q + q_{air}. \end{aligned}$$

In general, air flows are described on the space-time domain $\Omega \times \mathscr{T} \subseteq \mathbb{E}^3 \times \mathbb{R}$ by the mass density $\rho_\star$, the velocity $\boldsymbol{v}_\star$, the temperature $T_\star$, the internal stress tensor $\boldsymbol{S}_\star$, the external non-fiber-dynamic forces $\boldsymbol{f}_\star$ (e.g. the gravitational force), the fiber-dynamic forces $\boldsymbol{f}_{jets}$, the internal energy $e_\star$, the thermal conductivity $\boldsymbol{q}_\star$, the external non-fiber-dynamic heat sources $q_\star$ and the fiber-dynamic heat sources q_{jets}, that satisfy the conservation equations for mass, linear momentum and energy:

$$\begin{aligned} \partial_t \rho_\star + \nabla \cdot (\rho_\star \boldsymbol{v}_\star) &= 0, \\ \partial_t(\rho_\star \boldsymbol{v}_\star) + \nabla \cdot (\boldsymbol{v}_\star \otimes \rho_\star \boldsymbol{v}_\star) &= \nabla \cdot {\boldsymbol{S}_\star}^\top + \boldsymbol{f}_\star + \boldsymbol{f}_{jets}, \\ \partial_t(\rho_\star e_\star) + \nabla \cdot (\rho_\star e_\star \boldsymbol{v}_\star) &= \boldsymbol{S}_\star : \nabla \boldsymbol{v}_\star - \nabla \cdot \boldsymbol{q}_\star + q_\star + q_{jets}. \end{aligned}$$

To map the interactions between flow and fibers, the above equations have to be combined into a coupled system in a suitable manner.

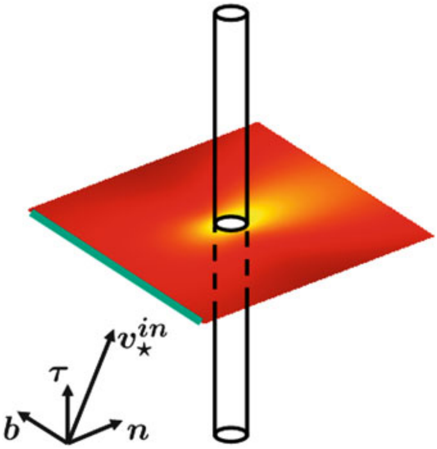


Fig. 2 Flow around a cylinder with tangent $\boldsymbol{\tau}$ and far-field (inflow) velocity $\boldsymbol{v}_\star^{in}$

In the research of fiber-flow interactions the general action-reaction principle plays, of course, a central role, which states that every force causes an equal "counter force" that acts on the cause of the drag. This means in our case, the effect of the flow on the fibers causes an opposite reaction of the fibers on the flow, and vice versa. Thereby we must move to a "generalized" action-reaction principle in terms of the flow influence and the fiber influence as external factors: In a weak formulation, the two equations

$$\int_{I_V(t)} \boldsymbol{f_{air}}(s,t)\mathrm{d}s = -\int_V \boldsymbol{f_{jet}}(\boldsymbol{x},t)\mathrm{d}\boldsymbol{x},$$

$$\int_{I_V(t)} q_{air}(s,t)\mathrm{d}s = -\int_V q_{jet}(\boldsymbol{x},t)\mathrm{d}\boldsymbol{x}$$

hold for all volumes $V \subseteq \Omega$ and $I_V(t) = \{s \in [0,l(t)]\,;\, \boldsymbol{r}(s,t) \in V\}$.

3 One-Way Coupling

In a one-way coupling the effect of the fibers on the flow with respect to the acting forces and the thermal influence is neglected: $\boldsymbol{f_{jets}} = \boldsymbol{0}$ and $q_{jets} = 0$. For the air drag and the air heat flow, the existing approaches [3], use an air drag model $\mathscr{F}^{OW}$ and a heat model $\mathscr{Q}^{OW}$ that rely on fiber data and flow data, which we express symbolically by Ψ for the fibers and $\Psi_\star$ for the flow. The common features of these models are asymptotical and experimental air drag and heat exchange studies considering far-field information of a flow around an object, see e.g. the air drag model for a curved fiber by Marheineke and Wegener [3] or the fundamental studies for a flow passing an infinitely long cylinder by Oseen, Lamb [8] and Tomotika et al. [9, 10] (see Fig. 2). However, these models are practically evaluated with local information of the flow at the fibers. Specifically, the air drag and the thermal influence are defined as follows: $\boldsymbol{f}_{\boldsymbol{air}}(s,t) = \mathscr{F}^{OW}(\Psi(s,t), \Psi_\star(\boldsymbol{r}(s,t),t))$ and $q_{air}(s,t) = \mathscr{Q}^{OW}(\Psi(s,t), \Psi_\star(\boldsymbol{r}(s,t),t))$.

Using the concept of one-way coupling many industrial problems can be satisfactorily resolved. But neglecting the "reaction" of the fibers on the flow is not adequate in cases with thousands of fibers in bundles or curtains, such as in rotational spinning (Fig. 1) where the fibers have a clear pulling effect on the flow [5, 11].

4 Two-Way Coupling

We want to remedy the deficiencies of the one-way coupling by the concept of two-way coupling, which is based on the action-reaction principle and measures the "reaction" of the fibers on the flow, too.

Under the assumption that there exist an air drag model $\boldsymbol{f}_{\boldsymbol{air}} = \mathscr{F}^{TW}(\cdot\,,\,\cdot\,,\Psi,\Psi_\star)$ and an air heat model $q_{air} = \mathscr{Q}^{TW}(\cdot\,,\,\cdot\,,\Psi,\Psi_\star)$ for the description of the acting forces and the thermal influence of the flow on the fibers in a two-way coupling, we can calculate the "counter force" $\boldsymbol{f}_{\boldsymbol{jets}}$ and the "counter heat source" q_{jets} of the fibers on the flow, so that the generalized action-reaction principle is satisfied, where δ is the Dirac delta distribution:

$$\boldsymbol{f}_{\boldsymbol{jets}}(\boldsymbol{x},t) := -\int_{I_V(t)} \mathscr{F}^{TW}(s,t,\Psi,\Psi_\star)\delta(\boldsymbol{x}-\boldsymbol{r}(s,t))\mathrm{d}s,$$

$$q_{jets}(\boldsymbol{x},t) := -\int_{I_V(t)} \mathscr{Q}^{TW}(s,t,\Psi,\Psi_\star)\delta(\boldsymbol{x}-\boldsymbol{r}(s,t))\mathrm{d}s.$$

A naive approach would like to combine the existing models for one-way coupling with the presented generalized action-reaction principle. Due to the Dirac delta shaped source terms, we obtain singularities on flow side at the fiber points. This is shown in Fig. 3 using the example of an infinitely long cylinder (cf. Fig. 2). The models, which are applied to the local flow information at the fibers cannot deal with these singularities and produce on both sides (fiber and flow) serious problems. Moreover, the used near-field information differ from the theoretically required far-field information. Therefore, this approach fails.

An approach with averaging strategies solves the problems. Thereby the naive approach is extended to the point that averaged flow data is used, concretely: The flow data $\langle\Psi_\star\rangle_R$ is averaged in an appropriate domain with typical length L relative to the fiber diameter d, so $R := L/d$. The averaged local velocity as near-field information is then mapped to the correct inflow velocity as far-field information by a function $\boldsymbol{C}$ (see Fig. 3). By such a modification of the air force model for the one-way coupling, we obtain an adequate air force model for the two-way coupling: $\mathscr{F}^{TW}(s,t,\Psi,\Psi_\star) := \mathscr{F}^{OW}(\Psi(s,t), \boldsymbol{C}(\langle\Psi_\star\rangle_R(\boldsymbol{r}(s,t),t); R))$.

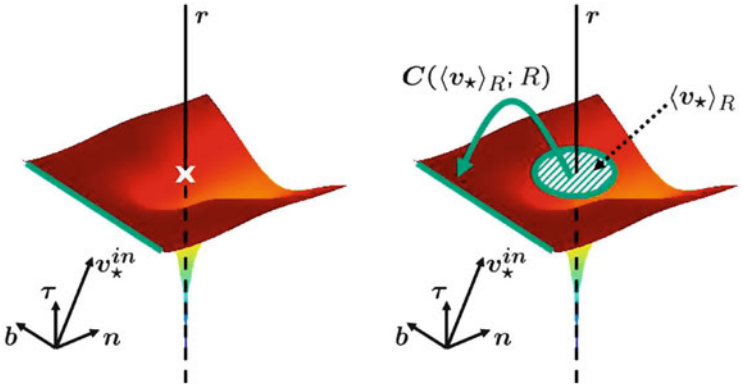


Fig. 3 Two-way coupling: arising singularities in the naive approach (*left*) and appropriate approach with averaging strategies (*right*)

Obviously wanted is this function $\boldsymbol{C}$.
We determine the function $\boldsymbol{C}$ for the case of an infinitely long cylinder and circular averaging areas. An analytical study provides at least information about the inverse function of $\boldsymbol{C}$ with respect to the orthonormal basis $(\boldsymbol{n}, \boldsymbol{b}, \boldsymbol{\tau})$ of normal, binormal and tangent, which results from the fiber's orientation $\boldsymbol{\tau}$ and the inflow velocity $\boldsymbol{v}_\star^{\boldsymbol{in}}$: If the velocity is divided in the normal, binormal and tangential components, it can be shown that the normal average velocity depends only on the normal inflow velocity. For symmetry reasons, the binormal component vanishes. The tangential average velocity depends on the normal inflow velocity and linearly on the tangential inflow velocity. That is, there exist functions $\tilde{f}$ and $\tilde{g}$ such that

$$\begin{aligned}
\langle v_{\star n}\rangle_R &= \boldsymbol{C}_n^{-1}(\boldsymbol{v}_\star^{\boldsymbol{in}}; R) = \tilde{f}(v_{\star n}^{in}; R),\\
\langle v_{\star b}\rangle_R &= \boldsymbol{C}_b^{-1}(\boldsymbol{v}_\star^{\boldsymbol{in}}; R) = 0,\\
\langle v_{\star \tau}\rangle_R &= \boldsymbol{C}_\tau^{-1}(\boldsymbol{v}_\star^{\boldsymbol{in}}; R) = \tilde{g}(v_{\star n}^{in}; R) v_{\star \tau}^{in}.
\end{aligned}$$

This is exploited to determine the function $\boldsymbol{C}$, where only $f = \tilde{f}^{-1}$ and $h := \tilde{g} \circ \tilde{f}^{-1}$ have to be found as functions of the normal inflow velocity:

$$\begin{aligned}
v_{\star n}^{in} &= \boldsymbol{C}_n(\langle \boldsymbol{v}_\star\rangle_R; R) = f(\langle v_{\star n}\rangle_R; R)\\
v_{\star b}^{in} &= \boldsymbol{C}_b(\langle \boldsymbol{v}_\star\rangle_R; R) = 0\\
v_{\star \tau}^{in} &= \boldsymbol{C}_\tau(\langle \boldsymbol{v}_\star\rangle_R; R) = \langle v_{\star \tau}\rangle_R / h(\langle v_{\star n}\rangle_R; R)
\end{aligned}$$

Considering simulation results, function classes are determined for the unknown quantities:

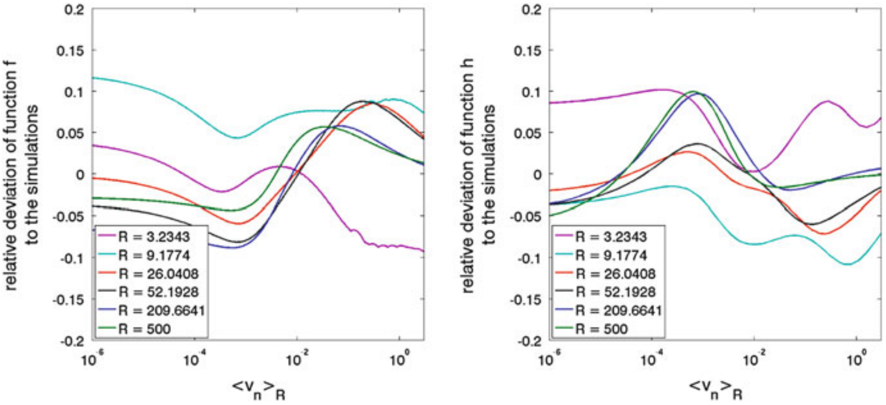


Fig. 4 Relative deviation of the approximated functions f (*left*) and h (*right*) to simulation results

$$f(\langle v_{\star n}\rangle_R; R) = \langle v_{\star n}\rangle_R \left(\frac{\alpha_1 R^{\alpha_2}}{\alpha_3 \langle v_{\star n}\rangle_R^{\alpha_4} + 1} + 1\right),$$

$$h(\langle v_{\star n}\rangle_R; R) = 1 - \frac{\beta_1 + \beta_2 \log(R)}{\beta_3 R^{\beta_4} \langle v_{\star n}\rangle_R^{\beta_5} + 1}.$$

The required parameters are determined via parameter identification by minimizing an relative ℓ^2 error:

$$\boldsymbol{\alpha} = (9.7785, -0.6130, 7.1976, 0.4977),$$

$$\boldsymbol{\beta} = (0.9585, -0.0942, 0.4230, 0.7778, 0.5705).$$

The deviation of the approximated functions f and h to the simulation results is at most about 10 % as it is shown in Fig. 4 for different radii. This is remarkably good because, for example, the range of function f extends over several orders of magnitude. Also, the limit behavior to 0 and ∞ is mapped correctly by the functions.

The modeled function $\boldsymbol{C}$ is generally applied to the air drag $\mathscr{F}^{TW}$. Particularly for small averaging areas and low relative fiber-flow velocities, $\boldsymbol{C}$ clearly differs from the identity map $\boldsymbol{I}$. For large averaging areas or high velocities, however, the difference is not significant such that we can simplify $\boldsymbol{C} = \boldsymbol{I}$. This is also shown in our studies on the approach with averaging strategies in the rotational spinning process of [5, 11] (Fig. 1). There, the flow equations are solved by a finite volume method. Investigating whether the grid cells are appropriate as averaging areas, Fig. 5 visualizes the results. Critically small averaging areas are directly at the spinning nozzles. However, because of the fortunate fact that there the velocities are high, the deviation of function $\boldsymbol{C}$ to the identity map $\boldsymbol{I}$ is acceptable in magnitude. The $\mathscr{L}^2(\Omega)$ error is only about 5.4 %. This justifies the simplification $\boldsymbol{C} = \boldsymbol{I}$ that has been applied in [5, 11].

Ongoing work deals with a transfer of the strategy to the heat exchange.

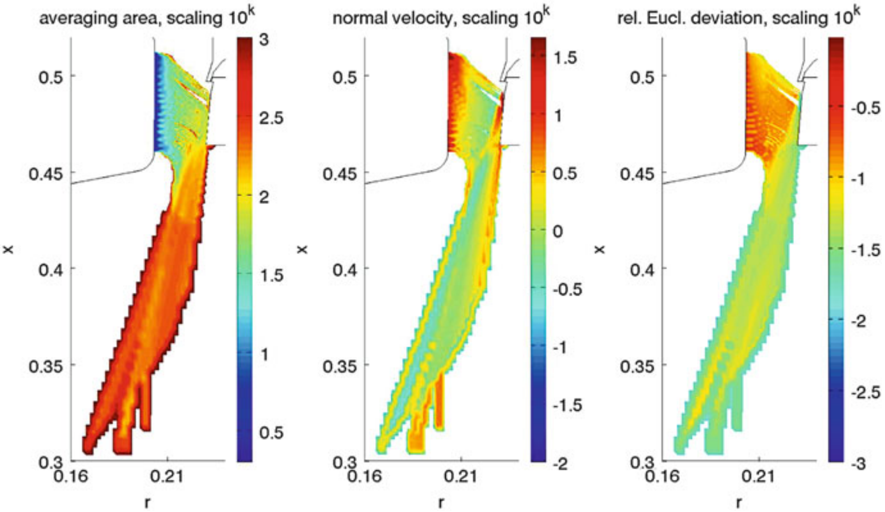


Fig. 5 *From left to right*: R, $\langle v_n - v_{\star n}\rangle_R$, $\|\boldsymbol{C} - \boldsymbol{I}\|_2$ in process of Fig. 1 (2d cut due to rotational symmetry)

Acknowledgements This work has been supported by Bundesministerium für Bildung und Forschung, Verbundprojekt ProFil, 03MS606.

References

1. Klar, A., Marheineke, N., Wegener, R.: Hierarchy of mathematical models for production processes of technical textiles. Z. Ang. Math. Mech. **89**(12), 941–961 (2009)
2. Mark, A.: A novel immersed-boundary method for multiple moving and interacting bodies. Ph.D. thesis, Chalmers University of Technology (2007)
3. Marheineke, N., Wegener, R.: Modeling and application of a stochastic drag for fibers in turbulent flows. Int. J. Multiphase Flow **37**, 136–148 (2011)
4. Marheineke, N., Wegener, R.: Fiber dynamics in turbulent flows: general modeling framework. SIAM J. Appl. Math. **66**(5), 1703–1726 (2006)
5. Marheineke, N., Liljo, J., Mohring, J., Schnebele, J., Wegener, R.: Multiphysics and multimethods problem of rotational glass fiber melt-spinning. Int. J. Numer. Anal. Model. B **3**(3), 330–344 (2012)
6. Arne, W., Marheineke, N., Wegener, R.: Asymptotic transition from Cosserat rod to string models for curved viscous inertial jets. Math. Models Methods Appl. Sci. **21**(10), 1987–2018 (2011)
7. Antman, S.: Nonlinear Problems of Elasticity. Springer, New York (2004)
8. Lamb, H.: On the uniform motion of a sphere through a viscous field. Philos. Mag. **6**, 113–121 (1911)
9. Tomotika, S., Aoi, T.: An expansion formula for the drag on a circular cylinder moving through a viscous fluid at small Reynolds number. Q. J. Mech. Appl. Math. **4**, 401–406 (1951)

10. Tomotika, S., Aoi, T., Yosinobu, H.: On the forces acting on a circular cylinder set obliquely in a uniform stream at low values of Reynolds number. Proc. R. Soc. Lond. A **219**, 233–244 (1953)
11. Arne, W., Marheineke, N., Schnebele, J., Wegener, R.: Fluid-fiber-interactions in rotational spinning process of glass wool manufacturing. J. Math. Ind. **1**, 2 (2011)

Efficient Simulation of Random Fields for Fiber-Fluid Interactions in Isotropic Turbulence

Florian Hübsch, Nicole Marheineke, and Raimund Wegener

Abstract In some processes for spinning synthetic fibers the filaments are exposed to highly turbulent flows to achieve a high degree of stretching. The quality of the resulting fabric is thus determined essentially by the turbulent fiber-fluid interactions. Due to the required fine resolution, direct numerical simulations fail. Therefore we model the flow fluctuations as random field in $\mathbb{R}^4$ on top of a k-ϵ turbulence description and describe the interactions in the context of slender-body theory as one-way-coupling with a corresponding stochastic aerodynamic drag force on the fibers. Hereby we exploit the special covariance structure of the random field, namely isotropy, homogeneity and decoupling of space and time. In this work we will focus on the construction and efficient simulation of the turbulent fluctuations assuming constant flow parameters and give an outlook on applications.

1 Introduction

A modeling framework for the dynamics of slender fibers in turbulent flows was developed in [6] and further extended in [7]. It is based on a k-ϵ description of the turbulent flow, considering the actual velocity as sum of a mean and a fluctuating part $\mathbf{u} = \overline{\mathbf{u}} + \mathbf{u}'$. Whereas $\overline{\mathbf{u}}$ is computed explicitly, the fluctuations $\mathbf{u}'$ are only

F. Hübsch (✉)
TU Kaiserslautern, Paul-Ehrlich-Straße 31, 67663 Kaiserslautern, Germany
e-mail: huebsch@itwm.fhg.de

N. Marheineke
FAU Erlangen-Nürnberg, Cauerstr. 11, 91508 Erlangen, Germany
e-mail: marheineke@am.uni-erlangen.de

R. Wegener
Fraunhofer ITWM, Fraunhofer-Platz 1, 67663 Kaiserslautern, Germany
e-mail: wegener@itwm.fhg.de

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_16,

characterized by the kinetic turbulent energy $\mathrm{k} = \frac{1}{2}\mathbb{E}(\langle \mathbf{u}', \mathbf{u}' \rangle)$ and the dissipation $\epsilon = \nu\mathbb{E}(\|\nabla\mathbf{u}'\|_F^2)$ with flow viscosity ν, expectation $\mathbb{E}$ and Frobenius norm $\|\cdot\|_F$. The fluctuations are modeled as centered homogeneous Gaussian random fields whose covariance structure obeys Kolmogorov's isotropy assumptions and the requirements of the k-ϵ model and can be expressed by two scalar-valued functions, i.e. energy spectrum and temporal correlation, see [7] for details. In this work we deal with the efficient sampling of the fluctuations, assuming constant flow parameters $\overline{\mathbf{u}}, \mathrm{k}, \epsilon, \nu$ for simplicity.

Throughout this paper we use bold-faced letters for vector- and matrix-valued quantities. By $\langle .,. \rangle$ and $\|\cdot\|$ we denote the Euclidean inner product and norm, respectively. Moreover we distinguish between dimensional and dimensionless quantities, writing the last in a kursiv style. We make the fluctuations $\mathbf{u}'$ dimensionless using the typical turbulent length $\mathrm{k}^{3/2}/\epsilon$ and time k/ϵ

$$\mathbf{u}'(\mathbf{x}, \mathrm{t}) = \mathrm{k}^{1/2}\boldsymbol{u'}\left(\frac{\epsilon}{\mathrm{k}^{3/2}}\mathbf{x}, \frac{\epsilon}{\mathrm{k}}\mathrm{t}; \frac{\epsilon}{\mathrm{k}^2}\nu\right)$$

with

$$\mathbf{x} = \frac{\mathrm{k}^{3/2}}{\epsilon}\boldsymbol{x} \quad \mathrm{t} = \frac{\mathrm{k}}{\epsilon}t, \quad \nu = \frac{\mathrm{k}^2}{\epsilon}\zeta.$$

The dimensionless viscosity ζ enters the model via the consistency with the k-ϵ-description, see [6, 7].

2 Construction of Velocity Fluctuation Field

There are some possibilities for the simulation of homogeneous and isotropic Gaussian vector random fields with given covariance function or given spectral function, respectively, see, e.g., [4]. We construct the random fields so that they can be simulated efficiently and evaluated at a given point on demand. Our starting point is the centered, $\mathbb{R}^3$-valued Gaussian random field $\boldsymbol{u'} = (\boldsymbol{u'}(\boldsymbol{x}, t))_{(\boldsymbol{x},t)\in\mathbb{R}^4}$ with covariance function

$$\mathbb{E}\left(\boldsymbol{u'}(\boldsymbol{x_1}, t_1) \otimes \boldsymbol{u'}(\boldsymbol{x_2}, t_2)\right) = \boldsymbol{\gamma}(\boldsymbol{x_1} - \boldsymbol{x_2} - \overline{\boldsymbol{u}}(t_1 - t_2)) \exp\left(-\frac{(t_1 - t_2)^2}{2\tau_l^2}\right)$$

with $\tau_l = 0.212$ and dimensionless mean velocity $\overline{\boldsymbol{u}} = \overline{\mathbf{u}}/\mathrm{k}^{1/2}$. Here $\boldsymbol{\gamma} : \mathbb{R}^3 \to \mathbb{R}^{3\times 3}$ is implicitly given by its Fourier transform

$$\boldsymbol{s_\gamma}(\boldsymbol{\kappa}) = \frac{1}{8\pi^3}\int_{\mathbb{R}^3} \exp(-i\langle\boldsymbol{\kappa}, \boldsymbol{x}\rangle)\boldsymbol{\gamma}(\boldsymbol{x})\, d\boldsymbol{x} = \frac{1}{4\pi}\frac{E(\|\boldsymbol{\kappa}\|)}{\|\boldsymbol{\kappa}\|^2}\left(\boldsymbol{I} - \frac{1}{\|\boldsymbol{\kappa}\|^2}\boldsymbol{\kappa}\otimes\boldsymbol{\kappa}\right) \tag{1}$$

with identity $\boldsymbol{I}$ and energy spectrum $E : \mathbb{R}_0^+ \to \mathbb{R}_0^+$. An appropriate choice of E ensures the almost sure differentiability of the realizations of $\boldsymbol{u}'$. In the following we use the model of [7] depending on the dimensionless viscosity $\zeta = \frac{\epsilon\nu}{\mathrm{k}^2}$. In order to split up the covariance function into spatial and time parameter, we define a new random field $\boldsymbol{\eta} = (\boldsymbol{\eta}(\boldsymbol{x}, t))_{(\boldsymbol{x},t)\in\mathbb{R}^4}$ by

$$\boldsymbol{\eta}(\boldsymbol{x}, t) = \boldsymbol{u}'(\boldsymbol{x} + \overline{\boldsymbol{u}}t, t).$$

Then, the so defined field has the covariance function

$$\boldsymbol{K}_{\boldsymbol{\eta}}(\boldsymbol{x_1}, t_1, \boldsymbol{x_2}, t_2) = \mathbb{E}(\boldsymbol{\eta}(\boldsymbol{x_1}, t_1) \otimes \boldsymbol{\eta}(\boldsymbol{x_2}, t_2)) = \boldsymbol{\gamma}(\boldsymbol{x_1} - \boldsymbol{x_2}) \exp\left(-\frac{(t_1 - t_2)^2}{2\tau_l^2}\right).$$

As we can regain $\boldsymbol{u}'$ easily from $\boldsymbol{\eta}$ via

$$\boldsymbol{u}'(\boldsymbol{x}, t) = \boldsymbol{\eta}(\boldsymbol{x} - \overline{\boldsymbol{u}}t, t)$$

we focus on the construction of $\boldsymbol{\eta}$.

2.1 Construction of η

In the following we assume the existence of all occurring stochastic processes and random fields as we construct them later on. Let $\psi = (\psi(t))_{t\in\mathbb{R}}$ be a centered stochastic process with covariance function

$$\mathbb{E}(\psi(t_1)\psi(t_2)) = \exp\left(-\frac{(t_1 - t_2)^2}{2\tau_l^2}\right)$$

and $\boldsymbol{\xi} = (\boldsymbol{\xi}(\boldsymbol{x}))_{\boldsymbol{x}\in\mathbb{R}^3}$ a centered, $\mathbb{R}^3$-valued random field with covariance function

$$\mathbb{E}\left(\boldsymbol{\xi}(\boldsymbol{x_1}) \otimes \boldsymbol{\xi}(\boldsymbol{x_2})\right) = \boldsymbol{\gamma}(\boldsymbol{x_1} - \boldsymbol{x_2}).$$

Let us further assume that ψ and $\boldsymbol{\xi}$ are stochastically independent. If we define a random field $\tilde{\boldsymbol{\eta}}$ by

$$\tilde{\boldsymbol{\eta}}(\boldsymbol{x}, t) = \boldsymbol{\xi}(\boldsymbol{x})\psi(t)$$

then $\tilde{\boldsymbol{\eta}}$ has the desired covariance function $\boldsymbol{K}_{\boldsymbol{\eta}}$. As we are interested in a Gaussian field, we consider for $M \in \mathbb{N}$ random fields $\tilde{\boldsymbol{\eta}}_M = (\tilde{\boldsymbol{\eta}}_M(\boldsymbol{x}, t))_{(\boldsymbol{x},t)\in\mathbb{R}^4}$ of the form

$$\tilde{\boldsymbol{\eta}}_M(\boldsymbol{x}, t) = \frac{1}{\sqrt{M}} \sum_{l=1}^{M} \tilde{\boldsymbol{\eta}}^{(l)}(\boldsymbol{x}, t)$$

in which $\tilde{\boldsymbol{\eta}}^{(1)}, \dots, \tilde{\boldsymbol{\eta}}^{(M)}$ are independent copies of $\tilde{\boldsymbol{\eta}}$. The central limit theorem ensures the convergence in distribution

$$\tilde{\boldsymbol{\eta}}_M(\boldsymbol{x}, t) \xrightarrow{d} \mathscr{N}\left(0, \boldsymbol{K}_\eta(\boldsymbol{x}, t, \boldsymbol{x}, t)\right) = \mathscr{N}\left(0, \frac{2}{3}\boldsymbol{I}\right)$$

for every $(\boldsymbol{x}, t) \in \mathbb{R}^4$ as M tends to infinity. So in order to construct $\tilde{\boldsymbol{\eta}}$ respectively $\tilde{\boldsymbol{\eta}}_M$ we focus on the construction of $\boldsymbol{\xi}$ and ψ.

2.2 Spatial Field $\boldsymbol{\xi}$

In this subsection we exploit the special structure of the spectral function $\boldsymbol{s}_{\boldsymbol{\gamma}}$ (1) of the spatial field $\boldsymbol{\xi}$. Let $\boldsymbol{w} = (\boldsymbol{w}(t))_{t\in\mathbb{R}}$ be a centered, homogeneous and $\mathbb{R}^3$-valued stochastic process with spectral function $\boldsymbol{s}_{\boldsymbol{w}}(\kappa) = E(|\kappa|)/2 \cdot \boldsymbol{I}$, i.e. its components w_i, $i \in \{1, 2, 3\}$, are uncorrelated processes with the same spectral function $s_{w_i}(\kappa) = E(|\kappa|)/2$. Moreover, let $\boldsymbol{z}$ be a uniformly distributed random vector on the unit sphere $S^2 = \{\boldsymbol{x} \in \mathbb{R}^3 : \|\boldsymbol{x}\| = 1\}$. Then, under the assumption that $\boldsymbol{w}$ and $\boldsymbol{z}$ are independent, the random field $\boldsymbol{\xi}$ that is defined by

$$\boldsymbol{\xi}(\boldsymbol{x}) = (\boldsymbol{I} - \boldsymbol{z} \otimes \boldsymbol{z})\boldsymbol{w}(\langle \boldsymbol{x}, \boldsymbol{z}\rangle)$$

has the spectral function $\boldsymbol{s}_{\boldsymbol{\gamma}}$ given by (1) and hence the desired covariance function $\boldsymbol{\gamma}$, [2, 5]. As the components w_i are uncorrelated it is sufficient to focus on the construction of one component w_i. This can be done in the following manner [4]: As $E(\kappa) \geq 0$ for all $\kappa \geq 0$ and $\int_{\mathbb{R}} s_{w_i}(\kappa)\, \mathrm{d}\kappa = \int_0^\infty E(\kappa)\, \mathrm{d}\kappa = 1$, the function s_{w_i} is a continuous probability density on $\mathbb{R}$. Choosing a random variable R with this probability density and two standard normally distributed random variables X and Y such that X, Y, R are stochastically independent, the $\mathbb{C}$-valued process $(\tilde{w}(t))_{t\in\mathbb{R}}$ defined by

$$\tilde{w}(t) = Z \exp(iRt), \quad Z = X + iY,$$

has the covariance function

$$\mathbb{E}\left(\tilde{w}(t_1)\, \overline{\tilde{w}}(t_2)\right) = \mathbb{E}\left(\exp(iR(t_1 - t_2))\right) \mathbb{E}(Z\overline{Z}) = 2\int_{\mathbb{R}} \exp(i\kappa(t_1 - t_2))\, s_{w_i}(\kappa)\, \mathrm{d}\kappa$$

and hence the spectral function $2s_{w_i}$. By taking its real or imaginary part we obtain a $\mathbb{R}$-valued process with the desired spectral function s_{w_i}. The so defined process $w_i = \mathrm{Re}(\tilde{w})$ respectively $w_i = \mathrm{Im}(\tilde{w})$ has almost surely differentiable realizations and hence the same holds for $\boldsymbol{\xi}$.

2.3 Time Process ψ

The time process ψ can be constructed with the same methods as for w_i. We define a new process $\tilde{\psi}$ by $\tilde{\psi}(t) = \psi(\tau_l t)$ having the covariance function

$$\mathbb{E}\left(\tilde{\psi}(t_1)\tilde{\psi}(t_2)\right) = \exp\left(-\frac{(t_1 - t_2)^2}{2}\right)$$

and hence the spectral function

$$s_{\tilde{\psi}}(\omega) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\omega^2}{2}\right).$$

As $s_{\tilde{\psi}}$ is the probability density of the standard normal distribution, we take three independent, standard normally distributed random variables R, X, Y and set $\tilde{\psi}(t) = Z\exp(iRt)$ with $Z = X + iY$. Then, the process $\psi(\cdot) = \mathrm{Re}(\tilde{\psi}(\cdot/\tau_l))$ or $\psi(\cdot) = \mathrm{Im}(\tilde{\psi}(\cdot/\tau_l))$ has the desired covariance function and almost surely differentiable realizations.

3 ODE Model for Fiber Spinning Due to Turbulence

In [7] the dynamics of a slender fiber in a turbulent flow is modeled by help of a stochastic drag force in an one-way coupling. The dimensionless force $\boldsymbol{f}$ depends on the relative velocity between flow and fiber and on the fiber tangent. Instead of the complex PDE fiber model that contains inner stresses we use here a system of first order ODEs in time for fiber position $\mathbf{r}$, velocity $\mathbf{v}$ and elongation e. In $\boldsymbol{f}$ we approximate the tangent (space derivative) by the direction of the fiber velocity $\mathbf{v}/\|\mathbf{v}\|$. Moreover, we propose an evolution equation for the elongation that is motivated from the stationary situation where $e = \|\mathbf{v}\|/\mathrm{v}_0$ with initial velocity v_0 (e.g. at the spinning nozzle). The resulting model (2) describes the path and behavior of a single fiber point whose motion is exclusively driven by a turbulent flow.

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{dt}}\mathbf{r} = \mathbf{v} \qquad (2)\\
&\frac{\mathrm{d}}{\mathrm{dt}}\mathbf{v} = e^{3/2}\,\mathrm{a}\,\boldsymbol{f}\left(\frac{\mathbf{v}}{\|\mathbf{v}\|}, \frac{1}{\sqrt{e}}\frac{\mathbf{u}(\mathbf{r},\mathrm{t}) - \mathbf{v}}{\mathrm{b}}\right)\\
&\frac{\mathrm{d}}{\mathrm{dt}}e = \frac{1}{\mathrm{v}_0}e^{3/2}\,\mathrm{a}\left\|\boldsymbol{f}\left(\frac{\mathbf{v}}{\|\mathbf{v}\|}, \frac{1}{\sqrt{e}}\frac{\mathbf{u}(\mathbf{r},\mathrm{t}) - \mathbf{v}}{\mathrm{b}}\right)\right\|\\
&\mathbf{r}(0) = \mathbf{r_0},\ \mathbf{v}(0) = \mathrm{v}_0\boldsymbol{\tau}_\mathbf{0},\ e(0) = 1,
\end{aligned}$$

with

$$\mathrm{a} = \frac{4}{\pi}\frac{\rho\nu^2}{\rho_\mathrm{F}\mathrm{d}_0^3}, \qquad \mathrm{b} = \frac{\nu}{\mathrm{d}_0},$$

containing fiber and flow informations (fiber density ρ_F, initial diameter d_0, flow density ρ and viscosity ν).

4 Simulation Results

The ODE model (2) for the fiber dynamics allows for space- and time-dependent flows. But so far the construction of our random field $\boldsymbol{u'}$ expects constant flow parameters, a generalization is in work. For the forthcoming simulations we use typical values of a spinning process. We consider a flow field that is directed vertically downwards with

$$\bar{\mathbf{u}} = -10^2\boldsymbol{e_1}\left[\frac{m}{s}\right], \quad \mathrm{k} = 10^3\left[\frac{m^2}{s^2}\right], \quad \epsilon = 5\cdot 10^6\left[\frac{m^2}{s^3}\right],$$
$$\nu = 1.5\cdot 10^{-5}\left[\frac{m^2}{s}\right], \quad \rho = 1\left[\frac{kg}{m^3}\right], \quad M = 50,$$

with $\boldsymbol{e_1} = [1,0,0]^T$. The fiber is initialized with

$$\mathbf{r_0} = \mathbf{0}\,[m], \quad \boldsymbol{\tau_0} = -\boldsymbol{e_1}, \quad \mathrm{v}_0 = 10^{-2}\left[\frac{m}{s}\right],$$
$$\mathrm{d}_0 = 4\cdot 10^{-4}\,[m], \quad \rho_F = 7.33\cdot 10^2\left[\frac{kg}{m^3}\right],$$

and simulated for the time interval $[0,\mathrm{T}]$ with $\mathrm{T} = 2\cdot 10^{-3}$ [s]. Figure 1 shows the trajectory of the fiber point. To get an impression of the impact of the turbulent drag force we study the elongations. Figure 2 shows the estimated probability density of $e(\mathrm{T})$ for a Monte-Carlo simulation with 1,000 replications. We get a mean of $2.4\cdot 10^5$, in comparison the result without turbulence is approximately 10^4. The simulation result raises hope that the proposed strategy is capable of predicting the large elongations that are observed in turbulent spinning processes (like melt-blown) in experiments. So far, numerical simulations fail but they neglect the fluctuations.

Further work [3] deals with the extension of the random field sampling to realistic settings with space- and time-dependent flow parameters. Moreover, we plan to introduce an appropriate PDE-Cosserat model for the spinning of a viscous jet with inner strains, e.g. [1].

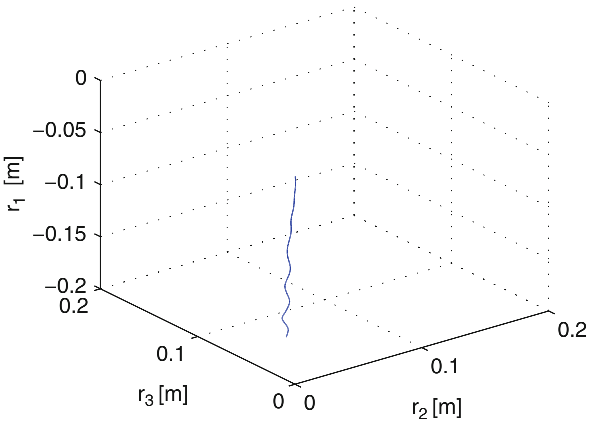


Fig. 1 Trajectory of fiber point

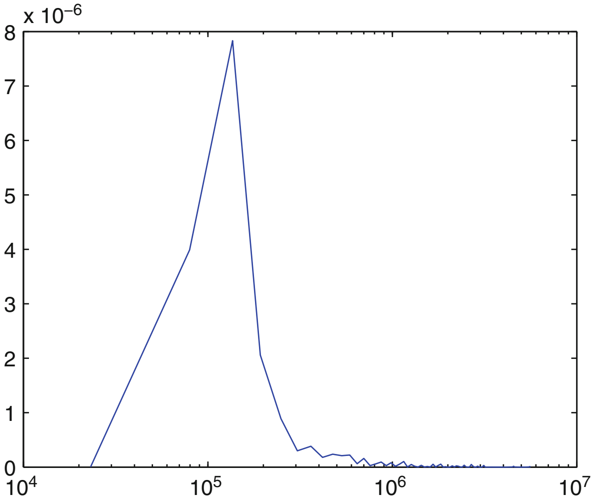


Fig. 2 Estimated probability density of $e(\mathrm{T})$

Acknowledgements Author F.H. acknowledges the financial support by the Fraunhofer Innovationszentrum Applied System Modeling, Kaiserslautern.

References

1. Arne, W., Marheineke, N., Wegener, R.: Asymptotic transition of Cosserat rod to string models for curved viscous inertial jets. Math. Models Methods Appl. Sci. **21**(10), 1987–2018 (2011)
2. Elliott, F., Majda, A.: A new algorithm with plane waves and wavelets for random velocity fields with many spatial scales. J. Comput. Phys. **117**, 146–162 (1995)
3. Hübsch, F., Marheineke, N., Ritter, K., Wegener, R.: Random field sampling for a simplified model of melt-blowing considering turbulent velocity fluctuations. J. Stat. Phys. **150**(6), 1115–1137 (2013)

4. Kurbanmuradov, O., Sabelfeld, K.: Stochastic spectral and fourier-wavelet methods for vector gaussian random fields. Monte Carlo Methods Appl. **12**, 395–446 (2006)
5. Majda, A.: Random shearing direction models for isotropic turbulent diffusion. J. Stat. Phys. **75**(5–6), 1153–1165 (1994)
6. Marheineke, N., Wegener, R.: Fiber dynamics in turbulent flows: general modeling framework. SIAM J. Appl. Math. **66**, 1703–1726 (2006)
7. Marheineke, N., Wegener, R.: Modeling and application of a stochastic drag for fibers in turbulent flows. Int. J. Multiphase Flow **37**, 136–148 (2011)

On Stability of a Concentrated Fiber Suspension Flow

Uldis Strautins

Abstract Linear stability analysis of a fiber suspension flow in a channel domain is performed using a modified Folgar-Tucker equation. Two kinds of potential instability are identified: one is associated with overcritical Reynolds number and another is associated with certain perturbations in fiber orientation field and is present for any Reynolds numbers. The second type of instability leads to initially growing transient perturbations in the microstructure. It is shown that both types of instability lead to instability of the bulk velocity field. As for the perturbed Orr-Sommerfeld eigenvalues, the presence of fibers increases the stability region; the stability region increases with growing C_i and decreases with growing S_0 in the modified Folgar-Tucker model.

1 Fiber Suspension Flows

Injection molding and compression molding are efficient manufacturing techniques for processing short fiber reinforced thermosoftening plastics. The material is heated and mixed in a barrel and injected into a mold as a fiber suspension. Upon solidification of the matrix a part is obtained with anisotropic material properties, which depend strongly on the microstructure characterized by fiber orientation and concentration. The microstructure is coupled to the bulk flow of the suspension and can vary considerably in a typical sample. In order to obtain a part with prescribed properties, the mold has to be designed appropriately. Simulations play a major role in the design process [8].

Stability of the flow can considerably influence the outcome, e.g., the surface roughness of the part. Most often one studies the hydrodynamical instability, which

U. Strautins (✉)
Department of Physics and Mathematics, University of Latvia, Riga, Latvia
e-mail: uldis.strautins@lu.lv

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_17,

depends on the Reynolds number and can lead to turbulence. The transition from laminar to turbulent regime can be conveniently studied by examining stability of simple flows wrt. small perturbations such that the governing equations can be linearized around the base flow. Most theoretical studies show that the presence of fibers stabilizes the flow, delaying the onset of turbulence. Dean's flow has been considered in [5] assuming that all fibers orient tangentially to the streamlines. More recent publications use the Folgar-Tucker equation for evolution of the microstructure. Stability analysis of Taylor-Couette flow has been reported in [3]. Pressure driven channel flow using the natural closure approximation has been considered in [10]. The implications of non-linear effects due to perturbations of finite magnitude was the object of study in [9].

Another, less studied type of instability is independent on the Reynolds number; it is associated with certain perturbations of the fiber orientation field leading to prominent transient behavior characterized by an initial growth of the perturbations in both the microstructure and bulk velocity, see e.g. [4]. The main goal of this study is to explain this behaviour using a linear stability analysis and to better understand the stability results demonstrated in the works cited above. We have chosen a modified Folgar-Tucker equation derived for concentrated suspension flows [4]; the classical Folgar-Tucker model is a special case thereof. Our results suggest that for highly concentrated suspensions (high N_p) the two types of instability cannot be clearly separated, so both should be considered in the design process.

We illustrate this phenomenon for the planar Poiseulle flow, i.e., pressure gradient driven flow through a domain between two stationary walls. Linear stability analysis showing the relation between the wave number of a perturbation to its growth rate leads to a generalized eigenvalue problem (GEP) for a differential algebraic system to be solved numerically for a range of parameter values—Reynolds number, wave number of the perturbation and the model parameters C_I and U_0.

2 Models

In hydrodynamic limit fiber suspensions are modeled as non-Newtonian fluids. The conservation of mass and momentum equations are

$$\nabla \cdot u = 0,$$

$$\frac{\partial u}{\partial t} + u \cdot \nabla u = -\rho^{-1} \nabla p + \mathrm{Re}^{-1} \triangle u + \nabla \cdot \tau,$$

where u is bulk velocity, p is pressure, ρ is density, Re is the Reynolds number (based on the viscosity of the matrix), and τ is the extra stress associated with the presence of fibers. According to the Dinh-Armstrong model, it is given by

$$\tau = N_p \nabla u : a^{(4)},$$

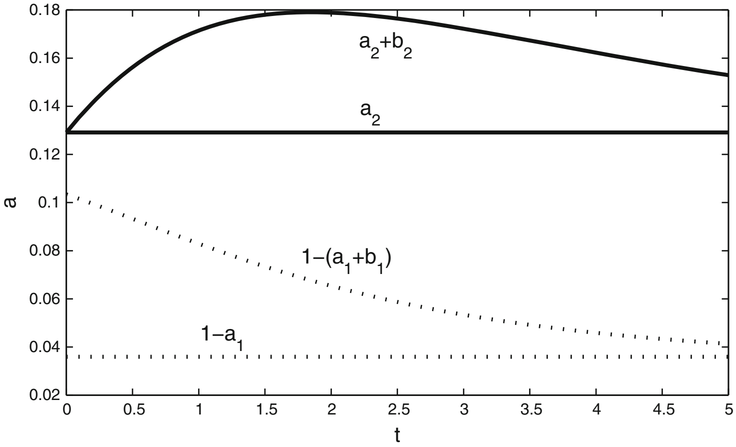


Fig. 1 Solution of (1) for a perturbation $a + b$ of the stationary state $a = (a_1, a_2)$. The flow is a shear flow $\frac{\partial u_1}{\partial y} = 1$, parameters are $C_i = 0.01$ and $S_0 = 0$ (Folgar-Tucker model). Note that the perturbation of the off-diagonal component (*solid line*) initially grows

where N_p is the dimensionless particle number (depends on concentration) and $a^{(4)}$ is the fourth-order orientation tensor.

A fiber suspension is called concentrated if $n_f l_f^2 d_f > 1$, where n_f denotes fiber number density, l_f and d_f is the length and diameter of a fiber, see [1, 6]. The tensor $a^{(4)}$ is approximated in terms of the second order orientation tensor $a^{(2)} = \begin{pmatrix} a_1 & a_2 \\ a_2 & 1-a_1 \end{pmatrix}$ by means of a closure approximation. The following model has been introduced in [4] for evolution of $a^{(2)}$:

$$\begin{aligned} \frac{D}{Dt}a^{(2)} &= a^{(2)} \cdot M + M^\top \cdot a^{(2)} - (M + M^\top) : a^{(4)} \\ &+ \dot{\gamma} \left\{ C_i(I - 3a^{(2)}) + S_0(a^{(2)} \cdot a^{(2)} - a^{(2)} : a^{(4)}) \right\}. \end{aligned} \tag{1}$$

Here $M = \frac{\lambda+1}{2}\nabla u + \frac{\lambda-1}{2}\nabla^\top u$ is effective velocity gradient, $\lambda = \frac{r_a^2-1}{r_a^2+1}$, aspect ratio $r_a = l_f/d_f$, $\dot{\gamma}$ is the scalar shear rate, C_i is orientational diffusivity constant and S_0 is a constant quantifying the excluded volume effect in the suspension.

The present work was carried out using the simple quadratic closure approximation $a^{(4)}_{\text{quad}} = a^{(2)} \otimes a^{(2)}$. We note that this approach can be carried out for arbitrary closure approximations with similar results.

We close this section by demonstrating that for a shear flow, small perturbations around the stationary point of (1) can grow, see Fig. 1. The off-diagonal component a_2 is the most unstable, and exactly this component contributes to the extra stress τ for a shear flow: $\tau = N_p a_2 \frac{\partial u_1}{\partial y} a^{(2)}$.

3 Linear Stability Analysis

In this work we consider the 2D case in which a stream-function can be used. This approach is justified for Newtonian fluids by the Squire theorem, however, most non-Newtonian fluids do require treatment of the full 3D stability problem. The extension to 3D is straight forward, see e.g. [3] for the pipe flow.

Let us consider a channel domain. Let the x axis point in the direction of the flow, y axis in the normal direction of the walls, so that the walls are defined by $y = -1$ and $y = 1$. The base flow is a pressure-driven Poiseuille flow with parabolic profile

$$u_0 = [U_0;\ 0]^\top = [1 - y^2;\ 0]^\top.$$

The corresponding field of orientation tensors can readily be computed from (1). We look for quasi-stationary solutions of the form

$$\begin{aligned}
\tilde{u} &= U_0(y) + u_1'(y)e^{i\alpha(x-ct)},\\
\tilde{v} &= u_2'(y)e^{i\alpha(x-ct)},\\
\tilde{a}^{(2)} &= a(y) + b(y)e^{i\alpha(x-ct)},\\
\tilde{\psi} &= \psi(y) + \phi(y)e^{i\alpha(x-ct)},\\
\tilde{\tau} &= \tau(y) + \tau'(y)e^{i\alpha(x-ct)},
\end{aligned}$$

where ψ_0 is the base stream function, ϕ is the perturbation of the base stream function, $\alpha \in \mathbb{R}$ is the wavenumber and c is a generalized complex eigenvalue, so that $\alpha\mathrm{Im}(c)$ determines the rate of growth of the perturbation.

Linearizing and rewriting the momentum equation in terms of ϕ and eliminating the pressure leads to

$$\begin{aligned}
i\alpha[(U_0 - c)(\partial_y^2 - \alpha^2) - \partial_y^2 U_0]\phi - \mathrm{Re}^{-1}(\partial_y^2 - \alpha^2)^2\phi&\\
= (\partial_y^2 + \alpha^2)\tau_{12}' + i\alpha\partial_y(\tau_{11}' - \tau_{22}')&,
\end{aligned} \tag{2}$$

where ∂_y^k is the k-th order derivative operator wrt y, with the boundary conditions

$$\phi(-1) = \partial_y\phi(-1) = \phi(1) = \partial_y\phi(1) = 0. \tag{3}$$

The components of the extra stress τ_f are computed in a similar way by plugging the perturbations in (1) and ignoring terms of higher order:

$$\tau' = N_p\left[(\nabla u' : a)a + (\nabla u : b)a + (\nabla u : a)b\right] \tag{4}$$

We also get a linear algebraic system for the perturbation of orientation field. Denoting the components $a = \begin{pmatrix} a_1 & a_2 \\ a_2 & 1-a_1 \end{pmatrix}$, $b = \begin{pmatrix} b_1 & b_2 \\ b_2 & -b_1 \end{pmatrix}$, $\lambda = -i\alpha c$, it reads:

$$(A - \lambda I) \cdot b = r \tag{5}$$

where the coefficients are

$$\begin{aligned}
A_{11} &= -i\alpha U_0 - 2a_2 m - 2\dot{\gamma} C_i + \dot{\gamma} S_0 \left[6a_1(1-a_1) - 1 - 2a_2^3\right], \\
A_{12} &= 2\left[m(a_1 - 1) + \dot{\gamma} S_0 a_1 (2a_2 - 1)\right], \\
A_{21} &= \dot{\gamma} S_0 (4a_1 - 2) a_2 - m, \\
A_{22} &= -i\alpha U_0 + 4a_2 m + 2\dot{\gamma} C_i + \dot{\gamma} S_0 \left[a_1^2 + (1-a_1)^2 + 6a_2^2 - 1\right], \\
r_1 &= \left[2a_1 n_{11} + 2(1-a_1) n_{22} + 2(n_{21} + n_{22}) a_2\right] a_1, \\
r_2 &= (1-a_1) n_{21} - \left[2a_1 n_{11} + 2(1-a_1) n_{22} + 2(n_{21} + n_{12}) a_2\right] a_2 + a_1 n_{12}.
\end{aligned}$$

Here we have denoted the gradient of the perturbation of velocity $\nabla u'$ by n, i.e., $n_{11} = i\alpha\phi'$, $n_{12} = \alpha^2\phi$, $n_{21} = \phi''$, $n_{22} = -i\alpha\phi'$, and $m = M_{21}$.

Equations (2)–(5) form a GEP for a differential-algebraic system: find complex $\lambda = -i\alpha c$ such that the system (2)–(5) admits nontrivial solutions. By solving (5) for b and plugging the result into (4), it can be reduced to a nonlinear GEP for a fourth order equation.

4 Numerics

The method of choice for discretizing Orr-Sommerfeld type equations is using a pseudospectral method on a Chebyshev grid, e.g., a Chebyshev-tau method [2,7].

By discretizing the full system (2)–(5), we obtain a linear GEP for a differential-algebraic system of equations:

$$\begin{pmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & D_{22} & D_{23} \\ C_{31} & D_{32} & D_{33} \end{pmatrix} \cdot \begin{pmatrix} \phi \\ b_1 \\ b_2 \end{pmatrix} = \lambda \begin{pmatrix} B & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{pmatrix} \cdot \begin{pmatrix} \phi \\ b_1 \\ b_2 \end{pmatrix} \tag{6}$$

where the blocks denoted by D are diagonal. This fact allows to easily eliminate b_1 and b_2 obtaining a non-linear GEP for ϕ alone:

$$\left[C_{11} - \lambda B + \left(C_{12}\ C_{13}\right) \cdot \begin{pmatrix} D_{22} - \lambda I & D_{23} \\ D_{32} & D_{33} - \lambda I \end{pmatrix}^{-1} \cdot \begin{pmatrix} C_{21} \\ C_{31} \end{pmatrix}\right] \phi = 0. \tag{7}$$

The inverse matrix is block diagonal and elements are rational functions of λ.

The nonlinear GEP (7) has been solved in [10], and the results suggest that the computed eigenvalues are perturbations of the eigenvalues for $N_p = 0$ with eigenvectors that have vanishing fiber orientation components $b_1 = b_2 = 0$, which we call the Orr-Sommerfeld eigenvalues. Since (7) is equivalent to (6), it has the same number of generalized eigenvalues, namely, three times more than the size of vector ϕ. Therefore, (7) for $N_p > 0$ has generalized eigenvalues which are not perturbations of the Orr-Sommerfeld eigenvalues. These extra eigenvalues have eigenvectors with $\phi = 0$ for $N_p = 0$, however, when $N_p > 0$, the eigenvectors also have a nonzero ϕ component and thus are not spurious but are relevant to the stability of the flow itself.

We prefer to solve the full system (2)–(5) because of the linearity—this allows employing the QZ decomposition to compute all the eigenvalues of the discretized problem at once.

5 Results and Conclusions

Left panel of Fig. 2 shows that there are two families of eigenvalues for $N_p = 0$: pure Orr-Sommerfeld ones (circles) and pure fiber orientation ones (dots without circles); the latter ones populate two parabolae. The right panel shows a slight perturbation $N_p = 10^{-4}$; note that both populations of eigenvalues mix together and cannot be separated in ones associated with perturbation of velocity field and perturbation of fiber orientation.

There are two types of instability of fiber suspension flow. The one is associated with sufficiently high Reynolds numbers. As a rule, the perturbed Orr-Sommerfeld eigenvalues move to the left in the complex plane as the particle number N_p increases, thus the stability region increases. This can be explained by the associated increase in *effective viscosity* of the suspension. In the case of a shear flow, the increase in effective viscosity is proportional to the off-diagonal component of $a^{(2)}$ which increases with growing C_i and decreases with growing S_0. The off-diagonal component a_2 effectively controls the critical Reynolds number.

The other kind of instability is associated with perturbations of $a^{(2)}$. If the equilibrium orientation state is perturbed to decrease the component a_1, some fibers have to make a rotation by almost full 180° according to Jeffery's model to return to the equilibrium state; this increases the off-diagonal component of $a^{(2)}$ and strongly influences the stress distribution over the suspension. This instability is independent on Reynolds number and present even for creeping flows. More involved nonlinear stability analysis in the spirit of [9] is required to demonstrate the transient nature of this second kind of instability. It should also be kept in mind when implementing Folgar-Tucker like models fully coupled to the flow.

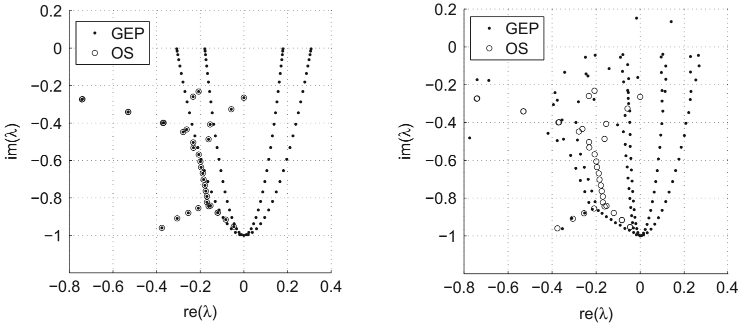


Fig. 2 Orr-Sommerfeld eigenvalues and the full set of eigenvalues for Re $= 5770$. *Left*: $N_p = 0$. The parabolae fill up densely with eigenvalues as the number of discretization points grows. *Right*: perturbation $N_p = 10^{-4}$

Acknowledgements This work was supported by ESF research project 2009/0223/1DP/1.1.1.2.0/09/APIA/VIAA/008.

References

1. Doi, M., Edwards, S.F.: The Theory of Polymer Dynamics. Clarendon Press, Oxford (1986)
2. Dongarra, J.J., Straughan, B., Walker, D.W.: Chebyshev tau-QZ algorithm methods for calculating spectra of hydrodynamic stability problems. Appl. Numer. Math. **22**, 399–434 (1996)
3. Gupta, V.K., Sureshkumar, R., Khomami, B., Azaiez, J.: Centrifugal instability of semidilute non-Brownian fiber suspensions. Phys. Fluids **14**, 1958–1971 (2002)
4. Latz, A., Strautins, U., Niedziela, D.: Comparative numerical study of two concentrated fiber suspension models. J. Nonnewton. Fluid Mech. **165**, 764–781 (2010)
5. Nsom, B.: Stability of fiber suspension flow in curved channel. J. Phys. II Paris **6**, 1483–1492 (1996)
6. Petrie, C.J.S.: The rheology of fibre suspensions. J. Nonnewton. Fluid Mech. **87**, 369–402 (1999)
7. Treffethen, N.L.: Spectral Methods in MATLAB. SIAM, Philadelphia (2000)
8. Tucker, C.L., Advani, S.G.: Processing of short-fiber systems. In: Advani, S.G. (ed.) Flow and Rheology in Polymer Composites Manufacturing, pp. 147–202. Elsevier, Amsterdam (1994)
9. Wan, Z., Lin, J., Xiong, H.: On the non-linear instability of fiber suspensions in a poiseuille flow. Int. J. Non-linear Mech. **43**, 898–907 (2008)
10. Zhenjiang, Y., Jianzhong, L., Zhaosheng, Y.: Hydrodynamic instability of fiber suspensions in channel flows. Fluid Dyn. Res. **34**, 251–271 (2004)

Microstructure Simulation of Paper Forming

Erik Svenning, Andreas Mark, Lars Martinsson, Ron Lai, Mats Fredlund, Ulf Nyman, and Fredrik Edelvik

Abstract This work presents a numerical framework designed to simulate the early paper forming process. This process is complex and includes strong fluid-structure interaction and complex geometries. The fluid flow solver IBOFlow, employs immersed boundary methods to simulate the flow around the fibers without the necessity of a boundary conforming grid. The fibers are approximated as slender beams with an elliptic cross section and modeled with the Euler-Bernoulli beam equation. A penalty based contact model is implemented. Finally, the potential of the framework is illustrated with an example.

E. Svenning • A. Mark (✉) • F. Edelvik
Fraunhofer-Chalmers Centre, Chalmers Science Park, SE-412 88 Gothenburg, Sweden
e-mail: erik.svenning@fcc.chalmers.se; andreas.mark@fcc.chalmers.se; fredrik.edelvik@fcc.chalmers.se

L. Martinsson
Albany International, Box 510, SE-301 80 Halmstad, Sweden
e-mail: lars.martinsson@albint.com

R. Lai
Eka Chemicals, SE-445 80 Bohus, Sweden
e-mail: ron.lai@akzonobel.com

M. Fredlund
Stora Enso Packaging, Box 9090, SE-650 09 Karlstad, Sweden
e-mail: mats.fredlund@storaenso.com

U. Nyman
Tetra Pak Packaging Solutions AB, Ruben Rausings Gata, SE-221 86 Lund, Sweden
e-mail: ulf.nyman@tetrapak.com

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_18,

1 Introduction

Paper forming is the process where a fiber suspension flows through a forming fabric, resulting in gradual build up of a fiber web. The motion of the fibers during this process step has a large influence on the final paper quality. Understanding this process is therefore valuable for the development of improved paper products. Simulation of paper forming is, however, challenging due to the complex fluid flow, the large structural displacements of the fibers and the strong fluid-structure coupling. To gain a deeper understanding of the paper forming process, the flow through the forming fabric and the gradually evolving fiber web needs to be resolved.

2 Numerical Method

In the present work, the finite volume based, incompressible Navier-Stokes software IBOFlow (Immersed Boundary Octree Flow Solver) is used to simulate the fluid flow. The immersed boundary methods developed by Mark et al. [1, 2] are used to model the flow through the fiber web and the forming fabric. A finite element discretization of the Euler-Bernoulli beam equation is used to describe the fiber motion and geometric nonlinearities are accounted for with the co-rotational formulation described by Crisfield et al. [3] and Nour-Omid and Rankin [4, 5]. The contacts are modeled with a penalty method. Elastic/inelastic contacts are taken into account as suggested by Harmon [6] and friction is treated with a regularization described by Wriggers [7]. A description of the simulation framework as well as a validation of the fluid structure coupling can be found in [8, 9].

The forming fabrics are described by triangulations generated from SEM images provided by Albany International. The fibers are modeled as slender beams with hollow elliptical cross section, allowing for different lengths, widths and shapes of the fibers.

3 Results

Microstructure simulations of paper forming are performed on a representative volume element containing a piece of the forming fabric. A pressure drop is applied over the domain and fibers are continuously injected at the inlet. As a result of the pressure drop, the fluid starts to flow through the forming fabric, so that the fibers are transported towards the fabric, where they start to form a fiber web as shown in Fig. 1. Initially the whole pressure drop takes place over the forming fabric. However, as the fiber web forms, the pressure drop over the fiber web gradually increases as shown in Fig. 2.

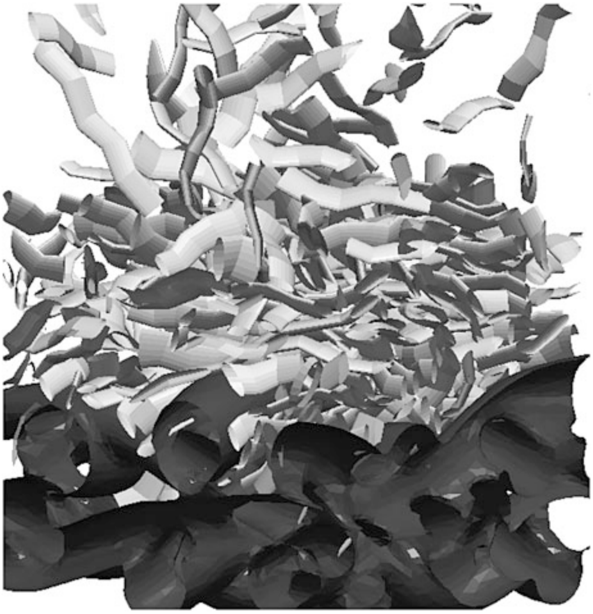


Fig. 1 Forming fabric and fiber web. Forming fabric geometry courtesy of Albany International

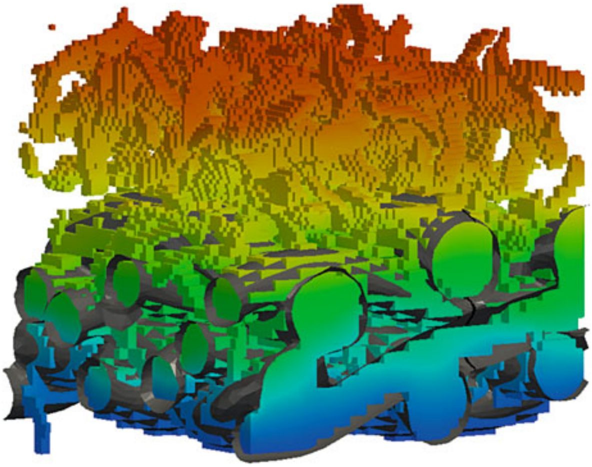


Fig. 2 Pressure drop over the forming fabric and the fiber web. Forming fabric geometry courtesy of Albany International

4 Conclusions

Microstructure simulations of paper forming allow relevant output data such as the orientation and distribution of fibers to be studied. The influence of different forming fabric geometries as well as different pulp properties can be investigated. The simulations can therefore provide a deeper understanding of the process conditions affecting the final paper quality.

Acknowledgements This work was performed within the Innovative Simulation of Paper (ISOP) project and funded in part by Akzo Nobel, Albany International, Stora Enso and Tetra Pak, and in part by forsk (ngpannefreningen's Foundation for Research and Development). The support is gratefully acknowledged.

References

1. Mark, A., van Wachem, B.: Derivation and validation of a novel implicit second-order accurate immersed boundary method. J. Comput. Phys. **227**, 6660–6680 (2008)
2. Mark, A., Rundqvist, R., Edelvik, F.: Comparison between different immersed boundary conditions for simulation of complex fluid flows. Fluid Dyn. Mater. Process. **7**(3), 241–258 (2011)
3. Crisfield, M., Galvanetto, U., Jelenić, G.: Dynamics of 3-d co-rotational beams. Comput. Mech. **20**, 507–519 (1997)
4. Nour-Omid, B., Rankin, C.: The use of projectors to improve finite element performance. Comput. Struct. **30**, 257–267 (1988)
5. Nour-Omid, B., Rankin, C.: Finite rotation analysis and consistent linearization using projectors. Comput. Methods Appl. Mech. Eng. **93**, 353–384 (1991)
6. Harmon, D.: Robust, efficient and accurate contact algorithms. Ph.D. thesis, Columbia University (2010)
7. Wriggers, P.: Computational Contact Mechanics. Springer, Berlin (2006)
8. Mark, A., Svenning, E., Rundqvist, R., Edelvik, F., Glatt, E., Rief, S., Wiegmann, A., Fredlund, M., Lai, R., Martinsson, L., Nyman, U.: Microstructure simulation of early paper forming using immersed boundary methods. TAPPI J. **10**(11), 23–30 (2011)
9. Svenning, E., Mark, A., Edelvik, F., Glatt, E., Rief, S., Wiegmann, A., Martinsson, L., Lai, R., Fredlund, M., Nyman, U.: Multiphase simulation of fiber suspension flows using immersed boundary methods. Nordic Pulp Pap. Res. J. **27**, 184–191 (2012)

Three-Dimensional Fiber Lay-Down in an Industrial Application

Johannes Maringer, Axel Klar, and Raimund Wegener

Abstract In this work we present fiber lay-down models that enable an efficient simulation of nonwoven structures. The models describe the form of deposited fibers with help of stochastic differential equations. The model parameters have to be estimated from more complex models in combination with measurements of the resulting nonwoven. We discuss the adaptation of a three-dimensional model on the basis of a typical industrial problem.

1 Introduction

In the manufacturing of technical textiles, thousands of individual slender fibers are overlapped to form random fiber webs yielding nonwoven materials. A typical method of production is the melt-spinning process, where the fibers are generated by extrusion of melted polymer through narrow nozzles. Then the fibers are stretched and spun until they solidify due to cooling air streams. These highly turbulent air flows account for swirling of the fibers before they lay-down on a moving conveyor belt. The resulting fiber web eventually passes through several process steps of reworking and reinforcing. The quality of the final nonwoven can be measured in terms of homogeneity, basis weight and permeability. An objective in industry is the simulation of the deposited fiber web and its optimization with respect to the desired characteristics. A mathematical model describing the fiber dynamics in turbulent air flows has been derived in [1, 2] and provides the basis for the

J. Maringer • A. Klar
TU Kaiserslautern, Gottlieb-Daimler Str. 48, 67663 Kaiserslautern, Germany
e-mail: johannes-maringer@web.de; klar@mathematik.uni-kl.de

R. Wegener (✉)
Fraunhofer ITWM, Fraunhofer-Platz 1, 67663 Kaiserslautern, Germany
e-mail: raimund.wegener@itwm.fraunhofer.de

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_19,

software tool FIDYST,[1] that enables among others the simulation of the lay-down process. Since this approach is computationally expensive, surrogate models have been developed in [3–7] as supportive methods. Combined with the computation of a few representative fibers with FIDYST, these surrogate models help after their calibration to simulate a whole virtual fiber web. In this work we consider the application of our models to a real industry problem from Oelikon Neumag. Besides FIDYST computations, wherein information about machine geometry and problem setting are included, we use image processing data from CT-scans of the resulting nonwoven to make up the adaptation of the three dimensional surrogate model.

2 The Models

We give a brief overview of the surrogate fiber lay-down models. The common approach is to model directly the fiber web on the transport belt instead of its complex antecedents.

2.1 The 2D Model

The original version has been introduced in [3]

$$\begin{aligned} d\boldsymbol{\xi}_t &= \boldsymbol{\tau}(\alpha_t)\,dt \\ d\alpha_t &= -\nabla V(\boldsymbol{\xi}_t)\cdot\boldsymbol{\tau}^{\perp}(\alpha_t)\,dt + A\,dW_t\ . \end{aligned} \tag{1}$$

Here, the arc-length parametrized curve $\boldsymbol{\xi} : \mathbb{R}_0^+ \to \mathbb{R}^2$ represents one deposited fiber. The tangent is normalized by $\boldsymbol{\tau}(\alpha) = [\cos\alpha, \sin\alpha]^T$. The drift term in the second equation models the coiling behavior of the fiber, where $\boldsymbol{\tau}^{\perp}(\alpha) = \frac{d\boldsymbol{\tau}(\alpha)}{d\alpha}$ and V is a suitable potential. The one-dimensional Wiener process with constant noise $A \in \mathbb{R}_0^+$ expresses the stochastic force, i.e. the effect of the turbulent air flows. The moving conveyor belt can be easily included in (1) as an additional reference curve,

$$d\boldsymbol{\xi}_t = \boldsymbol{\tau}(\alpha_t)\,dt + v\boldsymbol{e}_1\,dt\ ,$$

where $v = v_{belt}/v_{in}$ defines the ratio between belt speed and spinning speed of the fiber, compare [5]. The image of the fiber on the belt, denoted by $(\boldsymbol{\eta}_t)_{t\geq 0}$, is then obtained by $\boldsymbol{\eta}_t = \boldsymbol{\xi}_t - vt\boldsymbol{e}_1$. To obtain more realistic and smoother fiber paths we can further replace the Wiener process by an Ornstein-Uhlenbeck process, see [4] for a similar model,

[1] FIDYST: **Fi**ber **Dy**namics **S**imulation **T**ool developed at Fraunhofer ITWM, Kaiserslautern.

$$\begin{aligned} d\boldsymbol{\xi}_t &= \boldsymbol{\tau}(\alpha_t)\,dt \\ d\alpha_t &= -\nabla V(\boldsymbol{\xi}_t)\cdot\boldsymbol{\tau}^{\perp}(\alpha_t)\,dt + \kappa_t\,dt \\ d\kappa_t &= -\lambda\kappa_t\,dt + \mu\,dW_t\ , \end{aligned} \tag{2}$$

where $\lambda > 0$ is inversely related to the stiffness of the fiber and $\mu \in \mathbb{R}_0^+$ is another noise parameter.

2.2 The 3D Model

As established in [6, 7], the 2D models (1) and (2) can be extended to three dimensions. The natural arisen isotropic 3D formulations have been modified to take account for physical constraints like the impenetrably conveyor belt leading to anisotropic fiber orientations. The basic 3D model is given by

$$\begin{aligned} d\boldsymbol{\xi}_t &= \boldsymbol{\tau}(\alpha_t,\theta_t)\,dt \\ d\alpha_t &= -p_t\,dt + \frac{A}{\sin\theta_t}\,dW_t \\ d\theta_t &= -q_t\,dt + \frac{1}{2}A^2\cot\theta_t\,dt + A\sqrt{B}\,d\tilde{W}_t\ . \end{aligned} \tag{3}$$

while the smooth 3D model reads

$$\begin{aligned} d\boldsymbol{\xi}_t &= \boldsymbol{\tau}(\alpha_t,\theta_t)\,dt \\ d\alpha_t &= -p_t\,dt + \tfrac{\kappa_t}{\sin\theta_t}\,dt \\ d\theta_t &= -q_t\,dt + B\nu_t\,dt \\ d\kappa_t &= [p_t\nu_t\cos\theta_t - \kappa_t\nu_t\cot\theta_t - \lambda\kappa_t]\,dt + \mu\,dW_t \\ d\nu_t &= [-p_t\kappa_t\cos\theta_t + \kappa_t^2\cot\theta_t - B\lambda\nu_t]dt + \sqrt{B}\mu\,d\tilde{W}_t \end{aligned} \tag{4}$$

with abbreviations

$$\begin{aligned} p_t &:= p(\boldsymbol{\xi}_t,\alpha_t,\theta_t) := \tfrac{1}{B+1}\tfrac{1}{\sin\theta_t}\nabla V(\boldsymbol{\xi}_t)\cdot\boldsymbol{n_1}(\alpha_t) \\ q_t &:= q(\boldsymbol{\xi}_t,\alpha_t,\theta_t) := \tfrac{B}{B+1}\nabla V(\boldsymbol{\xi}_t)\cdot\boldsymbol{n_2}(\alpha_t,\theta_t)\ . \end{aligned}$$

Here, $(W_t)_{t\geq 0}$ and $(\tilde{W}_t)_{t\geq 0}$ denote independent one-dimensional Wiener processes. Furthermore, the tangent and the orthonormal vectors are expressed by $\boldsymbol{\tau}(\alpha,\theta) = [\cos\alpha\sin\theta, \sin\alpha\sin\theta, \cos\theta]^T$, $\boldsymbol{n}_1(\alpha) = \frac{1}{\sin\theta}\partial_\alpha\boldsymbol{\tau}(\alpha,\theta)$ and $\boldsymbol{n}_2(\alpha,\theta) = \partial_\theta\boldsymbol{\tau}(\alpha,\theta)$. The anisotropy of the fiber is represented by the weighting parameter $B \in [0,1]$. We note, that by the case $B = 0$ the respective 2D models (1), (2) are included. For more information and motivations we refer to the stated sources. Eventually, the moving conveyor belt can be analogously incorporated via $d\boldsymbol{\xi}_t = \boldsymbol{\tau}(\alpha_t, \theta_t)\,dt + v\boldsymbol{e}_1\,dt$.

3 Application

In the following we want to adapt our models in combination with FIDYST simulations to real nonwoven production processes. The process data stem from a pilot plant by the company Oerlikon Neumag. These are used to set up a FIDYST computation. Besides, pieces of the resulting nonwoven, corresponding to the process configuration, have been cut out and analyzed in CT-scans. The image processing data give indication of the fiber orientation in the nonwoven and should complement the FIDYST information with regard to the third dimension.

3.1 Parameter Estimation

Our aim is the calibration of the smooth 3D model (4) with moving conveyor belt. Therefore, we need estimations of μ, λ and B as well as the shape of the potential V, which should be of the form

$$V(\boldsymbol{\xi}) = \tilde{V}(\xi_1, \xi_2) + \Phi(\xi_3), \quad \tilde{V}(\xi_1, \xi_2) = \frac{\xi_1^2}{2\sigma_1^2} + \frac{\xi_2^2}{2\sigma_2^2},$$

where σ_1, σ_2 represent the deposition ranges of the fiber on the belt, and Φ is a confining potential taking account of the resistant belt, such that $\xi_3 \in (0, d_f)$. The nonwoven thickness d_f is an uncertain magnitude, that is hardly determinable. Thus we consider a range of potential thicknesses as multiples of the fiber diameter, that constitutes obviously a lower bound for d_f.

Basically, we follow the proposed strategy from [7], i.e. the 2D influenced parameters $(\sigma_1, \sigma_2, \mu, \lambda)$ can be estimated from FIDYST, whereas the remaining B is meant to be obtained from CT-scans. We make use of the equilibrium state to the situation of a non-moving transport belt, that is the solution of the stationary Fokker-Planck equation associated to (4) and is explicitly given by

$$p(\boldsymbol{\xi}, \theta, \kappa, \nu) = C\,\tilde{p}(\xi_1, \xi_2, \kappa)e^{-\Phi(\xi_3)}(\sin\theta)^{\frac{1}{B}}e^{-\frac{\nu^2}{\mu^2/\lambda}},$$

compare [7], with respective 2D solution

$$\tilde{p}(\xi_1, \xi_2, \kappa) = \tilde{C}e^{-\tilde{V}(\xi_1,\xi_2)}e^{-\frac{\kappa^2}{\mu^2/\lambda}},$$

where $C, \tilde{C}$ are normalization constants.

Estimating the mentioned 2D parameters, we resort to the heuristic approach from [5]. Therefore let $\mathscr{D} = (\boldsymbol{D}_1, \ldots, \boldsymbol{D}_N)$, $\boldsymbol{D}_i = (\boldsymbol{\eta}_{t_i}, \alpha_{t_i}, \kappa_{t_i}), i \in \{1, \ldots, N\}$ be equidistantly discrete process points with $\triangle t = t_{i+1} - t_i$ and consider a slightly different functional of characteristic properties than stated in [5],

$$\mathscr{F}(\mathscr{D}) = \left(\mathscr{S}(\xi_{t,1}), \mathscr{S}(\xi_{t,2}), \mathscr{S}(\kappa_t), \mathscr{K}(\kappa_t)\right),$$

where we define for an accordingly discretized stochastic process $(X_t)_{t\in\{t_1,\ldots,t_N\}}$,

$$\mathscr{S}(X_t) := \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(X_{t_i} - \frac{1}{N}\sum_{j=1}^{N} X_{t_j}\right)^2}$$

$$\mathscr{K}(X_t) := \max_{k\in\{1,\ldots,\bar{k}\}} \sqrt{\frac{\sum_{i=1}^{N-k}(X_{t_{i+k}} - X_{t_i})^2}{k(N-k)\triangle t}}, \; \bar{k} \ll N.$$

The respective angles α_{t_i} and curvatures κ_{t_i} can be reconstructed from the fiber points $\boldsymbol{\eta}_{t_i}$ by finite differences. We recall that $\boldsymbol{\xi}_{t_i} = \boldsymbol{\eta}_{t_i} + v t_i \boldsymbol{e}_1$ holds. Then we search for the optimal calibration $\boldsymbol{P} = (\sigma_1, \sigma_2, \frac{\mu}{\sqrt{2\lambda}}, \mu)$ of the surrogate model (2) with moving belt and potential $\tilde{V}$, denoted by $\mathscr{D}_{sur}(\boldsymbol{P})$, more precisely, we have to solve the minimization problem

$$\boldsymbol{P}^* = \operatorname{argmin}_{\boldsymbol{P}} \|\mathscr{F}(\mathscr{D}_{sur}(\boldsymbol{P})) - \mathscr{F}(\mathscr{D}_{fid})\|_2\,,$$

where $\mathscr{D}_{fid}$ indicates the data sample received from a FIDYST simulation. This can be solved by a relaxated quasi Newton method, compare [5]. We note that $\mathscr{F}$ is an almost perfect estimator for $\boldsymbol{P}$, if the fiber process is close to equilibrium $\tilde{p}$, i.e. for adequately large data sample and small speed ratio v.

Next, we consider the image processing data of the CT-scans. These contain information about the orientations of the fibers in space, i.e. the density distribution $B(\alpha, \theta)$ of the spherical polar angles, that determine the tangents at the fibers. However, this density has to be distinguished from the one obtained from our model (4), in the following denoted by $M(\alpha, \theta)$, since the CT-scan does not allow to differentiate the pathway of a single fiber. It holds $B(\alpha, \theta) = \frac{1}{2}\big(M(\alpha, \theta) + M(\alpha + \pi, \pi - \theta)\big)$, where $\alpha \in \mathbb{R}/2\pi\mathbb{Z}$ and $\theta \in (0, \frac{\pi}{2}]$. Here we take the accurate alignment of the nonwoven, in particular the direction $\boldsymbol{e}_3$, for granted. Then under reasonable symmetry and periodicity assumptions on the angle distribution of our model, in detail, $M(\pi + \alpha, \theta) = M(\pi - \alpha, \theta)$ and $M(\alpha, \pi - \theta) = M(\alpha, \theta)$, we deduce that $B_\alpha(\alpha, \theta)$ is π-periodic and has extrema in $\frac{\pi}{2}$ and π. Here the subscripted α denotes the respective marginal density distribution. Moreover, we expect that the reworking steps, in particular the stretching of the fibers during the bonding process, might largely influence $B_\alpha(\alpha, \theta)$ in terms of the amplitude of the extrema. Nevertheless, we presume that these impacts are less affecting $B_\theta(\alpha, \theta)$.

Based on this consideration, the parameter B should be adjusted demanding the equality of the variances $V(\theta; M_\theta(\alpha,\theta)) = V(\theta; B_\theta(\alpha,\theta))$. Again, under the assumption that the fiber process is close to its stationary state, we avail ourselves of the density function $(\sin\theta)^{1/B}$ corresponding to $M_\theta(\alpha,\theta)$ yielding the desired B.

In principle, we are now able to simulate a virtual fiber web by simultaneous usage of the surrogate model to a large number of fibers, where we neglect influence of fiber-fiber-contact. The respective reference points determined by the positions of the nozzles have to be added as constants to the $\xi_{t,1}$-process and $\xi_{t,2}$-process.

3.2 Example from Industrial Problem

In the following we want to apply our theoretical considerations to the industrial test case from Oerlikon Neumag. On the basis of parameters specifying the process configuration, i.e. machine geometry and prevailing air flows, a corresponding FIDYST simulation has been realized. The FIDYST computation comprises positions of 13 filaments on the belt, each with around 19,350 data points. The belt speed is $v_{belt} = 0.633$ m/s and the spinning velocity $v_{in} = 79.4$ m/s, consequently we are situated close to the non-moving case. The space discretization is given by $\triangle t = 0.001$ m. Our algorithm produces as the optimal choice of 2D parameter

$$\boldsymbol{P}^* = (0.0050, 0.0049, 1077, 49096)$$

where for checking purposes the associated functional values read

$$\mathscr{F}(\mathscr{D}_{fid}) = (0.0075, 0.0072, 1183, 53376)$$
$$\mathscr{F}(\mathscr{D}_{sur}(\boldsymbol{P})) = (0.0072, 0.0070, 1188, 53504)\ .$$

In Fig. 1 we illustrate a comparison of fibers gained from FIDYST and the calibrated surrogate model (2). In fact, they show qualitatively similar lay-down structures which confirm our approach. Furthermore, the measured distributions of the angles $B(\alpha,\theta)$ from the CT-scan comply well with our theoretical considerations above. We observe the nearly symmetric profile of $B_\alpha(\alpha,\theta)$ with a large amplitude, compare Fig. 2. On the contrary, both FIDYST and the surrogate models show an almost uniform distribution, which is due to the small speed ratio v and the description of the fibers without the involvement of reworking steps. The variance of the analyzed θ-distribution can be determined as $V(\theta; B_\theta(\alpha,\theta)) = 0.32$ that corresponds to parameter $B = 0.398$, see Fig. 2. We are now in the position to simulate a fiber web with help of (4), if the fleece thickness d_f is suitable chosen.

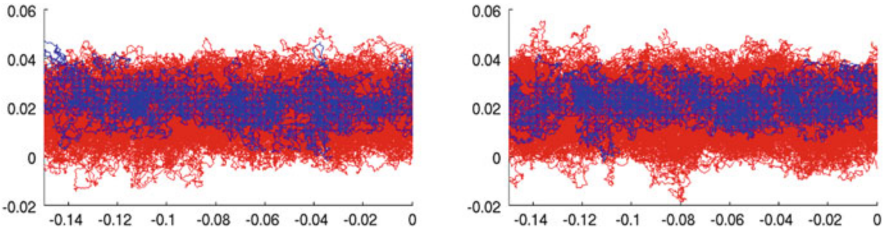


Fig. 1 Simulated fibers (*red*) with one emphasized filament (*blue*). *Left*: FIDYST, *right*: surrogate

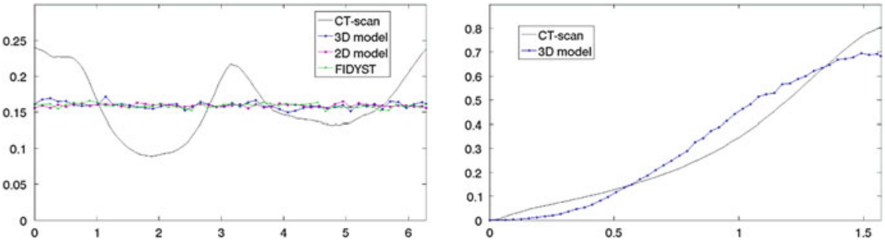


Fig. 2 Angular distribution, *left*: α, *right*: θ. Comparison of CT-scan and calibrated surrogate models and FIDYST

4 Conclusion and Outlook

We presented the application of a 3D surrogate fiber lay-down model to an industrial problem. The parameters are identified on the basis of experimental data. The calibrated model enables the efficient simulation of a whole virtual fiber web. Their usage for evidence of the quality of the corresponding real nonwoven, however, require further modifications of the model, particularly with regard to the impenetrability of the fibers. This is examined in further studies.

Acknowledgements This work has been supported by Bundesministerium für Bildung und Forschung (BMBF), Verbundprojekt ProFil, 03MS606. Moreover, we thank the colleagues from the departments Transport Processes and Image Processing at Fraunhofer ITWM.

References

1. Marheineke, N., Wegener, R.: Fiber dynamics in turbulent flows: general modeling framework. SIAM J. Appl. Math. **66**(5), 1703–1726 (2006)
2. Marheineke, N., Wegener, R.: Modeling and application of a stochastic drag for fibers in turbulent flows. Int. J. Multiphase Flow **37**, 136–148 (2011)
3. Götz, T., Klar, A., Marheineke, N., Wegener, R.: A stochastic model and associated Fokker-Planck equation for the fiber lay-down process in nonwoven production processes. SIAM J. Appl. Math. **67**(6), 1704–1717 (2007)

4. Herty, M., Klar, A., Motsch, S., Olawsky, F.: A smooth model for fiber lay-down processes and its diffusion approximations. Kinet. Relat. Models **2**(3), 480–502 (2009)
5. Klar, A., Marheineke, N., Wegener, R.: Hierarchy of mathematical models for production processes of technical textiles. Z. Angew. Math. Mech. **89**(12), 941–961 (2009)
6. Klar, A., Maringer, J., Wegener, R.: A smooth 3d model for fiber lay-down in nonwoven production processes. Kinet. Relat. Models **5**, 97–112 (2012)
7. Klar, A., Maringer, J., Wegener, R.: A 3d model for fiber lay-down in nonwoven production processes. Math. Models Methods Appl. Sci. **22**(9), 1250020 (2012)

3d Modeling of Dense Packings of Bended Fibers

Hellen Altendorf and Dominique Jeulin

Abstract For the simulation of fiber systems, there exist several stochastic models: systems of straight non overlapping fibers, systems of overlapping bending fibers, or fiber systems created by sedimentation. However, there is a lack of models providing dense, non overlapping fiber systems with a given random orientation distribution and a controllable level of bending. We present in this paper the recently developed stochastic model that generalizes the force-biased packing approach to fibers represented as chains of balls. The starting configuration is a boolean system of fibers modeled by random walks, where two parameters in the multivariate von Mises-Fisher orientation distribution control the bending. The points of the random walk are associated with a radius and the current orientation. The resulting chains of balls are interpreted as fibers. The final fiber configuration is obtained as an equilibrium between repulsion forces avoiding crossing fibers and recover forces ensuring the fiber structure. This approach can provide high volume fractions up to 72 %. Furthermore, we study the efficiency of replacing the boolean system by a more intelligent placing strategy, before starting the packing process. Experiments show that a placing strategy is highly efficient for intermediate volume fraction.

1 Introduction

The increasing interest in fibrous materials (as glass or carbon fiber reinforced composites) necessitates the development of quantitative methods of characterization [1, 2]. The macroscopic properties of these materials are highly influenced by the geometry of the fiber component, in particular by the direction distribution. With virtual material design material properties can be optimized by adapting the

H. Altendorf (✉) • D. Jeulin
Mathématiques et Systèmes, Centre de Morphologie Mathématique, Mines Paris Tech, Fontainebleau, France
e-mail: hellen.altendorf@mines-paristech.fr; dominique.jeulin@mines-paristech.fr

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_20,

direction distribution respectively. To this end we need a parametric stochastic model for the fiber structure. Most of the existing approaches model fibers as cylinders (dilated Poisson line process [3], random sequential absorption [4, 5] for cylinders or deposition of cylinder [6, 7]), which limits the material to straight fiber segments as e.g. glass fibers. However, carbon fiber reinforced polymers or non-woven with high fiber volume fraction and non-overlapping, bending fibers request more complex stochastic models.

We present in this paper the recently developed hardcore fiber model [8] with a controllable level of bending and high volume fractions realizing given orientation distributions. For this purpose, random walks are used to create realistic bending fibers, represented as chains of balls. The created fibers are placed randomly in a softcore system as a boolean model. With force-biased fiber packing, we achieve a non-overlapping configuration of the fiber system. Two kinds of forces are applied on the ball centers: repulsion and recover forces. The repulsion force arises in case of a fiber overlap and displaces the balls to an independent position. The recover force maintains the fiber structure and orientation. Volume fraction around 50 % could be achieved for several input parameters. In our experiments, the maximal volume fraction was 72 % for z preferred orientation distribution and an aspect ratio of 9.

In this paper, we replace the boolean model by a placing strategy similar to the idea of the random sequential adsorption (RSA). Optimally, this addition decreases the overlap, and therefore also the amount of iterations necessary in the packing process, which leads to lower computation times. Experiments show that a placing strategy is highly efficient for intermediate volume fraction, whereas for very low or high volume fractions it is not necessary.

2 Fiber Model

The stochastic model considered in this paper is based on ball chains, initiated from a random walk and packed to a hardcore system with a force-biased approach. A fiber in the stochastic model is presented as a sequence of balls $P = \{p_1, \ldots, p_l\}$ with $p_i = (x_i, \mu_i, r_i) \in \mathbb{R}^3 \times S^2 \times \mathbb{R}^+$, consisting of the coordinate of the ball center $x_i \in \mathbb{R}^3$, an orientation $\mu_i \in S^2$ and a radius $r_i \in \mathbb{R}^+$. The orientation describes the local fiber orientation and the radius describes the local fiber radius. The main fiber orientation is chosen from a global orientation distribution defined for the system. We propose the β-distribution (see [9] or [10]) with a global parameter $\beta \in \mathbb{R}^+ \backslash \{0\}$.

A ball chain is created by a random walk starting from a random point in a cubic window with periodic boundary conditions. The orientation assigned to the first ball μ_1 is initiated with the main fiber orientation, chosen from the global orientation distribution. The orientation assigned to the i-th ball is distributed with the multivariate von Mises-Fisher distribution (see [11–13]). The parameters of this distribution are two preferred directions and their reliability parameters κ_1 and κ_2. In our case, the preferred directions are the main fiber orientation μ_1 and the last chosen orientation μ_{i-1}. The level of bending is defined by the reliability parameters.

The radius r_i could be chosen from any distribution. In this paper, we have chosen a fix radius for the system. The coordinates of the i-th ball are then defined by $x_i = x_{i-1} + \frac{r_i}{2}\mu_i$. This approach defines a overlapping system of bended fibers.

In a second step, we apply a force-biased approach, to achieve a hardcore configuration of the fiber system. Force-biased algorithms on spheres were introduced in [14] and statistically analyzed in [15]. The forces in our approach were inspired by the energy reducing models known from molecular dynamics [14] and describe the necessary displacement of the balls to relax the system. They do not act like mechanical forces. The algorithm works stepwise: In every step, forces are calculated according to the recent configuration and the balls are displaced with respect to their forces.

Two kinds of forces are applied to the ball centers: repulsion and recover forces. The repulsion force arises in case of a fiber overlap and displaces the balls to an independent position. The recover force maintains the fiber structure. It keeps the distance and the angles between a ball and its neighbors, allowing only small deviation. The force conserving distances simulates springs between neighbor ball centers. The angle force simulates open springs between neighbor connections, which allows straightening of the fiber, but preserves fibers to bend in a clew. Both recover forces are provided with an initiating friction, which assures the stabilization of the packing process. The new configuration at the end of one step is defined by the displaced ball centers according to the sum of all forces.

Figure 1 shows realizations of the presented model with varying input parameters for the fiber aspect ratio χ, the number of objects n and the main orientation distribution. The parameters and the achieved volume fraction V_V are given below each realization. For more details on the stochastic model see [8]. Parameters for the stochastic model can be estimated from a separated fiber system as shown in [16].

In the following, we describe a configuration of the fiber system by a list of sphere chains $P = \{p_{1,1}, p_{1,2}, \ldots, p_{1,l_1}, p_{2,1}, \ldots, p_{n,l_n}\}$ with $p_{j,i} = (x_{j,i}, \mu_{j,i}, r_{j,i}) \in \mathbb{R}^3 \times S^2 \times \mathbb{R}^+$. The fiber index is indicated with j and the balls in one fiber are ordered by the index i.

3 Joining with RSA Approach

In the above described stochastic model, there is no strategy in placing the fibers in the initial configuration. Still, the initial placement influences highly the time of stabilization in the packing process. We make use of the idea in the random sequential adsorption algorithm (RSA, see [5]) to create the initial configuration.

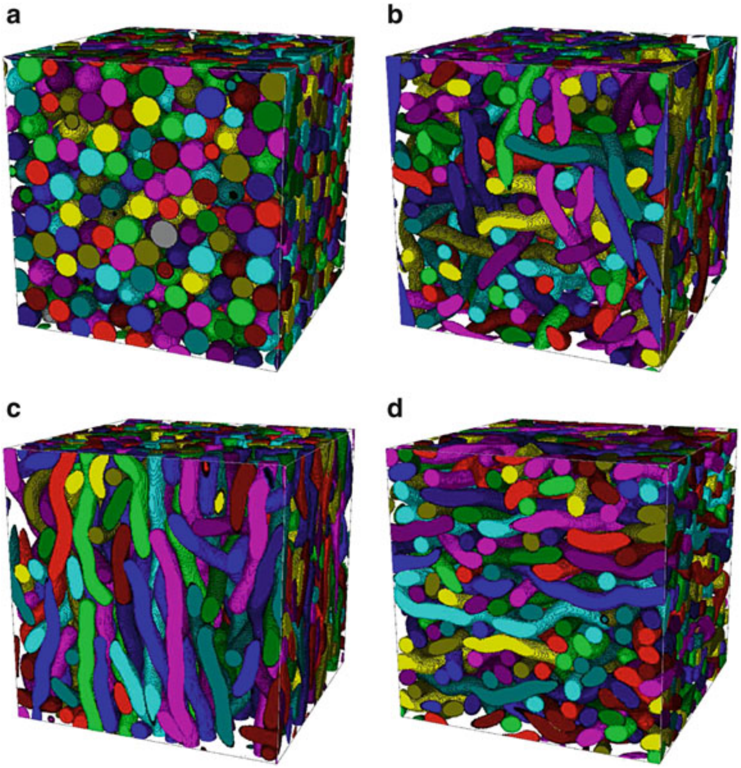


Fig. 1 Realizations for packed fiber systems. Common parameters are as follows: window side length $s = 100$ and bending parameters $\kappa_1 = 10$ and $\kappa_2 = 100$. (**a**) $\chi = 1$, $n = 1{,}146$, $V_V = 57.27\,\%$, (**b**) $\chi = 33$, $n = 90$, $V_V = 50.35\,\%$, isotropic orientation ($\beta = 1$), (**c**) $\chi = 17.67$, $n = 170$, $V_V = 49.60\,\%$, orientation in z-direction ($\beta = 0.1$), (**d**) $\chi = 17.67$, $n = 170$, $V_V = 49.95\,\%$, orientation in xy-plane ($\beta = 10$)

The RSA algorithm was originally invented for sphere packing, but can be generalized to any kind of objects. In a first step, a finite set of objects is created in a stochastic process. In the second step, the objects are iteratively inserted in the scene of interest with well defined boundary conditions. For every object, which should be inserted in the scene, we randomly choose new placements and place the object as soon as a placement is found without any collision with the already inserted objects. To assure chosen characteristic distributions of the object (as for example size or orientation distribution) it is important that the object is not recreated, but only displaced. Otherwise, the last object inserted would surely have smaller size and would align to the existing structure (in the case of elongated objects).

In the case of fiber systems, the RSA algorithm combined with cylindrical objects has the disadvantage, that only low volume fractions can be achieved. However, it may serve to create a more intelligent and less overlapping initial configuration for our stochastic model. The criterion to evaluate a placement for a sphere chain

$p_m = \{p_{m,1}, \ldots, p_{m,l_m}\}$ in a scene with $m-1$ already inserted fibers $p_1, \ldots, p_{m-1}$ is either the maximal overlap or the sum of overlaps of the spheres of the fiber p_m with the already inserted fibers. We define these two criteria as following:

$$C_{\max}(p_m \mid p_1, \ldots, p_{m-1}) = \max_{i,j,k}\{\vartheta(p_{m,i}, p_{k,j})\} \tag{1}$$

$$C_{\text{sum}}(p_m \mid p_1, \ldots, p_{m-1}) = \sum_{i,j,k} \vartheta(p_{m,i}, p_{k,j}) \tag{2}$$

with $1 \leq i \leq l_m$, $1 \leq k \leq m-1$, $1 \leq j \leq l_k$, and $\vartheta(p,q)$ describing the overlap of two spheres with centers x_p, x_q and radii r_p, r_q:

$$\vartheta(p,q) = \max\{r_p + r_q - |x_q - x_p|,\ 0\}. \tag{3}$$

We displace the objects as long as the criterion is higher than the global limit ϑ_0, which serves also as stop criterion for the packing process. After a certain volume fraction the probability to exceed this limit is quite low (if not the packing process would not be necessary). Therefore, we fix a certain number of placements (in the experiments prepared for this paper we tested 10 and 100). We insert the object at those placement, having the lowest value of the evaluation criterion.

We tested the computation time for the fiber packing with 10 and 100 steps for the initial configuration versus the standard boolean configuration. The model is realized in a unit cube with periodic boundary conditions. The fibers are isotropically oriented and have an aspect ratio of 9 and a volume of 1 % of the cube volume. This implies that the volume fraction of the fiber system in percent equals the amount of fibers. The curvature parameters are chosen as $\kappa_1 = 10$ and $\kappa_2 = 100$.

Figure 2 shows the computation times for the different approaches and Fig. 3 the ratio of the placing and packing time versus the packing time without strategy for the initial configuration. The experiment runs the realization of the stochastic model 100 times and averages the extracted characteristics. We observe that the strategy is not very efficient for low volume fractions, as in this case the probability of overlap is very low, thus already the first random placing has low overlap and even in the case of overlap, there is enough free space to displace the fiber in only few steps. The influence of the placing strategy rises with the volume fraction until about 50 %, where the strategy of placing results in a more evenly placed systems and decreases local overcrowded areas. When the volume fraction rises over 50 %, the influence of the placing strategy decreases again and may even vanish for very dense systems. This effect has two reasons. First, the process stops without success after a certain amount of iterations, which represents a fix time. That means, for all unsuccessfully finished jobs, that occur often for the packing with placing strategy, we assume a too low computation time, thus the ratio of computation times is not appropriate. This theory is supported by Fig. 4, which shows that the placing strategy increases the probability to successfully finish the packing process for high volume fractions. Secondly, for a higher volume fraction, we have also a higher amount of fibers,

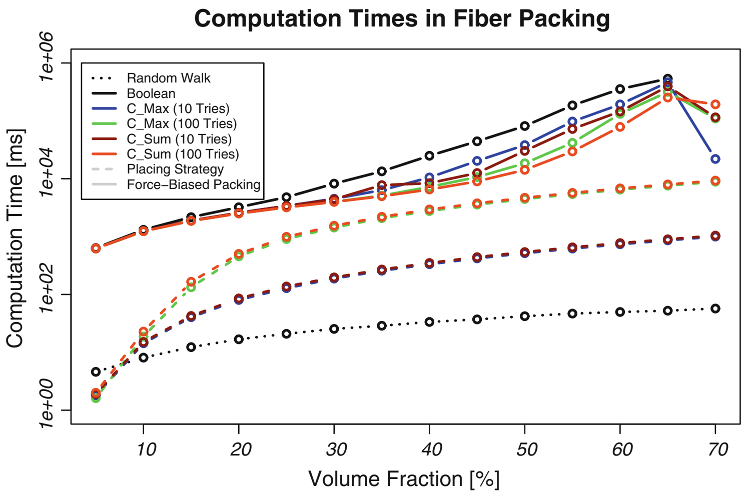


Fig. 2 Computation times for the object generation

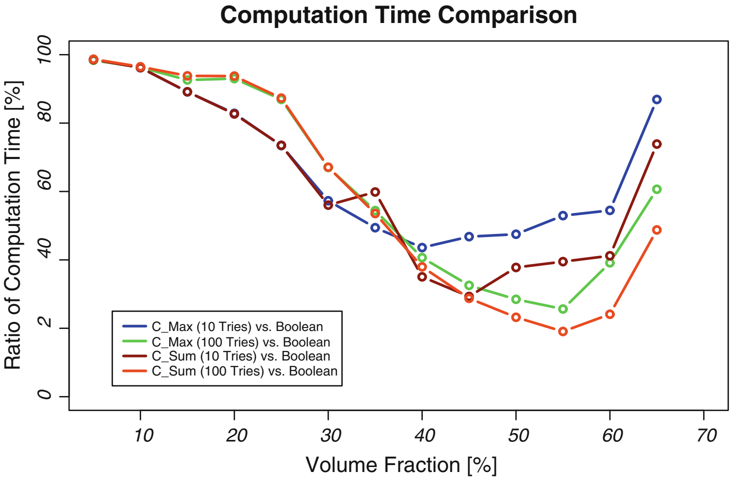


Fig. 3 Ratio between the times of fiber packing with and without placing strategy

which may be naturally placed more evenly over the image (according to the law of large numbers). Additionally, the placing strategy gets ineffective as with the high density, as for the last fibers, there exits no "good" placement any more. Thus the last fibers are placed randomly and the configurations with or without strategy are equally distributed in space.

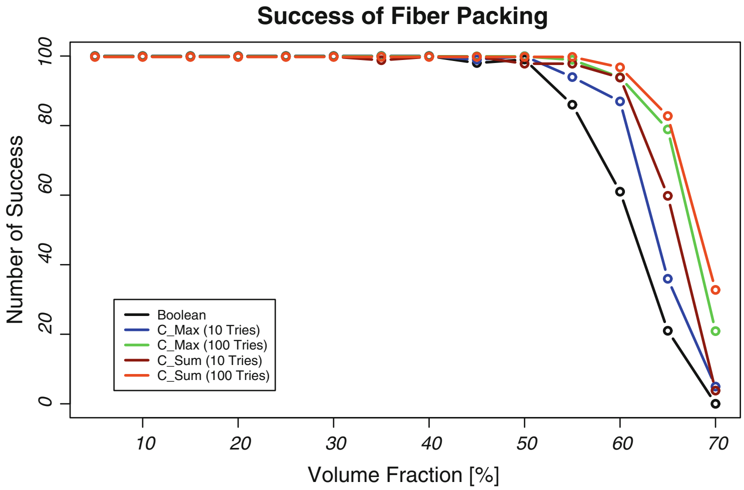


Fig. 4 Probability of success for the fiber packing process

4 Conclusion

We have presented an algorithm generating bending hardcore fibers, with given orientation, radius, and length distributions. We evaluated an intelligent placing strategy based on overlap criteria, and conclude that the strategy is most efficient for mean volume fractions around 50 %. In the future, we will include further recover forces to be able to realize more restricted orientation distributions.

References

1. Altendorf, H., Jeulin, D.: 3d directional mathematical morphology for analysis of fiber orientations. Image Anal. Stereol. **28**, 143–153 (2009)
2. Altendorf, H., Jeulin, D.: Fiber separation from local orientation and probability maps. In: Wilkinson, M.H.F., Roerdink, J.B.T.M. (eds.) ISMM 2009 Abstract Book, pp. 33–36. University of Groningen, Groningen (2009)
3. Matheron, G.: Random Sets and Integral Geometry. Series in Probability and Mathematical Statistics. Wiley, London (1974)
4. Widom, B.: Random sequential addition of hard spheres to a volume. J. Chem. Phys. **44**(10), 3888 (1966)
5. Feder, J.: Random sequential adsorption. J. Theor. Biol. **87**(2), 237–254 (1980)
6. Provatas, N., Haataja, M., Asikainen, J., Majaniemi, S., Alava, M., Ala-Nissila, T.: Fiber deposition models in two and three spatial dimensions. Colloids Surf. A Physicochem. Eng. Asp. **165**(1–3), 209–229 (2000)
7. Coelho, D., Thovert, J.-F., Adler, P.M.: Geometrical and transport properties of random packings of spheres and aspherical particles. Phys. Rev. E **55**(2), 1959–1978 (1997)

8. Altendorf, H., Jeulin, D.: Random-walk-based stochastic modeling of three-dimensional fiber systems. Phys. Rev. E **83**(4), 041804 (2011)
9. Schladitz, K., Peters, S., Reinel-Bitzer, D., Wiegmann, A., Ohser, J.: Design of acoustic trim based on geometric modeling and flow simulation for non-woven. Comput. Mater. Sci. **38**(1), 56–66 (2006)
10. Ohser, J., Schladitz, K.: 3d Images of Materials Structures – Processing and Analysis. Wiley VCH, London (2009)
11. Karkkainen, S., Nyblom, J., Miettinen, A., Turpeinen, T., Pötschke, P.: A stochastic shape model for fibres with an application to carbon nanotubes. In: 10th European Congress of Stereology and Image Analysis (2008)
12. Ko, D.: Robust estimation of the concentration parameter of the von Mises-Fisher distribution. Ann. Stat. **20**, 917–928 (1992)
13. Banerjee, A., Dhillon, I.S., Ghosh, J., Sra, S.: Clustering on the unit hypersphere using von Mises-Fisher distributions. J. Mach. Learn. Res. **6**, 1345–1382 (2005)
14. Mościński, J., Bargieł, M.: The force-biased algorithm for the irregular close packing of equal hard spheres. Mol. Simul. **3**(4), 201–212 (1989)
15. Bezrukov, A., Bargieł, M., Stoyan, D.: Statistical analysis of simulated random packings of spheres. Part. Part. Syst. Charact. **19**, 111–118 (2002)
16. Altendorf, H., Jeulin, D.: Stochastic modeling of a glass fiber reinforced polymer. In: Wilkinson, M.H.F., Roerdink, J.B.T.M. (eds.) ISMM 2011 Abstract Book. University of Groningen, Groningen (2011)

Part IV
Flow

Overview

At the ECMI Conference 2012 several advances in the large field of fluid dynamics were presented, focusing for example on fluid-structure interactions, two-phase flows, thin films, boundary layer flows and turbulence. They spanned the whole range from modeling, analysis to simulation and optimization. The nine contributions in this section *Flow* provide a detailed description and envisage solution of dedicated applications.

Fluid-structure interactions is the core issue of the first two papers. The coupling of fluid and structure as well as time-dependent (moving) domains make the simulation challenging. In *Simulation of a Rubber Beam Interacting with a Two-Phase Flow in a Rolling Tank* Erik Svenning et al. demonstrate the applicability of an immersed boundary method to couple a finite volume based Navier-Stokes solver with a finite element based structural mechanics solver for large deformations. They use the approach for the benchmark simulation of an elastic rubber beam in a rolling tank partially filled with oil, yielding good agreements with experiments. For a simplified formulation of fluid-structure interactions Julia Niemeyer and Bernd Simeon analyze the coupling condition and the effect of a moving fluid on the numerical solution in *Modelling of a Simplified Fluid-Structure Interaction Formulation*. They apply a semidiscretization with finite elements in time. The time integration is performed implicitly, whereas the coupling conditions are enforced explicitly by means of corresponding constraint equations in a differential-algebraic system.

Why do Guinness bubbles sink? This question is answered by Cathal P. Cummins et al. in *Sinking Bubbles in Stout Beers*. They show that the circulation in a container with a bubbly liquid (e.g. a glass of stout beer) is determined by the container's shape. Another kind of two-phase flow is topic in Andrew Gordon and Michael Vynnycky's work *Analysis of Two-Phase Flow in the Gas Diffusion Layer of a Polymer Electrolyte Fuel Cell*. Considering a two-phase (gas/liquid) flow in the porous gas diffusion layer on the cathode and a water transport in the fuel cell, they investigate asymptotically the dependency of the degree of water saturation on the liquid phase relative permeability and the behavior when the gas diffusion layer is hydrophilic.

The formation of an air gap at the mould-metal interface in continuous casting has a detrimental effect on the efficiency of the process. Due to the complexity of three-dimensional numerical simulations, a quasi-analytical model is derived by Michael Vynnycky using asymptotic methods. The model captures the essential characteristics and allows for a full coupling between the thermal and mechanical features. The influence of the process parameters on the onset of air gap formation is studied in *A Criterion for Air-Gap Formation in Vertical Continuous Casting: The Effect of Superheat*.

In *Moulding Contact Lenses* Ellen Murphy and William T. Lee model the moulding process of a monomer-based fabric by help of a thin film approximation and investigate the role of curvature, surface tension and motion in the formation of defects.

In *Enhanced Water Flow in Carbon Nanotubes and the Navier Slip Condition* Tim G. Myers sets up a model for the water flow in carbon nanotubes which contains a depletion layer with reduced viscosity near the wall. Whereas in the limit of large tubes it shows no enhancement, for smaller tubes the model predicts enhancement that increases as the tube radius decreases. Moreover, the model provides a physical interpretation of the classical Navier slip condition and explains why slip-lengths may be greater than the tube radius.

The last two contributions are concerned with numerical simulations and turbulence models. The paper *Flow Field Numerical Research in a Low-Pressure Centrifugal Compressor with Vaneless Diffuser* is focused on the capabilities and constraints of the steady-state numerical simulations for an accurate prediction of the flow through a compressor stage, therefore Alexey Frolov discusses different discretization schemes and turbulence models. In region of high flow rates the steady results turn out to be in good agreement with experiments, whereas for low flow rates the unsteady effects dominate the flow behavior. In *Large Eddy Simulation of Boundary-Layer Flows over Two-Dimensional Hills* Ashvinkumar Chaudhari et al. perform Large Eddy Simulations for turbulent boundary layer flows over two-dimensional hills or ridges of different slopes and compare the results (mean velocity, flow separation, turbulence quantities) with wind tunnel experiments.

Nicole Marheineke

Simulation of a Rubber Beam Interacting with a Two-Phase Flow in a Rolling Tank

Erik Svenning, Andreas Mark, and Fredrik Edelvik

Abstract The aim of this paper is to present and validate a modeling framework that can be used for simulation of industrial applications involving fluid structure interaction with large deformations. Fluid structure interaction phenomena involving elastic structures frequently occur in industrial applications such as rubber bushings filled with oil, the filling of liquid in a paperboard package or a fiber suspension flowing through a paper machine. Simulations of such phenomena are challenging due to the strong coupling between the fluid and the elastic structure. In the literature, this coupling is often achieved with an Arbitrary Lagrangian Eulerian framework or with smooth particle hydrodynamics methods. In the present work, an immersed boundary method is used to couple a finite volume based Navier-Stokes solver with a finite element based structural mechanics solver for large deformations. The benchmark of an elastic rubber beam in a rolling tank partially filled with oil is simulated. The simulations are compared to experimental data as well as numerical simulations published in the literature. 2D simulations performed in the present work agree well with previously published data. Our 3D simulations capture effects neglected in the 2D case, showing excellent agreement with previously published experiments. The good agreement with experimental data shows that the developed framework is suitable for simulation of industrial applications involving fluid structure interaction. If the structure is made of a highly elastic material, e.g. rubber, the simulation framework must be able to handle the large deformations that may occur. Immersed boundary methods are well suited for such applications, since they can efficiently handle moving objects without the need of a body-fitted mesh. Combining them with a structural mechanics solver for large deformations allows complex fluid structure interaction problems to be studied.

E. Svenning • A. Mark (✉) • F. Edelvik
Fraunhofer-Chalmers Centre, Chalmers Science Park, SE-412 88 Gothenburg, Sweden
e-mail: erik.svenning@fcc.chalmers.se; andreas.mark@fcc.chalmers.se; fredrik.edelvik@fcc.chalmers.se

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_21,

1 Introduction

Numerical simulations of highly elastic structures deforming in a free surface flow are challenging since the fluid-structure coupling is strong. The geometrically nonlinear response of the structure and the need to accurately resolve the free surface further increases the complexity of the simulations. The coupling between the fluid and the structure can be handled in different ways. A popular approach is the Arbitrary Lagrangian Eulerian (ALE) method [1], where the grid is deformed when the structure moves. Simulations with Smooth Particle Hydrodynamics (SPH) [2, 3] and Particle Finite Element Methods (PFEM) [4] are also reported in the literature. Immersed Boundary Methods (IBM) allow the flow around deforming objects in the flow to be resolved without the need of a body-fitted mesh. IBMs are therefore well suited for Fluid Structure Interaction (FSI) applications with large structural displacements. The original IBM developed by Peskin [5] was explicitly formulated and only first-order accurate in space. Majumdar et al. [6] developed a more stable method, which is implicitly formulated and second-order accurate in space. However, this method suffers from problems with mass conservation and pressure oscillations. To resolve these issues, Mark et al. [7, 8] developed a second-order accurate hybrid IBM. The IBM developed by Mark et al. has been used in several applications. It has been validated for simulation of fiber suspension flows with elastic fibers in [9], it was used to study Jeffery orbits in [10] and it was applied to FSI with heat transfer in [11].

FSI simulations can be performed in a monolithic or a partitioned way. Using a monolithic approach implies that all equations are solved simultaneously in the same matrix. In the partitioned approach, the different equations are solved separately and coupling algorithms are employed. Using the partitioned approach without coupling iterations between the fluid and the structure solutions is attractive in terms of computational efficiency. However, this approach often results in instabilities due to the added mass effect if the simulation time is long enough [12]. Gauss-Seidel iterations as well as quasi-Newton [13] techniques have been proposed to deal with these problems.

The aim of this paper is to present and validate a modeling framework that can be used for simulation of FSI in industrial applications. To achieve this, the partitioned approach with Gauss-Seidel iterations is used. The fluid-structure coupling is handled with the IBM developed by Mark et al. [8] and the structure is modeled as a St. Venant-Kirchhoff material, thus taking large deformations into account.

2 Theory

In the present work, a finite volume discretization on a Cartesian octree grid is used to solve the Navier-Stokes equations. A finite element discretization in total Lagrangian formulation is used to predict the motion of the structure. The fluid and structure models together with the FSI coupling are described in the following.

2.1 Fluid Model

The motion of an incompressible fluid is modeled by the Navier-Stokes equations:

$$\nabla \cdot \vec{u} = 0\,, \tag{1}$$

$$\rho_f \frac{\partial \vec{u}}{\partial t} + \rho_f \, \vec{u} \cdot \nabla \vec{u} = -\nabla p + \mu \nabla^2 \vec{u}\,, \tag{2}$$

where $\vec{u}$ is the fluid velocity, ρ_f is the fluid density, p is the pressure and μ is the dynamic viscosity. The finite volume method is used to solve the Navier-Stokes equations. The equations are solved in a segregated way and the SIMPLEC method derived in [14] is used to couple the pressure and the velocity fields. All variables are stored in a co-located arrangement and the pressure weighted flux interpolation proposed in [15] is used to suppress pressure oscillations. Two-phase flows are handles with the Volume Of Fluid (VOF) method, where an additional transport equation for the volume fraction is solved. A Cartesian octree grid is used for the spatial discretization of the fluid domain, that allows dynamic refinements around moving objects in the flow. The Backward Euler scheme is used for the temporal discretization.

2.2 Structure Model

The strong form of the equations of motion for an elastic solid is given by

$$\nabla \cdot \sigma + \rho \, \vec{b} - \rho \, \vec{a} = \vec{0}, \tag{3}$$

where σ is the Cauchy stress, $\vec{b}$ is the volume force and $\vec{a}$ is the acceleration, ρ denotes density and $\nabla \cdot$ is the divergence operator. Equation (3) can be expressed in terms of the Second Piola-Kirchhoff stress S by exploiting the relation between the Cauchy stress and the Second Piola-Kirchhoff stress

$$\sigma = J^{-1} \, F \cdot S \cdot F^T, \tag{4}$$

where F is the deformation gradient and $J = \det F$.

In the present work, large deformations are taken into account and St. Venant-Kirchhoff elasticity is assumed, with a strain energy potential given by [16]

$$\Psi = \frac{1}{2} \lambda \, (tr \, E)^2 + \mu \, E : E, \tag{5}$$

where E is the Green strain tensor and λ and μ are material parameters.

The Finite Element Method (FEM) is used to discretize Eq. (3). A total Lagrangian formulation is used and the nonlinear system of equations is solved with Newton's method. Newmark's time stepping scheme is used for the temporal discretization. Twenty-node hexahedral elements are used in the simulations presented in this paper.

2.3 Fluid-Structure Coupling

FSI simulations can be performed in a monolithic or a partitioned way. Using a monolithic approach implies that all equations are solved simultaneously in the same matrix. In the partitioned approach, the different equations are solved separately and coupling algorithms are employed. In the present work, the partitioned approach is employed and the simulations are performed without coupling iterations when possible. Gauss-Seidel iterations are used when necessary for stability reasons.

The mirroring IBM [8] is used to model the presence of solid objects in the flow by imposing the velocity of the solid as a Dirichlet boundary condition on the fluid. The boundary conditions are imposed in mirroring points defined on the interface between the fluid and the structure. The method is implicity formulated and second order accurate in space [8]. The force exerted on the solid by the fluid is computed by numerically integrating the traction vector over the fluid-solid interface.

3 Results and Discussion

In this section, numerical results for a benchmark case are presented and compared to previously published data from experiments [17, 18] and simulations [19]. The case considered is a rolling tank partially filled with oil. In the version considered in the present work, a flexible beam is clamped at the bottom of the tank. The tank is forced to rotate around the y-axis in point A as shown in Fig. 1, causing the oil inside the tank to move and interact with the beam. The tank is 0.609 m wide and 0.3445 m high. The length of the beam, which is equal to the oil depth, is 0.1148 m. The thickness of the beam in the x-direction is 4 mm. In the experiments reported in [17, 18], the tank thickness in the y-direction is 39 mm and the beam thickness in the y-direction is 33.2 mm, thus leaving a gap of 2.9 mm between the beam and the walls with normal in the y-direction. The oil is a sunflower oil with a density of 900 kg/m^3 and a viscosity of 45 mPa s. The second fluid in the tank is air at ambient conditions. The beam is made of a rubber material with a density of 1,100 kg/m^3, Young's modulus $E = 6$ MPa and Poisson's ratio $\nu = 0.49$.

The tank has two holes in the upper wall, so that zero pressure can be prescribed there. When 2D simulations are performed, symmetry boundary conditions are used on the faces with normal in the y-direction and no slip conditions are enforced on

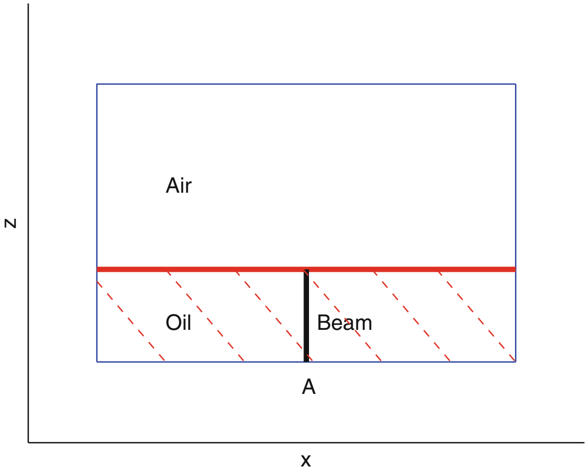


Fig. 1 Domain of the rolling tank case: the part of the domain marked with *dashed red lines* is filled with oil, the rest is filled with air. The beam is clamped to the tank in point A and an electric motor forces the tank to rotate around the y-axis in this point

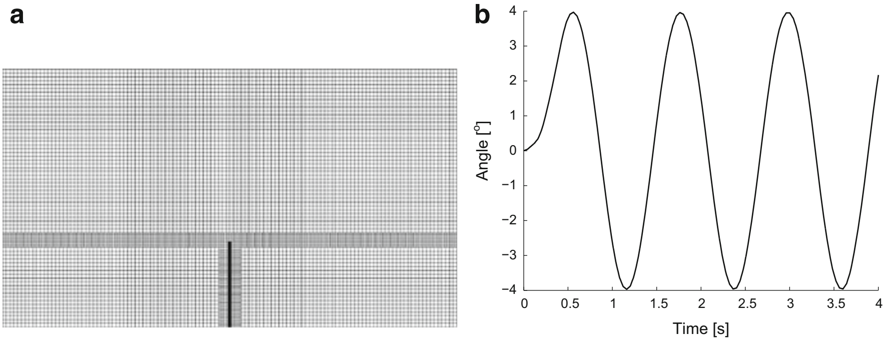


Fig. 2 (**a**) Baseline grid. (**b**) Temporal history of the rotation angle

the remaining walls. When 3D simulations are performed, no slip is enforced on all walls. The beam is clamped at the point A. When 2D simulations are performed, all nodes of the solid mesh are locked in the y-direction, leading to a plane strain assumption.

The baseline 2D grid, denoted grid 1, is shown in Fig. 2a. The grid is refined by halving the cell size and one refinement is added around the beam and the oil-air surface. The baseline grid consists of approximately 12,400 fluid cells and 100 solid elements. The number of solid elements remains constant during a simulation, but the number of fluid cells changes slightly due to the adaptive grid refinements. The fluid is discretized on an octree grid with cubic cells and the structure is meshed with 20-node hexahedral elements. The tank rotates around the point A and the temporal

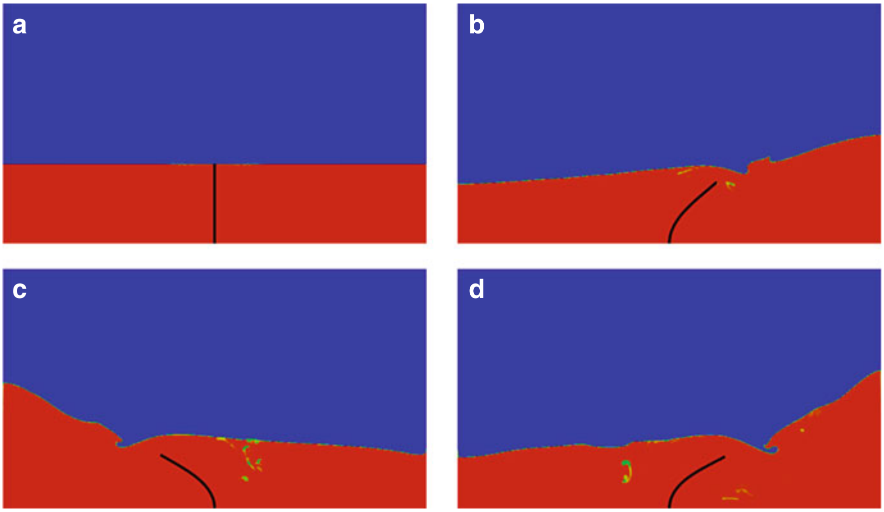


Fig. 3 Volume fraction: *Blue* corresponds to $\alpha = 0$ (air) and *red* corresponds to $\alpha = 1$ (oil). (**a**) $t = 0$ s. (**b**) $t = 1.25$ s. (**c**) $t = 1.85$ s. (**d**) $t = 2.5$ s

history of the rotation angle is shown in Fig. 2b. Numerical data for the history of the angle is available in [18]. Note that the tank motion is harmonic with period $T = 1.21$ s except at the first few tenths of a second, where a transient behavior can be seen.

In the simulations, the gravitation vector was rotated instead of rotating the whole domain. The centrifugal forces, arising from the fact that the simulation is performed in an accelerating coordinate system, have been neglected. This is justified because the angular velocity of the motion is small. As will be seen, good results are obtained with this approximation.

Four seconds of physical time are simulated, covering three full periods of the beam motion. Figure 3 shows snapshots from a 2D simulation at different time steps. The angular frequency of the forced rotation is close to the eigenfrequency of the system and therefore the waves grow larger with time. The beam undergoes large deformation due to the interaction with the fluid.

The displacement of the beam tip, measured in a coordinate system moving with the tank, is shown in Fig. 4. The agreement with the experimental data presented in [18] and the simulations in [19] is very good. The differences between the results obtained with grid 1 (12,400 fluid cells and 100 solid elements), grid 2 (53,200 fluid cells and 784 solid elements) and grid 3 (157,000 fluid cells and 3,136 solid elements) are small, indicating that grid convergence has been obtained. Figure 5 shows the displacement predicted with grid 2 for three different time steps. The differences are small, indicating that the time step is sufficiently short.

To investigate whether the differences between the 2D simulation and the experimental data originate from neglected 3D effects, 3D simulations were performed.

Fig. 4 Tip displacement of a beam in a rolling tank: grid convergence and comparison with reference data

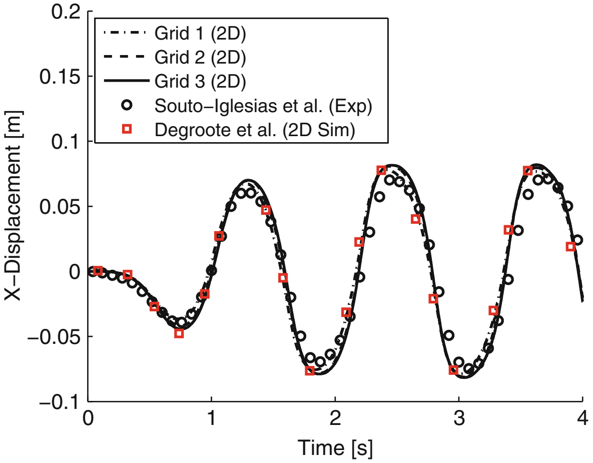
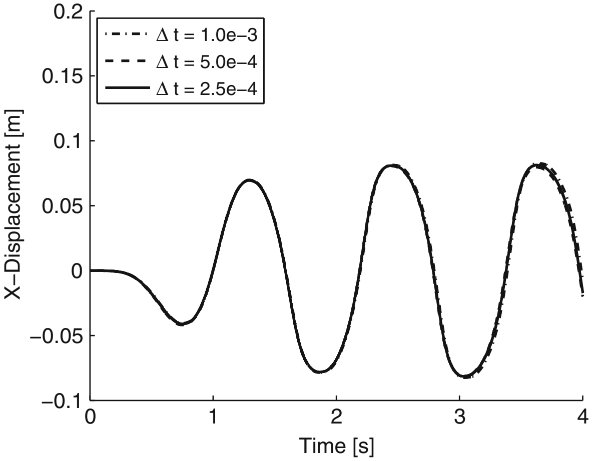


Fig. 5 Temporal convergence of tip displacement



This is indeed the case as shown in Fig. 6, where the 2D simulation and the experiments are compared to a 3D simulation with a cell size roughly corresponding to grid 2. The agreement between the 3D simulation and the experiment is excellent. It can be noted that the 2D simulation slightly overpredicts the amplitude, while the 3D simulation captures the amplitude very well. This is probably an effect of the walls with normal in the y-direction. The friction between the fluid and these walls will dissipate kinetic energy from the fluid, thus decreasing the amplitude of the motion. This effect is not captured in a 2D simulation, where symmetry (free slip) boundary conditions are applied to the walls with normal in the y-direction. The 3D effects are clearly visible in Fig. 7, that shows the beam and the oil-air interface.

The 2D simulations presented in Fig. 5 were performed without coupling iterations. However, Gauss-Seidel iterations were used in the 3D simulation to get a stable solution.

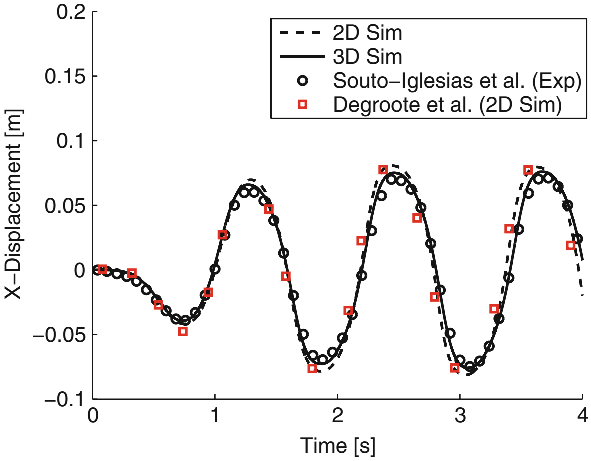


Fig. 6 Tip displacement of a beam in a rolling tank: comparison between our simulations and results from the literature

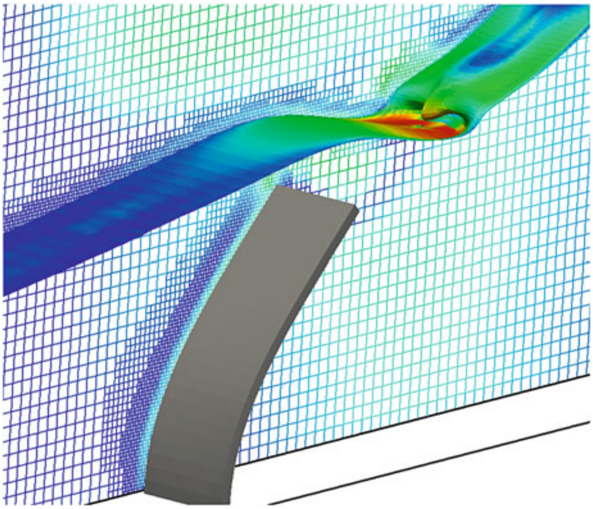


Fig. 7 Snapshot from the 3D simulation. The interface and the grid are colored by the fluid velocity and 3D effects are clearly visible on the oil-air interface

4 Conclusions

A framework for simulation of FSI has been developed and validated. Combining the IBM with the VOF method allows adaptive grid refinements around the structure and the oil-air surface without deterioration of the mesh quality. Using this method to couple the Navier-Stokes solver with a structural dynamics solver for large deformations results in a robust framework that allows complex three-dimensional FSI applications to be studied. The good agreement with previously published data demonstrates the accuracy of the method.

Acknowledgements This work was supported in part by the Sustainable Production Initiative and the Production Area of Advance at Chalmers. The support is gratefully acknowledged.

References

1. Hu, H., Patankar, N., Zhu, M.: Direct numerical simulation of fluid-solid systems using arbitrary lagrangian-eulerian technique. J. Comput. Phys. **169**, 427–462 (2001)
2. Monaghan, J.: Smoothed particle hydrodynamics. Rep. Prog. Phys. **68**, 1703–1759 (2005)
3. Cummins, S., Rudman, M.: An sph projection method. J. Comput. Phys. **152**, 584–607 (1999)
4. Onate, E., Idelsohn, S., Pin, F.D., Aubry, R.: The particle finite element method. An overview. Int. J. Comput. Methods **1**, 267–307 (2004)
5. Peskin, C.: Numerical analysis of blood flow in the heart. J. Comput. Phys. **25**, 220–252 (1977)
6. Majumdar, S., Iaccarino, G., Durbin, P.: Rans solvers with adaptive structured boundary non-conforming grids. Technical report, Center for Turbulence Research (2001). Annual Research Briefs
7. Mark, A., van Wachem, B.: Derivation and validation of a novel implicit second-order accurate immersed boundary method. J. Comput. Phys. **227**, 6660–6680 (2008)
8. Mark, A., Rundqvist, R., Edelvik, F.: Comparison between different immersed boundary conditions for simulation of complex fluid flows. Fluid Dyn. Mater. Process. **7**(3), 241–258 (2011)
9. Mark, A., Svenning, E., Rundqvist, R., Edelvik, F., Glatt, E., Rief, S., Wiegmann, A., Fredlund, M., Lai, R., Martinsson, L., Nyman, U.: Microstructure simulation of early paper forming using immersed boundary methods. TAPPI J. **10**(11), 23–30 (2011)
10. Svenning, E., Mark, A., Edelvik, F., Glatt, E., Rief, S., Wiegmann, A., Martinsson, L., Lai, R., Fredlund, M., Nyman, U.: Multiphase simulation of fiber suspension flows using immersed boundary methods. Nord. Pulp Pap. Res. J. **27**, 184–191 (2012)
11. Mark, A., Svenning, E., Edelvik, F.: An immersed boundary method for simulation of flow with heat transfer. Int. J. Heat Mass Transf. **56**, 424–435 (2013)
12. Forster, C., Wall, W., Ramm, E.: Artificial added mass instabilities in sequential staggered coupling of nonlinear structures and incompressible viscous flows. Comput. Methods Appl. Mech. Eng. **196**, 1278–1293 (2007)
13. Degroote, J., Bathe, K., Vierendeels, J.: Performance of a new partitioned procedure versus a monolithic procedure in fluid-structure interaction. Comput. Struct. **87**, 793–801 (2009)
14. Doormaal, J.V., Raithby, G.: Enhancements of the simple method for predicting incompressible fluid flows. Numer. Heat Transf. **7**, 147–163 (1984)
15. Rhie, C., Chow, W.: Numerical study of the turbulent flow past an airfoil with trailing edge separation. AIAA J1 **21**, 1527–1532 (1983)
16. Bonet, J., Wood, R.: Nonlinear Continuum Mechanics for Finite Element Analysis. Cambridge University Press, Cambridge (1997)
17. Botia-Vera, E., Bulian, G., Lobovsky, L.: Three sph novel benchmark test cases for free surface flows. In: 5th ERCOFTAC SPHERIC Workshop on SPH Applications (2010)
18. Sphercic benchmarks. http://canal.etsin.upm.es/ftp/SPHERIC_BENCHMARKS/ (2012)
19. Degroote, J., Souto-Iglesias, A., van Paepegem, W., Annerel, S., Bruggeman, P., Vierendeels, J.: Partitioned simulation of the interaction between an elastic structure and free surface flow. Comput. Methods Appl. Mech. Eng. **199**, 2085–2098 (2010)

Modelling of a Simplified Fluid-Structure Interaction Formulation

Julia Niemeyer and Bernd Simeon

Abstract A simplified formulation of the fluid-structure interaction problem is presented in order to analyze the coupling conditions and the effect of a moving fluid domain on the numerical solution. The resulting one-dimensional model equations are discretized by the finite element method in space and then solved by implicit timestepping schemes, with the coupling conditions explicitly enforced by means of corresponding constraint equations in a differential-algebraic formulation. First numerical results indicate an influence of the moving fluid mesh on the stability properties of commonly used time integrators.

1 Fluid-Structure Interaction in a Nutshell

Fluid-Structure Interaction problems arise in nearly all engineering fields where the motion of an elastic structure and the flow of a circulating fluid affect each other. The mathematical problem is described by Cauchy's equations in the solid part and by the Navier–Stokes equations in the fluid part.

Let $\Omega^t \subset \mathscr{R}^n, n = 2, 3$ be a bounded domain with $\Omega_t = \Omega^t_f \cup \Omega^t_s$, $t \geq t_0$, and $[t_0, T]$ the considered time interval. Here, Ω^t_f, Ω^t_s denote the fluid and solid subdomains, respectively. The interface boundary Γ^t_I is then given by $\Gamma^t_I := \Omega^t_f \cap \Omega^t_s$.

Assuming an elastic and nearly incompressible body, the deformation of the solid part is described by the displacement field d and the pressure p_s. The balance equations in the reference configuration Ω^0_s read

J. Niemeyer (✉) • B. Simeon
Felix-Klein Centre of Mathematics, Technical University of Kaiserslautern
Paul-Ehrlich-Straße 31, 67663 Kaiserslautern, Germany
e-mail: niemeyer@mathematik.uni-kl.de; simeon@mathematik.uni-kl.de

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_22,

$$\rho_s \ddot{d} - \operatorname{div}\, P(d, p_s) = f_s \quad \text{in } \Omega_s^0 \tag{1a}$$

$$g(d, p_s) = 0 \quad \text{in } \Omega_s^0 \tag{1b}$$

with the second Piola-Kirchhoff stress tensor

$$P(d, p_s) = \frac{\partial W}{\partial \nabla d} \tag{2}$$

and a strain energy function W that defines the material behaviour. This model of a solid allows for large deformation and hyperelastic material laws.

The flow of an incompressible Newtonian fluid is described by the velocity field u and the pressure p_f, and the corresponding equations are in general formulated in the Eulerian framework. To couple the fluid equations with the equations of motion (1) we introduce the *Arbitrary Lagrangian Eulerian* (ALE) formulation of the Navier–Stokes equations [9, 10].

Let Ω_f^0 denote the reference configuration, then the ALE map is defined by

$$\mathscr{A}_t : \Omega_f^0 \times [t_0, T] \to \Omega_f^t,\ (\xi, t) \mapsto \mathscr{A}_t(\xi) =: x(\xi, t), \tag{3}$$

and the domain velocity w at a reference point $\xi \in \Omega_f^0$ is given by

$$w(x, t) := \frac{\partial \mathscr{A}_t}{\partial t}|_\xi . \tag{4}$$

A more detailed description can be found in [4]. The Navier–Stokes equations on a moving domain read

$$\rho_f \dot{u} + \rho_f (u - w) \cdot \nabla u - \operatorname{div} \sigma = \rho_f f_f \qquad \text{in } \Omega_f^t \tag{5a}$$

$$\operatorname{div} u = 0 \qquad \text{in } \Omega_f^t \tag{5b}$$

with stress tensor

$$\sigma = -p_f I + \rho_f \nu \left(\nabla u + \nabla u^T\right). \tag{6}$$

To close the coupled system, we need to solve an additional partial differential equation for the domain velocity that describes the domain movement. This could be done by considering the domain as an elastic solid and solving the equations of elastodynamics [5]. Another approach is to use the *harmonic extension* or the biharmonic extension [11]. However, since we are interested in the effect of the moving grid on the stability properties of the time integration schemes, we will assume a known ALE map in the rest of this paper.

Of particular interest are the coupling conditions between the solid and the fluid part. These interface equations on Γ_I^t are given as

$$\sigma \cdot n_f = P \cdot n_s, \quad u = \dot{d}, \quad d = \mathscr{A}_t \tag{7}$$

and stand for the equality of the stresses in normal direction and for the equality of the displacement and velocity fields at the interface.

2 Simplified Fluid-Structure Interaction Formulation

We are interested in studying the effects of the ALE formulation on the properties of the usual time integration schemes, inspired by the work of [1,6] in an advection-diffusion framework without the interaction with a solid. To simplify the equations, we will consider linear models in both subdomains and restrict ourselves to the one-dimensional case. The extension to higher dimensional models with $n = 2, 3$ is straightforward.

In our simplified formulation, a wave equation models the deformation in the solid part, and the fluid motion is replaced by a linear advection-diffusion equation where we also introduce the ALE formulation as in [6]. Let $\Omega^t \subset \mathscr{R}$ with $\Omega^t = \Omega^t_s \cup \Omega^t_f$ and consider the time interval $[t_0, T]$. The coupled system can be formulated as

$$\ddot{d} - \kappa_s \Delta d = 0 \qquad \text{in } \Omega^t_s \tag{8a}$$

$$\dot{u} - \kappa_f \Delta u + \nu \nabla u - w \nabla u = 0 \qquad \text{in } \Omega^t_f \tag{8b}$$

$$\nabla u \cdot n_f = \nabla d \cdot n_s \qquad \text{on } \Gamma^t_I \tag{8c}$$

$$u = \dot{d} \qquad \text{on } \Gamma^t_I \tag{8d}$$

$$d = \mathscr{A}_t \qquad \text{on } \Gamma^t_I \tag{8e}$$

with a given ALE map $\mathscr{A}_t$ and the domain velocity w as in (4).

In the next step, we apply a linear finite element method to discretize the simplified FSI problem (8) in space. Let $\Omega^t_h = \Omega^t_{h,s} \cup \Omega^t_{h,f}$ be the space-discrete domain. The space-discrete solution variables are given by $d_h(x,t) := \sum_{i=1}^{\mathscr{N}_s} d_i(t)\varphi^i_s(x)$ and $u_h(x,t) := \sum_{i=1}^{\mathscr{N}_f} u_i(t)\varphi^i_f(x)$ where φ^j_i, $j \in \{s, f\}$, denotes the i-th finite element basis function and $\mathscr{N}_j$, $j \in \{s, f\}$, the number of knots in the subspaces, respectively.

When setting up the finite-dimensional discretized analogue of (8), the interface conditions require particular attention. While the Neumann condition (8c) is directly included in the weak problem formulation, the Dirichlet conditions can be either treated as explicit constraints or implicitly enforced by means of eliminating the corresponding degrees of freedom. We choose here a differential-algebraic approach and employ Lagrange multipliers to enforce the constraints [2]. Introducing the discrete Lagrange multiplier Λ for the coupling conditions (8d) and (8e) [3], we end up with a differential-algebraic system

$$M(t)\dot{z}_h(t) + K(t, w_h)z_h(t) + B^T \Lambda(t) = 0 \qquad \text{in } \Omega^t_h \tag{9a}$$

$$B z_h(t) - b(t) = 0 \qquad \text{in } \Omega^t_h \tag{9b}$$

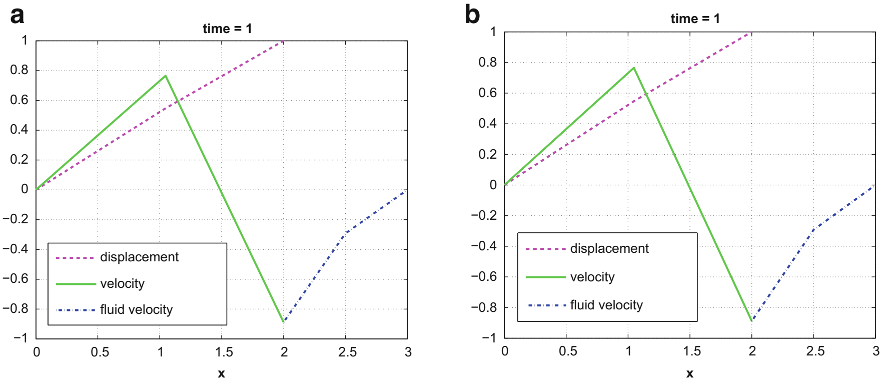


Fig. 1 Solution at time $t = 1$ with $\Delta t = 0.1$. (**a**) DAE solution. (**b**) ODE solution

with unknowns $z_h = (d_h, u_h)^T$ and Λ. The coupling conditions are formulated by means of a linear constraint with matrix B and inhomongeneity $b(t)$. Moreover, the discrete domain velocity $w_h := \frac{\partial \mathscr{A}_{h,t}}{\partial t}$ shows up in the discretized advection term, with $\mathscr{A}_{h,t}$ describing the space-discrete ALE map.

Due to the full rank of B, the index of the DAE (8) is two. Integrating differential algebraic systems with an index greater than one like (9a) may lead to numerical difficulties. As a remedy, it is possible to reduce the index by differentiating the algebraic constraints and solving for the Lagrange multiplier as $\Lambda(z_h(t), t)$. Upon inserting this expression into the semi-discrete system, one obtains an ordinary differential equation

$$M(t)\dot{z}_h(t) + K(t, w_h)z_h(t) + B^T \Lambda(z_h(t), t) = 0. \tag{10}$$

Because of the linearity of the constraints, no drift off from the original constraints shows up as long as the initial values are consistent and $b(t)$ is a linear function.

3 Numerical Results

Let $\Omega^0 = [0, 2]$ be the considered bounded domain with subdomains $\Omega_s^0 = [0, 1]$ and $\Omega_f^0 = [1, 2]$. The time interval is $[0, \pi]$. As mentioned above Eq. (8) are discretized in space using the linear finite element method. We choose the implicit Euler scheme and the midpoint rule to integrate the semi-discrete system. The tested ALE maps are

$$\mathscr{A}_t^1(\xi) = \xi + t, \qquad \mathscr{A}_t^2(\xi) = \xi + 10t, \qquad \mathscr{A}_t^3(\xi) = \xi + \sin(t) \tag{11}$$

and the resulting solution is displayed in Fig. 1. A time integration scheme is said to be stable if the eigenvalues of the increment function lie in the stability region of the

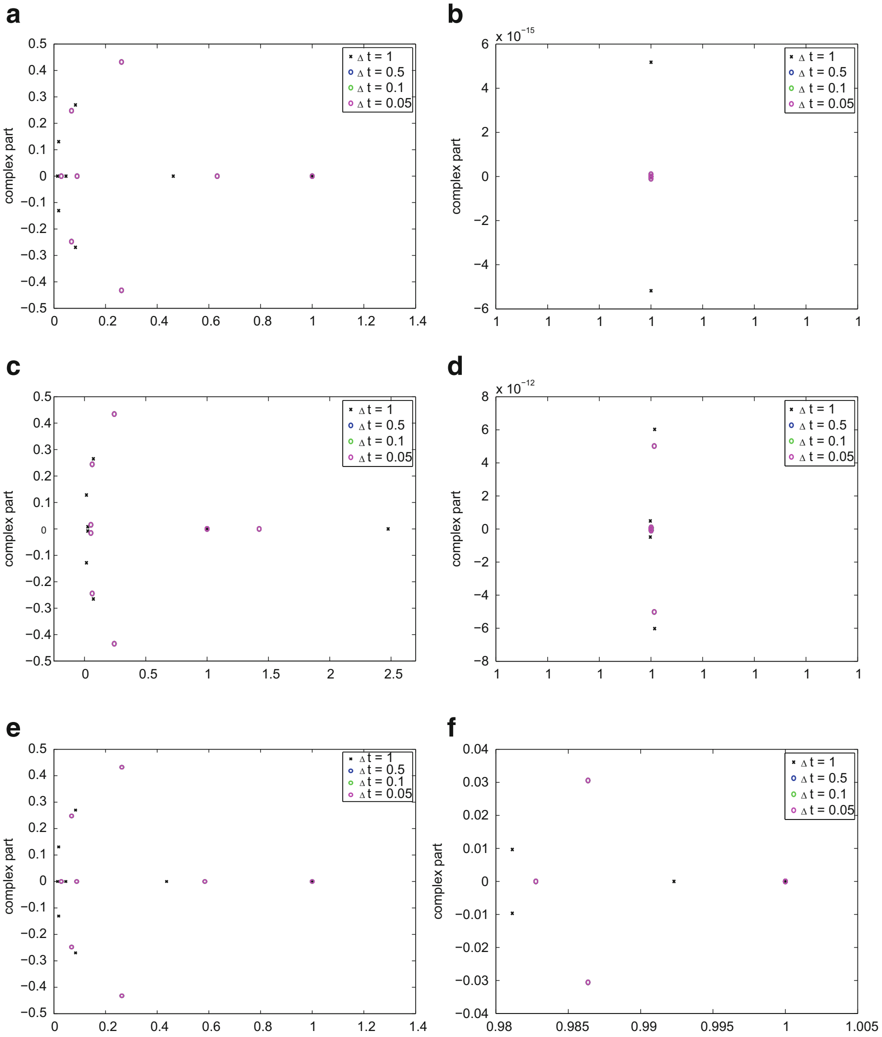


Fig. 2 Eigenvalues at time $t = 1$ with different Δt. (**a**) Implicit Euler, $\mathscr{A}_t^1$. (**b**) Midpoint rule, $\mathscr{A}_t^1$. (**c**) Implicit Euler, $\mathscr{A}_t^2$. (**d**) Midpoint rule, $\mathscr{A}_t^2$. (**e**) Implicit Euler, $\mathscr{A}_t^3$. (**f**) Midpoint rule, $\mathscr{A}_t^3$

integrator [8]. Therefore we plotted the eigenvalues in the complex plane according to different used time steps Δt. As one can see in Fig. 2b, d, f using the midpoint rule as a time integrator the eigenvalues are bounded for every used time step and every choice of the ALE map (11). While integrating the ODE using the implicit Euler scheme the eigenvalues are only bounded for every time step using the ALE map $\mathscr{A}_t^1$ and $\mathscr{A}_t^3$ which lead to a small domain velocity w compared to the time step Δt, Fig. 2a, e. In contrast the disposal of $\mathscr{A}_t^2$ cause a growth of the real part of the eigenvalues.

4 Conclusion

We have formulated a simplified linear fluid-structure interaction problem where the coupling is expressed using the Lagrange multiplier technique. This results in an index-two differential-algebraic system. To derive the corresponding ordinary differential system, the algebraic constraints are differentiated and written as a function of the solution variable.

In the solution plots Fig. 1 we see no drift in the original constraints as stated in Sect. 2. The look at the eigenvalues give us some hints of stability problems in commonly known unconditionally stable time integration schemes as the implicit Euler scheme as stated in [1, 6, 7] in the context of the geometric conservation law. Therefore a more detailed analysis of the influence of the coupling conditions on the stability properties of the time integrators seems to be worthwhile.

Acknowledgements Supported within the BMBF project: SNiMoRed, Multidisziplinäre Simulation, nichtlineare Modellreduktion und proaktive Regelung in der Fahrzeugdynamik.

References

1. Boffi, D., Gastaldi, L.: Stability and geometric conservation laws for ale formulations. Comput. Methods Appl. Mech. Eng. **193**, 4717–4739 (2004)
2. Brezzi, F., Fortin, M.: Mixed and Hybrid Finite Element Methods. Springer, New York (1991)
3. Doerfel, M.R., Simeon, B.: Fluid-structure-interaction: acceleration of strong coupling by preconditioning of the fixed-point iteration. In: Proceedings of the ENUMATH 2011 Conference (2011)
4. Donea, J., Huerta, A.: Finite Element Methods for Flow Problems. Wiley, New York (2003)
5. Farhat, C., Lesoinne, M., Maman, N.: Mixed explicit/implicit time integration of coupled aeroelastic problems: three-field formulation, geometry conservation and distributed solution. Int. J. Numer. Methods Fluids **21**, 807–835 (1995)
6. Formaggia, L., Nobile, F.: A stability analysis for the arbitrary Lagrangian Eulerian formulation with finite elements. East West J. Numer. Math. **7**, 105–132 (1999)
7. Gastaldi, L.: A priori error estimates for the arbitrary Lagrangian Eulerian formulation with finite elements. East West J. Numer. Math. **9**, 123–156 (2001)
8. Hairer, E., Norsett, S., Wanner, G.: Solving Ordinary Differential Equations I. Springer, New York (1987)
9. Nobile, F.: Numerical approximation of fluid-structure interaction problems with applications to haemodynamics. Ph.D. thesis, Department of Mathematics, école Polytechnique Fédèrale de Lausanne, Switzerland (2001)
10. Wall, W.A.: Fluid-Struktur-Interaktion mit stabilisierten Finiten-Elementen. Ph.D. thesis, Institut für Baumechanik, Universität Stuttgart (1999)
11. Wick, T.: Fluid-structure interactions using different mesh motion techniques. Comput. Struct. **89**, 1456–1467 (2011)

Sinking Bubbles in Stout Beers

Cathal P. Cummins, Eugene S. Benilov, and William T. Lee

Abstract Anyone who has ever tried Guinness or another stout beer knows that the bubbles in the glass appear to sink. This suggests that they are driven by a downward flow, the velocity of which exceeds the upward velocity of the bubble due to the Archimedean force. The existence of such a flow near the wall of the glass implies that there must be an upward flow somewhere in the interior. The mechanism of such a circulation is, however, unclear. In this work, we demonstrate that the circulation in a glass of stout—or any other container with a bubbly liquid—is determined by the container's shape. If it narrows downwards (as the stout glass does), the circulation is directed downwards near the wall and upwards in the interior. If the container widens downwards, the circulation is opposite to that described above.

1 Introduction

Bubbles in liquids normally float up due to the Archimedean force—yet those in so-called stout beers appear to go down. Such counter-intuitive phenomena rarely occur in our everyday life, challenging the curiosity of both scientists and lay people.

Interestingly, even though the effect of bubbles sinking in Guinness is widely known and that the bubbles/liquid interaction in stouts has been examined before [1], no explanation of this puzzling phenomenon has been put forward so far. In this work, we shall first describe the properties of Guinness as a two-phase medium and explain the basic mechanism that drives bubbles in Guinness downwards.

C.P. Cummins (✉) • E.S. Benilov • W.T. Lee
MACSI, University of Limerick, Limerick, Ireland
e-mail: cathal.cummins@ul.ie; eugene.benilov@ul.ie; william.lee@ul.ie

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_23,

2 Properties of Stout Beers

We shall model Guinness by a liquid of density ρ_l and viscosity μ_l , with randomly distributed bubbles of gas of density ρ_g and viscosity μ_g. For a temperature of 6 °C (recommended for consumption of Guinness by its producer Diageo) and normal atmospheric pressure, we have

$$\begin{aligned} \rho_l &= 1{,}007\,\mathrm{kg\,m^{-3}} & \qquad \mu_l &= 2.06 \times 10^{-3}\,\mathrm{Pa\,s} \\ \rho_g &= 1.223\,\mathrm{kg\,m^{-3}} & \qquad \mu_l &= 0.017 \times 10^{-3}\,\mathrm{Pa\,s} \end{aligned}$$

where the former values have been measured by ourselves and verified against the extrapolation formula of [3]. To check whether the bubble shapes differ from spheres, we introduce the Bond number

$$\mathrm{Bo} = \frac{\rho_l g d_b^2}{\sigma}$$

where d_b is the bubbles' characteristic diameter, σ is the surface tension of the liquid/gas interface, and g is the acceleration due to gravity. Assuming $d_b = 122\,\mu\mathrm{m}$ (as reported in [1]) and $\sigma = 0.745\,\mathrm{N\,m^{-1}}$ (which corresponds to water/air interface), we obtain Bo $\approx$ 0.002 which is sufficiently small to assume that bubbles in Guinness are spherical.

Note also that Guinness (as well as the vast majority of "real" liquids) contains a lot of surfactants, which make the bubbles behave as rigid spheres. This allows one to estimate the characteristic bubble velocity u_b using the Stokes formula for a rigid sphere,

$$u_b = \frac{\rho_l - \rho_b g d_b^2}{18\mu_l} \approx 3.96\,\mathrm{mm\,s^{-1}}.$$

Estimating the corresponding Reynolds number

$$\mathrm{Re} = \frac{\rho_l u_b d_b}{\mu_l} \approx 0.24,$$

confirms that the Stokes formula yields a qualitatively correct value for u_b. Furthermore, given that u_b is much smaller than the speed of sound, the gas can be treated as incompressible.

Finally, we introduce the void fraction, f, i.e. the gas's share of the volume of the liquid/gas mixture. For canned Guinness, $f \approx 0.05$ (see [1]), whereas for draught Guinness served in pubs, $f \approx 0.1$ (according to our own measurements). Note, however, that, traditionally, bartenders first fill, say, 80 % of the glass and wait until it has fully settled (i.e. all the bubbles have gone out of the liquid into the foamy head), after which they would fill the glass full. Thus, when Guinness is served to the customer, the void fraction can be estimated as $f \approx 0.02$, which is the value used in this work.

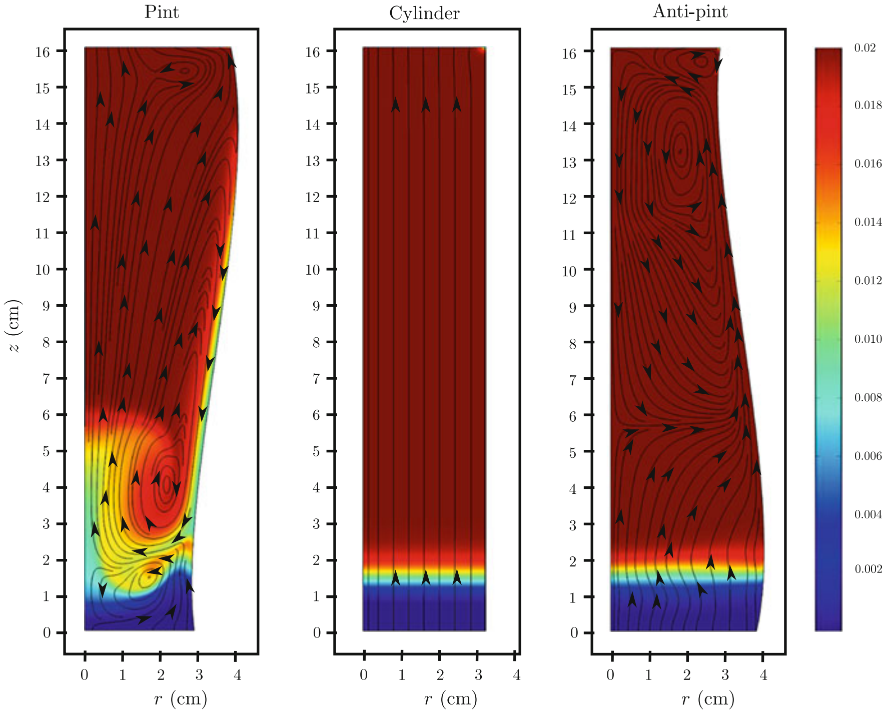


Fig. 1 Numerical simulations of bubbly flows for the pint, a cylinder, and an anti-pint. The *curves* show the streamlines for the bubbles, the *color* shows the void fraction f. The snapshots displayed correspond to $t = 4$ s. Observe the region of reduced f near the wall of the pint (the near-wall region of increased f in the anti-pint is not visible in this figure, but can be observed in Fig. 2)

3 Mathematical Modelling

Since we attempt to explain the downflow of bubbles in Guinness by the geometry of the container and not by a physical effect, we shall use the standard model for bubbly flows included in the COMSOL Multiphysics package, based on the finite element method. We shall not discuss this model's physical foundations, as they are described in detail in [2], but mention only that it assumes that the bubbles are all of the same size. In view of the problem's axial symmetry, the axi-symmetric version of the model is used.

Two geometries were examined (see Fig. 1): a pint and an "anti-pint", i.e. the pint turned upside-down. In both cases the initial distribution of bubbles was uniform. The results of typical simulations are shown in Fig. 1. One can see that an elongated vortex arises near the sloping part of the pint container, resulting in a downflow of bubbles along the wall. A similar vortex also exists in the anti-pint, but it rotates in the opposite direction and, thus, causes an upward flow.

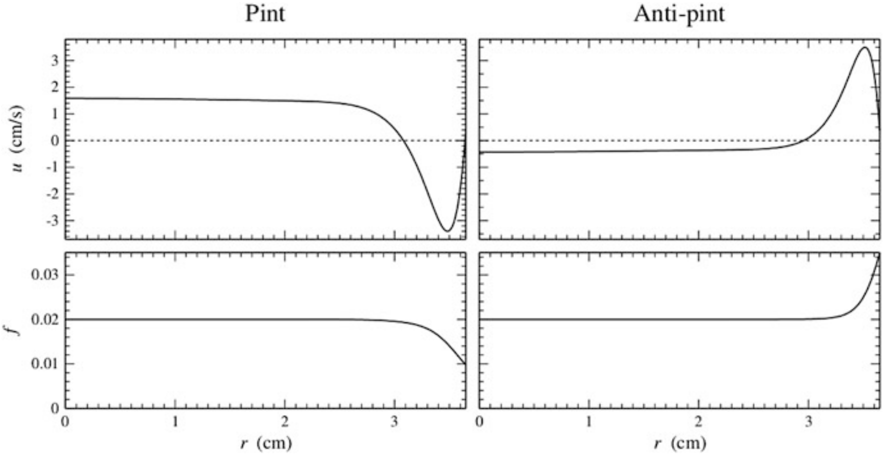


Fig. 2 The half-height cross-sections of the vertical velocity u and the void fraction f for the pint and anti-pint geometries (these graphs correspond to the (r, z) diagrams shown in Fig. 1). The *dotted lines* in the upper panels separate the regions of upward/downward flow

The latter results can be explained using the same kinematic argument as those for the pint geometry: if the container widens downwards, bubbles travel towards the wall (as illustrated in Fig. 1 (right)). This increases the near-wall density of bubbles and, thus, the upward drag applied to the liquid, resulting in an upward flow.The above argument, for both pint and anti-pint, is corroborated by the cross-sections of the bubble density and velocity shown in Fig. 2.

References

1. Robinson, M., Fowler, A.C., Alexander, A.J., O'Brien, S.B.G.: Waves in Guinness. Phys. Fluids **20**(6), 067101 (2008)
2. Sokolichin, A., Eigenberger, G., Lapin, A.: Simulation of buoyancy driven bubbly flow: established simplifications and open questions. AIChE J. **50**(1), 24–45 (2004)
3. Zhang, Y., Xu, Z.: *Fizzics* of bubble growth in beer and champagne. Elements **4**(1), 47–49 (2008)

Analysis of Two-Phase Flow in the Gas Diffusion Layer of a Polymer Electrolyte Fuel Cell

Andrew Gordon and Michael Vynnycky

Abstract The last decade has seen a proliferation of modelling activity on the polymer electrolyte fuel cell (PEFC); an important subset of this activity is the modelling of the two-phase (gas/liquid) flow that occurs in the porous gas diffusion layer (GDL) on the cathode (Djilali, Energy 32:269–280, 2007; Gurau and Mann, SIAM J. Appl. Maths 70:410–454, 2009). The prevailing approach employs a generalized form of Darcy's law, which has been widely used over the last several decades to analyze the movement of oil and water in soils and porous rock (Bear, Dynamics of Fluids in Porous Media, American Elsevier, New York, 1972). Applied to water transport in fuel cells, the Darcy model characterizes the response of the porous material by the capillary pressure, the gas and liquid phase relative permeabilities, and the effective gas diffusion coefficient, all of which depend on the fraction of the local pore volume occupied by liquid water; an additional feature is that the porous medium can be either hydrophobic or hydrophilic. The majority of approaches have, however, been primarily numerical, which has obscured some of the properties of the model. Here, using asymptotic methods, we extend earlier work (Vynnycky, Appl. Math. Comp. 189:1560–1575, 2007) to demonstrate how the degree of water saturation depends on the liquid phase relative permeability, as well as how the model behaves when the GDL is only just hydrophilic.

1 Introduction

The last decade has seen a proliferation of modelling activity on the polymer electrolyte fuel cell (PEFC); an important subset of this activity is the modelling of the two-phase flow that occurs in the gas diffusion layer (GDL) on the cathode [3,4].

A. Gordon (✉) • M. Vynnycky
MACSI, University of Limerick, Limerick, Ireland
e-mail: andrew.gordon@ul.ie; michael.vynnycky@ul.ie

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_24,

The prevailing approach employs a generalized form of Darcy's law, which has been widely used over the last several decades to analyze the movement of oil and water in soils and porous rock [1]. Applied to water transport in fuel cells, the Darcy model characterizes the response of the porous material by the capillary pressure, the gas and liquid phase relative permeabilities, and the effective gas diffusion coefficient, all of which depend on the fraction of the local pore volume occupied by liquid water. Although there are by now numerous other models for two-phase flow in the gas diffusion layer (GDL), this one is still frequently used [9] and therefore merits closer scrutiny.

In this paper, we provide asymptotic analysis and a numerical solution for the simplest possible model for this situation, by considering a 1D steady state isothermal model for the GDL.

2 Model Equations

We employ the version of multi-fluid model formulation given in [11]. There is essentially no difference between this version and that used by other authors, as was shown numerically in [11]; it does, however, remove the more ad hoc nature of the way that the inter-phase transfer is treated, thereby making the analysis mathematically more transparent. We consider an isothermal GDL of thickness H in which there is two-phase flow. As in all earlier papers on two-phase flow in the GDL of a PEFC, we assume negligible quantities of oxygen and nitrogen in the liquid phase.

Starting with steady state conservation equations for both gas and liquid phases, we have

$$\frac{dn_{\mathrm{O_2}}^{(\mathrm{g})}}{dy} = 0, \quad \frac{dn_{\mathrm{N_2}}^{(\mathrm{g})}}{dy} = 0, \quad \frac{dn_{\mathrm{H_2O}}^{(\mathrm{g})}}{dy} = -\dot{m}_{\mathrm{H_2O}}, \quad \frac{d}{dy}\left(\rho^{(\mathrm{l})} v^{(\mathrm{l})}\right) = \dot{m}_{\mathrm{H_2O}}, \tag{1}$$

where $n_i^{(\mathrm{g})}$ denotes the mass flux for species i, $\rho^{(\mathrm{l})}$ denotes the liquid density, $v^{(\mathrm{l})}$ denotes the liquid velocity and $\dot{m}_{\mathrm{H_2O}}$ is the interface mass transfer of water between the gas and liquid phase. Adding the equations in (1) together gives

$$\frac{d}{dy}\left(\rho^{(\mathrm{g})} v^{(\mathrm{g})} + \rho^{(\mathrm{l})} v^{(\mathrm{l})}\right) = 0, \tag{2}$$

where $\rho^{(\mathrm{g})}$ is the gas mixture density and $v^{(\mathrm{g})}$ is the mass-averaged velocity of the gas phase; this eliminates $\dot{m}_{\mathrm{H_2O}}$, so that only the first two equations in (1) and Eq. (2) need be considered, although $\dot{m}_{\mathrm{H_2O}}$ can be computed a posteriori if necessary [11]. In turn, we have

$$\begin{aligned} v^{(g)} &= -\frac{\kappa\kappa_{rel}^{(g)}(s)}{\mu^{(g)}}\frac{dp^{(g)}}{dy}, \quad v^{(l)} = -\frac{\kappa\kappa_{rel}^{(l)}(s)}{\mu^{(l)}}\frac{dp^{(l)}}{dy}, \\ n_i{}^{(g)} &= \rho^{(g)}\omega_i^{(g)}v^{(g)} + j_i^{(g)}, \quad i = H_2O, O_2, N_2; \end{aligned} \tag{3}$$

in (3), $\kappa_{rel}^{(g)}(s)$ is the gas relative permeability, $\kappa_{rel}^{(l)}(s)$ is the liquid relative permeability, $\mu^{(g)}$ and $\mu^{(l)}$ are, respectively, the gas and liquid phase dynamic viscosities, $\omega_i^{(g)}$ is the mass fraction and $j_i^{(g)}$ describes the diffusion-driven transport. For the latter, we use

$$j_i^{(g)} = -\rho^{(g)}\omega_i^{(g)}\gamma^{3/2}(1-s)\sum_{j=H_2O,N_2,O_2}\tilde{D}_{ij}\left(\frac{dx_j^{(g)}}{dy} + \frac{x_j^{(g)} - \omega_j^{(g)}}{p^{(g)}}\frac{dp^{(g)}}{dy}\right),$$

where $\tilde{D}_{ij}$ is the (i, j)-component of the multicomponent Fick diffusivity matrix—which is modified by $(1-s)$ to take account of the liquid phase and the porosity of the GDL, γ- and $x_j^{(g)}$, denoting the mole fraction of species j, is related to $\omega_i^{(g)}$ by $x_i^{(g)} = M^{(g)}\omega_i^{(g)}/M_i, \quad i = H_2O, N_2, O_2$; in turn, this introduces the relative molecular weights of nitrogen, oxygen and water (M_{N_2}, M_{O_2} and M_{H_2O}, respectively) and the mixture molecular weight, $M^{(g)}$, given by $M^{(g)} = M_{H_2O}x_{H_2O}^{(g)} + M_{N_2}x_{N_2}^{(g)} + M_{O_2}x_{O_2}^{(g)}$. For a ternary system, these are related to the multicomponent Maxwell-Stefan diffusivities (D_{ij}) through, for $i, j, k = H_2O, N_2, O_2$,

$$\tilde{D}_{ii} = \frac{\frac{(\omega_j+\omega_k)^2}{x_i D_{jk}} + \frac{\omega_j^2}{x_j D_{ik}} + \frac{\omega_k^2}{x_k D_{ij}}}{\frac{x_i}{D_{ij}D_{ik}} + \frac{x_j}{D_{ij}D_{jk}} + \frac{x_k}{D_{ik}D_{jk}}}, \quad \tilde{D}_{ij} = -\left(\frac{\frac{\omega_i(\omega_j+\omega_k)}{x_i D_{jk}} + \frac{\omega_j(\omega_j+\omega_k)}{x_j D_{ik}} - \frac{\omega_k^2}{x_k D_{ij}}}{\frac{x_i}{D_{ij}D_{ik}} + \frac{x_j}{D_{ij}D_{jk}} + \frac{x_k}{D_{ik}D_{jk}}}\right),$$

$i \neq j \neq k$. Thence, for gases at low density, the multicomponent Maxwell-Stefan diffusivities, D_{ij}, can be replaced with the binary diffusivities, $\hat{D}_{ij}$, for all pairs of species in the mixture; explicit expressions for these, based on the Chapman-Enskog theory, can be found in [2]. In addition to the above, we must have

$$x_{O_2}^{(g)} + x_{N_2}^{(g)} + x_{H_2O}^{(g)} = 1;$$

there is no differential equation for $x_{H_2O}^{(g)}$, since it takes the saturation value, i.e.

$$\begin{aligned} &x_{H_2O}^{(g)} = p_{H_2O}^{sat}(T)/p^{(g)}, \\ &\text{where} \quad p_{H_2O}^{sat}(T) = 10^{\left(2.8206+0.02953t-9.1837\times10^{-5}t^2+1.4454\times10^{-7}t^3\right)}, \end{aligned}$$

with T as the temperature and $t = T - 273.15$.

Constitutive relations are required for $\rho^{(g)}$, $\kappa_{rel}^{(g)}$ and $\kappa_{rel}^{(l)}$. For $\rho^{(g)}$, we use the ideal gas law,

$$\rho^{(g)} = p^{(g)} M^{(g)} / \mathscr{R} T,$$

where $\mathscr{R}$ is the universal gas constant. We do not yet take any particular forms for $\kappa_{rel}^{(g)}(s)$ and $\kappa_{rel}^{(l)}(s)$, other than to require that $\kappa_{rel}^{(g)}(0) = 1$, $\kappa_{rel}^{(g)}(1) = 0$, $\kappa_{rel}^{(l)}(0) = 0$, $\kappa_{rel}^{(l)}(1) = 1$. A further relation is required to relate $p^{(g)}$ and $p^{(l)}$; this is done via the capillary pressure, p_c, which is itself a function of the saturation. By definition, $p_c = p_{nw} - p_w$, where p_{nw} and p_w denote the pressures of the non-wetting and wetting phases, respectively [1,5]; so,

$$p_c = \begin{cases} p^{(g)} - p^{(l)} & \text{if} \quad \theta_c < 90^o \\ p^{(l)} - p^{(g)} & \text{if} \quad \theta_c > 90^o \end{cases},$$

where θ_c is the contact angle for water in the GDL.

The boundary conditions are: at $y = 0$,

$$\rho^{(g)} v^{(g)} + \rho^{(l)} v^{(l)} = \frac{i_c}{4F} \{2(1 + 2\alpha) M_{H_2O} - M_{O_2}\}, \quad n_{O_2}^{(g)} = -\frac{M_{O_2} i_c}{4F}, \quad n_{N_2}^{(g)} = 0, \tag{4}$$

where α is the number of water molecules dragged through the membrane by each proton, i_c is the current density and F is Faraday's constant (96,487 C mol^{-1}); at $y = H$,

$$x_{O_2}^{(g)} = x_{O_2}^{in}, \quad s = 0, \quad p^{(g)} = p^{out}. \tag{5}$$

In particular, i_c is normally chosen to be of the form

$$i_c = (1 - s)\, i_c^*, \quad \text{with} \quad i_c^* = \frac{i_0 p^{(g)} x_{O_2}^{(g)}}{p_{ref}^{(g)}} \left(2^{(T-273)/10}\right) \exp\left(\frac{F \eta_c}{2\mathscr{R} T}\right),$$

where i_c^* is current density that one would expect in the absence of liquid water [6,7], i_0 is the exchange current density, η_c is the cathodic overpotential and p_{ref} is a reference pressure, which we set to be p^{out}.

3 Nondimensionalization and Analysis

We write

$$Y = y/H, \quad V^{(g)} = \frac{v^{(g)}}{[v^{(g)}]}, \quad V^{(l)} = \frac{v^{(l)}}{[v^{(l)}]}, \quad P^{(g)} = \frac{p^{(g)} - p^{out}}{\Delta P}, \quad P^{(l)} = \frac{p^{(l)} - p^{out}}{\Delta P},$$

$$P_c = \frac{p_c}{[p_c]}, \quad \varrho^{(g)} = \frac{\rho^{(g)}}{\left[\rho^{(g)}\right]}, \quad \mathscr{M}^{(g)} = \frac{M^{(g)}}{\left[M^{(g)}\right]}, \quad J_i^{(g)} = \frac{j_i^{(g)}}{\left[\rho^{(g)}\right][v^{(g)}]},$$

$$\hat{\mathscr{D}}_{ij} = \frac{\hat{D}_{ij}}{[D]}, \quad \tilde{\mathscr{D}}_{ij} = \frac{\tilde{D}_{ij}}{[D]}, \quad I_c = \frac{i_c}{[i]}, \quad \tilde{\eta}_c = \frac{\eta_c}{[\eta_c]},$$

where $\Delta P = [p_c] = \sigma/\kappa^{1/2}$, $[v^{(g)}] = D/H$, and

$$[v^{(l)}] = \frac{\left[\rho^{(g)}\right][v^{(g)}]}{\left[\rho^{(l)}\right]}, \quad [i] = i_0\left(2^{(T-273)/10}\right)\exp\left(\frac{F\,[\eta_c]}{2\mathscr{R}T}\right);$$

note that $[\eta_c]$ varies between 0 and around 0.6 V. The governing equations are now, for $0 \le Y \le 1$,

$$\frac{d}{dY}\left(\varrho^{(g)}\omega_i^{(g)}V^{(g)} + J_i^{(g)}\right) = 0, \text{ for } \quad i = \mathrm{O_2}, \mathrm{N_2}, \quad \frac{d}{dY}\left(\varrho^{(g)}V^{(g)} + V^{(l)}\right) = 0, \tag{6}$$

where $\varrho^{(g)} = [M]\,p^{out}\left(1 + \Pi P^{(g)}\right)\mathscr{M}^{(g)}/\left[\rho^{(g)}\right]RT$, with $\Pi = \Delta P/p^{out}$, and

$$J_i^{(g)} = -\varrho^{(g)}\omega_i^{(g)}\gamma^{3/2}(1-s)\sum_{j=\mathrm{H_2O,N_2,O_2}}\tilde{\mathscr{D}}_{ij}\left(\frac{dx_j^{(g)}}{dY} + \frac{\Pi\left(x_j^{(g)} - \omega_j^{(g)}\right)}{1 + \Pi P^{(g)}}\frac{dP^{(g)}}{dY}\right).$$

Also,

$$\widetilde{Ca}V^{(g)} = -\kappa_{rel}^{(g)}(s)\frac{dP^{(g)}}{dY}, \quad \widetilde{Ca}V^{(l)} = -\chi\kappa_{rel}^{(l)}(s)\frac{dP^{(l)}}{dY}, \quad P_c(s) = \pm\left(P^{(g)} - P^{(l)}\right), \tag{7}$$

where $\chi = \rho^{(l)}\left[\mu^{(g)}\right]/\left[\rho^{(g)}\right]\mu^{(l)}$ as well as $\widetilde{Ca} = Ca\left(H/\kappa^{1/2}\right)$ hold, with $Ca(= \left[\mu^{(g)}\right][v^{(g)}]/\sigma)$. Combining the second equation in (6) and the three equations in (7) eliminates $V^{(l)}$ and $P^{(l)}$ to give

$$\frac{d}{dY}\left(\widetilde{Ca}\left\{\varrho^{(g)} + \chi\frac{\kappa_{rel}^{(l)}(s)}{\kappa_{rel}^{(g)}(s)}\right\}V^{(g)} \pm \chi\kappa_{rel}^{(l)}(s)P_c'(s)\frac{ds}{dY}\right) = 0; \tag{8}$$

in (8), $'$ denotes differentiation with respect to s, whereas the ± sign refers to the case $\theta_c \lesseqgtr 90^{\mathrm{o}}$, respectively.

As for the boundary conditions, Eqs. (4) and (5) give, at $Y = 0$ and 1, respectively,

$$V^{(l)} + \varrho^{(g)}V^{(g)} = \frac{\Omega}{4}\left\{2(1+2\alpha)\mathscr{M}_{\mathrm{H_2O}} - \mathscr{M}_{\mathrm{O_2}}\right\}I_c,$$

$$\varrho^{(g)}\omega_{\mathrm{O_2}}^{(g)}V^{(g)} + J_{\mathrm{O_2}}^{(g)} = -\frac{\mathscr{M}_{\mathrm{O_2}}\Omega I_c}{4}, \quad \varrho^{(g)}\omega_{\mathrm{N_2}}^{(g)}V^{(g)} = -J_{\mathrm{N_2}}^{(g)},$$

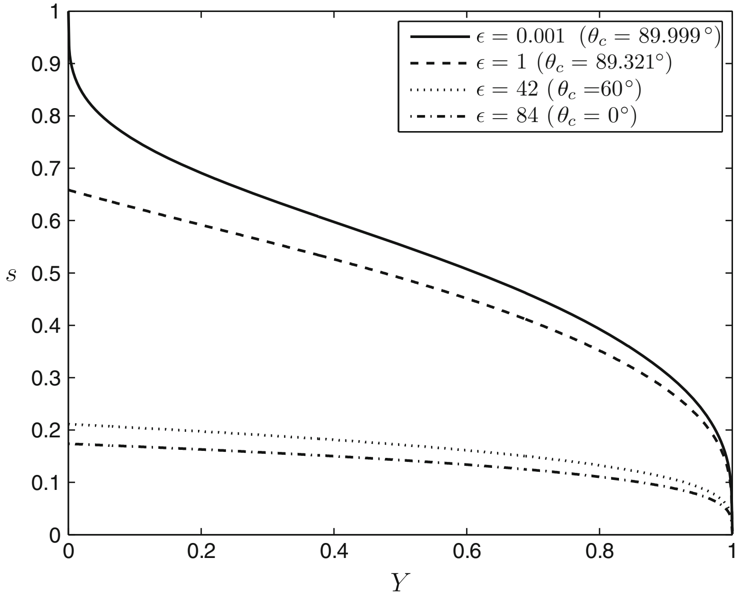


Fig. 1 s vs. Y for different values of $\epsilon(\theta_c)$

$$x_{\mathrm{O}_2}^{(\mathrm{g})} = x_{\mathrm{O}_2}^{in}, \quad s = 0, \quad P^{(\mathrm{g})} = 0,$$

where $I_c = (1-s)(1+\varPi P^{(\mathrm{g})})x_{\mathrm{O}_2}^{(\mathrm{g})}$ and $\varOmega = [i]\left[M^{(\mathrm{g})}\right]/F\left[\rho^{(\mathrm{g})}\right]\left[v^{(\mathrm{g})}\right]$.

With $H \sim 3 \times 10^{-4}$ m , $[D] \sim 10^{-5}\,\mathrm{m^2\,s^{-1}}$, $\kappa \sim 10^{-12}\,\mathrm{m^2}$, $\left[\rho^{(\mathrm{g})}\right] \sim 1\,\mathrm{kg\,m^{-3}}$, $\left[\rho^{(\mathrm{l})}\right] \sim 10^3\,\mathrm{kg\,m^{-3}}$, $\left[\mu^{(\mathrm{g})}\right] \sim 10^{-5}\,\mathrm{kg\,m^{-1}\,s^{-1}}$, $\left[\rho^{(\mathrm{l})}\right] \sim 10^3\,\mathrm{kg\,m^{-3}}$, $p^{out} = 10^5$ Pa, $\sigma = 0.07\,\mathrm{N\,m^{-2}}$, $T = 333$ K, we find that $[v^{(\mathrm{g})}] \sim 3\times10^{-2}\,\mathrm{m\,s^{-1}}$, $\varDelta P \sim 7\times10^4$ Pa, leading to $\widetilde{Ca} \sim 10^{-3}$, $\quad \varPi \sim 0.7$, $\quad \chi \sim 1.25$; with $0 \lesssim [\eta_c] \lesssim 0.6$ V, we have, in addition, that $10^{-3} \lesssim \varOmega \lesssim 10^2$. Furthermore, with

$$P_c(s) = \gamma^{1/2}\cos\theta_c \begin{cases} 1.417(1-s) - 2.120(1-s)^2 + 1.263(1-s)^3 & \text{if } \theta_c < 90°, \\ 1.417s - 2.120s^2 + 1.263s^3 & \text{if } \theta_c > 90°, \end{cases}$$

as in [8, 10], and since, typically, $\kappa_{rel}^{(\mathrm{l})} = s^n$ $(n > 0)$, Eq. (8) indicates that $s \sim \widetilde{Ca}^{1/(n+1)}$ if $\epsilon := \cos\theta_c/\widetilde{Ca} \sim 1$; hence, the size of s is related to the power in the expression for $\kappa_{rel}^{(\mathrm{l})}$. However, if $\epsilon \ll 1$, an entirely different structure emerges, as is demonstrated in Fig. 1 via numerical solutions to the governing equations for $\gamma = 0.3$, $n = 3$, $\theta_c < 90°$, $\eta_c = 0.6$ V, $T = 333$ K and with $\kappa_{rel}^{(\mathrm{g})} = (1-s)^3$. In particular, we see that as θ_c approaches 90°, $s \sim O(1)$; consequently, it appears that, for decreasing hydrophilicity, not only is oxygen transport to the catalyst increasingly hindered, but water blockage, rather than oxygen depletion, may even be the leading reason for limiting current.

Acknowledgements The authors acknowledge the support of the Mathematics Applications Consortium for Science and Industry (www.macsi.ul.ie) funded by the Science Foundation Ireland Mathematics Initiative Grant 06/MI/005.

References

1. Bear, J.: Dynamics of Fluids in Porous Media. American Elsevier, New York (1972)
2. Bird, R.B., Stewart, W.E., Lightfoot, E.N.: Transport Phenomena, 2nd edn. Wiley, New York (2002)
3. Djilali, N.: Computational modelling of polymer electrolyte membrane (PEM) fuel cells: challenges and opportunities. Energy **32**, 269–280 (2007)
4. Gurau, V., Mann, J.A.: A critical overview of computational fluid dynamics multiphase models for proton exchange membrane fuel cells. SIAM J. Appl. Maths **70**, 410–454 (2009)
5. Nam, J.H., Kaviany, M.: Effective diffusivity and water-saturation distribution in single- and two-layer PEMFC diffusion medium. Int. J. Heat Mass Transf. **46**, 4595–4611 (2003)
6. Natarajan, D., Nguyen, T.V.: A two-dimensional, two-phase, multicomponent, transient model for the cathode of a proton exchange membrane fuel cell using conventional gas distributors. J. Electrochem. Soc. **148**, A1324–A1335 (2001)
7. Natarajan, D., Nguyen, T.V.: Three-dimensional effects of liquid water flooding in the cathode of a PEM fuel cell. J. Power Sources **115**, 66–80 (2003)
8. Pasaogullari, U., Wang, C.Y.: Liquid water transport in gas diffusion layer of polymer electrolyte fuel cells. J. Electrochem. Soc. **151**, A399–A406 (2004)
9. Qin, C., Rensink, D., Fell, S., Hassanizadeh, S.M.: Two-phase flow modeling for the cathode side of a polymer electrolyte fuel cell. J. Power Sources **197**, 136–144 (2011)
10. Udell, K.S.: Heat transfer in porous media heated from above with evaporation, condensation and capillary effects. J. Heat Transf. **105**, 485–492 (1983)
11. Vynnycky, M.: On the modelling of two-phase flow in the cathode gas diffusion layer of a polymer electrolyte fuel cell. Appl. Math. Comput. **189**, 1560–1575 (2007)

A Criterion for Air-Gap Formation in Vertical Continuous Casting: The Effect of Superheat

Michael Vynnycky

Abstract The formation of an air gap at the mould-metal interface in continuous casting has long been known to have a detrimental effect on the efficiency of the process, and has therefore attracted many attempts at mathematical modelling. While many efforts consist of complex three-dimensional numerical simulations of the phenomenon, a sequence of recent papers by the present author has used asymptotic techniques to derive a quasi-analytical model that captures the essential characteristics. The model allows for full two-way coupling between the thermal and mechanical problems: the formation of the air gap affects the heat transfer, whilst the heat transfer affects the stresses that lead to the formation and evolution of the air gap. In this contribution, earlier numerical results for the case of superheat—when the molten metal temperature is greater than the melting temperature—are complemented by an analysis of the criterion that predicts how the onset of air-gap formation depends on process parameters: the mould temperature, the casting speed and the superheat itself.

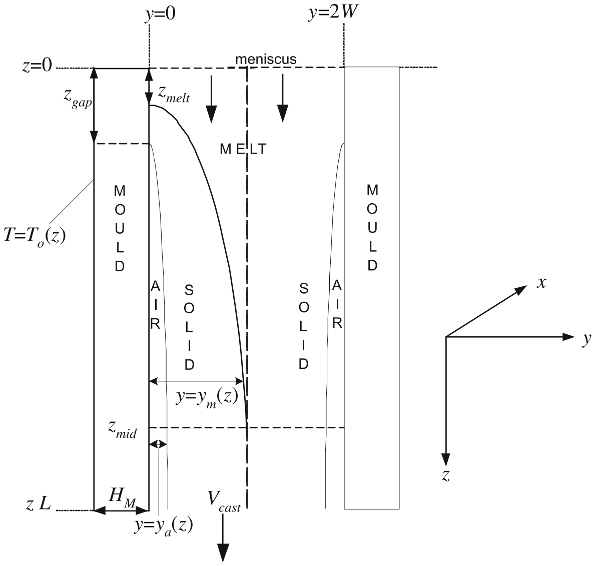
1 Introduction

Air-gap formation in the industrial continuous casting of metals and metal alloys has long been recognized as having an adverse effect on process efficiency. A schematic of the situation is given in Fig. 1, which shows molten metal, typically copper, aluminium or steel alloys, passing vertically downwards through a cooled mould, solidifying and being withdrawn at casting speed, V_{cast}. In descending from the meniscus, there is typically first a region where liquid metal is in contact with the

M. Vynnycky (✉)
Mathematics Applications Consortium for Science and Industry (MACSI), Department of Mathematics and Statistics, University of Limerick, Limerick, Ireland
e-mail: michael.vynnycky@ul.ie

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_25,

Fig. 1 2D schematic of air gap formation



mould wall, followed by a region where the solidified shell is in contact; after this, at $z = z_{gap}$, an air gap begins to form between the solidified shell and the mould wall. Eventually, at some location $z = z_{mid}$, complete solidification occurs at the centreline. In particular, the formation of the air gap prohibits effective heat transfer between the mould and shell, leading to longer solidification lengths and requiring supplementary process design considerations, such as mould tapering.

In view of the detrimental effect that the air gap has on process efficiency, mathematical models of varying degrees of complexity have been derived to describe the phenomenon. Early models for predicting the onset of air-gap formation were analytical [1–4]; most subsequent models [5–11] have been solely numerical. However, whilst able to capture the thermomechanical interaction of gap formation and evolution, such models are computationally expensive, unwieldy and do not give a qualitative understanding of the air-gap's dependence on different operating parameters, or indeed whether it is possible to avoid air-gap formation completely. An exception to all of the above are recent models [12–15] that use asymptotic methods; however, the majority of these assumed that the incoming metal was at melting temperature, T_{melt}. In [13], which was for the case of non-zero superheat, i.e. the incoming molten metal temperature, T_{cast}, was greater than T_{melt}, the resulting equations were integrated numerically, but no details were given as regards the initial stages of solidification and air-gap formation; these details were, however, given in [12, 14] for the case of zero superheat, and the purpose of this paper is to complement the numerical results in [13]. Interestingly, the fact that $T_{cast} > T_{melt}$ not only leads to completely different results, but also to results are not foreseeable from the analysis for $T_{cast} = T_{melt}$.

2 Model Equations

Due to space constraints, we omit the dimensional form of the model equations, which can be found in [13], but move directly to the dimensionless form. Setting $y = WY$ and $z = LZ$, we have

$$\widetilde{Pe}_l \frac{\partial \theta_l}{\partial Z} = \frac{\partial^2 \theta_l}{\partial Y^2}, \quad \widetilde{Pe}_s \frac{\partial \theta_s}{\partial Z} = \frac{\partial^2 \theta_s}{\partial Y^2}, \tag{1}$$

where $\widetilde{Pe}_l$ and $\widetilde{Pe}_s$ are reduced Péclet numbers, given by $\widetilde{Pe}_l = \rho c_l V_{cast} W^2 / k_l L$, $\widetilde{Pe}_s = \rho c_s V_{cast} W^2 / k_s L$, where ρ is solid and liquid metal density and $(k_j)_{j=l,s}$ and $(c_j)_{j=l,s}$ are thermal conductivity and specific heat capacity, respectively. θ_l and θ_s are, respectively, the dimensionless liquid and solid temperatures and are related to the actual temperatures, T_l and T_s, by $\theta_j = (T_{melt} - T_j)/\Delta T$ for $j = l, s$, where ΔT is a temperature scale that will be specified shortly. The boundary conditions for θ_l and θ_s are then

$$\theta_s = \theta_l = 0, \quad \frac{\partial \theta_s}{\partial Y} - \left(\frac{\kappa_s}{\kappa_l}\right) \frac{\partial \theta_l}{\partial Y} = -\frac{\widetilde{Pe}_s}{St} \frac{dY_m}{dZ} \quad \text{at} \quad Y = Y_m(Z), \tag{2}$$

$$\frac{\partial \theta_l}{\partial Y} = \kappa_l \left(\theta_l - \theta_o(Z)\right) \quad \text{at} \quad Y = 0, \quad \text{for} \quad 0 \le Z < Z_{melt}, \tag{3}$$

$$\frac{\partial \theta_s}{\partial Y} = \kappa_s \left(\theta_s - \theta_o(Z)\right) \quad \text{at} \quad Y = 0, \quad \text{for} \quad Z_{melt} \le Z < Z_{gap}, \tag{4}$$

$$\frac{\partial \theta_s}{\partial Y} = \frac{\kappa_s}{(1 + Y_a(Z))} \left(\theta_s - \theta_o(Z)\right) \quad \text{at} \quad Y = \delta Y_a(Z), \quad \text{for} \quad Z > Z_{gap}, \tag{5}$$

$$\frac{\partial \theta_l}{\partial Y} = 0 \quad \text{at} \quad Y = 1 \quad \text{for} \quad 0 \le Z \le Z_{mid}, \tag{6}$$

$$\frac{\partial \theta_s}{\partial Y} = 0 \quad \text{at} \quad Y = 1 \quad \text{for} \quad Z_{mid} \le Z \le 1, \tag{7}$$

where $Z_{melt} = z_{melt}/L$, $Z_{gap} = z_{gap}/L$, $Z_{mid} = z_{mid}/L$ and $St (= c_s \Delta T / \Delta H_f)$ is the Stefan number, with $\kappa_l = k_M W / k_l H_M$, $\kappa_s = k_M W / k_s H_M$, $\delta = k_{air} H_M / k_M W$. In Eq. (2), Y_m is the dimensionless location of the solid-liquid interface, whilst Y_a in Eq. (5) is the scaled dimensionless air-gap thickness. In Eqs. (3)–(5), θ_0 is the dimensionless temperature at the outer surface of the mould, and is related to the experimentally measurable temperature, $T_o(z)$, by $\theta_o = (T_{melt} - T_o)/\Delta T$. Typically, T_o decreases with Z, and it is convenient to use it in defining an appropriate temperature scale: we take $\Delta T = T_{melt} - T_o^{\min}$, where $T_o^{\min} = \min\left(T_o(z) | z \ge 0\right)$. With Z acting as a time-like variable, the initial conditions are

$$\theta_l = \theta_{cast} \quad \text{at} \quad Z = 0, \quad \theta_s = 0 \quad \text{at} \quad Z = Z_{melt}, \quad Y_m(Z_{melt}) = 0, \tag{8}$$

where $\theta_{cast} = \left(T_{melt} - T_{cast}\right) / \Delta T$.

Although most of the dimensionless model parameters are of $O(1)$, further simplification is possible because typically $\delta \ll 1$; thus boundary condition (5) can be taken at $Y = 0$.

3 Analysis

In this paper, we are primarily concerned with determining for which combinations of process variables the air gap is more likely to form; for the corresponding problem in [12] with $\theta_{cast} = 0$, it was found that an air gap was more likely to form if $\widetilde{Pe}_s \dot{\theta}_o(0) > \kappa_s^2 St \theta_o^2(0)$, where the dot denotes differentiation with respect to Z. Thus, from now on, we focus on $0 \leq Z \leq Z_{gap}$; the solution for $Z > Z_{gap}$ was obtained numerically in [13], but no study was ever carried as regards whether there is a criterion that corresponds to the one given above when $\theta_{cast} \neq 0$. In particular, Z_{gap} is given by the solution to

$$-P_0 - P_1 Z_{melt} + \int_{Z_{melt}}^{Z_{gap}} \dot{\Sigma}\left(0, Z'\right) dZ' = 0, \tag{9}$$

where

$$\dot{\Sigma}(Y, Z) = \left(\frac{1}{1-\nu}\right)\left(\dot{\theta}_s - \frac{1}{Y_m(Z)} \int_0^{Y_m(Z)} \dot{\theta}_s dY'\right),$$

with ν as the Poisson ratio, $P_0 = p_0/E\alpha\Delta T (> 0)$ and $P_1 = \rho g L/E\alpha\Delta T (> 0)$; here, p_0 is the pressure at the meniscus, E is the Young's modulus, α is the thermal expansion coefficient and g is the gravitational acceleration. An indicator of whether an air gap forms is the sign of $\dot{\Sigma}\,(0, Z)$ for $Z > Z_{melt}$: if it is positive, it is evident that Eq. (9) will have a solution for Z_{gap}. Since the air gap often forms just a short distance after solidification first occurs, it is therefore instructive to consider the analysis for $\zeta := Z - Z_{melt} \ll 1$, where series expansions for θ and Y_m in terms of ζ ought still to be valid. On setting $\theta_s = Y_m(\zeta) F\,(\zeta, \eta)$, $\eta = Y/Y_m\,(\zeta)$, the second equation in (1) becomes

$$\widetilde{Pe}_s Y_m \left(\dot{Y}_m F + Y_m F_\zeta - \eta \dot{Y}_m F_\eta\right) = F_{\eta\eta}, \tag{10}$$

with the boundary conditions becoming

$$F_\eta = \kappa_s \left(Y_m F - \theta_o(Z)\right) \quad \text{at} \quad \eta = 0, \tag{11}$$

$$F = 0, \;\; \left(F_\eta\right)_{\eta=1} - \left(\frac{\kappa_s}{\kappa_l}\right)\left(\frac{\partial \theta_l}{\partial Y}\right)_{Y=Y_m(Z)} = -\widetilde{Pe}_s St^{-1} \dot{Y}_m \quad \text{at} \quad \eta = 1. \tag{12}$$

Now, in the transformed coordinates, $\dot{\Sigma}$ is given by

$$\dot{\Sigma}\left(\eta,\zeta\right)=\frac{1}{1-\nu}\left(\dot{\theta}_s-\int_0^1\dot{\theta}_s d\eta'\right),\quad \text{where}\quad \dot{\theta}_s=\dot{Y}_m F+Y_m F_\zeta-\eta\dot{Y}_m F_\eta. \tag{13}$$

As $\zeta\to 0$, a self-consistent boundary-value problem is obtained if $Y_m(\zeta)\sim\zeta^{3/2}$; note that this result, for which $\theta_{cast}\neq 0$ and $Z_{melt}>0$, has only been found recently [16, 17], whereas the result for the case when $\theta_{cast}=0$, which leads to $Z_{melt}=0$ and $Y_m(\zeta)=\lambda\zeta+o(\zeta)$, has been known since much earlier [18]. To proceed, we write

$$F=F_0\left(\eta\right)+\zeta F_1\left(\eta\right)+o\left(\zeta\right)+\ldots,\quad Y_m(\zeta)=\lambda_1\zeta^{3/2}+o\left(\zeta^{3/2}\right)+\ldots, \tag{14}$$

which suggests, at first sight, that $\dot{\Sigma}\left(0,\zeta\right)\sim\zeta^{1/2}$ in Eq. (9); note also that, as demonstrated in [16, 17], λ_1 is a strictly positive constant whose value is given by $\lambda_1=4St\partial^2\theta_l/\partial Y^2\left(0,Z_{melt}\right)/3\pi^{1/2}$, with Z_{melt} such that $\theta_l(0,Z_{melt})=0$, i.e. the distance from the inlet at which solidification first starts. Now, at ζ^0 and ζ^1, we have

$$F_{0\eta\eta}=0,\quad \text{subject to}\quad F_{0\eta}(0)=-\kappa_s\theta_o(Z_{melt}),\qquad F_0(1)=0, \tag{15}$$

$$F_{1\eta\eta}=0,\quad \text{subject to}\quad F_{1\eta}(0)=-\kappa_s\dot{\theta}_o(Z_{melt}),\qquad F_1(1)=0, \tag{16}$$

respectively; thus,

$$F_0(\eta)=\kappa_s\theta_o(Z_{melt})\left(1-\eta\right),\quad F_1=\kappa_s\dot{\theta}_o(Z_{melt})\left(1-\eta\right). \tag{17}$$

The forms of F_0 and F_1 mean that the first contribution to $\dot{\Sigma}\left(0,\zeta\right)$ is at $O\left(\zeta^{3/2}\right)$; more exactly,

$$\dot{\Sigma}\left(0,\zeta\right)\sim\frac{1}{2(1-\nu)}\lambda_1\left(5F_1\left(0\right)-\int_0^1\left(5F_1-3\eta F_{1\eta}\right)d\eta'\right)\zeta^{3/2}, \tag{18}$$

which can be simplified to give

$$\dot{\Sigma}\left(0,\zeta\right)\sim\frac{1}{2(1-\nu)}\lambda_1\kappa_s\dot{\theta}_o(Z_{melt})\zeta^{3/2}. \tag{19}$$

So, an air gap is more likely to form if $\dot{\theta}_o\left(Z_{melt}\right)>0$, indicating that an air gap will always form; interestingly, this is considerably different to the result in [12] for the case when $\theta_{cast}=0$.

Acknowledgements The author acknowledges the support of the Mathematics Applications Consortium for Science and Industry (www.macsi.ul.ie), funded by the Science Foundation Ireland (SFI) Mathematics Initiative grant 06/MI/005 and SFI grant 12/IA/1683.

References

1. Savage, J.: A theory of heat transfer and air gap formation in continuous casting molds. J. Iron Steel Inst. **198**, 41–47 (1962)
2. Richmond, O., Tien, R.H.: Theory of thermal stresses and air-gap formation during the early stages of solidification in a rectangular mold. J. Mech. Phys. Solids **19**, 273–284 (1971)
3. Kristiansson, J.O.: Thermal stresses in the early stage of the solidification of steel. J. Therm. Stresses **5**, 315–330 (1982)
4. Tien, R.H., Richmond, O.: Theory of maximum tensile stresses in the solidifying shell of a constrained regular casting. J. Appl. Mech. **49**, 481–486 (1982)
5. Grill, A., Sorimachi, K., Brimacombe, J.K.: Heat flow, gap formation and break-outs in the continuous casting of steel slabs. Metall. Mater. Trans. B **7B**, 177–189 (1976)
6. Kelly, J.E., Michalek, K.P., O'Connor, T.G., Thomas, B.G., Dantzig, J.A.: Initial development of thermal and stress fields in continuously cast steel billets. Metall. Mater. Trans. A **19A**, 2589–2602 (1988)
7. Bellet, M., Decultieux, F., Menai, M., Bay, F., Levaillant, C., Chenot, J.L., Schmidt, P., Svensson, I.L.: Thermomechanics of the cooling stage in casting processes: three-dimensional finite element analysis and experimental validation. Metall. Mater. Trans. B **27B**, 81–99 (1996)
8. Huespe, A.E., Cardona, A., Fachinotti, V.: Thermomechanical model of a continuous casting process. Comput. Methods Appl. Mech. Eng. **182**, 439–455 (2000)
9. Kim, K.Y.: Analysis of gap formation at mold-shell interface during solidification of aluminium alloy plate. ISIJ Int. **43**, 647–652 (2003)
10. Li, C., Thomas, B.G.: Thermomechanical finite-element model of shell behavior in continuous casting of steel. Metall. Mater. Trans. B **35B**, 1151–1172 (2004)
11. Sun, D., Annapragada, S.R., Garimella, S.V., Singh, S.K.: Analysis of gap formation in the casting of energetic materials. Numer. Heat Transf. **51**, 415–444 (2007)
12. Vynnycky, M.: An asymptotic model for the formation and evolution of air gaps in vertical continuous casting. Proc. R. Soc. A **465**, 1617–1644 (2009)
13. Vynnycky, M.: A mathematical model for air-gap formation in vertical continuous casting: the effect of superheat. Trans. Ind. Inst. Met. **62**, 495–498 (2009)
14. Vynnycky, M.: Air gaps in vertical continuous casting in round moulds. J. Eng. Maths **68**, 129–152 (2010)
15. Vynnycky, M.: On the role of radiative heat transfer in air gaps in vertical continuous casting. Appl. Math. Model. **37**, 2178–2188 (2013)
16. Mitchell, S.L., Vynnycky, M.: An accurate finite-difference method for ablation-type Stefan problems. J. Comput. Appl. Math. **236**, 4181–4192 (2012)
17. Vynnycky, M., Mitchell, S.L.: On the solution of Stefan problems with delayed onset of phase change. In: Proceedings of the 7th International Conference on Heat Transfer, Fluid Mechanics and Thermodynamics, Antalya, pp. 1404–1410 (2010)
18. Carslaw, H.S., Jaeger, J.C.: Conduction of Heat in Solids, 2nd edn. Oxford University Press, New York (1959)

Moulding Contact Lenses

Ellen Murphy and William T. Lee

Abstract The moulding process in the manufacture of a certain monomer-based product, is modelled using the thin film approximation with the aim of reducing defects in which the mould is partially filled. A simple model neglecting curvature of the moulds is considered first. This assumption is verified by a polar coordinate model that investigates the effects of curvature of the dynamics of the fluid. We investigate the role of surface tension and horizontal motion of the lower mould in the formation of defects.

1 Introduction

A stage of the manufacture of a certain product consists of filling a mould with a viscous fluid and pushing a second mould down on top of the first [1]. This action squeezes the fluid out between the two moulds, with the desired effect being the complete filling of the space between the moulds with fluid. However, this does not always occur. In some cases, the fluid flows asymmetrically out of the gap, resulting in partially filled moulds. This is highly undesirable and results in the rejection of these specimens. The aim of this study is to determine the factors contributing to asymmetrical flow and to develop recommendations for its avoidance.

E. Murphy • W.T. Lee (✉)
MACSI, Department of Mathematics and Statistics, University of Limerick, Limerick, Ireland
e-mail: ellen.murphy@ul.ie; william.lee@ul.ie

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_26,

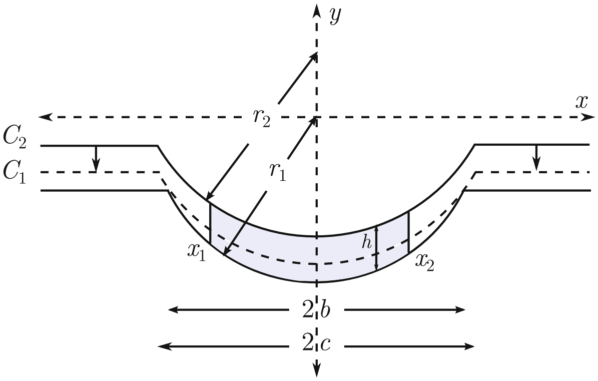


Fig. 1 Dimensional geometry of the moulds and positions of the fluid boundaries (not to scale)

2 Mathematical Model

The height between the two moulds is much less than the length of the moulds. The fluid in question is highly viscous and is assumed to be Newtonian. Analysis shows that the modified Reynolds number for this problem is small and so the thin film approximation is appropriate. The system is axisymmetric and, for simplicity, is modelled in two dimensions.

The setup consists of a lower mould, C_1, and an upper mould, C_2, as shown in Fig. 1 (not to scale). Both moulds have the form of truncated hemispheres joined to flat, horizontal sides. Fluid is placed in the lower mould and the upper mould is then pushed vertically down on to the lower mould at a constant speed. The circular part of the lower mould has a greater radius of curvature than the upper mould. Assuming the system to be axially symmetric, the problem is modelled in 2-d. The height of the fluid between moulds, h, is small relative to b, that is $h \ll b$ and so the dynamics of the fluid are modelled with the thin film equations. Assuming no surface tension effects and knowing the height of the fluid between the moulds, the system reduces to two ODEs which determine the evolution of the fluid boundaries. Of particular interest was whether or not the curvature of the moulds influenced the dynamics of the monomer. To check this, the system was first modelled in cartesian coordinates, where the curvature of the moulds was neglected. This was compared to a polar model, which included curvature effects.

2.1 Numerical Results and Conclusions

Figure 2 displays the numerical solutions for the two systems of ODEs. Results from both the simple cartesian model and the polar model are shown. As can be seen in the figures, there is no qualitative difference between the two sets of results. This

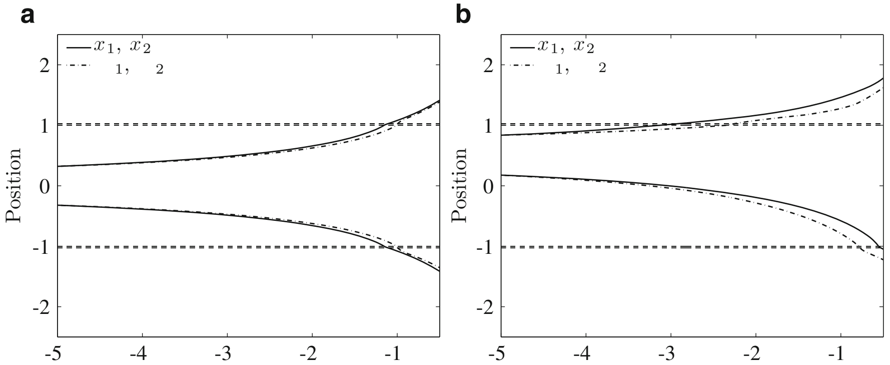


Fig. 2 Evolution of the fluid boundaries for the polar model, overlaid with the simple case. (**a**) $x_{\mathrm{com}}(t_0) = 0$. (**b**) $x_{\mathrm{com}}(t_0) = 0.5$

validates the use of the "flat" model and also helps to explain why the monomer remains off-centred once it begins like so. From the perspective of the fluid, the moulds appear flat, therefore when the upper mould is pushed down on the fluid it is squeezed out in both directions from its initial position. There is no mechanism available for it to overcome its off-centredness.

References

1. Chapman, S.J., Cribbin, L., Dunne, A., Dellar, P.J., Fowkes, N., Lapin, V., Lee, W.T., Murphy, E., Power, O., Vazquez, C., Sweatman, W.L.: Monomer flow for contact lens manufacture. In: O'Brien, S., O'Sullivan, M., Hanrahan, P., Lee, W.T., Mason, J., Charpin, J., Korobeinikov, A. (eds.) Proceedings of the Seventy Fifth European Study Group with Industry (2010)

Enhanced Water Flow in Carbon Nanotubes and the Navier Slip Condition

Tim G. Myers

Abstract A possible explanation for the enhanced flow in carbon nanotubes is given using a mathematical model that includes a depletion layer with reduced viscosity near the wall. In the limit of large tubes the model predicts no noticeable enhancement. For smaller tubes the model predicts enhancement that increases as the radius decreases. An analogy between the reduced viscosity and slip-length models shows that the term slip-length is misleading and that on surfaces which are smooth at the nanoscale it may be thought of as a length-scale associated with the size of the depletion region and viscosity ratio. The model therefore provides a physical interpretation of the classical Navier slip condition and explains why "slip-lengths" may be greater than the tube radius.

1 Introduction

The classical model for flow in a circular cylindrical pipe is described by the Hagen-Poiseuille equation

$$u_{HP} = -\frac{p_z R^2}{4\mu}\left(1 - \frac{r^2}{R^2}\right) \tag{1}$$

where $u_{HP}(r)$ is the velocity in the z direction, p_z is the pressure gradient along the pipe, R is the radius and μ the fluid viscosity. The corresponding flux is given by

T.G. Myers (✉)
Centre de Recerca Matemàtica, Campus de Bellaterra Edifici C, 08193 Barcelona, Spain

Departament de Matemàtica Aplicada I, Universitat Politécnica de Catalunya, Barcelona, Spain
e-mail: tmyers@crm.cat

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_27,

$$Q_{HP} = 2\pi \int_0^R r u_{HP} \, dr = -\frac{\pi R^4 p_z}{8\mu} \,. \tag{2}$$

In carbon nano-tubes (CNT) it is well documented that the flow is enhanced and the true value of the flux is significantly higher than this classical value.

A popular approach to explain this enhancement is to introduce a slip-length into the mathematical model, that is, the no-slip boundary condition $u(R) = 0$ is replaced by

$$u(R) = -L_s \left. \frac{\partial u}{\partial r} \right|_{r=R} \tag{3}$$

where L_s is the slip-length. This leads to modified velocity and flux expressions

$$u_{slip} = -\frac{R^2 p_z}{4\mu} \left[1 - \frac{r^2}{R^2} + \frac{2L_s}{R} \right] \qquad Q_{slip} = Q_{HP} \left(1 + \frac{4L_s}{R} \right) , \tag{4}$$

hence any magnitude of enhancement can be accounted for by using an appropriate value for L_s.

Assuming fluid slip at the wall the value of the velocity at the channel wall is positive: the slip length is defined as the distance the velocity profile must be extrapolated beyond the wall to reach zero [1]. In general the slip length is significantly smaller than the thickness of the bulk flow [2]. For example, Tretheway and Meinhart [3] carry out experiments on water flow in a coated microchannel of width 30 μm and find a slip length of 1 μm. In 1–2 μm channels Choi et al. [4] determine values of the order 30 nm. However, in CNTs Whitby et al. [5] quote lengths of 30–40 nm for experiments in pipes of 20 nm radius. Holt et al. [6] and Majumder et al. [7] quote slip lengths on the order of microns for their experiments with nanometer size pores.

The high values of slip-length in CNT studies have led some authors to question the validity of the slip modified Hagen-Poiseuille model [8, 9]. An alternative explanation to the slip-length is based on the fact that CNTs are hydrophobic [10–12]. The strength of attraction between the water molecules is greater than the attraction between the hydrophobic solid and the water [13,14]. Indeed it was mainly experiments performed with hydrophobic surfaces that supported early arguments for a slip boundary condition [2]. It has been postulated that hydrophobicity may result in gas gaps, depletion layers or the formation of vapour: experimentally this may be interpreted as "apparent" slippage, see [15].

Obviously any depletion layer must be small. Experiments and simulations have shown that the fluid viscosity is in close agreement with its bulk value down to separations of about ten molecular diameters [2]. For CNTs the fluid properties typically vary within an annular region approximately 0.7 nm from the wall [8, 16, 17].

Consequently, in the following work we will investigate a mathematical model for flow including a region of low viscosity near the tube wall. In light of the results quoted in [14, 18] we will assume the theory is not valid for films below ten molecular diameters thickness. This limit is also imposed through the validity of the continuum assumption, for example the MD simulations of [19] shows results that coincide with a continuum model for a pipe radius of ten molecular diameters.

2 Mathematical Model

Consider a pipe of cross-section R, occupied by two fluids. In the bulk flow region, defined by $0 \leq r \leq \alpha$, we impose a viscosity μ_1. In the annular region near the wall, defined by $\alpha \leq r \leq R$, we impose a viscosity $\mu_2 < \mu_1$. The assumption of two regions with different viscosities leads to what is commonly termed a bi-viscosity model in the non-Newtonian flow literature. In the following analysis there is uncertainty about the values to choose for viscosity and the distance α. If we define the position of the transition $\alpha = R - \delta$ then, based on previous studies of water in CNTs we will choose $\delta = 0.7\,\text{nm}$. However, experiments show that the slip length is a measure of hydrophobicity [4, 20–22] and so for other liquid–solid systems the value of δ may differ.

For unidirectional pressure driven flow through a circular pipe the appropriate mathematical model is

$$\frac{\mu_1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial u_1}{\partial r}\right) = \frac{\partial p}{\partial z} \qquad 0 \leq r \leq \alpha\,, \quad \frac{\mu_2}{r}\frac{\partial}{\partial r}\left(r\frac{\partial u_2}{\partial r}\right) = \frac{\partial p}{\partial z} \qquad \alpha \leq r \leq R\,. \tag{5}$$

Appropriate boundary conditions are

$$\left.\frac{\partial u_1}{\partial r}\right|_{r=0} = 0 \qquad\qquad u_2(R,z) = 0\,, \tag{6}$$

which represent symmetry at the centreline and no-slip at the solid boundary. At the interface between the fluids, $r = \alpha$, there is continuity of velocity and shear stress

$$u_1 = u_2 \qquad\qquad \mu_1\frac{\partial u_1}{\partial r} = \mu_2\frac{\partial u_2}{\partial r}\,. \tag{7}$$

The velocity expressions are then

$$u_1 = \frac{p_z}{4\mu_1}(r^2-\alpha^2) - \frac{p_z}{4\mu_2}(R^2-\alpha^2) \qquad\qquad u_2 = \frac{p_z}{4\mu_2}(r^2-R^2)\,. \tag{8}$$

The flux Q_μ is defined as the sum of fluxes in the two regions

$$Q_\mu = -\frac{\pi\alpha^4 p_z}{8\mu_1}\left[1-\frac{2\mu_1}{\mu_2}\left(1-\frac{R^2}{\alpha^2}\right)\right]-\frac{\pi\alpha^4 p_z}{8\mu_2}\left(1-\frac{R^2}{\alpha^2}\right)^2 \tag{9}$$

$$= Q_{HP}\frac{\alpha^4}{R^4}\left[1+\frac{\mu_1}{\mu_2}\left(\frac{R^4}{\alpha^4}-1\right)\right]. \tag{10}$$

The flow rate enhancement is defined as

$$\epsilon_\mu = \frac{Q_\mu}{Q_{HP}} = \frac{\alpha^4}{R^4}+\frac{\mu_1}{\mu_2}\left(1-\frac{\alpha^4}{R^4}\right). \tag{11}$$

For the slip model the corresponding enhancement is

$$\epsilon_{slip} = 1+\frac{4L_s}{R}. \tag{12}$$

3 Model Validation

To verify whether this model gives reasonable results we consider the experiments of Whitby et al. [5]. Their flow enhancement indicates a slip length of 30–40 nm for pipes of radius 20 nm. Setting $L_s = 35$ nm, $R = 20$ nm determines their enhancement factor as $\epsilon_{slip} = 8$. Rearranging the expression for ϵ_μ gives

$$\mu_2 = \mu_1\left[\frac{R^4-\alpha^4}{\epsilon_\mu R^4-\alpha^4}\right]. \tag{13}$$

To obtain the same enhancement we set $\epsilon_\mu = 8$ and also take $\alpha = R-\delta = 19.3$ nm to find $\mu_2 = 0.018\mu_1$. So, the current model will provide an enhancement factor of 8 with an average viscosity in the depletion layer approximately 0.02 times that of the bulk flow. It is interesting to note that the viscosity of oxygen is also approximately 0.02 that of water, so this value supports the depletion layer theory. Thomas et al. [23] find $\epsilon_{slip} \approx 32$ nm when $R = 3.5$ nm, taking $\mu_2 = 0.018\mu_1$ Eq. (11) indicates $\epsilon_\mu \approx 33.2$ nm.

To clarify the behaviour of the current model we set $\alpha = R-\delta$. Since ϵ_μ is simply a quartic in α we may expand and rearrange the expression to find

$$\epsilon_\mu = 1+\frac{4\delta}{R}\left(\frac{\mu_1}{\mu_2}-1\right)\left[1-\frac{3}{2}\frac{\delta}{R}+\left(\frac{\delta}{R}\right)^2-\frac{1}{4}\left(\frac{\delta}{R}\right)^3\right], \tag{14}$$

which is a monotonically decreasing function of R. This is in accordance with the findings of Thomas and McGaughey [8] that the enhancement factor decreases with increasing tube radius. Noting that the reduced viscosity model requires two distinct regions, hence $R \geq \delta$, the limit to the enhancement predicted by the current theory is determined by setting $R = \delta$, $\mu_2/\mu_1 = 0.018$ and $\delta = 0.7\,\text{nm}$ to give $\epsilon_\mu \approx 50$: Whitby et al. [5] predict an enhancement of up to 45 times theoretical predictions.

Equation (15) also allows us to make further inference about the model behaviour and its relation to the slip model. If we compare the above expression with that for ϵ_{slip} we may define the slip length in terms of the thickness of the depletion layer and the viscosity ratio

$$L_s = \delta\left(\frac{\mu_1}{\mu_2} - 1\right)\left[1 - \frac{3}{2}\frac{\delta}{R} + \left(\frac{\delta}{R}\right)^2 - \frac{1}{4}\left(\frac{\delta}{R}\right)^3\right]. \tag{15}$$

Further, noting that $\mu_1/\mu_2 \gg 1$, we can identify three distinct regimes:

1. For sufficiently wide tubes, $(\delta/R)(\mu_1/\mu_2) \ll 1$, then by Eq. (14) $\epsilon_\mu \approx 1$. There is no noticeable flow enhancement and the no-slip boundary condition will be sufficient.
2. For moderate tubes, $(\delta/R)(\mu_1/\mu_2)$ is order 1 but $\delta/R \ll 1$ then

$$\epsilon_\mu \approx 1 + \frac{4\delta}{R}\frac{\mu_1}{\mu_2}. \tag{16}$$

3. For very small tubes, δ/R is order 1, then the full expression for ϵ_μ is required.

Note, numerous papers report constant slip-lengths between 20 and 40 nm when $R \in$ "some nanometers up to several hundred nanometers", see [20] for example. Thomas et al. [23] suggest L_s varies with R for $R \in [1.6, 5]\,\text{nm}$.

4 Discussion

The motivation behind this paper was to explain the unrealistically large slip-lengths reported in nanotubes. The mathematical model developed shows that the flow enhancement can be plausibly related to a reduced viscosity model, where the viscosity in the depletion region is always much lower than in the bulk. In pipes with a radius greater than the depletion layer thickness the model indicates that the flow can only be enhanced by an order of magnitude (around 50), not orders as reported in some papers. The term slip-length may be considered misleading, in fact it appears to be a length-scale proportional to the product of the viscosity ratio and the width of the depletion region. This length-scale is a property of the fluid–solid system and remains approximately constant, down to very small radius tubes.

In a wider context the reduced viscosity model provides one possible explanation for the Navier slip boundary condition on a hydrophobic solid surface that is smooth down to the nanoscale (and hence an explanation for flow enhancement). In other systems there may well be different mechanisms to explain the slip boundary condition, for example on rough surfaces one would expect the slip length to be determined by the roughness height-scale.

Acknowledgements The author gratefully acknowledges the support of this research through the Marie Curie International Reintegration Grant *Industrial applications of moving boundary problems* Grant no. FP7-256417 and Ministerio de Ciencia e Innovación Grant MTM2011-23789.

References

1. Denn, M.M.: Extrusion instabilities and wall slip. Annu. Rev. Fluid Mech. **33**, 265–287 (2001)
2. Neto, C., Evans, D.R., Bonaccurso, E., Butt, H.-J., Craig, V.S.J.: Boundary slip in Newtonian liquids: a review of experimental studies. Rep. Prog. Phys. **68**, 2859–2897 (2005)
3. Tretheway, D.C., Meinhart, C.D.: Apparent fluid slip at hydrophobic microchannel walls. Phys. Fluids **14**(3), L9–L12 (2002)
4. Choi, C.-H., Westin, J.A., Breuer, K.S.: Apparent slip flows in hydrophilic and hydrophobic microchannels. Phys. Fluids **15**(10), 2897–2902 (2003)
5. Whitby, M., Cagnon, L., Thanou, M., Quirke, N.: Enhanced fluid flow through nanoscale carbon pipes. Nano Lett. **8**(9), 2632–2637 (2008)
6. Holt, J.K., Park, H.G., Wang, Y., Stadermann, M., Artyukhin, A.B., Grigoropoulos, C.P., Noy, A., Bakajin, O.: Fast mass transport through sub-2-nanometer carbon nanotubes. Science **312**, 1034 (2006)
7. Majumder, M., Chopra, N., Andrews, R., Hinds, B.J.: Enhanced flow in carbon nanotubes. Nature **438**, 44 (2005)
8. Thomas, J.A., McGaughey, A.J.H.: Reassessing fast water transport through carbon nanotubes. Nano Lett. **8**(9), 2788–2793 (2008)
9. Verweij, H., Schillo, M.C., Li, J.: Fast mass transport through carbon nanotube membranes. Small **3**(12), 1996–2004 (2007)
10. Werder, T., et al.: Molecular dynamics simulation of contact angles of water droplets in carbon nanotubes. Nano Lett. **1**, 697–702 (2001)
11. Hummer, G., Rasaiah, J.C., Noworyta, J.P.: Water conduction through the hydrophobic channel of a carbon nanotube. Nature **414**(8), 188–190 (2001)
12. Noya, A., et al.: Nanofluidics in carbon nanotubes. NanoToday **2**(6), 22–29 (2007)
13. Vinogradova, O.I.: Slippage of water over hydrophobic surfaces. Int. J. Miner. Process. **56**, 31–60 (1999)
14. Eijkel, J.C.T., van den Berg, A.: Nanofluidics: what is it and what can we expect from it? Microfluid. Nanofluid. **1**, 249–267 (2005)
15. Myers, T.G.: Why are slip lengths so large in carbon nanotubes? Microfluid. Nanofluid. **10**, 1141–1145 (2011)
16. Joseph, S., Aluru, N.R.: Why are carbon nanotubes fast transporters of water? Nano Lett. **8**(2), 452–458 (2008)
17. Thomas, J.A., McGaughey, A.J.H.: Density, distribution, and orientation of water molecules inside and outside carbon nanotubes. J. Chem. Phys. **128**, 084715 (2008)
18. Verdaguer, A., Sacha, G.M., Bluhm, H., Salmeron, M.: Molecular structure of water at interfaces: wetting at the nanometer scale. Chem. Rev. **106**, 1478–1510 (2006)
19. Travis, K.P., Todd, B.D., Evans, D.J.: Departure from Navier-Stokes hydrodynamics in confined liquids. Phys. Rev. E **55**(4), 4288–4295 (1997)
20. Cottin-Bizonne, C., Cross, B., Steinberger, A., Charlaix, E.: Boundary slip on smooth hydrophobic surfaces: intrinsic effects and possible artifacts. Phys. Rev. Lett. **94**, 056102 (2005)

21. Alexeyev, A.A., Vinogradova, O.I.: Flow of a liquid in a nonuniformly hydrophobized capillary. Colloids Surf. A Physicochem. Eng. Aspects **108**, 173–179 (1996)
22. Zhu, Y., Granick, S.: Rate-dependent slip of Newtonian liquids at smooth surfaces. Phys. Rev. Lett. **87**, 96105 (2001)
23. Thomas, J.A., McGaughey, A.J.H., Kuter-Arnebeck, O.: Pressure-driven water flow through carbon nanotubes: insights from molecular dynamics simulation. Int. J. Therm. Sci. **49**, 281–289 (2010)

Flow Field Numerical Research in a Low-Pressure Centrifugal Compressor with Vaneless Diffuser

Alexey Frolov, Rudolf Izmaylov, and Denis Voroshnin

Abstract This work demonstrates the results of the first phase of the problem that is aimed at the numerical investigation of such unsteady effects as the precursor stall and the rotating stall in the stage with a vaneless diffuser of a centrifugal compressor. This paper is focused on the capabilities and constraints of the steady-state numerical simulations for an accurate prediction of the flow through the compressor stage. Numerical simulations were carried out in NUMECA FINE/TURBO 8.9.1 for a single blade passage. The results were validated through a comparison with the experimental data at the diffuser inlet and outlet. The results of numerical simulations using different discretization schemes and turbulence models predicted different flow structure. The results obtained with the second order discretization agree with the experiments for the steady-state case in the region of high flows rates. In the area of low flow rates the unsteady effects significantly influence the flow leading to poor predictions. An analysis of an influence of the geometry model and the grid resolution on the convergence is required to predict the satisfactory agreement with experiment.

1 Introduction

State-of-the-art gas compressors often have to run far away from their design point maintaining safe and reliable operation. At off-design conditions instabilities like stall and surge can lead to the significant decrease of reliability or even to the

A. Frolov (✉) • R. Izmaylov
Saint Petersburg State Polytechnical University, Saint Petersburg, Russian Federation
e-mail: frolovalex@lamm.spbstu.ru; ira1239@gmail.com

D. Voroshnin
Numeca Russia, Saint Petersburg, Russian Federation
e-mail: d.voroshnin@rescent.ru

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_28,

destruction of the whole rig. The reason for this behavior is the internal structure of the flow field, namely, vortices formation and development with a decrease of the flow rate. The development and behavior of such unsteady effects as precursor stall and rotating stall have been experimentally investigated in sufficient detail [1]. However, high cost and lack of completeness of the experimental study prevents the identification of the internal nature of these phenomena. Therefore there is a need for a detailed numerical investigation of the flow field structure with a comparison to the experimental data to be carried out to clarify the reasons for the stall and surge.

The flow in a stage of the centrifugal compressor is highly three-dimensional, spatially non-uniform and intrinsically unsteady. To provide the inside look at the unsteady effects by the numerical simulation, a full annulus should be considered under the study for both impeller and diffuser in the transient mode. However, this technique is high-cost in computer resources and time. Thus this paper is dedicated to the analysis of capabilities and constraints of the steady-state numerical simulations of the stalling regimes in the centrifugal compressor. The main objective is to analyze averaged characteristics of the stage.

2 Main Section: Numerical Research

2.1 Problem Description

A geometry model of the single-stage centrifugal compressor with the vaneless diffuser was created at the Compressor Department of the Saint-Petersburg State Polytechnical University (LPI) [2]. The intermediate stage under the investigation consists of the impeller (frequency of rotation: $n = 6{,}944\,\text{rpm}$; diameter: $D_2 = 275\,\text{mm}$) with $Z = 16$ blades (inlet angle: $\beta_{in} = 34.4°$; outlet angle: $\beta_{out} = 48.9°$) and vaneless diffuser (inlet diameter: $D_3 = 1.047D_2$; outlet diameter: $D_4 = 1.44D_2$). The compressor is designed to operate at flow coefficient $\phi_2 = 0.275$ (flow rate $G = 0.425\,\text{kg/s}$). The maximum flow rate in the experiment is at $\phi_2 = 0.4$ ($G = 0.65\,\text{kg/s}$) and the strong unsteady stalling effects appear at $\phi_2 = 0.2$ ($G = 0.325\,\text{kg/s}$).

Both instant and averaged experimental data is available [2]. Averaged data was obtained with traditional measurements on slow varying parameters. The unsteady experimental data was obtained with high frequency pressure pick-ups and hot-wire techniques. The averaged data is available for the different flow rates at the different cross sections of the stage namely the diffuser inlet and outlet (Fig. 1).

2.2 Modeling Details

The simulation was carried out using the NUMECA FINE/Turbo software with the EURANUS block-structures solver. The solver applies a CFD code based on a

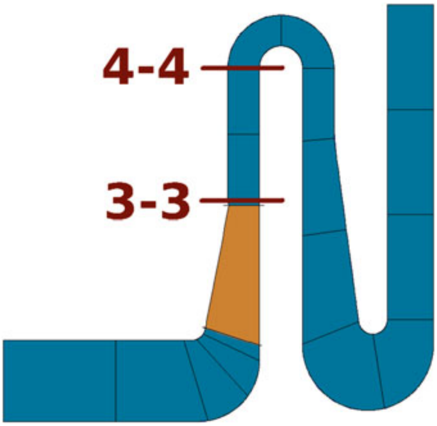


Fig. 1 Meridional view of the compressor stage (3-3—diffuser inlet, 4-4—diffuser outlet)

3D steady compressible, finite volume scheme to solve Reynolds-Averaged Navier-Stokes (RANS) equations, namely continuity (1), momentum conservation (2), and energy conservation (3) equations, in a conservative formulation [3]:

$$\frac{\partial \rho}{\partial t} + \frac{\partial \left(\rho u_j\right)}{\partial x_j} = 0, \tag{1}$$

$$\frac{\partial \left(\rho u_i\right)}{\partial t} + \frac{\partial \left(\rho u_i u_j\right)}{\partial x_j} = -\frac{\partial p}{\partial x_i} + \frac{\partial \tau_{ij}}{\partial x_j} + F_i, \quad i = 1, 2, 3, \tag{2}$$

$$\frac{\partial \left(\rho E\right)}{\partial t} + \frac{\partial \left(\rho E u_j\right)}{\partial x_j} = -\frac{\partial \left(p u_j\right)}{\partial x_j} + \frac{\partial \left(q_j + u_i \tau_{ij}\right)}{\partial x_j} + W_f, \tag{3}$$

where ρ is the fluid density, u_j are the velocity components, $E = e + \frac{1}{2} u_i u_i$ is the total energy, p is the static pressure, τ_{ij} are the stress tensor components, F_i are the external forces components, $W_f = \rho \mathbf{F} \cdot \mathbf{u}$ is the work performed by external forces, $q_j = k \frac{\partial T}{\partial x_j}$ are the heat flux components, k is the laminar thermal conductivity.

These equations are solved in a rotational reference frame, which leads to the presence of additional terms (i.e. Coriolis force) in these equations. Spalart-Allmaras (SA), standard k-ε (KE) and shear-stress transport (SST) turbulence models were used for turbulence closure.

Different spatial discretization techniques were under the investigation, namely second order cell-centered and first and second order upwind discretization schemes. An explicit four-stage Runge-Kutta scheme and local time-stepping technique were used for fictitious time iteration. The convergence of the solver was accelerated by the enlarged Courant number and the application of a multi-grid acceleration procedure with an increased number of smoothing steps on the coarse grid levels. The convergence criterion is that a global residual is less than 10^{-6}.

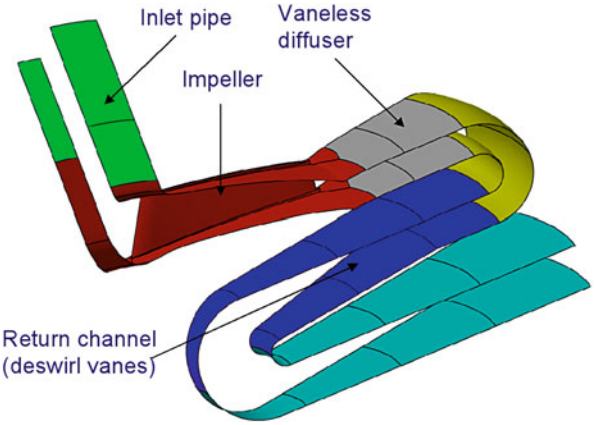


Fig. 2 3D view of the compressor stage

The steady-state simulations were carried out to verify the turbulence models and to make some tuning for the unsteady simulations. Hub and shroud leakage flows between impeller and casing are neglected. Cover disk friction is neglected as well.

2.3 Numerical Details

The intermediate compressor stage is modeled as a set of four components: an inlet pipe, an impeller, a vaneless diffuser and a vaneless return channel (Fig. 2). The model of the stage was built in CAD software and transferred into the NUMECA/AUTOGRID specialized block-structured mesh generator. The mesh generated for a computational domain of the single blade passage of the stage is depicted in Fig. 3 (the number of cells is about 900,000). The wall cell width was chosen to be 10^{-6} m so that values of dimensionless wall distance were kept below $y^+ \leq 0.5$ on all the solid boundaries.

The boundary conditions were set in accordance with the 1D characteristic theory. Total pressure, temperature, and velocity components were imposed at the inlet. At the outlet an averaged static pressure was imposed at high flow rates and self-adaptive mass flow rate at other operating conditions. Non-slip boundary conditions were applied on solid boundaries. Matching periodicity boundary conditions were applied on peripheral periodic boundaries. To create an initial solution for the simulation on the finest grid level, the simulations were successively carried out on other grid levels starting from the coarsest one. The calculated solution of each operating point was used as the initial solution for the next higher pressure operating point.

All steady-state calculations were carried out on a single workstation with following characteristics: Intel Core i7-950 processor (3.06 GHz), 8 Gb RAM, Linux x64 operating system. Calculation time was about 18 h to converge on the specified mesh, and 1 week to obtain all the characteristic points.

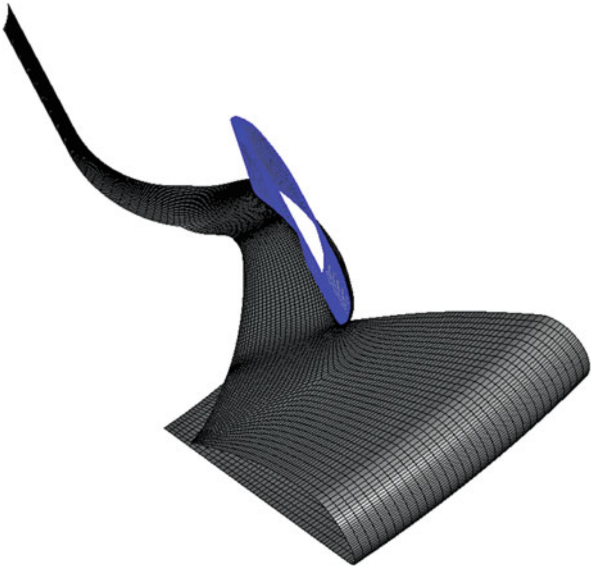


Fig. 3 Mesh generated for a single blade passage

2.4 Results and Discussion

The steady-state simulations were performed using different turbulence models and discretization schemes. The comparison of the calculated pressure characteristic (a dependence of the pressure coefficient from the flow coefficient) with the first order upwind discretization schemes for the different turbulence models to the experimental values is depicted in Fig. 4. All the turbulence models predicted similar pressure characteristic shapes which are plain even in the area of the strong unsteady effects $\phi_2 < 0.2$ ($G < 0.325$ kg/s). Moreover, the obtained total pressure coefficient values are underpredicted in the region of the high flow rates. This is likely to mean high losses due to the dissipation effect of the first order schemes. It should be noted that the overall prediction of the characteristic curve shape in comparison to the shape of experimental curve is unsatisfactory. Thus, the first-order accuracy is not sufficient to reproduce the stationary characteristic of the compressor stage regardless of the turbulence model.

The comparison of the pressure characteristic obtained by the second order central and upwind discretization schemes for the Spalart-Allmaras (SA) and k-ε (KE) turbulence models to experimental values is depicted in Fig. 5. It should be noted that there were strong convergence issues for the low flow rates ($\phi_2 < 0.2$), which could be the result of strong instabilities in the flow reproduced by the second-order discretization. The second-order discretization schemes predicted the pressure characteristic curve shape much better than the first-order schemes. Due to neglect of the hub and shroud leakages overprediction of the pressure coefficient in the area of the high flows rates is quite expectable.

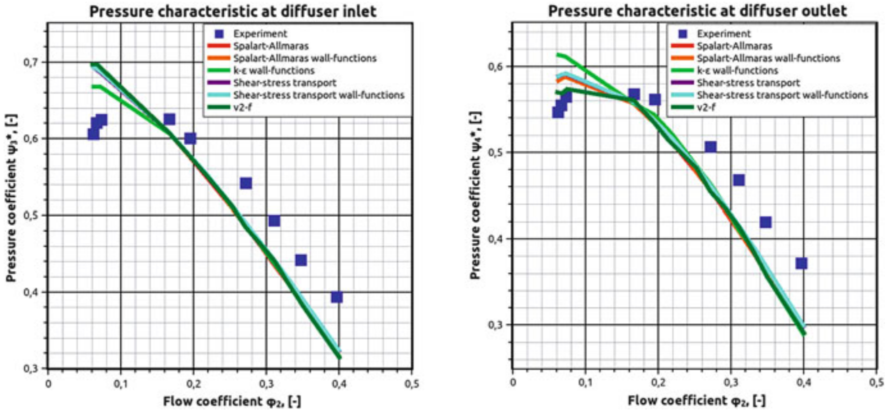


Fig. 4 Pressure characteristic for the first-order schemes (*left*—diffuser inlet, *right*—diffuser outlet)

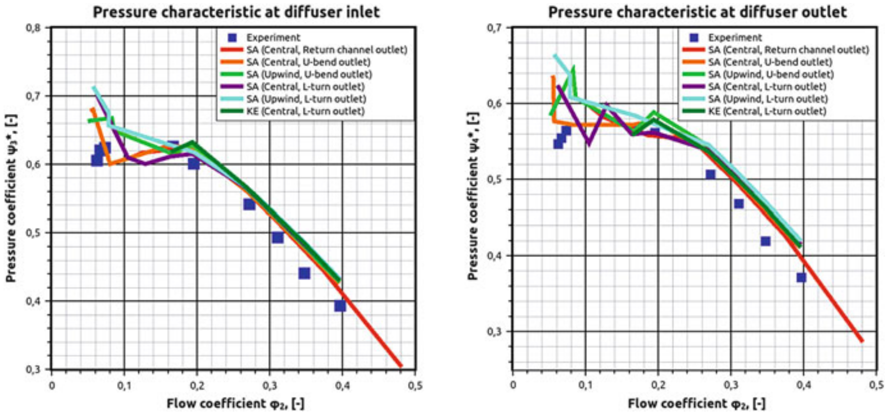


Fig. 5 Pressure characteristic for the second-order schemes (*left*—diffuser inlet, *right*—diffuser outlet)

Formation of a large vortex structure on the shroud with the decrease of the flow rate from $\phi_2 = 0.275$ to $\phi_2 = 0.1$ (from $G = 0.425$ to $G = 0.175$ kg/s) is depicted in Fig. 6. A severe difference between the flow fields predicted by the first and the second order discretization techniques is noticeable. The predicted vortex starts to form when the flow rate falls below $\phi_2 = 0.4$ ($G = 0.65$ kg/s) only with the second-order discretization, and subsequently, this vortex occupies a significant area of the channel.

A detailed analysis of the convergence issues occurring at low flow rates with second-order discretization was performed. Different outlet locations were investigated first (return channel inlet and outlet, full U-bend and L-turn), but the nature of convergence retained. Then return channel vanes were added to the

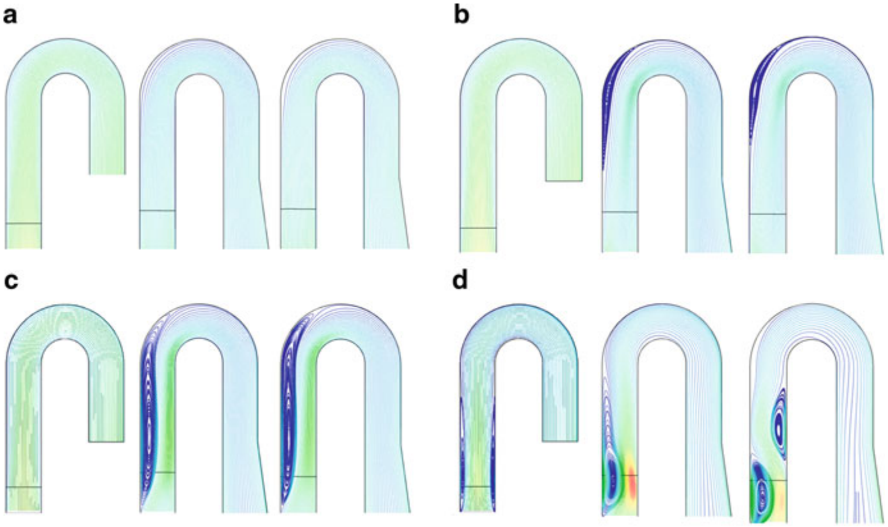


Fig. 6 Flow paths colored by velocity magnitude for different flow rates (*left*—first order upwind, *center*—second order central, *right*—second order upwind). (**a**) Flow rate: $\varphi_2 = 0.275$. (**b**) Flow rate: $\varphi_2 = 0.2$. (**c**) Flow rate: $\varphi_2 = 0.175$. (**d**) Flow rate: $\varphi_2 = 0.1$

geometry model to significantly diminish the circumferential velocity component at the outlet. However, the nature of convergence retained. Only the decrease of the number of meridional flow paths and cells around the blade profile provided desirable convergence. Thus, the lack of convergence is due to the strong vortices near blade profile at low flow rates. Therefore unsteady simulations should be performed for the accurate resolution of these vortices.

3 Conclusions

The results of performed simulations are highly dependent on the order of discretization. The first order accuracy is unacceptable due to the numerical dissipation effects. The second order accuracy performs good agreement with the experimental data and is capable of reproducing vortices development. In the area of the low flow rates the large vortices formed in the interblade passage lead to the lack of convergence. Additional steady-state simulations should be carried out on different grids to determine convergence parameters. Unsteady simulations of the flow field for full annulus geometry of the compressor should be carried out then.

Acknowledgements The authors thank Numeca Russia Company for funding the work and for valuable technical support.

References

1. Izmaylov, R.: Numerical modeling of unsteady flow phenomena in a centrifugal compressor stage. Compressor Pneumatics **5**, 10–16 (2011)
2. Kononov, S.: Investigation of unsteady processes in centrifugal compressor for developing diagnostics of unstable regimes. Ph.D. thesis, Leningrad, LPI (1985)
3. Numeca International: Numeca Fine/Turbo User Manual 8.9. Numeca International, Belgium (2011). http://www.numeca.com

Large Eddy Simulation of Boundary-Layer Flows over Two-Dimensional Hills

Ashvinkumar Chaudhari, Antti Hellsten, Oxana Agafonova, and Jari Hämäläinen

Abstract Large Eddy Simulations (LES) are performed for turbulent boundary-layer flows over two-dimensional (2D) hills or ridges of two different slopes at Reynolds number equal to 3,120 based on the hill height and the free stream velocity. The surface of the hill is assumed to be aerodynamically smooth. The hill height is considerably smaller than the boundary-layer depth. The hill models used in this study are the same as those used in the RUSHIL wind tunnel experiment carried out by Khurshudyan et al. (United States Environmental Protection Agency Report, EPA-600/4-81-067, 1981) and LES results are compared with the wind tunnel measurements. This study focuses on the overall flow behaviour changes as a function of the hill slope. The results of the mean velocity, the flow separation, and the turbulence quantities are discussed in the paper. It is shown that LES produces overall satisfactory results on the turbulent flow over the 2D hills. Especially for less steep hill, the flow behaviour is well predicted by LES.

A. Chaudhari (✉) • J. Hämäläinen
Center of Computational Engineering and Integrated Design (CEID), Lappeenranta University of Technology (LUT), PO Box 20, 53851 Lappeenranta, Finland
e-mail: Ashvinkumar.Chaudhari@lut.fi; Jari.Hamalainen@lut.fi

A. Hellsten
Finnish Meteorological Institute (FMI), PO Box 503, 00101 Helsinki, Finland
e-mail: Antti.Hellsten@fmi.fi

O. Agafonova
Department of Mathematics and Physics, Lappeenranta University of Technology (LUT), PO Box 20, 53851 Lappeenranta, Finland
e-mail: Oxana.Agafonova@lut.fi

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_29,

1 Introduction

The modelling of a wind flow over complex terrains containing, e.g. hills, ridges, forests, and lakes is of great interest in wind energy applications, as it can help in locating and optimizing the wind farms. Computational Fluid Dynamics (CFD) has become a popular technique during the last few decades. Due to inherent unsteady phenomena of wind flow over complex terrain, it can be difficult to model by the Reynolds-Averaged Navier-Stokes (RANS) approach. Thus, unsteady simulation approaches, most importantly Large Eddy Simulation (LES), are often more suitable for this kind of flows. This research is oriented towards LES for the atmospheric flows over complex terrains. However, a systematic study of the boundary layer flow over an idealized hilly terrains is a necessary step towards better understanding of the flow over realistic complex terrains. It is therefore desirable to first validate LES results against the wind tunnel measurements to get confidence on our LES approach, and that is the subject of this paper. In this paper, LES are carried out for the turbulent boundary layer flows over aerodynamically smooth two-dimensional (2D) hills with two different slopes. So far, several studies have been reported on the flow over hills as well as series of hills using RANS and LES approaches [1,4,5,7,8]. Turbulent flow over a steep hill contains relatively complex mean-flow characteristics such as separation and reattachment. As the flow passes over the hill, a recirculation region can be formed behind the hill and the turbulence is enhanced in the wake region. Thus, it is important to detect the influence of different hill shapes on overall flow behaviour over hilly terrains. The focus of this paper is on the changes in the mean velocity field and in the turbulence intensity as a function of the hill slope. It is shown that the present LES produces reasonably realistic results on the turbulent flow over the 2D hills. Moreover, the prediction of the flow separation and reattachment-length for the steeper hill is closer to the measurements than the other numerical studies reported in the past for the same hill geometry.

2 Numerical Model and Computational Details

The generic hill geometries used here are the same as those used in the RUSHIL wind tunnel experiment carried out by Khurshudyan et al. [6]. In this study, two 2D hills with different width to height ratios are studied. The hill height H is fixed to 0.117 m in both cases but the hill half length a is varied from $3H$ to $5H$ as shown in Fig. 1. The shapes of the hills are defined by the parametric formulae given in [1,4]. These two hills are named here as Hill3 and Hill5 according to their a/H ratios and their corresponding maximum hill slopes are 26° and 16°, respectively. The depth of the boundary layer δ is assumed to be 1 m, i.e. $\delta = 8.55H$. The total wind-wise (horizontal) length of the computational domain L_x is set to 5.34 m and the width of domain in the cross-wind direction (z) is set to one boundary layer depth, i.e. $L_z = \delta = 1$ m. In the present flows, the frictional Reynolds number Re_τ based on the friction velocity u_τ and δ is equal to 1,187, which is by far high enough to sustain

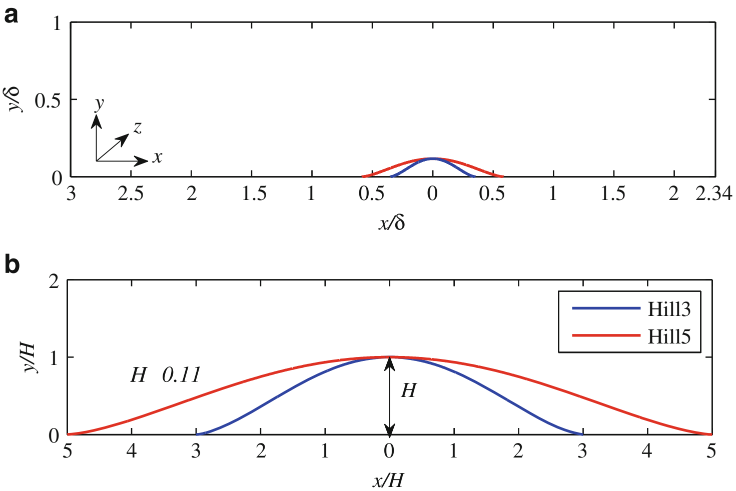


Fig. 1 (**a**) Side-view of the computational domains, (**b**) closer look on the hill shapes (Hill3 and Hill5)

fully turbulent flow. The grid resolution in the vertical direction Δy is varied from 0.0004 to 0.0379 m, corresponding to $y_1^+ \approx 0.5$. The wind-wise grid resolutions Δx is non-uniform with relatively finer grid on the hill surface corresponding to the average value $\Delta x_{avg} = 0.01948$ m. The cross-wind grid resolution Δz is fixed to 0.01587 m. The whole computational grid consists of 275 × 121 × 64 hexahedron cells for both Hill3 and Hill5.

LES directly resolves the large turbulent eddies by the computational grid whereas the eddies smaller than the grid size need to be modelled using sub-grid-scale (SGS) model. The filtered continuity and momentum equations for incompressible flow as given by [2] are time-integrated numerically using the second order implicit method and discretized in space using the bounded central difference scheme. The commercial finite-volume-based software ANSYS Fluent 13.0 is employed with the Smagorinsky-Lilly SGS-model [2].

Two different simulations for two hills are run for $t = 40$ s with all quantities are time averaged over the last 30 s. It was checked in the Hill3 case that the flow statistics were almost converged after 30 s of time averaging. In addition to the time averaging, the results are also averaged over the homogeneous cross-wind direction.

2.1 Boundary Conditions

The logarithmic mean-velocity profile

$$u = \frac{u_\tau}{\kappa} \ln\left(\frac{y}{y_0}\right) \tag{1}$$

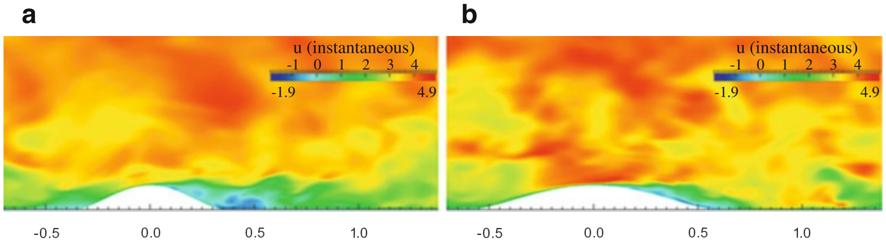


Fig. 2 Instantaneous wind-wise velocity contours. (**a**) Hill3. (**b**) Hill5

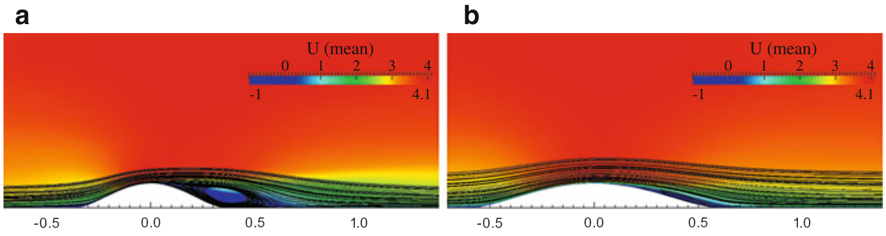


Fig. 3 Mean wind-wise velocity contours together with mean streamlines. (**a**) Hill3. (**b**) Hill5

is used at the inlet boundary. Here u_τ is the friction velocity, $\kappa = 0.41$ is the Von Karman constant, and $y_0 = 0.000157\,\mathrm{m}$ is the ground roughness length. The outflow boundary condition [2] is used at the outlet boundary. Periodic boundary conditions are set in the cross-wind direction and the symmetry condition is used on the top boundary. No-slip condition is set on the lower boundary, i.e. the ground surface. On the inflow boundary, artificial perturbations are generated using so called random 2D vortex method [2] leading to constant turbulence intensity of 12 %. The perturbation field is superimposed to the mean velocity profile via a vorticity field. The Reynolds number Re_H based on H and the free stream velocity U_∞ is equal to 3,120.

3 Results and Discussions

The LES results are compared with the hot-wire measurements of the RUSHIL wind-tunnel experiment [6]. Figure 2 shows the instantaneous wind-wise velocity distributions on xy planes of Hill3 and Hill5. Figure 3 shows the mean wind-wise velocity distributions together with mean streamlines at the lower part of xy planes of Hill3 and Hill5. The upstream flow is found almost fully developed shortly after it enters the domain at $x = -23.5H$ mostly owing to the artificial inflow turbulence. In the closer proximity of the hill it gets influenced by the presence of hill downstream. A very small volume of reversed flow is found at the upwind base of both hills $x = -a$. The streamlines in Fig. 3 shows the major recirculation,

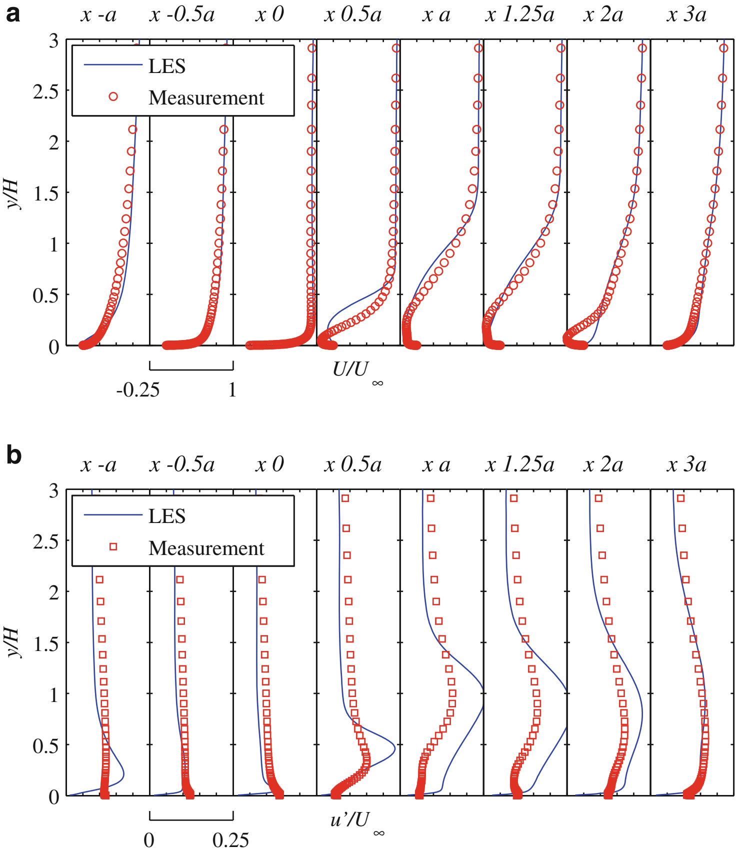


Fig. 4 Vertical profiles of mean wind-wise velocity U/U_∞ and turbulence intensity u'/U_∞ compared with measurements for Hill3, $a = 3H$. (**a**) Mean velocity U/U_∞. (**b**) Turbulence intensity u'/U_∞

and the flow separation and reattachment locations. After the reattachment the flow gradually redevelops toward downstream.

Figure 4a, b show the mean wind-wise velocity U/U_∞ and the turbulence intensity u'/U_∞ profiles compared with the measurements of Hill3, respectively. According to Fig. 4a, the LES mean-flow profiles agree reasonably well with the measurements in case of Hill3. However, the reattachment point is predicted at $x = 5.75H$ which is somewhat more upstream than the measurement reattachment location $x = 6.5H$. Castro and Apsley [4] performed RANS simulation for flow over the same hill (Hill3) using a modified $k-\epsilon$ turbulence model and predicted the reattachment point between $x = 4.1H - 5H$ [4]. Allen and Brown [1] performed LES for Hill3 and reported the reattachment point at $x = 3.6H$ [1]. Thus, our reattachment-length prediction for Hill3 is closer to the measurement than the other

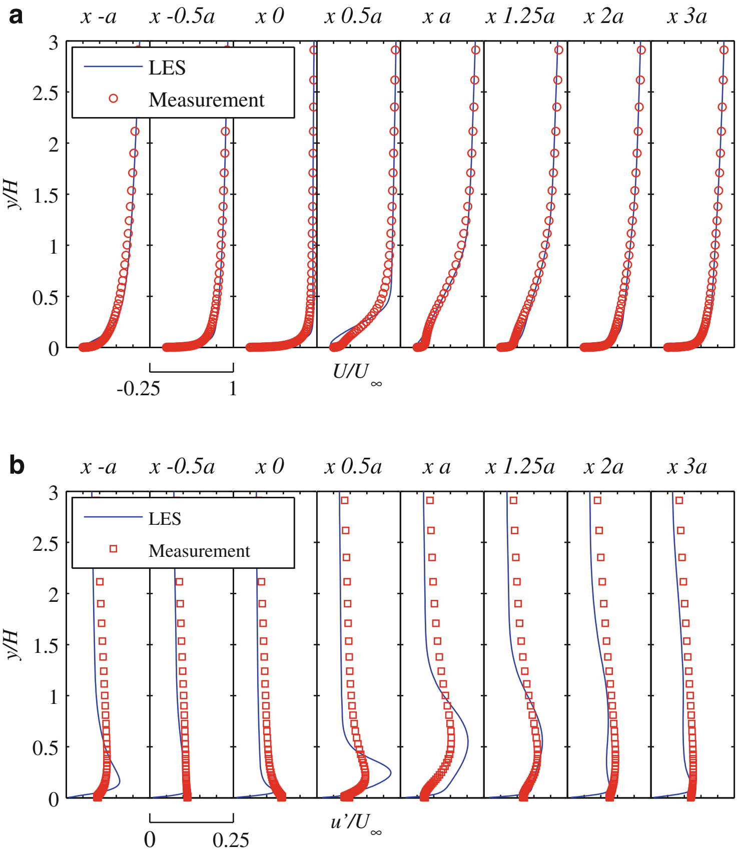


Fig. 5 Vertical profiles of mean wind-wise velocity U/U_∞ and turbulence intensity u'/U_∞ compared with measurements for Hill5, $a = 5H$. (**a**) Mean velocity U/U_∞. (**b**) Turbulence intensity u'/U_∞

numerical studies reported in the past. The turbulence intensity is found slightly higher than the measured values in the separated region (see Fig. 4b). On the other hand, the Reynolds number Re_H of the present flow is much smaller than the wind-tunnel value. Tamura et al. [8] carried out LES for a slightly different 2D hill ($a = 2.5H$) with $Re_H = 4{,}550$, and the present results also have qualitatively good agreement with their LES results as well as the wind tunnel measurements by [3].

Figure 5a, b show the mean wind-wise velocity U/U_∞ and turbulence intensity u'/U_∞ profiles compared with the measurements of Hill5, respectively. LES results for Hill5 have better agreement with the measured profiles than those of Hill3. From Fig. 5a, it seems that there is no mean flow separation according to the measurements but during the wind tunnel experiment, the instantaneous flow reversals were frequently observed through smoke visualization at the downwind base of Hill5.

However, in the average sense the flow remained attached [6], but present LES predicts a small flow separation on the lee side of the hill and flow reattaches quickly after the downwind base, i.e. at $x \approx 5.5H$. Tamura et al. [8] reported instantaneous flow separation after the hill summit even for more shallowed hill $a = 7.5H$ compared to Hill5 but also in that case the average flow remained attached [8]. In general, Hill5 case is more sensitive than Hill3 because of the lower slope and hence the flow being on the verge of separation. This means that the small changes in the upstream boundary layer may trigger separation and lead to a completely different flow over the lee side of the hill and downstream of it.

4 Conclusions

In this paper, we have carried out LES to investigate the turbulent boundary layer flows over two 2D hills with different width to height ratios and the results are compared with the RUSHIL wind tunnel measurements [6]. We have discussed the mean flow development, flow separation and reattachment due to change in a hill length. By comparing our results with [1, 4, 6, 8], it seems that LES produces reasonably realistic results for flow over the Hill3. To our knowledge, the present LES predicted the reattachment length more accurately than the previous studies for this particular hill geometry (Hill3). In the case of Hill5, LES results have even better agreement with measured profiles compared to Hill3. Actually most of the observed discrepancies between LES and the measured flow are likely owing to the uncertainties related to the artificially generated turbulence at the inflow boundary. Also, the lower Reynolds number of LES may be responsible for some differences.

Acknowledgements The authors would like to thank to the CSC-IT Center for Science Ltd., Espoo (Finland) for their valuable computational facilities and support. The work presented here is connected to the RENEWTECH project which aimed for development of wind power technology and business in South-East Finland and the work is funded by the European Regional Development Fund (ERDF).

References

1. Allen, T., Brown, A.: Large-eddy simulation of turbulent separated flow over rough hills. Boundary Layer Meteorol. **102**, 177–198 (2002)
2. Ansys: ANSYS Fluent 13.0 Theory Guide. Ansys, Inc., Canonsburg (2010)
3. Cao, S., Tamura, T.: Experimental study on roughness effects on turbulent boundary layer flow over a two-dimensional steep hill. J. Wind Eng. Ind. Aerodyn. **94**, 1–19 (2006)
4. Castro, I.P., Apsley, D.D.: Flow and dispersion over topography: a comparison between numerical and laboratory data for two-dimensional flows. Atmos. Environ. **31**, 839–850 (1997)
5. Griffiths, A., Middleton, J.: Simulations of separated flow over two-dimensional hills. J. Wind Eng. Ind. Aerodyn. **98**, 155–160 (2010)

6. Khurshudyan, L.H., Snyder, W.H., Nekrasov, I.V.: Flow and dispersion of pollutants over two-dimensional hills. United States Environmental Protection Agency, Report No. EPA-600/4-81-067 (1981)
7. Kim, J.J., Baik, J.J., Chun, H.Y.: Two-dimensional numerical modeling of flow and dispersion in the presence of hill and buildings. J. Wind Eng. Ind. Aerodyn. **89**, 947–966 (2001)
8. Tamura, T., Cao, S., Okuno, A.: LES study of turbulent boundary layer over a smooth and a rough 2D hill model. Flow Turbul. Combust. **79**, 405–432 (2007)

Part V
Medicine

Overview

This section contains four contributions dealing with industrial mathematics for medical applications. In a first contribution on *A Visual Representation of the Drug Input and Disposition Based on a Bayesian Approach*, Olivier Barrière et al. apply advanced mathematical tools to a practical problem: how to model the relation between the compliance to a drug prescription, i.e., the degree to which a patient correctly follows medical advice, and the drug disposition, i.e., the patient pharmacokinetics characteristics. Based on Bayesian theory, the authors develop a compliance spectrum to describe this relationship in a both intuitive and interactive way.

Magda Rebelo et al. develop in a second contribution on *Modelling a Competitive Antibody/Antigen Chemical Reaction that Occurs in the Fluorenscence Capillary-Fill Device* a mathematical model for a competitive chemical reaction between an antigen and a labelled antigen for antibody sites on a cell wall. This model consists of two coupled diffusion equations, equivalent to a pair of coupled singular integro-differential equations, which becomes both nonlinear and nonlocal via the boundary conditions. Numerical simulation results based on real data are obtained by a product integration method.

The third paper written by Thomas Martin Cibis and Nicole Marheineke on *Model-Based Medical Decision Support for Glucose Balance in ICU Patients: Optimization and Analysis* deals with the control of the glucose balance in intensive care unit (ICU) patients using an insulin therapy. More precisely, the authors both analyze and solve numerically the optimal control problem that arises if the simulation model GlucoSafe by Pielmeier et al. is used in this context. This model describes the temporal evolution of the blood glucose and insulin concentrations in the human body by help of a nonlinear dynamic system of first-order ordinary differential equations.

The last contribution on *Epileptic Seizures Diagnose using Kunchenko's Polynomials Template Matching* written by Oleg Chertov and Taras Slipets uses a template matching method based on Kunchenko's polynomials, a redundant dictionaries method, for electroencephalogram (EEG) signal processing.

Michael Günther

A Visual Representation of the Drug Input and Disposition Based on a Bayesian Approach

Olivier Barrière, Jun Li, and Fahima Nekka

Abstract Compliance to a drug prescription describes the degree to which a patient correctly follows medical advice. Poor compliance significantly impacts on the efficacy and safety of a planned therapy, which can be summed up by the dictum: "a drug only works if it's taken". However, the relationship between drug intake and pharmacokinetics (PK) is only partially known, especially the so-called inverse problem, concerned with the issue of retracing the patient compliance scenario using limited clinical knowledge. Based on the Bayesian theory, we develop a decision rule to solve this problem. Given an observed concentration, we determine, among all possible compliance scenarios, which is the most probable one. Using a simulation approach, we are able to judge the quality of this retracing process by measuring its global performance. Since the sampling concentration is the result of both patient compliance (drug input) and patient PK characteristics (drug disposition), two natural questions arise here: first, given two different sampling concentration values, can we expect the same performance of the retracing process? Second, how is this performance affected by the PK variability between individuals? For this, we here design an heatmap-style image, called Compliance Spectrum, that provides an intuitive and interactive way to evaluate the relationship between drug input and drug disposition along with their consequences on PK profile. The current work provides a solution to this inverse problem of compliance determination from a probability viewpoint and uses it as a base to build a visual representation of drug input and disposition.

O. Barrière • J. Li • F. Nekka (✉)
Faculté de Pharmacie, Université de Montréal, C.P. 6128, Succ. Centre-ville, Montréal (Québec), Canada H3C 3J7
e-mail: olivier.barriere@umontreal.ca; li@crm.umontreal.ca; fahima.nekka@umontreal.ca

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_30,

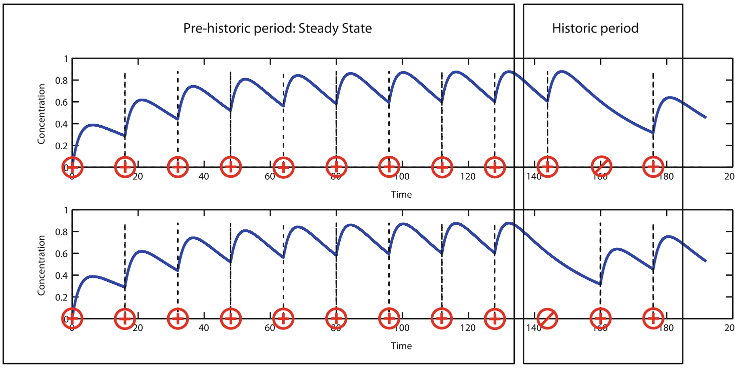


Fig. 1 Two different compliance scenarios over the last three doses are represented in the *upper* and *lower panels*, following a pre-historic period of perfect compliance

1 Bayesian Decision Approach for the Inverse Problem

Noncompliance to drugs generally involves errors in drug execution, such as missing or doubling prescribed doses, as well as deviations from nominal times. These errors in drug intake are complex and random in nature reflecting the involved psychological and societal factors. Linking compliance to drug exposure, usually recognized as the direct problem, has been so far the central topic. This stimulated many modeling and simulation efforts, with the purpose to establish a quantitative link between compliance and some drug related outcomes. This forward direction of the problem naturally raises the inverse version of reconstructing drug intake from limited clinical information. Compared to pharmacokinetics where deterministic compartmental approach are predominant, compliance has to be formulated using a probabilistic language [1,2].

1.1 Compliance Scenarios

Motivated by the information loss along the drug intake, we decomposed its time period into two parts, Fig. 1:

- The pre-historic period refers to drug events that happened long time ago as drug intake memory has been lost and are unlikely to be retraced. We thus assume the steady-state has been reached.
- This is followed by the historic period that precedes the patient's visit to the clinic, which we aim to retrace since the information from its dose events is still detectable.

As missing doses are the most frequent and influential on the issue of therapy, we only consider here scenarios where doses are either taken or missed on nominal

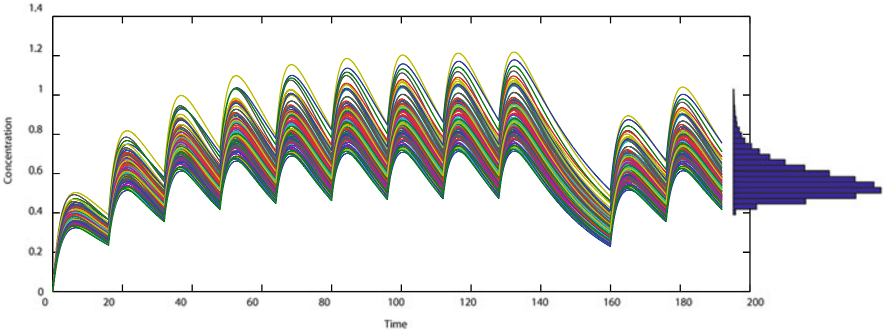


Fig. 2 Five hundred concentration tine courses are generated for the same combination of dosing events 110. The histogram on the *right* represents the probability density function of the final concentration

times. To represent these dose combinations, we adopt a binary system for their notation. For a historic period of length N, that is a total number of 2^N possible compliance scenarios, each compliance scenario ω_j is represented by a binary sequence of N digits, where 0 and 1 refer to missing and taken dose events, respectively. For instance, the combination $(1, 1, 0)$ (shortened to 110) has to be read from left to right: 1 dose taken yesterday, 1 dose taken the day before yesterday and one dose missed (0) 3 days ago. We treat the prior probabilities of combinations of dosing events as equiprobable. This choice is for sake of simplicity: any other probability distribution can be assumed, such as Binomial distribution or Markov chains.

1.2 Retracing Process Based on Bayesian Decision

Based on the Bayesian theory, we developed a decision rule to solve the inverse problem of compliance [3]. Given a Pop-PK model and an observed concentration C, we are able to determine, among all possible compliance scenarios ω_j, $j = 0, \cdots, 2^N - 1$, which is the most probable one $\widehat{\omega_j}$.

- First, using an approved Pop-PK model for a specific drug, we use Monte-Carlo simulations to get a whole range of concentration values for a population of virtual patients taking into account the distribution of the PK parameters and repeat this for all possible drug compliance scenarios, Fig. 2. We then estimate the different likelihoods of concentration at a specific sampling time given each compliance scenario: $p(C|\omega_j), \forall j = 0, \cdots, 2^N - 1$, Fig. 3.
- Next, based on the observed sampling concentration, we compute the posterior probabilities of each scenario using the Bayes rule: $P(\omega_j|C) = \frac{P(\omega_j)p(C|\omega_j)}{p(C)}, \forall j = 0, \cdots, 2^N - 1$, Fig. 4.

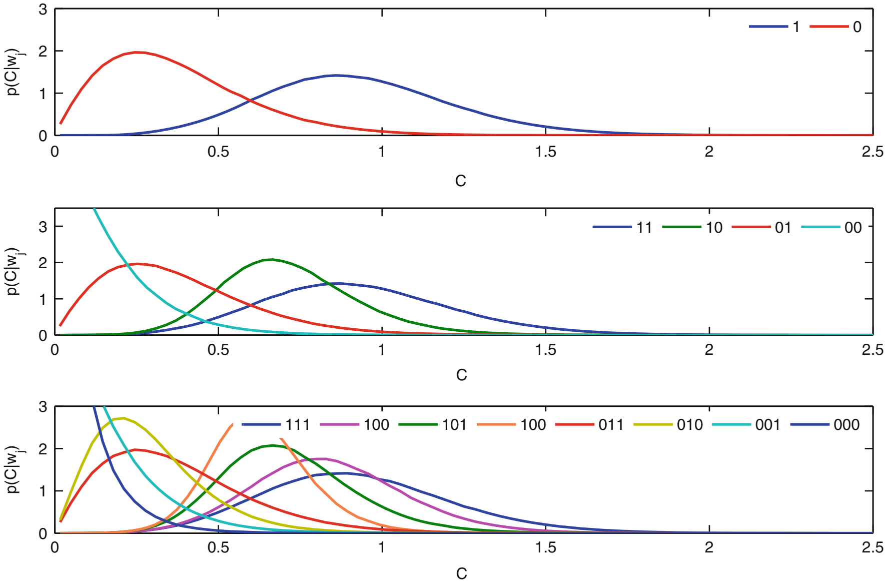


Fig. 3 Likelihood functions (probability density function) of the final concentration for scenarios ω_j: $p(C|\omega_j)$

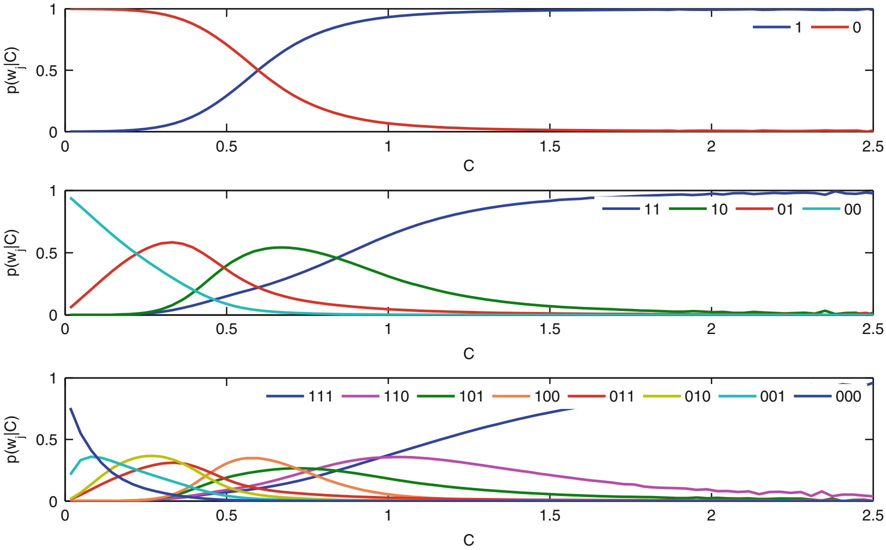


Fig. 4 Posterior probabilities of the various combinations of dosing events for a given sampling concentration value C: $P(\omega_j|C)$

- Finally, the most probable scenario is identified as the one with the largest posterior probability among all possible combinations given an observed drug concentration C at the sampling time, i.e. $\hat{j} = \arg\max_{j} \big(P(\omega_j|C)\big)$.

1.3 Performance of the Retracing Process

To judge the quality of the retracing process, we evaluate its success rate by comparing the estimated dosing events of a large number of virtual patients $\widehat{\omega_{j_i}}$ with the actual ones ω_{j_i}, based on the simulation of their sampling concentrations at a specific time. In the current study, the average number of the last scheduled doses correctly retraced during the historic period is used as a performance indicator. This indicator gives the average number of consecutive doses prior to the last sampling time that can be correctly retraced without interruption.

2 Compliance Spectrum

2.1 Challenges

The average number of the last scheduled doses correctly retraced is a reliable indicator to asses the performance of the decision rule. Nevertheless, this global value is based on the implicit assumption that the performance is the same for every patient: only one scalar value to asses the performance of the retracing process. *Does every concentration has the same odds of being correctly retraced?*

When there is no variability and no errors (deterministic case), every scenario leads to a different final concentration value. Thus, given the last sampling information as an input, there is one and only one possible scenario. On the other hand, when including the variability from the Pop-PK model (stochastic case) the problem becomes no more invertible since the same concentration value can come from multiple scenarios for different patients. *How is the transition from the deterministic case to the stochastic case?*

2.2 Performance Evaluation Broken Down by Concentration

To represent performance indicators for the different sampling concentrations values, we first split the range of possible concentrations into intervals, called bins. Patients' sampling concentrations will fall within one of these intervals. The average number of the last scheduled doses correctly retraced for patients whose sampling concentration falls within a specific interval is calculated and represents the performance for this interval.

2.3 Global Variability Coefficient: α

In the deterministic case, things happen exactly the way they would if we had only one typical patient with no observational error nor uncertainty. In the stochastic case, every patient has his own individual PK parameters which are often following log-normal distributions. The residual error can be modeled in different ways but also follows some probability distribution. The variance of these distributions (or their coefficient of variation) regulates their width: the larger the variability, the wider the distribution. We multiply this variability coefficient of all the distribution by a coefficient α. Therefore, if $\alpha = 0\,\%$, all the parameters are fixed to their typical value, there is no variability and we find the deterministic case again. If $\alpha = 100\,\%$ we get the actual variability of the published model and if $\alpha = 200\,\%$ we get twice the variability of the published model.

2.4 Construction of the Compliance Spectrum

The performance evaluation depends on two factors: the concentration (split into intervals) and the variability (handled by α). To obtain the Compliance Spectrum, the performance of the retracing process is reported in a 2D image with a heat-map format, where the horizontal axis represents the sampling concentration, the vertical axis is for the multiplicative factor α and the color expresses the performance.

2.5 Results

Taking various drug models, Fig. 5, we aim to retrace the last 2 days before sampling, which gives rise to 4 compliance scenarios. The Compliance Spectrum indicates that there are exactly four sampling concentrations, each being caused by a unique scenario as no variability is present. This corresponds to a unique solution of the inverse problem in the traditional meaning. These particular sampling values will be referred to as characteristic concentrations of the Compliance Spectrum. When variability is involved, we can notice that the traditional uniqueness of the solution is no more valid and in fact, one sampling value may originate from different scenarios. For small variability, we clearly see that the possible sampling concentrations are separated into four zones, each emerging from one characteristic concentration. The number of concentration zones corresponds to the number of scenarios being considered. In this situation, the concentrations observed in one

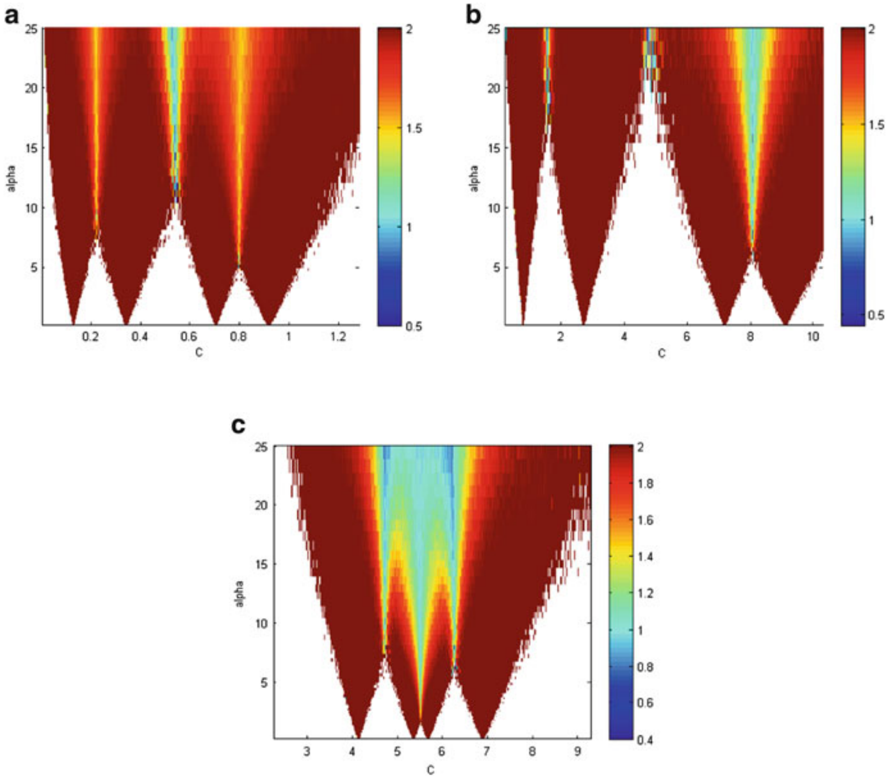


Fig. 5 Compliance spectra for three Pop-PK models: (**a**) and (**b**) one compartment models with different typical values, (**c**) two compartment model

zone can all be attributed to a single scenario. Outside of these zones, a sampling concentration is unlikely to be observed. Moreover, the size of these zones increases with variability, indicating that a larger range of sampling concentrations can be observed. Until a threshold variability, two adjacent zones will meet, making it difficult to attribute the sampling concentrations to a single scenario. From this merging point, the uniqueness of the solution has no meaning. As the variability increases, an increasing number of observed concentrations can be attributed to more than one scenario.

3 Conclusion and Perspective

To get a whole vision of the journey of a drug, when prescribed to the patient, the drug input should be given the same importance as drug disposition [4]. The Compliance Spectrum exhibits the interaction between drug input and drug disposition. These two processes, one being behavior related, thus active in nature and the other, physiology related and passive in nature, are put on the same level in order to help clarifying drug properties through extraction and exploitation of the hidden information. We provide here a direct picture of this drug intake-pharmacokinetics link. The rich information carried out by the Compliance Spectrum deserves to be thoroughly exploited. We have already identified the characteristic concentrations as the most readily exploitable feature. A deeper investigation of other properties of the compliance spectrum, either through shape or color, needs to be performed.

References

1. Li, J., Nekka, F.: A pharmacokinetic formalism explicitly integrating the patient drug compliance. J. Pharmacokinet. Pharmacodyn. **34**, 115–139 (2007)
2. Li, J., Nekka, F.: A probabilistic approach for the evaluation of pharmacological effect induced by patient irregular drug intake. J. Pharmacokinet. Pharmacodyn. **36**, 221–246 (2009)
3. Barriere, O., Li, J., Nekka, F.: A bayesian approach for the estimation of patient compliance based on the last sampling information. J. Pharmacokinet. Pharmacodyn. **38**(3), 333–351 (2011)
4. Harter, J., Peck, C.C.: Chronobiology: suggestions for integrating it into drug development. Ann. N. Y. Acad. Sci. **618**, 563–71 (1991)

Modelling a Competitive Antibody/Antigen Chemical Reaction that Occurs in the Fluorescence Capillary-Fill Device

Magda Rebelo, Teresa Diogo, and Sean McKee

Abstract A mathematical model in the form of two coupled diffusion equations is provided for a competitive chemical reaction between an antigen and a labelled antigen for antibody sites on a cell wall; boundary conditions are such that the problem is both nonlinear and nonlocal. This is then re-characterized as a pair of coupled singular integro-differential equations which is solved by a product integration method. Some numerical results based on real data are presented.

1 Introduction

This work is concerned with the development and analysis of a mathematical model to describe antibody/antigen chemical reactions occurring in the Fluorescence Capillary-Fill Device (FCFD). The FCFD is capable of detecting a particular disease provided the specific antibody produced by the human body is known. It consists of two plates of glass, separated by a narrow gap. A dissoluble reagent layer of antigen (or hapten) labelled with a fluorescent dye is affixed to the upper plate of the device while a specific antibody is immobilized on the lower plate. The cell is then filled,

M. Rebelo (✉)
Department of Mathematics, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Monte de Caparica, 2829-516 Caparica, Portugal

CEMAT-Instituto Superior Técnico, Lisbon, Portugal
e-mail: msjr@fct.unl.pt

T. Diogo
CEMAT-Department of Mathematics, Instituto Superior Técnico, Lisbon, Portugal
e-mail: tdiogo@math.ist.utl.pt

S. McKee
Department of Mathematics and Statistics, University of Strathclyde, Scotland, UK
e-mail: s.mckee@strath.ac.uk

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*, Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_31,

through capillary action, with a fluid which may or may not contain the (unlabelled) antigen. The objective is to determine whether this antigen is present and, if so, in what quantity. For instance, if a patient had a particular disease then (unlabelled) antigen would be present; otherwise, it would not be. The glass plates then act as a wave guide and a fluorescent beam is used to detect whether there is any (unlabelled) antigen (see Badley et al. [1]).

In this paper we focus on the competitive reaction between the antigen and the fluorescent antigen for the affixed antibodies. A mathematical model consisting of two coupled diffusion equations with nonlinear and nonlocal boundary conditions is obtained which can be re-characterized as a pair of coupled singular integro-differential equations. Another reformulation, as a system of four Volterra integral equations, is also considered in [3]. This work is a (considerable) extension of an earlier study of the non-competitive [2].

2 The Mathematical Model

Initially, the labelled antigen is wall-bound and it will be denoted by $X_F^{(b)}$. Upon dissolving, it shall be denoted by X_F. Furthermore, let X denote the unlabelled antigen and Y the specific antibody. Both X_F and X are free to diffuse in the solution, whereas the antibody Y is insoluble and remains on the lower plate where the antibody and the antigens react in the following way:

$$X + Y \underset{k_2}{\overset{k_1}{\rightleftharpoons}} XY; \qquad X_F + Y \underset{k_4}{\overset{k_3}{\rightleftharpoons}} X_F Y.$$

Thus on the lower plate labelled $(X_F Y)$ and unlabelled (XY) antigen-antibody molecules are created. On the other hand, the wall-bound antigen $X_F^{(b)}$ is treated as an independent species and we consider its dissolution as a further reaction $X_F^{(b)} \overset{k_5}{\rightarrow} X_F$. Note that the concentration of the labelled antigen on the side wall (i.e. $[X_F^{(b)}]$) is dissolved upon entry of the fluid possibly containing the unlabelled antigen; there is no recombination, so it is reasonable to consider the reaction as one way only. The parameters k_1, k_2 are the forward and backward reaction rates associated with the unlabelled antigen X; k_3, k_4 are the forward and backward reaction rates associated with the labelled antigen X_F and k_5 is the forward "reaction" rate associated with the wall-bound antigen $X_F^{(b)}$ (i.e. the rate at which $X_F^{(b)}$ dissolves). Let d denote the plate separation distance which is small compared with the size of the cell. Define the origin to be at some point on the upper plate and denote $x = d$ to be the corresponding point on the lower plate (see Fig. 1).

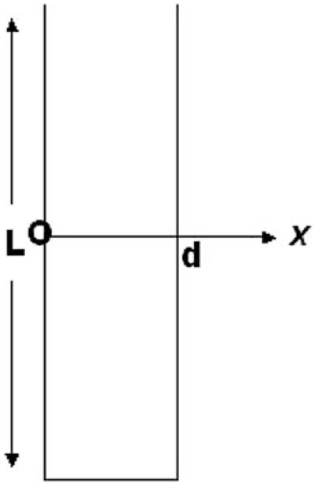


Fig. 1 Schematic diagram of a small cell

We denote by $[X]$, $[X_F]$, the concentrations (in moles/m^3) of X and X_F, respectively; $[XY]$, $[X_FY]$, the concentrations (in moles/m^2) of the complexes XY and X_FY, respectively, at $x = d$; $[X_F^{(b)}]$ denotes the concentrations (in moles/m^2) of $X_F^{(b)}$ at $x = 0$ and $[Y]$ denotes the concentrations (in moles/m^2) of the antibody Y at $x = d$. The variables $[X]$ and $[X_F]$ will now depend on x and t whereas $[XY]$, $[X_FY]$, $[Y]$ and $[X_F^{(b)}]$ only depend on t. Furthermore the initial concentrations of the antigens and antibody are given by $[X](x,0) = a$, $[X_F](x,0) = 0$, $x \in (0,d)$, $[X_F^{(b)}](0) = a_F$, $[Y](0) = c$. The parameters D and D_F denote the diffusion coefficients associated with X and X_F(m^2/s), respectively.

The one non-dimensional model which describes the competitive chemical reaction between an antigen (X) and a labelled antigen (X_F) for antibody (Y) sites on a cell wall is given by the following reaction-diffusion system with nonlinear boundary conditions:

$$\frac{\partial u}{\partial t}(x,t) = \delta \frac{\partial^2 u}{\partial x^2}(x,t), \quad \frac{\partial v}{\partial t}(x,t) = \frac{\partial^2 v}{\partial x^2}(x,t), \quad x \in (0,1), \quad t > 0, \tag{1}$$

subject to

$$u(x,0) = \mu,\ x \in (0,1), \tag{2}$$

$$v(x,0) = 0,\ x \in (0,1), \tag{3}$$

$$\frac{\partial u}{\partial x}(0,t) = 0, \quad t > 0, \tag{4}$$

$$\frac{\partial v}{\partial x}(0,t) = -\lambda\ \exp(-\lambda\, t), \quad t > 0, \tag{5}$$

$$\frac{\partial u}{\partial x}(1,t) = \gamma_1\, m\, (L_1 w_1(t) - (1 - w_1(t) - w_2(t))\, u(1,t)), \quad t > 0, \tag{6}$$

$$\frac{\partial v}{\partial x}(1,t) = \gamma_2\, m\, (L_2 w_2(t) - (1 - w_1(t) - w_2(t))\, v(1,t)), \quad t > 0, \tag{7}$$

together with the constraints

$$m\, w_1(t) + \int_0^1 u(x,t)dx = \mu, \quad t > 0, \tag{8}$$

$$m\, w_2(t) + \int_0^1 v(x,t)dx = 1 - \exp(-\lambda\, t), \quad t > 0, \tag{9}$$

where the dependent variables have been scaled as follows:

$$u(x',t') = \frac{d\,[X](x,t)}{a_F}, \quad v(x',t') = \frac{d\,[X_F](x,t)}{a_F},$$
$$w_1(t') = \frac{[XY](t)}{c}, \quad w_2(t') = \frac{[X_F Y](t)}{c}.$$

The other non-dimensional constants are given by (note time-scale ratios are abbreviated by (t-s r))

$$m = \frac{c}{a_F}, \quad \text{(molar ratio)}, \qquad \lambda = k_5 \frac{d^2}{D_F}, \qquad \text{(dissolution/diffusion t-s r)},$$

$$L_1 = \frac{d k_2}{a_F k_1}, \quad L_2 = \frac{d k_4}{a_F k_3}, \qquad \text{(reaction t-s r)},$$

$$E_1 = \frac{d^2}{D}\left(\frac{a_F}{d} k_1 + k_2\right), \quad E_2 = \frac{d^2}{D_F}\left(\frac{a_F}{d} k_3 + k_4\right), \quad \text{(diffusion/reaction t-s r)},$$

$$\delta = \frac{D}{D_F}, \quad \mu = \frac{a\, d}{a_F}, \quad \gamma_i = \frac{E_i}{1 + L_i}, \; i = 1, 2.$$

3 An Integro-Differential Equation Formulation

Taking Laplace transforms with respect to time of equations (1) and after some calculations we obtain that the solution $(u(x,t), v(x,t))$ at $x = 1$ satisfies

$$u(1,t) = \mu - m \int_0^t \frac{d w_1}{d\tau}(\tau) K(\delta(t-\tau)) d\tau, \quad t > 0, \tag{10}$$

$$v(1,t) = g(t) - m \int_0^t \frac{d w_2}{d\tau}(\tau) K(t-\tau) d\tau, \quad t > 0, \tag{11}$$

where $(w_1(t), w_2(t))$ is the solution of the two coupled integro-differential equations:

$$\frac{dw_1(t)}{dt} = \gamma_1\delta\left[(1 - w_1(t) - w_2(t))\left(\mu - m\int_0^t \frac{dw_1}{d\tau}(\tau)K(\delta(t-\tau))d\tau\right) - L_1 w_1(t)\right] \tag{12}$$

$$\frac{dw_2(t)}{dt} = \gamma_2\left[(1 - w_1(t) - w_2(t))\left(g(t) - m\int_0^t \frac{dw_2}{d\tau}(\tau)K(t-\tau)d\tau\right) - L_2 w_2(t)\right] \tag{13}$$

subject to the initial conditions $w_1(0) = w_2(0) = 0$, with

$$g(t) = 2\lambda\int_0^t \theta(1, t-s)\exp(-\lambda s)ds, \tag{14}$$

$$K(t) = \frac{1}{\sqrt{\pi t}}\left(1 + 2\sum_{n=1}^{\infty}\exp\left(-\frac{n^2}{t}\right)\right), \tag{15}$$

where θ is the *theta function*:

$$\theta(x,t) = \frac{1}{\sqrt{4\pi t}}\sum_{n=-\infty}^{\infty}\exp\left(-\frac{(x+2n)^2}{4t}\right), \quad -\infty < x < +\infty,\ t > 0.$$

Using this formulation of the problem it is possible derive small time asymptotic solutions (see [3]):

$$w_1(t) = \gamma_1\delta\mu t - \frac{4\mu\gamma_1^2 m\delta^{3/2}}{3\sqrt{\pi}}t^{3/2} + \mathscr{O}(t^2), \tag{16}$$

$$w_2(t) = \frac{2}{3}b_1\gamma_2 t^{3/2} - \frac{\sqrt{\pi}}{4}b_1 m\gamma_2 t^2 + \mathscr{O}(t^{5/2}), \tag{17}$$

$$u(1,t) = \mu - \frac{2m\gamma_1\mu\sqrt{\delta}}{\sqrt{\pi}}t^{1/2} + \gamma_1^2 m^2\delta\mu t + \mathscr{O}(t^{3/2}), \tag{18}$$

$$v(1,t) = b_1 t^{1/2} - \frac{\sqrt{\pi}}{2}b_1 m\gamma_2 t + \left(b_2 + \frac{2}{3}b_1 m^2\gamma_2\right)t^{3/2} + \mathscr{O}(t^2). \tag{19}$$

where $b_i \equiv b_i(\lambda)$ and are such that $g(t) = b_1 t^{1/2} + b_2 t^{3/2} + \mathscr{O}(t^{5/2}), \quad 0 < t \ll 1$ (for more details see [3]).

From (16) to (19), we observe that $w_1(t)$, $w_2(t)$, $u(1,t)$ and $v(1,t)$ have a singularity at $t = 0$.

4 A Numerical Method

We consider the product Euler method for the solution of the system (12)–(13). Define the uniform grid $I_h = \{t_i = ih, \quad 0 \leq i \leq N\}$, with stepsize $h = T/N$, on the interval $[0, T]$. On each subinterval $[t_j, t_{j+1}],\ j = 0,\ 1,\ \ldots, N-1$, we approximate $w_1(t)$ and $w_2(t)$ by their respective linear Lagrange polynomials and we obtain the scheme in the unknowns $(w_1^i, w_2^i),\ i = 1,\ 2,\ \ldots, N$,

$$\begin{cases} w_1^0 = 0, \quad w_2^0 = 0, \\ \dfrac{w_1^i - w_1^{i-1}}{h} = \gamma_1\, \delta \left(-L_1 w_1^i + (1 - w_1^i - w_2^i) \left(\mu - m \displaystyle\sum_{j=0}^{i-1} W_{i-j}(\delta)\, \frac{w_1^{j+1} - w_1^j}{h} \right) \right), \\ \dfrac{w_2^i - w_2^{i-1}}{h} = \gamma_2 \left(-L_2 w_2^i + (1 - w_1^i - w_2^i) \left(\tilde{g}(t_i) - m \displaystyle\sum_{j=0}^{i-1} W_{i-j}(1)\, \frac{w_2^{j+1} - w_2^j}{h} \right) \right), \end{cases} \tag{20}$$

where $w_k^i \approx w_k(t_i),\ k = 1, 2,\ i = 0, 1, \ldots, N$, with the quadrature weights given by

$$W_{i-j}(\delta) = \left(1 + 2 \sum_{n=1}^{l} \exp\left(-\frac{n^2}{\delta(t_i - t_j)} \right) \right) \int_{t_j}^{t_{j+1}} \frac{1}{\sqrt{\delta\pi(t_i - s)}} ds, \tag{21}$$

and $\tilde{g}(t_i)$ is an approximation of $g(t)$ at $t = t_i$, obtained by the product Euler method applied to (14) and given by

$$\begin{aligned} \tilde{g}(0) &= g(0) = 0 \\ \tilde{g}(t_i) &= \frac{\lambda}{\sqrt{\pi}} \sum_{j=0}^{i-1} \exp(-\lambda t_j) \sum_{n=-l}^{l} \exp\left(-\frac{(2n+1)^2}{4(t_i - t_j)} \right) \int_{t_j}^{t_{j+1}} \frac{1}{\sqrt{t_i - s}} ds, \\ & \qquad i = 1, 2, \ldots, N. \end{aligned}$$

Once we have computed the values w_1^i, w_2^i, the approximations u_i, v_i to $u(1, t_i), v(1, t_i)$, respectively, are given by the corresponding discretization of Eqs. (10) and (11), namely

$$\begin{aligned} u_i &= \mu - \frac{m}{h} \sum_{j=0}^{i-1} W_{i-j}(\delta)(w_1^{j+1} - w_1^j), \\ v_i &= \tilde{g}(t_i) - \frac{m}{h} \sum_{j=0}^{i-1} W_{i-j}(1)(w_2^{j+1} - w_2^j), \quad i = 1, 2, \ldots, N, \end{aligned} \tag{22}$$

with $u_0 = u(1, 0) = \mu, \quad v_0 = v(1, 0) = 0$ and $u_i \approx u(1, t_i),\ v_i \approx v(1, t_i),\ i = 0, 1, \ldots, N$.

Table 1 Data related with the Proteins (molecular weight $\simeq 10^5$) and used in numerical approximations

Dimensional parameters		Non-dimensional parameters	
D	$10^{-11}\,\mathrm{m}^2\,\mathrm{s}^{-1}$		
k_1	$10^5\,(\mathrm{moles})^{-1}\,\mathrm{s}^{-1}$	m	19,760.5
k_2	$10^{-4}\,\mathrm{s}^{-1}$	μ	0.001
a	1.67×10^{-6} moles m^{-3}	L_1	5.988×10^{-7}
D_F	$10^{-11}\,\mathrm{m}^2\,\mathrm{s}^{-12}$	L_2	5.988×10^{-7}
k_3	$10^5\,(\mathrm{moles})^{-1}\,\mathrm{s}^{-1}$	E_1	1,670
k_4	$10^{-4}\,\mathrm{s}^{-1}$	E_2	1,670
a_F	1.67×10^{-8} moles m^{-2}	δ	1
c	33×10^{-5} moles m^{-2}	λ	10^5
k_5	$10^4\,\mathrm{s}^{-1}$		

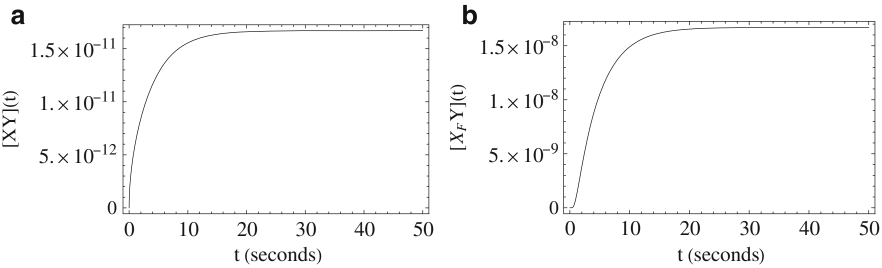


Fig. 2 Numerical approximation of the concentration of the two complexes. (**a**) Complex XY, $(XY)(t)$. (**b**) Complex $X_F Y$, $(X_F Y)(t)$

5 Numerical Results

In this section we present some numerical results for the initial-boundary value problem (1)–(9) with the data listed in Table 1.

In order to compute numerical approximations of w_1 and w_2 we consider algorithm (20) with stepsize $h = 1/1000$. The variables are then dimensionalized and numerical approximations of the concentrations of the complexes, $[XY](t)$ and $[X_F Y](t)$, and the concentrations of the labelled and unlabelled antigens at $x = d$, $[X](d,t)$ and $[X_F](d,t)$, are then determined. These are displayed in Figs. 2a, b, and 3a, b. In each figure, t denotes the time in seconds. From Fig. 2a, b we see that both $[XY]$ and $[X_F Y]$ grow monotonically at roughly the same speed to their respective (and rather different) asymptotic values, which they attain in approximately 30 s. The two orders of magnitude difference between $[XY]$ and $[X_F Y]$ would appear to be reflected in the two orders of magnitude difference between a and a_F. Figure 3a, b displays the antigen and the labelled antigen at the wall (i.e. $x = d$). One observes that $[X](d,t)$ drops initially as a result of the reaction and then grows to a peak due to diffusion (more rapidly than $[X_F](d,t)$) before reducing monotonically. The concentration $[X_F](d,t)$ is dissolved initially from the bound $X_F^{(b)}$. From Fig. 3b we see that $[X_F](d,t)$ grows to a peak (considerably smaller than $[X](d,t)$) before decreasing monotonically to zero in about 30 s. Thus, there is a small time delay

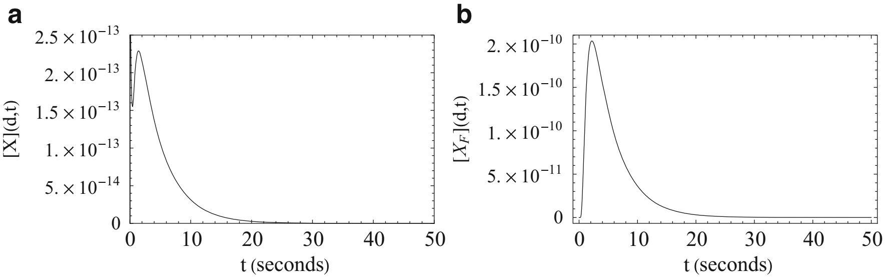


Fig. 3 Numerical approximation of the concentration of the labelled antigen at the wall side where the reaction takes place. (**a**) $x = d$, $(X)(d,t)$. (**b**) $x = d$, $(X_F)(d,t)$

while diffusion migrates the X_F molecules to $x = d$ whereupon there is an increase in $[X_F](d,t)$ before the reaction sets in.

Acknowledgements The authors acknowledge support from Fundação para a Ciência e a Tecnologia, Portugal, through Project PTDC/MAT/101867/2008.

References

1. Badley, R.A., Drake, R.A.L., Shanks, I.A., Smith, A.M., Stephenson, P.R.: Optical biosensors for immunoassays, the fluorescence capillary-fill device. Philos. Trans. R. Soc. Lond. Ser. B **316**, 143–160 (1987)
2. Jones, S., Jumarhon, B., McKee, S., Scott, J.A.: A mathematical model of a biosensor. J. Eng. Math. **30**, 321–337 (1996)
3. Rebelo, M., Diogo, T., McKee, S.: A mathematical treatment of the fluorescence capillary-fill device. SIAM J. Appl. Math. **71**(4), 1081–1112 (2012)

Model-Based Medical Decision Support for Glucose Balance in ICU Patients: Optimization and Analysis

Thomas Martin Cibis and Nicole Marheineke

Abstract Model-based medical decision support in terms of computer simulations and predictions gains increasing importance in health care systems worldwide. This work deals with the control of the glucose balance in ICU patients using an insulin therapy. The basis of our investigations is the simulation model GlucoSafe by Pielmeier et al. that describes the temporal evolution of the blood glucose and insulin concentrations in the human body by help of a nonlinear dynamic system of first-order ordinary differential equations. We aim at the theoretical analysis and numerical treatment of the arising optimal control problem.

1 Introduction

Glucose is a vitally important source of energy for the human body. The skeletal musculature, brain, central nervous system, etc. must always be adequately supplied with glucose. Too high or too low blood sugar levels are harmful and can even cause death. A healthy body regulates the blood sugar levels by itself, thereby the peptide hormone insulin plays a crucial role. It becomes problematic (dangerous for life) when the body has a resistance to insulin or an insulin deficiency, as it is for example the case in diabetic patients. ICU patients suffering from severe, sometimes life-threatening illnesses or injuries often show an impaired insulin sensitivity. Since (strongly) fluctuating blood sugar levels additionally hamper the healing process, these patients need to be strictly observed. Their metabolism of glucose is controlled from outside via the intake/medication of food and insulin yielding an increase or decrease, respectively. To guarantee an adequate control, many frequent blood glucose measurements and tests are manually performed in

T.M. Cibis (✉) • N. Marheineke
Department Mathematik, Friedrich-Alexander-Universität Nürnberg-Erlangen, Cauerstr. 11, 91058 Erlangen, Germany
e-mail: cibis@math.fau.de; marheineke@math.fau.de

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*, Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_32,

hospitals, which is obviously associated with large caring effort and hence high costs in terms of time and money. The number of people suffering from diseases of sugar increases steadily worldwide, and health care systems are already overloaded. Therefore, model-based medical decision support using computer simulations for (long-time) predictions and optimizations gains importance.

This work deals with the optimal control of the glucose balance. The basis is the bio-medical model GlucoSafe developed by Pielmeier et al. [1] in 2010. We perform a theoretical (mathematical) analysis of the model and propose an adequate and efficient numerical treatment.

2 Optimal Control Problem

The temporal evolutions of the glucose and insulin concentrations in the body of a patient are determined by a complex interaction where apart from the intake/medication of food and insulin also the activities of liver, kidneys, gut, muscles, central nervous system, brain, etc. play a role. From the bio-medical point of view there are several dependencies and effects that are not fully understood so far, e.g. the impact of the insulin saturation. Moreover, measurements are restricted. However, under simplifying assumptions and closure relations, Pielmeier et al. [1] developed a "grey" model (1) in form of a deterministic nonlinear dynamic system of first-order ordinary differential equations that contains patient-dependent as well as fixed parameters and functions, for details on the bio-medical background see [1–3] and on the mathematical formulation, exact definitions [4]. A graphical illustration of the underlying biological processes that are taken into account is given in Fig. 1.

$$\begin{aligned}
\frac{\mathrm{d}G}{\mathrm{d}t} &= \frac{w}{v_G}\left[E((P,G),i_\sigma) + d(D) + \pi(t)\right] \\
\frac{\mathrm{d}D}{\mathrm{d}t} &= -d(D) + \varepsilon(t) \\
\frac{\mathrm{d}I}{\mathrm{d}t} &= \frac{c}{v_I}(P - I) - (r_L + r_K)\,I + \frac{n + \xi(t)}{v_I} \\
\frac{\mathrm{d}P}{\mathrm{d}t} &= \frac{c}{v_P}(I - P) - r_E P
\end{aligned} \tag{1}$$

with $t \in \mathscr{T}$ compact time period and

$$E((P,G),i_\sigma) = h((P,G),i_\sigma) - a_R(G) - a_M((P,G),i_\sigma) - a_N(G).$$

The state variables are the glucose concentrations in the blood plasma G (to be controlled) and in the gut content D as well as the insulin concentrations in the blood plasma I and around the cells (the so-called peripheral compartment)

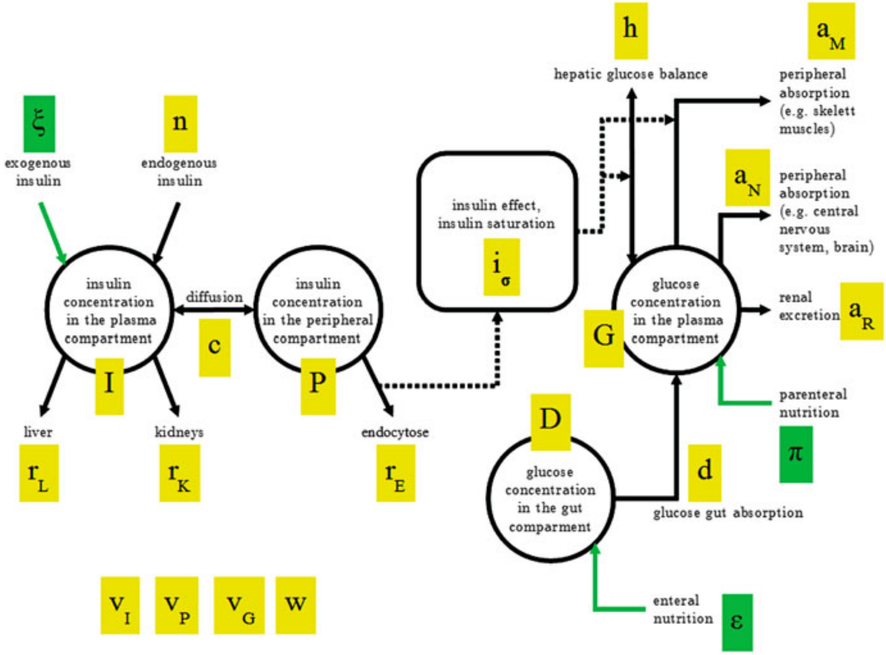

h
hepatic glucose balance
a_M
peripheral absorption (e.g. skelett muscles)
ξ
exogenous insulin
n
endogenous insulin
insulin effect, insulin saturation
i_σ
a_N
peripheral absorption (e.g. central nervous system, brain)
insulin concentration in the plasma compartment
diffusion
c
insulin concentration in the peripheral compartment
glucose concentration in the plasma compartment
G
renal excretion
a_R
I
P
parenteral nutrition
π
liver
kidneys
endocytose
D
r_L
r_K
r_E
glucose concentration in the gut comparment
d
glucose gut absorption
v_I
v_P
v_G
w
enteral nutrition
ε

Fig. 1 Illustration of the bio-medical model, in the style of Pielmeier et al.

P. Between the last two a difference-based diffusion process takes place. The controls are the intakes/medication of parenteral π and enteral nutrition ε as well as of exogenous insulin ξ, which we summarize as $\boldsymbol{u} = (\pi, \varepsilon, \xi)^\top$. The glucose balance of the liver h as well as the glucose absorption from the gut content d, of the skeletal musculature a_M, brain and central nervous system a_N are modeled as patient-independent functions, in contrast to the renal glucose excretion a_R that depend on the patient data (body weight w, size, age, gender and diabetic status). Further patient-dependent, but temporal constant parameters are the endogenous (post-hepatic) supply of insulin n, the rate of the insulin reduction in liver r_L, kidneys r_K and in the process of endocytosis r_E, the insulin diffusion constant c as well as the volumes of glucose blood plasma v_G, insulin blood plasma v_I and peripheral compartment v_P. The impact of the insulin enters (1) by the quantity $i_{\boldsymbol{\sigma}}$. Since the understanding of this biological process—involving impact sensitivity and saturation effect—is still rather limited, $i_{\boldsymbol{\sigma}}$ is expressed in terms of a non-negative parameter tuple $\boldsymbol{\sigma} \in \mathbb{R}^2$. This tuple is frequently adapted for each patient using a least-square parameter fit where the deviation of simulation results (1) to earlier measurements is minimized. Note that only the blood sugar G can be measured, but not D, I and P. This makes the initialization of (1) at a certain time t_0 inexact: in combination to a blood sugar measurement $G(t_0) = G_{meas}$, reference values for D, I, P at t_0 are taken from literature. The initial perturbation decreases over time due to the asymptotic stability of the model (see below). Thus, during the course of a treatment previous simulation results can be used as better initial guesses.

Considering the optimal control of G we solve the following constrained minimization problem,

$$\min_{(G,\boldsymbol{u})\in\mathfrak{Z}\times\mathfrak{S}} J(G,\boldsymbol{u}) \tag{2}$$

subject to

- G and $\boldsymbol{u}$ satisfy the dynamical system (1) with given initial values (not closer specified here)
- $\boldsymbol{u} \in U \subseteq \mathfrak{S}$, the set of admissible controls

As cost function we choose thereby

$$J(G,\boldsymbol{u}) = \frac{1}{2}\|G - G^*\|^2 + \frac{1}{2}\|\mathrm{diag}(\boldsymbol{\kappa})(\boldsymbol{u} - \boldsymbol{u}^*)\|^2,$$

where G^* is the target blood sugar level and $\boldsymbol{u}^*$ is the desired control based on bio-medical and economic reasons with weights $\boldsymbol{\kappa} \in (\mathbb{R}_0^+)^3$.

3 Analysis and Numerical Treatment

In this section we present a theoretical analysis of the model and propose an adequate numerical treatment.

The initial value problem (1) for the glucose balance is well-posed. For continuous controls it is resolvable in the classical sense. Existence and uniqueness holds according to the Picard-Lindelöf theorem for the sufficiently smooth model functions on the right-hand side of (1). However, bio-medical reasons require also non-continuous controls. Considering $\boldsymbol{u} \in L^\infty(\mathscr{T}, \mathbb{R}^3)$ with $\boldsymbol{u} \geq \mathbf{0}$, the system (1) has got a unique non-negative solution in the sense of Carathéodory for every choice of non-negative initial values. This stands in accordance with biological demands. In the following, we consider the space of piecewise constant non-negative bounded by a certain upper bound functions as set of admissible controls U. This is reasonable and sufficient for the application. The structure of the dynamical system allows the decoupling into a linear system for I and P, a Riccati equation for D and a nonlinear equation for G. The differential equations for I, P and D can be solved explicitly for the chosen U, for closed solution formulas see [4]. Moreover, for each steady state $\overline{G} > 0$ there exist a constant control $\overline{\boldsymbol{u}} \geq \mathbf{0}$ and steady states $\overline{I}, \overline{P}, \overline{D} \geq 0$, so that all together satisfy (1). In addition, it can be shown that this solution is asymptotically stable in all medically relevant cases for all possible patient data. So, the controllability of arbitrary stationary states is possible with $\mathfrak{S}$. The existence of optimal controls in the space $L^2(\mathscr{T}, (\mathbb{R}_0^+)^3)$ can be proven straight forward for (2), following the ideas and procedure prescribed in [5]. Uniqueness is lacked due to the non-linearity of (1); however, this is not necessary from a user point of view.

The ordinary differential system is not stiff such that the numerical computation of solutions can be performed by standard explicit Runge-Kutta methods with adaptive step size control. In particular, we use the method by Dormand/Prince [6]. The computational effort can be thereby reduced by a factor of two when the explicit solution formulas for D, I, P are used for the calculation for G. The optimal control problem (2) can be approached by direct and indirect methods, [5]. Thereby, it is advantageous to consider the associated equivalent reduced problem $\min_{\boldsymbol{u}} \tilde{J}(\boldsymbol{u})$, $\tilde{J}(\boldsymbol{u}) = J(G(\boldsymbol{u}), \boldsymbol{u})$. We have compared various direct and indirect methods. For the direct methods, a finite-dimensional optimization problem must be ultimately solved. The SQP method with numerically calculated gradients [7] turned out to be the best one with respect to the run time. As indirect methods, we tested conditional gradient method, gradient projection method and Newton-type methods. Regarding accuracy and computational efficiency, these methods cannot compete with the direct ones for this special problem (2); they are slower by a factor of about 50. The following simulation results are computed by MATLAB, version 7.7. Therefore the SQP method is implemented by MATLAB function `fmincon` with termination criteria: 10,000 function evaluations, 500 iterations, tolerance for variable/cost function of 10^{-12}.

4 Results and Discussion

The simulation results show that the model GlucoSafe leads to meaningful and interpretable results as long as we treat patients with a stabilized (non-fluctuating) blood sugar level. Figure 2 illustrates exemplarily the numerical results for G (red curve) in comparison to measured values (black crosses) for an arbitrary patient. From the bio-medical point of view the agreement is very satisfying since the measured values lie much closer than the acceptable area of 20 % deviation (green zone) would demand. In particular, the results for shorter time periods ($\leq$3.0 h) are much better than for long time periods. The reason for the worsening lies in the insulin effect i_σ which is actually a time- and patient-dependent function but modeled here by a simple parameter fit via $\boldsymbol{\sigma}$. Crucial for a reliable prediction of the long-time behavior is here a frequent adjustment of the parameter tuple $\boldsymbol{\sigma}$ to measurements. Figure 3 shows the temporal development of G for a forecast period of 3 h. Thereby, the desired blood sugar level is taken to be constant $G^* = 6.0$ mmol/l, [8], and the desired control $\boldsymbol{u}^* = \boldsymbol{0}$ with weighting factors $\boldsymbol{\kappa} = (\sqrt{10}, \sqrt{10}, 10^{-3})^\top$ in the cost functional J. As admissible controls, we have selected the set of non-negative functions $\boldsymbol{u}$ that are piecewise constant on the equidistant time grid with step size $\tau_\Delta = 1.0$ h and bounded from above by $(0.041 \text{ mmol/kg}, 0.026 \text{ mmol/kg}, 0.334 \text{ U})^\top/\text{min}$. A statistical validation of our predictions is not possible so far due to our relatively small sample size of

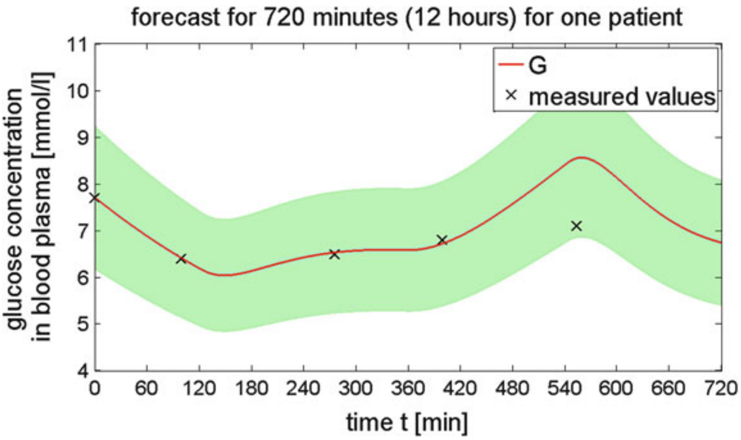


Fig. 2 Comparison of simulated blood glucose G with measured values for 12 h

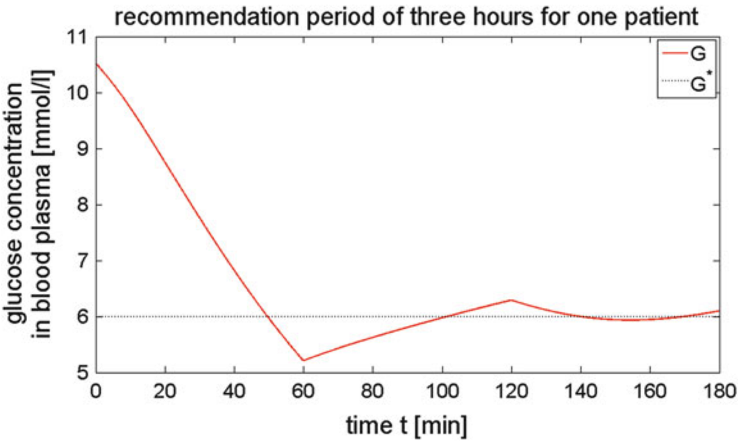


Fig. 3 Predication of the state G for a computed optimal control; desired state $G^* = 6.0$ mmol/l

data/measurements at hand. For a large clinical study the described methods have been implemented in a software tool by Ulrike Pielmeier and the glucose research group at the Center of Model-based Medical Decision Support, University Aalborg. This is recently applied and tested in hospitals.

5 Conclusion and Outlook

This work presented a theoretical analysis and numerical investigation of the biomedical model GlucoSafe used for the optimal control of the blood glucose via intake/medication of food and insulin. The simulation results are promising for ICU patients with non-fluctuating glucose levels. Improvements of the model lie surely in the concrete definition of the insulin function i_σ. But also the liver balance h and the endogenous insulin intake n that is assumed to be constant so far pose open research questions to bio-medical experts. A glucose-dependent n would imply a fully coupled dynamical system for all state variables. A interesting challenge from the mathematical point of view is the incorporation of uncertainties coming from the patient data and the measurements. This results in a stochastic control problem for which sensitivity/robustness and controllability have to be investigated.

Acknowledgements The authors thank Ulrike Pielmeier for her explanations of the bio-medical process and GlucoSafe. Moreover, the research team by Steen Andreassen, Center for Model-based Medical Decision Support, Aalborg University, is acknowledged for providing a sample of patient and treatment data.

References

1. Pielmeier, U., Andreassen, S., Nielsen, B.S., Chase, J.G., Haure, P.: A simulation model of insulin saturation and glucose balance for glycaemic control in ICU patients. Comput. Methods Programs Biomed. **97**, 211–222 (2010)
2. Arleth, T., Andreassen, S., Federici, M.O., Benedetti, M.M.: A model of the endogenous glucose balance incorporating the characteristics of glucose transporters. Comput. Methods Programs Biomed. **62**, 219–234 (2000)
3. van Cauter, E., Mestrez, F., Sturis, J., Polonsky, K.S.: Estimation of insulin secretion rates from c-peptide levels: comparison of individual and standard kinetic parameters for c-peptide clearance. Diabetes **41**, 368–377 (1992)
4. Cibis, T.M.: Optimale Steuerung des Glukosehaushalts bei intensiv-gepflegten Patienten. Master's thesis, Johannes Gutenberg-Universität Mainz (2010)
5. Tröltzsch, F.: Optimale Steuerung partieller Differentialgleichungen: Theorie, Verfahren und Anwendungen. Vieweg, Wiesbaden (2005)
6. Hanke-Bourgeois, M.: Grundlagen der Numerischen Mathematik und des Wissenschaftlichen Rechnens. Vieweg, Wiesbaden (2006)
7. Alt, W.: Nichtlineare Optimierung: Eine Einführung in Theorie, Verfahren und Anwendungen. Vieweg, Wiesbaden (2002)
8. Emminger, H.: Physikum EXAKT: Das gesamte Prüfungswissen für die 1. P. Thieme, Stuttgart (2003)

Epileptic Seizures Diagnose Using Kunchenko's Polynomials Template Matching

Oleg Chertov and Taras Slipets

Abstract The paper related to epilepsy's diagnosis as EEG analysis problem. Template matching method based on a Kunchenko's polynomials for EEG processing introduced. To demonstrate efficiency of method numeric experiment is given.

1 Theoretic Background

Epilepsy ranks the third place on the prevalence of neurological disease and occurs with a frequency of 0.5–1.5 %. It is a chronic neurological disorder that affects people of all ages, with 2–4 % lifetime illness risk. As the main method of epilepsy, diagnostics in clinical practice electroencephalography is used. In recent time its role in the epilepsy's diagnosis is becoming more important because diagnosis involves usage of a fairly long time electroencephalograms (EEG), in which patient's brain activity signs are fixed. There are several EEG phenomena types that suggest the epileptic activity, but most important is a complex type "sharp wave–slow wave".

With introduction of computer electroencephalography in clinical practice, new problems whose solution requires new methods of EEG investigation and analysis have appeared. Manual data analysis requires from doctor very durable, careful and painstaking work, which involves routine operations performing. Thus, epileptic activity search automation task as finding the "sharp wave–slow wave" complexes is very important.

Input data recognition and verification in medical diagnostics are one of the most actual engineering concepts in our days. One of such technique is template matching approach. The key idea is filtering special information (features) from input data set.

O. Chertov • T. Slipets (✉)
National Technical University of Ukraine "Kyiv Polytechnic Institute", Kyiv, Ukraine
e-mail: chertov@i.ua; taras.slipets@gmail.com

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_33,

There are several different approaches in template matching technique which can be divided into following groups: SIFT-methods, Oblique projection, SAD, Cross-correlation, and Approximation over Redundant Dictionaries (AORD).

Method based on Kunchenko's polynomials (KP) approximation in a space with generative function [2] can be classified as AORD method. The cornerstone stage for AORD methods is functions' dictionary $\{\phi_n < . >\}$ constructing. KP approach uses so-called generative transforms which are nonlinear one-dimensional functions. For instance, natural power transforms:

$$\phi_n < . >^n, n \in \mathbb{N}_0 \tag{1}$$

can be considered as generative transforms system. Applying (1) to some function called cardinal function or generative element we build generated functions system $\{\phi_n\}$. After that approximation polynomial for input signal is constructed. KP application for template matching to one-dimensional digital signal includes two basic steps [1]:

1. Approximation of input signal with modified Kunchenko's polynomials based on a template to be found;
2. Approximation efficiency estimation for different parts of input signal.

Modified KP (MKP) P^r_{mod} is used to approximate input signal. This approach differs from classic approximation via KP [2] in including generative element (in our case, a complex type "sharp wave–slow wave" as template is used) to generated functions (elements) system $\{\phi_\nu\}$. This inclusion gives possibility to deal with linear dependency between current part of input signal and template to be found [1]:

$$P^r_{mod} < e >= \sum_{\nu=0}^{r} \alpha_\nu \phi_\nu < e > \tag{2}$$

where r—polynomial's degree.

Coefficients $\alpha_v, v \neq 0$ can be found as solution of following linear equations system:

$$\sum_{k=0}^{r} \alpha_k F_{v,k} = F_{v,b}, v = \overline{1, r}, \tag{3}$$

where $F_{v,k}$—so-called centered correlants:

$$F_{v,k} \equiv \Psi_{v,k} - \frac{\Psi_{v,0} \cdot \Psi_{k,0}}{\|\varphi_0 [f(x)]\|^2} \tag{4}$$

$\Psi_{v,k}$ are called simple correlants:

$$\Psi_{v,k} \equiv \int_{d}^{a} \varphi_v \left[f(x) \right] \cdot \varphi_k \left[f(x) \right] dx \tag{5}$$

$f(x)$—generative function, ϕ_v—generative transforms system [2].

Coefficient must be equal to next expression:

$$\alpha_0 = \Psi_{b,0} - \sum_{v=1}^{r} \alpha_v \Psi_{v,0}. \tag{6}$$

Next cornerstone goal is efficiency coefficients calculation for different parts of input. For that purpose well-known "window-signal" technique using—the subset of points is selected from signal on each method's iteration. The support of selected signal's subset is equal to support of template's set of points. After that, modified KP (2) is constructed for selected part of input signal and polynomial's approximation to given template efficiency coefficient is calculated:

$$e_r = \pm \frac{\sum_{v=1}^{r} \alpha_0 F_{v,b}}{\left\| \varphi_b^c < e > \right\|^2}, \tag{7}$$

where $\varphi_b^c < e >$—centered main element (function). In case of all elements of free terms vector $\mathbf{F}_b$ from (3) are negative, we assume that considered signal's part contains inverted or distorted template and efficiency coefficient has minus sign.

2 Numeric Experiment

To demonstrate efficiency of introduced method real-data experiment was performed. Input data: patient's EEG with symptomatic epilepsy's focus in the anterior temporal lobe. Signal duration—630 s, resolution—256 Hz. Template used for matching complex "sharp wave–slow wave" is depicted on Fig. 1. It consist of 82 knots. For the Kunchenko's approximation polynomials effectogram's analysis, as measure of approximation estimation a threshold equal to 0.9 is taken. This value has been estimated experimentally according to the best correct-wrong detections ratio. Seizures search results are presented in Table 1.

Doctor identified 245 complexes in given signal. Using MKP approach 207 complexes was found, 28 wrong detections and 38 complexes were missed. Considered experiment illustrates that template matching algorithm based on Kunchenko's polynomials allows to diagnose (with several con-strains) epileptic seizures.

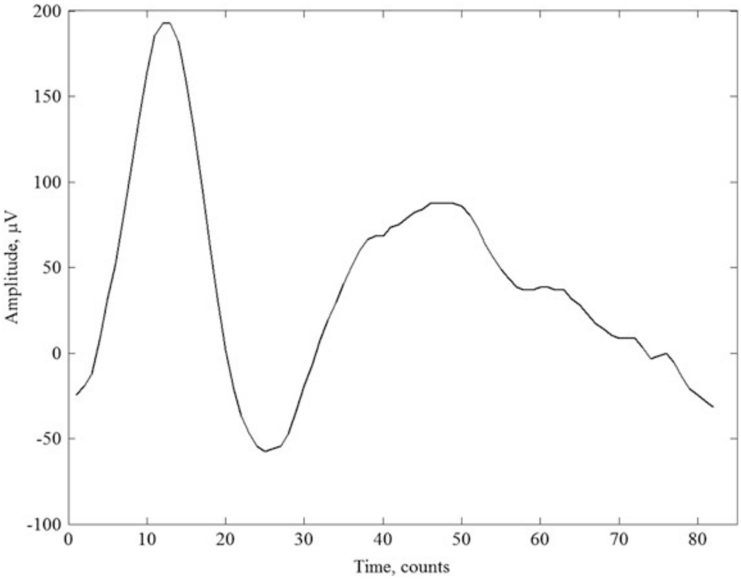


Fig. 1 Template for complex “sharp wave–slow wave”

Table 1 Experimental diagnosis results

Complex’s type	Found	Doctor’s diagnosis
Non-distorted	81 (93.1 %)	87
Distorted	107 (86.3 %)	124
Low-amplitude	10 (76.9 %)	13
Distorted low-amplitude	9 (42.8 %)	21
Total complex score	207 (84.4 %)	245

References

1. Chertov, O., Slipets, T.: Kunchenko polynomials for template matching. In: Zovko-Cihlar, B., Behlilovic, N., Hadzialic, M. (eds.) Proceedings of the 18th IEEE International Conference on Systems, Signals and Image Processing (IWSSIP), Sarajevo, 16–18 June 2011, p. 479. National Technical University of Ukraine “Kyiv Polytechnic Institute”, Ukraine (2011)
2. Kunchenko, Y.: Polynomial parameter estimation of close to Gaussian random variables. Shaker, Aachen (2002)

Part VI
Robotics and Automotive Industry

Overview

This section on *Robotics and Automotive Industry* contains six contributions addressing motion planning for robots, durability and sensitivity studies for automotive devices, systems and traffic.

The first two papers deal with motion planning of mechanical systems as they arise in robotics and other application areas, e.g. automotive systems, space mission design. In *Collision-Free Path Planning of Welding Robots* Chantal Landry et al. focus on collision-free path planning. Describing the robot dynamics by a system of ordinary differential equation and the objective function as the time to reach the final position, they set up an optimal control problem. Thereby, the collision avoidance criterion being a consequence of Farkas's lemma is incorporated as state constraint. The resulting model is solved by a sequential quadratic programming method where an active set strategy based on backface culling is added. Control problems for hybrid (i.e. mixed discrete and continuous) dynamical models are the topic of the paper *Motion Planning for Mechanical Systems with Hybrid Dynamics* by Kathrin Flaßkamp and Sina Ober-Blöbaum. Here, the motion planning is performed with respect to motion primitives that are collected in a library. A solution to a specific optimal control problem is then obtained by searching for the optimal sequence of concatenated primitives. The framework is extended to motions along invariant manifolds of the uncontrolled dynamics, e.g. trajectories on (un)stable manifolds of equilibria, and applied to an open-chain jointed robot.

Durability, reliability and sensitivity play an important role in the manufacturing of automotive devices and systems. In *Performance of Sensitivity Based NMPC Updates in Automotive Applications* Jürgen Pannek and Matthias Gerdts deal with the control of a half-car model under disturbances. They impose model predictive control without stabilizing terminal constraints or cost to generate a nominal solution and sensitivity updates to handle the disturbances. Stability of the resulting closed loop is guaranteed by a relaxed Lyapunov argument on the nominal system and Lipschitz conditions on the open loop change of the optimal value function and the stage costs. The proposed approach is real-time applicable and yields promising results. To optimize the chassis dynamics Johannes Michael and Matthias Gerdts take a closer look on modeling contact conditions as the contact force is directly related to handling characteristics of the automobile. In *Optimal Control in Proactive Chassis Dynamics: A Fixed Step Size Time-Stepping Scheme for Complementarity Problems* they propose a numerical scheme and compare the calculation of a quarter-car with a spring-damper road to wheel interaction to those resulting from the complementarity problem. The durability analysis and optimal design of tires and various mounts requires the efficient simulation of such contact problems. Since the complexity of high dimensional finite element models exceeds the applicability, Joachim Krencisznek and René Pinnau investigate model order reduction techniques in *Model Reduction of Contact Problems in Elasticity: Proper Orthogonal Decomposition for Variational Inequalities*. Considering a Signorini contact problem in variational inequality formulation they apply Proper Orthogonal Decomposition to compute an optimal projection subspace and discuss the reduced model's quality and efficiency an Encastre beam with contact.

The topic of the last paper is *Novel Updating Mechanisms for Stochastic Lattice-Free Traffic Dynamics*. Proposing a lattice-free model, Alexandros Sopasakis describes vehicle traffic on multi-lanes based on stochastic spin-flip and spin-exchange Arrhenius dynamic potentials. The solution is computed in real-time (even for large traffic streams) by a kinetic Monte-Carlo algorithm and compared with the ones of lattice-based (cellular automata) approaches.

Nicole Marheineke

Collision-Free Path Planning of Welding Robots

Chantal Landry, Matthias Gerdts, René Henrion, Dietmar Hömberg, and Wolfgang Welz

Abstract In a competitive industry, production lines must be efficient. In practice, this means an optimal task assignment between the robots and an optimal motion of the robots between their tasks. To be optimal, this motion must be collision-free and as fast as possible. It is obtained by solving an optimal control problem where the objective function is the time to reach the final position and the ordinary differential equations are the dynamics of the robot. The collision avoidance criterion is a consequence of Farkas's lemma. The criterion is included in the optimal control problem as state constraints and allows us to initialize most of the control variables efficiently. The resulting model is solved by a sequential quadratic programming method where an active set strategy based on backface culling is added.

1 Background

To be competitive, a car manufacturing must have efficient production lines. These lines are composed of robots and other machines grouped together in work cells. In each work cell a certain number of robots perform tasks on the same workpiece. An efficient production line is obtained when the total time taken by the robots

C. Landry (✉) • R. Henrion • D. Hömberg
Weierstrass Institute, Mohrenstr. 39, 10117 Berlin, Germany
e-mail: chantal.landry@wias-berlin.de; rene.henrion@wias-berlin.de; dietmar.hoemberg@wias-berlin.de

M. Gerdts
University of the Federal Armed Forces at Munich, Werner-Heisenberg-Weg 39, 85577 Neubiberg, Germany
e-mail: matthias.gerdts@unibw.de

W. Welz
Technical University of Berlin, Strasse des 17. Juni 136, 10623 Berlin, Germany
e-mail: wolfgang.welz@math.tu-berlin.de

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_34,

to complete all the tasks is as small as possible. This time is minimal when the following three points are optimized: (1) task assignment between the robots of the same work cell, (2) sequencing of the tasks of each robot, (3) path-planning of each robot avoiding collisions with obstacles. If we want to optimize the production, we cannot treat these three problems separately. The task assignment and the sequencing depend on the computation of the fastest collision-free trajectory of the robots between two tasks. However, because the computation of these trajectories between all pairs of task spots is expensive, estimated distances are used at first. The optimal path-planning of the robot is computed only when needed. Our algorithm to optimize production line is sketched as follows

Algorithm 1. 1. Find promising estimated tours for all robots
2. Calculate exact distance for these tours
3. Identify collisions for exact trajectories
4. Reoptimize: if collisions occur, find new tours and go to 2

Algorithm 1 is fully detailed in [8]. This paper is dedicated to the computation of the second step of Algorithm 1, whereas the first step is presented in [10] and the third step in [8].

2 Model

Let us consider a robot composed of p links which are connected by revolute joints. Let $q = (q_1, \dots, q_p)$ denote the vector of joint angles at the joints of the robot. Moreover, let the vector $v = (v_1, \dots, v_m)$ contain the joint angle velocities and $u = (u_1, \dots, u_m)$ describe the torques applied at the center of gravity of each link. The robot is asked to move as fast as possible from a given position to a desire location. Its motion is given in the Lagrangian form as follows

$$q'(t) = v(t) \;\text{ and }\; M(q(t))\, v'(t) = G(q(t), v(t)) + F(q(t), u(t)), \tag{1}$$

where $M(q)$ is the symmetric and positive definite mass matrix, $G(q, v)$ contains the generalized Coriolis forces and $F(q, u)$ is the vector of applied joint torques and gravity forces [1]. The function F is linear in u.

The motion of the robot must follow (1), but also be collision-free with the obstacles present in the workspace. For simplicity, let us assume that only one obstacle is present. To establish the collision avoidance condition, the robot and the obstacle are approximated by a union of convex polyhedra, see [4–6]. The approximation is denoted by P for the robot, by Q for the obstacle and are given by

$$\begin{aligned} P &= \cup_{i=1}^{p} P_i, &\quad \text{with } P_i &= \{x \in \mathbb{R}^3 \mid A^{(i)} x \le b^{(i)}\},\\ Q &= \cup_{j=1}^{q} Q_j, &\quad \text{with } Q_j &= \{x \in \mathbb{R}^3 \mid C^{(j)} x \le d^{(j)}\}, \end{aligned}$$

where $A^{(i)} \in \mathbb{R}^{p_i \times 3}$, $b^{(i)} \in \mathbb{R}^{p_i}$, $C^{(j)} \in \mathbb{R}^{q_j \times 3}$, $d^{(j)} \in \mathbb{R}^{q_j}$, and p_i and q_j are the number of faces in P_i and Q_j, respectively.

The robot P and the obstacle Q do not collide if and only if for each pair of polyhedra (P_i, Q_j), $i = 1, \dots, p$, $j = 1, \dots, q$, there exists a vector $w^{(i,j)} \in \mathbb{R}^{p_i + q_j}$ such that:

$$w^{(i,j)} \geq 0, \quad \begin{pmatrix} A^{(i)} \\ C^{(j)} \end{pmatrix}^T w^{(i,j)} = 0 \quad \text{and} \quad \begin{pmatrix} b^{(i)} \\ d^{(j)} \end{pmatrix}^T w^{(i,j)} < 0. \tag{2}$$

This is a direct consequence of Farkas's lemma. See [4] for more details.

The fastest trajectory of a robot is the solution of an optimal control problem, where the system of ordinary differential equations (ODE) are given by (1), see [1]. If an obstacle is present in the workspace, the collision avoidance is assured as soon as the vector $w^{(i,j)}$ of (2) is found at each time t and for all pairs of polyhedra. However, to be written as state constraints, the strict inequality in (2) has to be relaxed. Furthermore, since the robot moves, the matrices $A^{(i)}$ and the vectors $b^{(i)}$ evolve in time. Their evolution depends explicitly on $q(t)$. A complete formulation of $A^{(i)}(q(t))$ and $b^{(i)}(q(t))$ is given in [4]. Finally, the optimal control problem to find the fastest collision-free trajectory is given by:

Model 1. Find the final time t_f, the state variables $q, v : [0, t_f] \to \mathbb{R}^p$, and the controls $u : [0, t_f] \to \mathbb{R}^p$ and $w^{(i,j)} : [0, t_f] \to \mathbb{R}^{p_i + q_j}$, $i = 1, \dots, p$, $j = 1, \dots, q$ such that t_f is minimized subject to

1. the ordinary differential equations

$$q'(t) = v(t) \;\text{ and }\; v'(t) = M(q(t))^{-1} \left(G(q(t), v(t)) + F(q(t), u(t)) \right);$$

2. the state constraints

$$\begin{pmatrix} A^{(i)}(q(t)) \\ C^{(j)} \end{pmatrix}^T w^{(i,j)}(t) = 0, \qquad i = 1, \dots, p,\; j = 1, \dots, q; \tag{3}$$

$$\begin{pmatrix} b^{(i)}(q(t)) \\ d^{(j)} \end{pmatrix}^T w^{(i,j)}(t) \leq -\varepsilon, \qquad i = 1, \dots, p,\; j = 1, \dots, q; \tag{4}$$

3. the boundary conditions

$$R(q(0)) - R_0 = 0, \quad v(0) = 0, \quad R(q(t_f)) - R_f = 0 \quad \text{and} \quad v(t_f) = 0;$$

4. the box constraints $u_{\min} \leq u \leq u_{\max}$ and $0 \leq w^{(i,j)}$, $i = 1, \dots, p$, $j = 1, \dots, q$,

where $R(q)$ denotes the position of the barycenter of the last link of the robot and $R_0, R_f \in \mathbb{R}^m$ are given by the first step of Algorithm 1. The vectors u_{min} and u_{max} are also given and the relaxation parameter ε is positive and small.

Model 1 can be easily applied with several obstacles. It suffices to define control variables and to write (3)–(4) for each obstacle. Depending on the number of state constraints (3)–(4), the problem is inherently sparse since the artificial control variables $w^{(i,j)}$ do not enter the dynamics, the boundary conditions and the objective function of the problem, but only appear linearly in (3)–(4).

3 Numerical Method and Examples

The optimal control problem described in Model 1 is solved by using the software OC-ODE [3]. The method involves first discretizing the control problem and transforming it into a finite-dimensional nonlinear optimization problem. The control variables are approximated by B-splines of order 2 and the ordinary differential equations are integrated with the classical Runge-Kutta method of order 4. The resulting nonlinear optimization problem is then solved by a sequential quadratic programming method [2, 7]. As in [9] we use an Armijo type line-search procedure for the augmented Lagrangian function in our implementation. However, the resulting optimization problem contains a lot of constraints: at each time step of the control grid and for every pair of polyhedra $(P^{(i)}, Q^{(j)})$, four state constraints are defined [compare (3)–(4)]. To overcome this difficulty, we add an active set strategy based on the following observation: the state constraints are superfluous when the robot is far from the obstacle or moves in the opposite direction. The establishment of the active set strategy is fully detailed in [4].

A good initialization of the control variables u and $w^{(i,j)}$, $i = 1, \ldots, p$, $j = 1, \ldots, q$ can highly improve the convergence of the sequential quadratic programming method. In the first step of Algorithm 1, an estimated tour is computed. This tour is found by considering a grid on the workspace and applying a Dijkstra-like algorithm to find the shortest path which connects the starting position of the robot, R_0, to the final location, R_f. The shortest path is chosen such that the angles between two successive edges are minimized. If the shortest path is close to the straight line connecting R_0 to R_f, then u is initialized by solving the above optimal control problem without considering the obstacles, this new problem being far smaller and easier to solve than Model 1. If not, then the path $[R_0, R_f]$ is split into subpaths of the form $[R_0, R_I], [R_I, R_J], [R_J, R_f]$ where R_I and R_J are vertices on the grid where the change in the angle of the shortest path is high. The initial guess of u is then given by solving the optimal control problem without the state constraints on every subpath.

Once the initial guess of u is established at every time step t_k of the control grid $\{t_1, \ldots, t_N\}$, an estimate of $q(t_k)$, $k = 1, \ldots, N$ can be computed by solving the ordinary differential equations (1). This estimate allows us to initialize the remaining control variables $w^{(i,j)}$,$i = 1, \ldots, p$, $j = 1, \ldots, q$ by exploiting the collision avoidance condition (2). Indeed, the initial guess of $w^{(i,j)}$ at t_k is chosen as the solution of the following minimization problem:

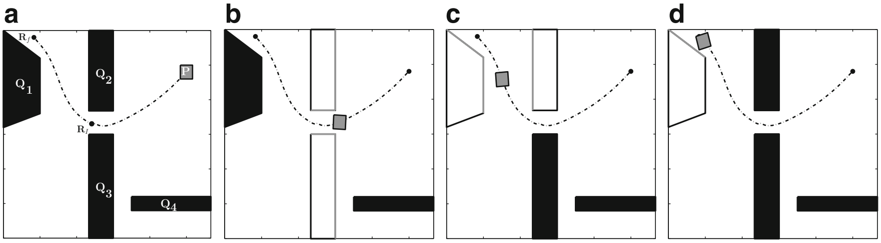


Fig. 1 Snapshots of the motion of the robot P moving to R_f and avoiding four obstacles. The visible obstacles are in *white*. (**a**) At t_1. (**b**) At t_{14}. (**c**) At t_{33}. (**d**) At t_{39}

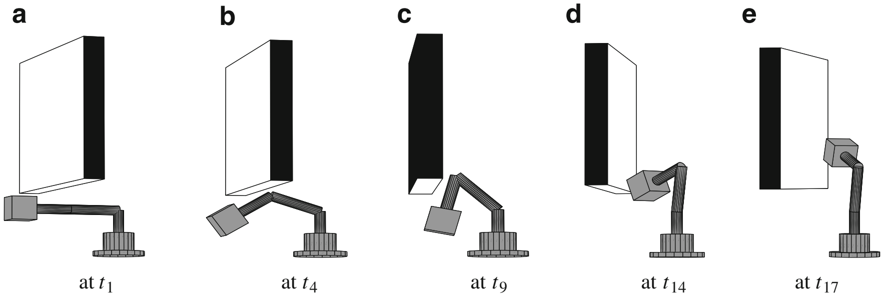


Fig. 2 Snapshots of the motion of the robot avoiding an obstacle. (**a**) At t_1. (**b**) At t_4. (**c**) At t_9. (**d**) At t_{14}. (**e**) At t_{17}

$$\min_{w} \begin{pmatrix} b^{(i)}(q_k) \\ d^{(j)} \end{pmatrix}^T w \text{ such that } \begin{pmatrix} A^{(i)}(q_k) \\ C^{(j)} \end{pmatrix}^T w = 0 \text{ and } w \geq 0,$$

where q_k is the approximation q at time t_k.

First, a two-dimensional numerical result is presented in Fig. 1. The robot is a square and four obstacles are present in the workspace. For the initialization of u, the middle point R_I was used. In Fig. 2 a robot composed by three links is moving around an obstacle without collision.

Acknowledgements This work has been supported by the DFG Research Center MATHEON Mathematics—for key technologies.

References

1. Diehl, M., Bock, H., Diedam, H., Wieber, P.: Fast Direct Multiple Shooting Algorithms for Optimal Robot Control, pp. 65–94. Springer, Heidelberg (2005)
2. Gerdts, M.: Optimal control of ordinary differential equations and differential-algebraic equations. Ph.D. thesis, Universität Bayreuth (2006)

3. Gerdts, M.: OC-ODE, Optimal Control of Ordinary-Differential Equations. Software User Manual (2010)
4. Gerdts, M., Henrion, R., Hömberg, D., Landry, C.: Path planning and collision avoidance for robots. Numer. Algebra Control Optim. **2**(3), 437–463 (2012)
5. Gilbert, E., Hong, S.: A new algorithm for detecting the collision of moving objects. In: Proceedings of IEEE International Conference on Robotics and Automation, vol.1, pp. 8–14 (1989)
6. Gilbert, E., Johnson, D.: Distance functions and their application to robot path planning in the presence of obstacle. IEEE J. Robot. Autom. **RA-1**, 21–30 (1985)
7. Gill, P., Murray, W., Saunders, M.: SNOPT: an SQP algorithm for large-scale constrained optimization. SIAM Rev. **47**, 99–131 (2005)
8. Landry, C., Henrion, R., Hömberg, D., Skutella, M., Welz, W.: Task assignment, sequencing and path-planning in robotic welding cells. In: 18th International Conference on Methods and Models in Automation and Robotics (MMAR), pp. 252–257 (2013)
9. Schittkowski, K.: On the convergence of a sequential quadratic programming method with an augmented Lagrangean line search function. Mathematische Operationsforschung und Statistik **14**, 197–216 (1983)
10. Skutella, M., Welz, W.: Route planning for robot systems. In: Hu, B., Morasch, K., Pickl, S., Siegle, M. (eds.) Operations Research Proceedings 2010, pp. 307–312. Springer, Berlin (2011)

Motion Planning for Mechanical Systems with Hybrid Dynamics

Kathrin Flaßkamp and Sina Ober-Blöbaum

Abstract Planning and optimal control of mechanical systems are challenging tasks in robotics as well as in many other application areas, e.g. in automotive systems or in space mission design. This holds in particular for hybrid, i.e. mixed discrete and continuous dynamical models. In this contribution, we present an approach to solve control problems for hybrid dynamical systems by motion planning with motion primitives. These canonical motions either origin from inherent symmetry properties of the systems or they are controlled maneuvers that allow sequencing of several primitives. The motion primitives are collected in a motion planning library. A solution to a specific optimal control problem can then be found by searching for the optimal sequence of concatenated primitives. Energy efficiency often forms an important objective in control applications. We therefore extend the motion planning framework by primitives that are motions along invariant manifolds of the uncontrolled dynamics, e.g. trajectories on (un)stable manifolds of equilibria. The approach is illustrated by an academic example motivated by an operating scenario of an open-chain jointed robot.

1 Introduction

Planning problems arise in many technical applications and typically one is interested in an optimal solution of the problem. Taking into account the dynamics of the technical system, e.g. an industrial robot, there has to be found a solution

K. Flaßkamp (✉)
Department of Mathematics, University of Paderborn, Warburger Str. 100, 33098 Paderborn, Germany
e-mail: kathrinf@math.uni-paderborn.de

S. Ober-Blöbaum
Institut für Wissenschaftliches Rechnen, Zellescher Weg 12–14, 01069 Dresden, Germany
e-mail: sina.ober-bloebaum@tu-dresden.de

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_35,

trajectory to the dynamic optimal control problem fulfilling in addition required start and final configurations (cf. e.g. [1]). Furthermore, the dynamics of a complex technical system has to be modeled by an interaction of continuous time and discrete event dynamics, thus by a hybrid system.

Optimal Control. An optimal control problem for a mechanical system with configuration manifold Q and states $x(t) = (q(t), \dot{q}(t)) \in TQ$ is defined by a cost functional $J(x,u) = \int_0^T C(x(t), u(t))\, dt$, that has to be minimized. Constraints are given by the system's dynamics, e.g. the *Euler-Lagrange equations*

$$\frac{\partial L}{\partial q}(q,\dot{q}) - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}}(q,\dot{q}) + f(q,\dot{q},u) = 0$$

with Lagrangian L and forces f depending on continuous time control inputs $u(t)$ and on $(q,\dot{q})$, by boundary conditions and typically further constraints on the states and controls. There exists a number of approaches for numerically solving optimal control problems (cf. e.g. [2] for an overview). For our computational example, we use DMOC (*Discrete Mechanics and Optimal Control*, [3]), a method that directly discretizes the problem such that a high dimensional constrained nonlinear optimization problem is obtained which can be solved e.g. by *sequential quadratic programming* (SQP, cf. e.g. [4]). Since these methods compute local optima only, it is necessary to provide good initial guesses for the optimal control method and it is beneficial to combine the method with global, e.g. planning techniques [5–7]. In [5], Frazzoli et al. present the approach for *motion planning with motion primitives* (cf. Sect. 2).

Hybrid Dynamics. The dynamic behavior of technical systems is typically modeled by systems of continuous time differential equations. However, for an appropriate description of complex behavior and interactions, discrete effects have to be additionally accounted for, leading to the general framework of *hybrid systems*. Considering mechanical systems, there is a number of origins for hybrid effects: a changing environment as well as varying internal parameters change the system's dynamics, obstacles lead to impacts or (de)coupling processes cause changes of the system's topology. Formally, a hybrid system can be defined by a finite family of continuous subsystems $\dot{x} = f_i(x,u)$, $i = 1,\dots,N$ (the vector fields origin from different Lagrangian L_i) defined on subsets $\mathscr{X}_i$ (*domains*) of a common state space and with the same control inputs. Switching between the subsystems is usually restricted by guards and reset maps (cf. e.g. [8]). Then, a hybrid trajectory consists of the continuous variables plus a discrete mode $d(t) \in \{1,\dots,N\}$ that defines which subsystem is active. The optimal control of hybrid systems is of great interest, since it includes an optimization of discrete and continuous variables leading to mixed-integer programming problems (cf. e.g. [9]).

The remainder of this paper is structured as follows: in Sect. 2, we introduce the different kinds of motion primitives and sketch the idea of motion planning with primitives. Extensions for an application to hybrid mechanical systems are presented in Sect. 3. Finally, in Sect. 4, the method is illustrated by an academic model of an open-chain jointed robot.

2 Motion Planning with Motion Primitives

The basic idea of motion planning with primitives (introduced in [5]) lies in exploiting the inherent symmetries of a system. Mechanical systems often inhabit *symmetries*, for example if they are *invariant* with respect to translations or rotations. Formally, this means that there exists a Lie group G with a left-action $\Phi_g^{TQ} : TQ \to TQ$, $g \in G$ on the state space which leaves the Lagrangian invariant, $L \circ \Phi_g^{TQ} = L$ for all $g \in G$. Then, we call two trajectories *equivalent*, if they are equal except for a symmetry transformation and a time shift. Symmetry helps to reduce the complexity of the motion planning library since it is sufficient to store one representative, a *motion primitive*, for all equivalent trajectories. Solving a motion planning problem corresponds to a search for the optimal sequence in this library represented by a maneuver automaton (cf. [5,6]).

Trim Primitives. A special kind of primitives is given by *trim primitives*, which are motions along the group orbits of G with a constant control value. Thus, the trajectories can be simply described by $(q,\dot{q})(t) = \Phi^{TQ}(\exp(\xi t), x_0)$, $u(t) \equiv u_0$ with ξ being an element of the Lie algebra corresponding to G, with the exponential map $\exp(\,\cdot\,)$ and some initial value x_0 (cf. [5,7] for details). In mechanical systems, trims are also known as relative equilibria and they are closely related to the conservation of momentum maps, the Noether theorem, and to symmetry reduction procedures (see [7]). For a spherical pendulum (Fig. 1), trims are horizontal rotations with constant velocity. In the lower half sphere, uncontrolled trims exist. A constant additive control can be chosen to create trims with arbitrary rotational velocities at any height.

Trajectories on (Un)stable Manifolds. The natural, i.e. uncontrolled dynamics of a mechanical system provide motions that can be of great interest when searching for energy efficient control maneuvers. In particular, trajectories on stable manifolds of hyperbolic unstable fixed points are promising candidates since a stable manifold contains all motions which tend to the corresponding equilibrium point (cf. e.g. [10]). The unstable manifold, in contrast, shows the direction of expansion from the equilibrium and is attractive. Formally, assuming $\bar{x} = (\bar{q}, 0)$ is an equilibrium of the system and $F_L(x,t)$ denotes the flow of the autonomous system defined by the Lagrangian L, the *local stable manifold* is given by

$$W^s_{\text{loc}}(\bar{x}) = \{x \in U \mid F_L(x,t) \to \bar{x} \text{ for } t \to \infty \text{ and } F_L(x,t) \in U \; \forall t \geq 0\}.$$

The global stable manifold W^s can be governed by the preimages of the flow on $W^s_{\text{loc}}(\bar{x})$. (For the unstable manifold W^u, the same holds in backward time ($t \leq 0$).)

Fig. 1 For a simple spherical pendulum, trim primitives are horizontal rotations, but (un)stable manifolds belongs to purely vertical motions. Thus, connecting maneuvers as motion primitives are required such that sequences of primitives can be found

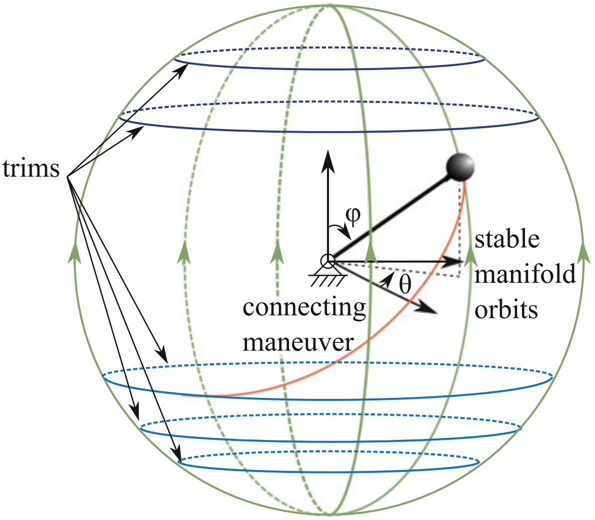
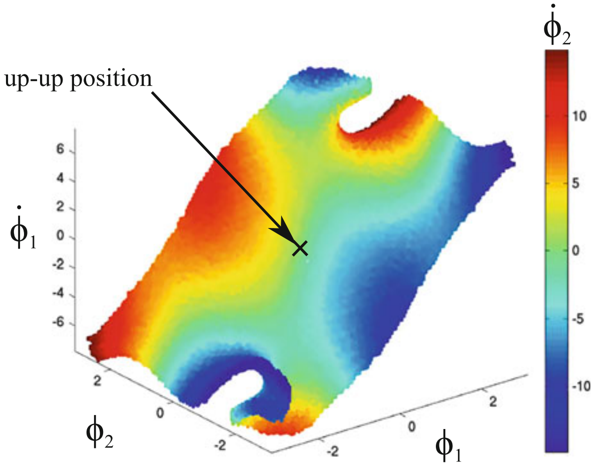


Fig. 2 Unstable manifold of the up-up equilibrium of a double pendulum (restricted to vertical motions). Motion primitives are generated by choosing trajectories with different time durations on the manifold



To compute such manifolds numerically, we use the method GAIO (*Global Analysis of Invariant Objects*, [11]), see Figs. 2 and 3 for single pendulum (cf. Fig. 1) and double pendulum (cf. Fig. 4) manifold examples. In [7], it is explained in detail how trajectories on manifolds can be chosen.

Connecting Maneuvers and Sequencing. Motion primitives of a third kind have to be computed to build up the motion planning library, namely short controlled maneuvers that connect trims with each other and trims to manifolds. This can be done for example by the optimal control method DMOC (see [3]). In Fig. 1, all three types of primitives for a simple spherical pendulum are sketched (we refer to [7] for a detailed description).

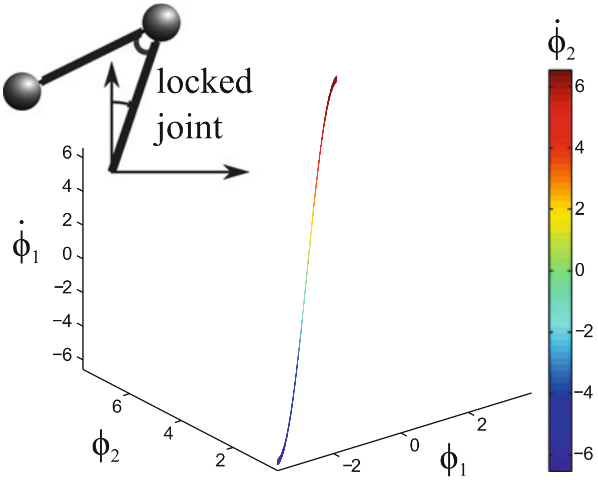


Fig. 3 The locked double pendulum (restricted to vertical motions) is a one degree of freedom system with a one-dimensional unstable manifold. For the numerical computations, the locking angle is set to 0.25π

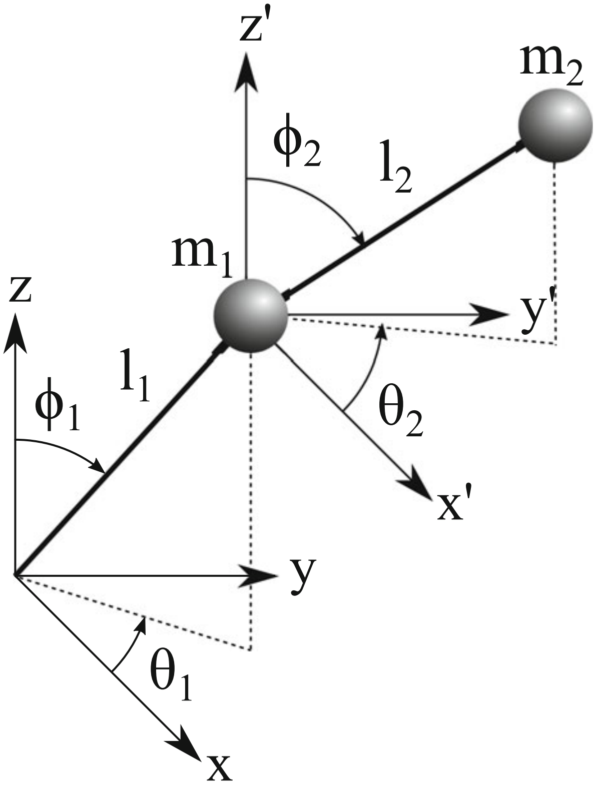


Fig. 4 Model of a double spherical pendulum with four degrees of freedom and chosen coordinates $(\phi_1, \phi_2, \theta_1, \theta_2)$. Actuation in both joints is assumed

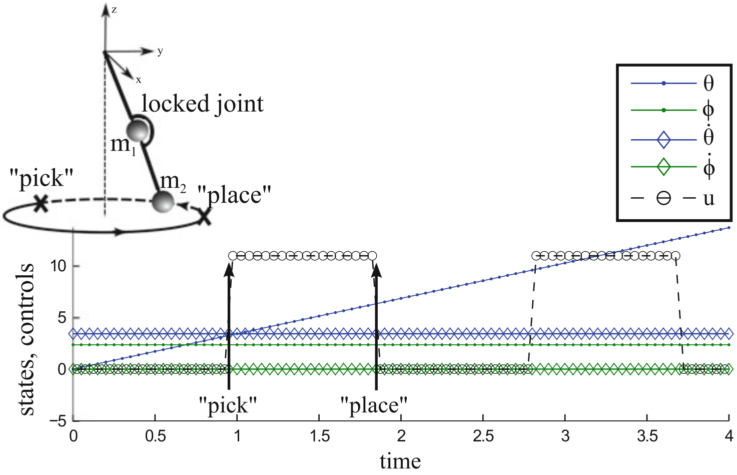


Fig. 5 Illustration and simulation of a hybrid trim for the pick and place scenario. The rotational velocity $\dot{\theta}$ is kept constant by the discontinuous, but piecewise constant control u

3 Motion Planning for Hybrid Mechanical Systems

In the motion planning framework, the hybrid properties of the system have to be accounted for: in the first time, when computing the primitives (restricted domains, limited time between switches), but also when searching for the optimal sequence in the library. In general, there is a need for hybrid control maneuvers, which connect primitives of the different continuous subsystems. For their computation additional constraints on the state space due to the guards have to be considered and an optimization of switching time has to be included (cf. Sect. 4 for illustrating examples).

Symmetries also occur in hybrid systems (cf. e.g. [12]). In the following, we restrict to a very specific case and assume that for two continuous subsystems, switching back and forth is allowed and the subsystems inhabit the same symmetry group G. We call a tuple of two pairs (ξ_1, u_1) and (ξ_2, u_2) a *hybrid trim*, if both are trims in their state spaces and if it holds that $x(t^-) = x(t^+)$, i.e. the state, in particular the velocity before and after switching is the same. In Fig. 5, an example is shown of a hybrid trim for a spherical pendulum which switches at the "pick" and "place" locations between two different modes (cf. Sect. 4 for a more detailed discussion.) By a hybrid control with switched constant control values, it is possible to generate a hybrid trim trajectory with constant horizontal velocity $\dot{\theta}$, i.e. in this example, we have $\xi_1 = \xi_2$ but $u_1 \neq u_2$.

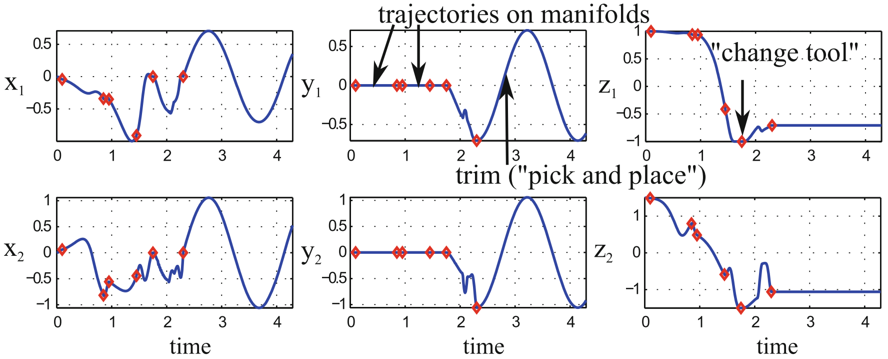


Fig. 6 Example solution sequence for the industrial robot scenario consisting of motion primitives. The scenario starts at the up-up position (subsystem f_1 active), is pushed to the unstable manifold (maneuver for f_1), than switches to a locked mode (f_2) and uses the corresponding unstable manifold to go downwards; after a short maneuver leaving the safety region (f_1), it rests at the down-down position to change the tool and finally steers (maneuver for f_1) to the rotational pick and place motion, which is a hybrid trim (switching between f_3 and f_4). The solution sequence is given in cartesian coordinates (cf. Fig. 4), *red dots* mark the switching between primitives

4 Example: An Academic Motion Planning Problem for an Industrial Robot

An open chain jointed robot as used e.g. in production facilities can be modeled—in an academic fashion allowing high simplifications on technical details—as a spherical pendulum. Thus, to illustrate the presented motion planning approach, we consider a double spherical pendulum with two-dimensional controls in both joints (cf. Fig. 4, $m_1 = 20\,\mathrm{kg}$, $m_2 = 8\,\mathrm{kg}$, $l_1 = 1\,\mathrm{m}$, $l_2 = 0.5\,\mathrm{m}$, $g = 9.81\,\mathrm{m/s^2}$). The Lagrangian and the equations of motion can be found e.g. in [7]; as the cost functional we chose the control effort modeled as $J(u) = \int_0^T u^2(t)dt$. The starting point for the scenario is the up-up position. The final condition is a periodic motion of the outstretched locked double pendulum, which is motivated by a *pick and place scenario* (a hybrid trim) assuming that m_2 is changed to 12 kg while the picked object is moved (cf. Fig. 5). Before heading to the final condition, the robot has to change the tool in the down-down equilibrium. Another kind of hybrid effect is brought into the problem by defining a safety region for $\phi_1 \in [\pm\pi/4, \pm\pi/2]$, where the second link has to be locked (cf. Fig. 3). To compute energy efficient control sequences, uncontrolled trajectories on the unstable manifolds (cf. Figs. 2 and 3) are used together with connecting control maneuvers. Thus, there are four different subsystems (labeled by their different vector fields for shortness): a double spherical pendulum (f_1), a locked double pendulum (f_2), and an outstretched locked pendulum with $m_2 = 8\,\mathrm{kg}$ (f_3) or $\overline{m}_2 = 12\,\mathrm{kg}$ (f_4). Figure 6 shows an example solution sequence for the motion planning problem.

In conclusion, this example shows that the motion planning with motion primitives method is particularly suited for an extension to hybrid systems: the flexibility of the method allows for incorporating *motion primitives from each continuous subsystem*, the computational effort of deriving an optimal hybrid solution is reduced by the motion planning library to finding a *hybrid optimal sequence*, and the method *exploits dynamical properties* which are present in hybrid as well as in ordinary mechanical systems. In future work, the approach has to be evaluated further by larger examples with more or different kinds of hybrid effects. Then, searching in the motion planning library will have to be performed by appropriate methods, e.g. sampling based road map algorithms (cf. e.g. [6]).

Acknowledgements This contribution was partially developed and published in the course of the Collaborative Research Centre 614 Self-Optimizing Concepts and Structures in Mechanical Engineering funded by the German Research Foundation (DFG) under grant number SFB 614.

References

1. Choset, H., Lynch, K.M., Hutchinson, S., Kantor, G.A., Burgard, W., Kavraki, L.E., Thrun, S.: Principles of Robot Motion: Theory, Algorithms, and Implementations. MIT Press, Cambridge (2005)
2. Binder, T., Blank, L., Bock, H.G., Bulirsch, R., Dahmen, W., Diehl, M., Kronseder, T., Marquardt, W., Schlöder, J.P., von Stryk, O.: Introduction to model based optimization of chemical processes on moving horizons. In: Grötschel, M., Krumke, S.O., Rambau, J. (eds.) Online Optimization of Large Scale Systems: State of the Art, pp. 295–340. Springer, Heidelberg (2001)
3. Ober-Blöbaum, S., Junge, O., Marsden, J.E.: Discrete mechanics and optimal control: an analysis. Control Optim. Calc. Var. **17**(2), 322–352 (2011)
4. Gill, P.E., Jay, L.O., Leonard, M.W., Petzold, L.R., Sharma, V.: An SQP method for the optimal control of large-scale dynamical systems. J. Comput. Appl. Math. **120**, 197–213 (2000). doi:10.1016/S0377-0427(00)00310-1
5. Frazzoli, E., Dahleh, M.A., Feron, E.: Maneuver-based motion planning for nonlinear systems with symmetries. IEEE Trans. Robot. **21**(6), 1077–1091 (2005)
6. Kobilarov, M.: Discrete geometric motion control of autonomous vehicles. Ph.D. thesis, University of Southern California (2008)
7. Flaßkamp, K., Ober-Blöbaum, S., Kobilarov, M.: Solving optimal control problems by exploiting inherent dynamical systems structures. J. Nonlinear Sci. **22**(4), 599–629 (2012)
8. Lygeros, J., Johansson, K.H., Simić, S.N., Zhang, J., Sastry, S.: Dynamical properties of hybrid automata. IEEE Trans. Automat. Control **48**(1), 2–17 (2003)
9. Buss, M., Glocker, M., Hardt, M., von Stryk, O., Bulirsch, R., Schmidt, G.: Nonlinear hybrid dynamical systems: modeling, optimal control, and applications. In: Engell, S., Frehse, G., Schnieder, E. (eds.) Modelling, Analysis, and Design of Hybrid Systems. Lecture Notes in Control and Information Sciences, vol. 279, pp. 311–335. Springer, Berlin (2002)
10. Guckenheimer, J., Holmes, P.: Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields. Applied Mathematical Sciences, vol. 42. Springer, New York (1983)
11. Dellnitz, M., Froyland, G., Junge, O.: The algorithms behind GAIO – set oriented numerical methods for dynamical systems. In: Fiedler, B. (ed.) Ergodic Theory, Analysis, and Efficient Simulation of Dynamical Systems, pp. 145–174. Springer, Berlin (2001)
12. Dellnitz, M., Hage-Packhäuser, S.: Stabilization via symmetry switching in hybrid dynamical systems. Discrete Continuous Dyn. Syst. Ser. B **16**(1), 239–263 (2011)

Performance of Sensitivity Based NMPC Updates in Automotive Applications

Jürgen Pannek and Matthias Gerdts

Abstract In this work we consider a half car model which is subject to unknown but measurable disturbances. To control this system, we impose a combination of model predictive control without stabilizing terminal constraints or cost to generate a nominal solution and sensitivity updates to handle the disturbances. For this approach, stability of the resulting closed loop can be guaranteed using a relaxed Lyapunov argument on the nominal system and Lipschitz conditions on the open loop change of the optimal value function and the stage costs. For the considered example, the proposed approach is realtime applicable and corresponding results show significant performance improvements of the updated solution with respect to comfort and handling properties.

1 Introduction

Within the last decades, model predictive control (MPC) has grown mature for both linear and nonlinear systems, see, e.g., [1–3]. Although analytically and numerically challenging, the method itself is attractive due to its simplicity and approximates an infinite horizon optimal control as follows: In a first step, a measurement of the current system state is obtained which in the second step is used to compute an optimal control over a finite optimization horizon. In the third and last step, a portion of this control is applied to the process and the entire problem is shifted forward in time rendering the scheme to be iteratively applicable.

Unfortunately, stability and optimality of the closed loop may be lost due to considering finite horizons only. To ensure stability of the resulting closed loop, one may impose terminal point constraints as shown in [4, 5] or Lyapunov type

J. Pannek (✉) • M. Gerdts
University of the Federal Armed Forces Munich, Werner-Heisenberg-Weg 39, 85577 Neubiberg, München, Germany
e-mail: juergen.pannek@unibw.de; matthias.gerdts@unibw.de

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_36,

terminal costs and terminal regions, see [6, 7]. A third approach uses a relaxed Lyapunov condition presented in [8] which can be shown to hold if the system is controllable in terms of the stage costs [9, 10]. Additionally, this method allows for computing an estimate on the degree of suboptimality with respect to the infinite horizon controller, see also [11, 12] for earlier works on this topic.

Here, we use an extension of the third approach to the case of parametric control systems and subsequent disturbance rejection updates. In particular, we focus on updating the MPC law via sensitivities introduced in [13]. Such updates have been analysed extensively for the case of open loop optimal controls, see, e.g, [14], but were also applied in the MPC closed loop context in [15, 16]. In order to avoid the usage of stabilizing Lyapunov type terminal costs and terminal regions and obtain performance results with respect to the infinite horizon controller, we utilize results from [16] in an advanced step setting, see, e.g., [17].

In the following, we present the considered half car model from [18, 19] and the imposed MPC setup. The obtained numerical results show that this approach is both realtime applicable and provides a cheap and yet significant performance improvement with respect to the comfort and handling objectives requested by our industrial partners.

2 Problem Setting

Throughout this work we consider the control systems dynamics of a half car which originate from [18, 19] and are slightly modified to incorporate active dampers, see Fig. 1 for a schematical sketch. The resulting second order dynamics read

$$\begin{aligned} m_1\ddot{x}_1 &= m_1 g + f_3 - f_1 & m_3\ddot{x}_3 &= m_3 g - f_3 - f_4 \\ m_2\ddot{x}_2 &= m_2 g + f_4 - f_2 & I\ddot{x}_4 &= \cos(x_4)(bf_3 - af_4) \end{aligned} \tag{1}$$

where the control enters the forces

$$\begin{aligned} f_1 &= k_1(x_1 - w_1) + d_1(\dot{x}_1 - \dot{w}_1) \\ f_2 &= k_2(x_2 - w_2) + d_2(\dot{x}_2 - \dot{w}_2) \\ f_3 &= k_3(x_3 - x_1 - b\sin(x_4)) + u_1(\dot{x}_3 - \dot{x}_1 - b\dot{x}_4\cos(x_4)) \\ f_4 &= k_4(x_3 - x_2 + a\sin(x_4)) + u_2(\dot{x}_3 - \dot{x}_2 + a\dot{x}_4\cos(x_4)) \end{aligned}$$

Here, x_1 and x_2 denote the centers of gravity of the wheels, x_3 the respective center of the chassis and x_4 the pitch angle of the car. The disturbances w_1, w_2 are connected via $w_1(t) = w(t)$, $w_2(t) = w(t - \Delta)$ and the control constraints $\mathbb{U} = [0.2\,\mathrm{kN\,s/m}, 5\,\mathrm{kN\,s/m}]^2$ limit the range of the active dampers. The remaining constants of the halfcar are displayed in Table 1.

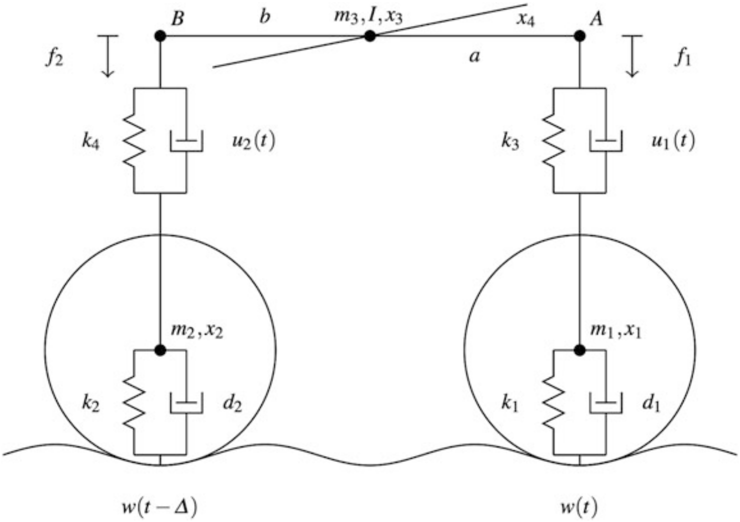


Fig. 1 Schematical sketch of a halfcar subject to road excitation w

Table 1 Parameters for the halfcar example

Name	Symbol	Quantity	Unit
Distance to joint	a, b	1	m
Mass wheel	m_1, m_2	15	kg
Mass chassis	m_3	750	kg
Inertia	I	500	kg m^2
Spring constant wheels	k_1, k_2	$2 \cdot 10^5$	kN/m
Damper constant wheels	d_1, d_2	$2 \cdot 10^2$	kN s/m
Spring constant chassis	k_3, k_4	$1 \cdot 10^5$	kN/m
Gravitational constant	g	9.81	m/s^2

3 MPC Algorithm

In order to design a feedback for the half car problem (1), we impose the cost functional

$$J_N(x, u, w) := \sum_{k=0}^{N-1} \mu_R F_R(k) + \mu_A F_A(k) \tag{2}$$

following ISO 2631 with horizon length $N = 5$. The handling objective is implemented via

$$F_R(k) := \sum_{i=1}^{2} \int_{kT}^{(k+1)T} \left(\frac{[k_i(x_i(t) - w_i(t)) + d_i(\dot{x}_i(t) - \dot{w}_i(t))] - F_i}{F_i} \right)^2 dt$$

with nominal forces

$$F_1 = (a \cdot g \cdot (m_1 + m_2 + m_3))/(a + b)$$
$$F_2 = (b \cdot g \cdot (m_1 + m_2 + m_3))/(a + b)$$

whereas minimizing the chassis jerk

$$F_A(k) := \int_{kT}^{(k+1)T} (m_3 \dddot{x}_3(t))^2 \, dt$$

is used to treat the comfort objective. Both integrals are equally weighted via $\mu_R = \mu_A = 1$ and are evaluated using a constant sampling rate of $T = 0.1\,\mathrm{s}$ during which the control are held constant, i.e. the control is implemented in a zero-order hold manner. The nominal disturbance $w(\cdot)$ and the corresponding derivates are computed from road profile measurements taken at a sampling rate of 0.002 s via a fast Fourier transformation (FFT).

For the resulting finite time optimal control problem, we denote a minimizer of (2) satisfying all constraints by $u^\star(\cdot, x, w)$. Since the control must be readily computed at the time instant it is supposed to be applied, $u^\star(\cdot, x, w)$ is computed in an advanced step setting, cf. [17]. To this end, the initial state x of the optimal control problem is predicted for a future time instant using the last known measurement and the intermediate control which is readily available from previous MPC iteration steps.

Since we want to apply sensitivity updates in case of measurement/prediction deviations and disturbances, we additionally precompute sensitivity information along the optimal open loop solution with respect to the predicted state $\partial u^\star/\partial x(\cdot, x, w)$ and the nominal disturbances $\partial u^\star/\partial w(\cdot, x, w)$. Then, once the nominal control $u^\star(\cdot, x, w)$ is to be applied, we use newly obtained state and disturbance information $\overline{x}$, $\overline{w}$ to update the control via

$$\overline{u}(\cdot, \overline{x}, \overline{w}) := u^\star(\cdot, x, w) + \begin{pmatrix} \frac{\partial u^\star}{\partial x}(\cdot, x, w) \\ \frac{\partial u^\star}{\partial w}(\cdot, x, w) \end{pmatrix}^\top \begin{pmatrix} \overline{x}(\cdot) - x(\cdot) \\ \overline{w}(\cdot) - w(\cdot) \end{pmatrix}, \tag{3}$$

see also [13, 14] for details on the computation and limitations of sensitivities.

For simplicity of exposition, we predict the initial state x using two sampling intervals T of the closed loop control. Note that although larger predictions are possible, robustness problems are more likely to occur since predicted and real solutions usually diverge, see, e.g., [20, 21].

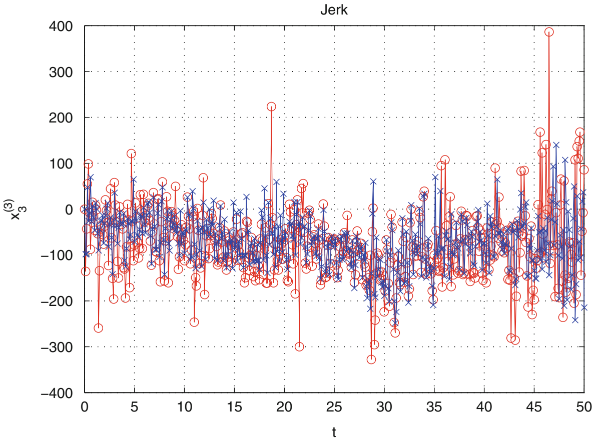


Fig. 2 Comparison plot for the chassis jerk using MPC with (*cross*) and without sensitivity update (*open circle*)

4 Numerical Results

During our simulations, we modified both the states of the system and the road profile measurements using a disturbance which is uniformly distributed in the interval $[-0.025\,\text{m}, 0.025\,\text{m}]$. For this setting, precomputation of $u^\star(\cdot, x, w)$, $\partial u^\star/\partial x(\cdot, x, w)$ and $\partial u^\star/\partial w(\cdot, x, w)$ required at maximum $0.168\,\text{s} < 2T = 0.2\,\text{s}$ which renders the scheme realtime applicable. As expected, the updated control law shows a better performance than the nominal control. The improvement cannot only be observed from Fig. 2, but also in terms of the closed loop costs: For the considered race track road data we obtain an improvement of approximately 8.2 % using the sensitivity update (3). Although this seems to be a fairly small improvement, the best possible result obtained by a full reoptimization reveals a reduction of approximately 10.5 % of the closed loop costs.

Note that due to the presence of constraints it is a priori unknown whether the conditions of the Sensitivity Theorem of [13] hold at each visited point along the closed loop. Such an occurrence can be detected online by checking for violations of constraints or changes in the control structure. Yet, due to the structure of the MPC algorithm, such an event has to be treated if one of the constraints is violated at open loop time instant $k = 1$ only which was not the case for our example.

Acknowledgements This work was partially funded by the German Federal Ministry of Education and Research (BMBF), grant no. 03MS633G.

References

1. Camacho, E., Bordons, C.: Model Predictive Control. Springer, Berlin (2004)
2. Rawlings, J.B., Mayne, D.Q.: Model Predictive Control: Theory and Design. Nob Hill Publishing, Madison (2009)
3. Grüne, L., Pannek, J.: Nonlinear Model Predictive Control: Theory and Algorithms. Springer, London (2011)
4. Keerthi, S., Gilbert, E.: Optimal infinite horizon feedback laws for a general class of constrained discrete-time systems: stability and moving horizon approximations. J. Optim. Theory Appl. **57**, 265–293 (1988)
5. Alamir, M.: Stabilization of Nonlinear Systems Using Receding-Horizon Control Schemes. Springer, London (2006)
6. Chen, H., Allgöwer, F.: A quasi-infinite horizon nonlinear model predictive control scheme with guaranteed stability. Automatica **34**(10), 1205–1218 (1998)
7. Mayne, D., Rawlings, J., Rao, C., Scokaert, P.: Constrained model predictive control: stability and optimality. Automatica **36**(6), 789–814 (2000)
8. Grüne, L., Rantzer, A.: On the infinite horizon performance of receding horizon controllers. IEEE Trans. Automat. Control **53**(9), 2100–2111 (2008)
9. Grüne, L.: Analysis and design of unconstrained nonlinear MPC schemes for finite and infinite dimensional systems. SIAM J. Control Optim. **48**, 1206–1228 (2009)
10. Grüne, L., Pannek, J., Seehafer, M., Worthmann, K.: Analysis of unconstrained nonlinear MPC schemes with varying control horizon. SIAM J. Control Optim. **48**(8), 4938–4962 (2010)
11. Shamma, J., Xiong, D.: Linear nonquadratic optimal control. IEEE Trans. Automat. Control **42**(6), 875–879 (1997)
12. Nevistić, V., Primbs, J.A.: Receding horizon quadratic optimal control: Performance bounds for a finite horizon strategy. In: Proceedings of the European Control Conference (1997)
13. Fiacco, A.: Introduction to Sensitivity and Stability Analysis in Nonlinear Programming. Academic, New York (1983)
14. Grötschel, M., Krumke, S., Rambau, J.: Online Optimization of Large Scale Systems. Springer, Heidelberg (2001)
15. Zavala, V.M., Biegler, L.T.: The advanced-step NMPC controller: optimality, stability and robustness. Automatica **45**(1), 86–93 (2009)
16. Pannek, J., Gerdts, M.: Robust stability and performance bounds for nmpc with abstract updates. In: Proceedings of the 4th IFAC Nonlinear Model Predictive Control Conference, pp. 311–316 (2012)
17. Findeisen, R., Allgöwer, F.: Computational delay in nonlinear model predictive control. In: Proceedings of the International Symposium on Advanced Control of Chemical Processes (2004)
18. Speckert, M., Dreßler, K., Ruf, N.: Undesired drift of multibody models excited by measured accelerations or forces. Tech. rep., ITWM Kaiserslautern (2009)
19. Popp, K., Schiehlen, W.: Ground Vehicle Dynamics. Springer, Heidelberg (2010)
20. Limon, D., Alamo, T., Raimondo, D.L., Bravo, J.M., Munoy de la Pena, D., Ferramosca, A., Camacho, E.F.: Input-to-state stability: an unifying framework for robust model predictive control, nonlinear model predictive control. In: Magni, L., Raimondo, D., Allgöwer, F. (eds.) Nonlinear Model Predictive Control: Towards New Challenging Applications. Lecture Notes in Control and Information Sciences, vol. 384, pp. 1–26. Springer, Heidelberg (2009)
21. Findeisen, R., Grüne, L., Pannek, J., Varutti, P.: Robustness of prediction based delay compensation for nonlinear systems. In: Proceedings of the 18th IFAC World Congress, Milan, pp. 203–208 (2011)

Optimal Control in Proactive Chassis Dynamics: A Fixed Step Size Time-Stepping Scheme for Complementarity Problems

Johannes Michael and Matthias Gerdts

Abstract This paper is about a fixed step size time-stepping scheme for the computation of solutions of complementarity problems. As we want to optimise chassis dynamics by solving optimal control problems, we took a closer look at modeling contact conditions. The latter are important, as the contact force is directly related to handling caracteristics of the automobile. This plays an important role particularly in certain driving situations, e.g. driving over a pothole. Hereafter the motivation for the development is carried out and the components of the scheme are explained. At the end we compare the calculation of a quartercar with a spring-damper road to wheel interaction to those resulting from the complementarity problem.

1 Aims and Setting

Our research addresses the development of real-time optimal control strategies to improve comfort and handling characteristics in automotives. The desired control is supposed to be proactive, i.e. it is calculated for an upcoming road segment and not by measuring wheel accelerations or other data via sensors. For this purpose we suppose to know the future road profile in a preview area, for example measured by laser sensors. To calculate the control we use a sensitivity approach by comparing the next road segment with parametrized comparandums. For these, the controls and sensitivities with respect to nominal road parameters are known and the applied control can be calculated by a sensitivity update from the nominal solution. In our case we use electro-rheological dampers as control devices, for whom the damper constants can be adjusted rapidly by a control current.

J. Michael (✉) • M. Gerdts
University of the Federal Armed Forces Munich, Werner-Heisenberg-Weg 39, 85577 Neubiberg, München, Germany
e-mail: johannes.michael@unibw.de; matthias.gerdts@unibw.de

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_37,

2 Modelling of a Quartercar

We consider the setting of a quatercar, see e.g. [3], consisting of two masses, one for the chassis and one for the wheel. For simplicity only movements in vertical direction are considered. The vector of the generalized coordinates consists of the two positions $q(t) = [q_1(t) \;\; q_2(t)]^\top$, where q_1 represents the wheel and q_2 the chassis position. The two masses are interconnected by a parallel spring-damper element, and the wheel is also connected to the road $w(t)$ by the same mechanical element.

3 Contact Dynamics as Complementarity Problem

As the control algorithm is supposed to act in crucial events, like a ride over edges, we investigate problems arising from contact mechanics in the case of loss of contact between wheel and road surface. To this end we replaced in the quartercar model the ground interaction by a masspoint without a fixed connection to the ground and a non penetration condition. In generalized coordinates these conditions lead to the following complementarity problem:

$$g(q(t)) \geq 0, \quad \lambda(t) \geq 0, \quad g(q(t)) \perp \lambda(t).$$

Here $g(q(t))$ represents the gap function between the wheel and the surface, $\lambda(t)$ is the amount of the contact force and the third condition ensures that there can only be a contact force, when the gap between wheel and surface is closed. More information about complementarity problems can be found in [1]. The dynamics of the contact condition can be written as first order differential equation in the following way.

$$\begin{aligned} \dot{q}(t) &= v(t), \\ M(q)\dot{v}(t) &= F(q(t), v(t)) + g'(q(t))^\top \lambda(t). \end{aligned}$$

All internal forces are summarized in $F(q(t), \dot{q}(t))$. The last term is the force vector of the contact force. To ensure a realistic behaviour we added another condition such that the elasticity of the tyre can be expressed. When an impact occurs there will be a rebound, that can be modeled via Newtons impact law, see e.g. [4].

$$g'(q(t))(v^+ + \varepsilon v^-) = 0$$

Here v^- and v^+ are defined for every contact time t_s as

$$v^- = \lim_{t \nearrow t_s} v(t), \quad v^- = \lim_{t \searrow t_s} v(t)$$

thus representing the left- and right-sided limit of the velocity at an impact. With this condition the amount of the force acting at an impact is defined by relating the velocity before and after contact by the elasticity coefficient $\varepsilon \in [0\,;1]$.

3.1 Problems with Fixed Step-Size Time-Stepping Schemes

To integrate the dynamics of the system with respect to time we need a suitable time-stepping scheme. Therefore we tried to use fixed step size schemes like the semi-implicit Euler scheme presented in [6]. This scheme is event driven, i.e. in every step one has to check if the complementarity condition is violated or not. But as we like to use the scheme in an optimization algorithm in future, we want to formulate the problem only by equality and inequality constraints. Especially if we want to use a fixed step size it is not possible to guarantee the desired behaviour at impact points if the impact time does not coincide with a discretization point.

3.2 A Fixed Step Size Time-Stepping Scheme for the Complementarity Problem

To overcome the problem above we reformulated the problem, such that the constraints consist only of equalities and inequalities. The presented scheme has a fixed step size h and the discretization points are written as $q^l = q(l \cdot h)$ and the other variables respectively. To determine if a contact occurs in the next time step we redefined the right side of the dynamics with the additional variable $\lambda(t)$ for the contact force as

$$\tilde{F}(q(t), v(t), \lambda(t)) = F(q(t), v(t)) + g'(q(t))^{\top}\lambda(t).$$

The actual time-stepping method stays the same as in [6]. It can be written as

$$q^{l+1} - q^l = hv^{l+1},$$
$$M(v^{l+1} - v^l) = h\tilde{F}(q^l, v^l, \lambda^l).$$

For simplification the mass matrix is supposed to be constant, what is true for the problem under consideration. The developed scheme consists of a preview step to calculate if there are violations of the state constraints at the next time step. Therefore we predict the state at the next discretization point $\tilde{q}^{l+1}$ using $\tilde{F}(q^l, v^l, 0)$, i.e. without any contact force:

$$\tilde{q}^{l+1} - q^l = h\tilde{v}^{l+1},$$

$$M(\tilde{v}^{l+1} - v^l) = h\tilde{F}(q^l, v^l, 0).$$

The following equalities and inequalities guarantee that the non-penetration and rebound condition are fulfilled:

$$\lambda^l \geq 0 \tag{1}$$

$$g(q^l) \geq 0 \tag{2}$$

$$g(\tilde{q}^{l+1})^\top \lambda^l \leq 0 \tag{3}$$

$$g'(q^l)^\top \lambda^l (v^{l+1} + \varepsilon v^l) = 0 \tag{4}$$

With the definition of (1) and (3) we obtain that if the predicted state is greater than zero the contact force has to be zero. But if $g(\tilde{q}^{l+1})$ is less than zero and the contact force is chosen to be zero, (2) would be violated in the next step, because then $\tilde{q}^{l+1} = q^{l+1}$ holds. Due to this, λ^l has to be chosen positive to guarantee feasibility. Considering (4) one obtains the amount of the contact force to ensure Newtons impact law. This time-stepping method will simulate the dynamics in a correct way if the step size is chosen small enough.

3.3 Numerical Results

In this section we present some simulation results concerning the comparison between the quartercar model with and without contact formulation. We simulated a ride over a step of height 10 cm with the constants taken from [5]. In Fig. 1 simulation results are depicted for both methods with the associated contact forces. The upper two plots depict the simulation when the wheel is modelled with a spring-damper element between its mass and the road. The lower depictions are the numerical results using the contact formulation with the preview step calculation and the conditions (1)–(4). In the left plots the solid line represents the wheel and the dashed one the chassis trajectory. One can see that in the first simulation the wheel does not behave as one would expect. It performs a rapid acceleration towards the road after the edge, due to the suddenly acting stressed spring in the model. At that instant of time the acting contact force, depicted in the right plot respectively, is negative, what cannot be true due to the assumption that the wheel does not glue to the road. In the second simulation using the contact formulation, the wheel behaves similar to a free falling ball, except disturbances due to the spring and damper forces by the connection to the chassis. The contact force is discontinuous, because the contact is lost at the edge and at every impact time impulsive forces occur to ensure the conditions above. After some jumps the wheel stays in contact with the road and also the contact force normalizes again to a continuous function.

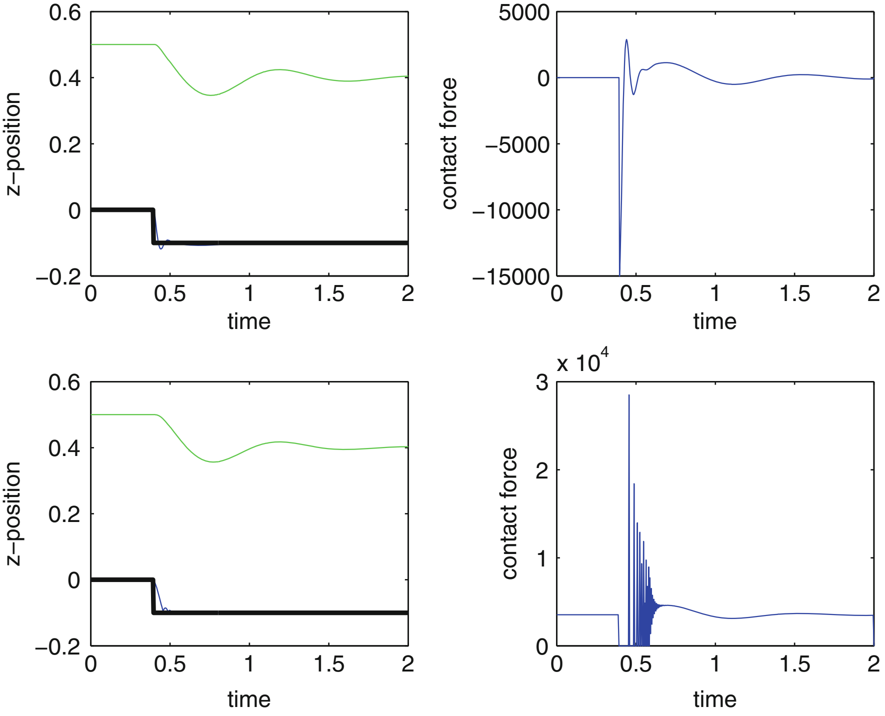


Fig. 1 Comparison between the spring-damper arrangement and the contact formulation

4 Outlook

The next step is to optimise the motion in the presence of contacts. First test examples done with the SQP algorithm snopt [2] regarding a bouncing ball work for one impact in the regarded time interval. We assume to figure out arising numerical difficulties and solve the optimal control problem to calculate control strategies and the corresponding sensitivities for crucial road situations. With this, the control can be calculated for upcoming disturbances with a sensitivity update in real-time.

Acknowledgements This work was partially funded by the German Federal Ministry of Education and Research (BMBF), grant no. 05M10WNA.

References

1. Brogliato, B.: Nonsmooth Mechanics. Models, Dynamics and Control. Communications and Control Engineering Series, 2nd edn. Springer, London (1999)
2. Gill, P.E., Murray, W., Saunders, M.A.: User's guide for snopt version 7 : software for large-scale nonlinear programming. Office **11**(1), 1–116 (2008)

3. Hac, A., Youn, I.: Optimal semi-active suspension with preview based on a quarter car model. In: American Control Conference, pp. 433–438 (1991)
4. Pfeiffer, F., Glocker, C.: Multibody Dynamics with Unilateral Contacts. Wiley-VCH Verlag GmbH, New York (2008)
5. Rettig, U., von Stryk, O.: Optimal and robust damping control for semi-active vehicle suspension. In: 5th EUROMECH Nonlinear Dynamics Conference (ENOC), Eindhoven, pp. 20–316 (2005)
6. Stewart, D.E.: Rigid-body dynamics with friction and impact. SIAM Rev. **42**(1), 3–39 (2000)

Model Reduction of Contact Problems in Elasticity: Proper Orthogonal Decomposition for Variational Inequalities

Joachim Krenciszek and René Pinnau

Abstract In this contribution a model order reduction method is applied to a Signorini contact problem. Due to the contact constraints classical linear reduction methods such as Craig–Bampton are not applicable. The Signorini contact problem is formulated as a variational inequality and Proper Orthogonal Decomposition (POD) is used to calculate an optimal projection subspace. Numerical results of the reduced model's quality and efficiency for an Encastre beam with contact are presented.

1 Content

In a lot of industrial processes e.g. in vehicle manufacturing the simulation of components that come in contact with each other or with the environment play a crucial role. For durability analysis and optimal design of tires and various mounts the numerical performance is an important issue. Hence the complexity of high dimensional finite element models exceeds the applicability of these uses and therefore model reduction has to be applied.

Classical model reduction techniques such as the Craig–Bampton method are used to reduce linear systems, but even with linear elasticity the contact inherits nonlinearity to the problem. The following paper presents a method to apply the nonlinear model reduction technique proper orthogonal decomposition to a problem in elasticity involving contact constraints.

J. Krenciszek (✉) • R. Pinnau
Fachbereich Mathematik, Technische Universität Kaiserslautern, Postfach 3049, 67653 Kaiserslautern, Germany
e-mail: krenciszek@mathematik.uni-kl.de; pinnau@mathematik.uni-kl.de

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*, Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_38,

2 Signorini Contact Problem

We will focus on the time-dependent Signorini contact problem. Consider an elastic solid body with constant density ρ that initially occupies the domain Ω. The boundary is partitioned into three parts: on Γ_D we have Dirichlet boundary condition with no displacement, Γ_N is subject to boundary forces f_N and Γ_C is subject to the Signori contact condition, that implies the possibility of contact. Volume forces f_V act on the whole solid.

Problem formulation: Find (u, σ) satisfying

$$\begin{aligned} \rho\ddot{u} - \operatorname{div}\sigma &= f_V \ \text{ in } \ \Omega \times (0,T) \\ n \cdot (\sigma n) &= f_N \ \text{ on } \ \Gamma_N \times (0,T) \\ u &= 0 \ \text{ on } \ \Gamma_D \times (0,T) \\ n \cdot u - g &\leq 0 \ \text{ on } \ \Gamma_C \times (0,T) \\ (n \cdot u - g)(n \cdot (\sigma n)) &= 0 \ \text{ on } \ \Gamma_C \times (0,T) \\ t \cdot (\sigma n) &= 0 \ \text{ on } \ \Gamma_C \times (0,T) \\ u(x,0) = u_0; \quad \dot{u}(x,0) &= u_1 \ \text{ in } \ \Omega \end{aligned}$$

To derive a variational formulation we define

$$\begin{aligned} V &= \{v \in [H^1(\Omega)]^n \ \mid \ v = 0 \ \text{ on } \ \Gamma_D\} \\ V_C &= \{v \in V \ \mid \ n \cdot v - g \leq 0 \ \text{ on } \ \Gamma_C\} \end{aligned}$$

Then the weak formulation reads: Find $u : [0,T] \to V_C$ satisfying

$$\int_\Omega \rho\ddot{u} \cdot (v - u) dx + \int_\Omega \sigma(u) : \epsilon(v - u) dx \geq F(v - u)$$

for all $v \in V_C$, where $\epsilon(u)$ is the strain tensor and $\sigma(u)$ is the stress tensor given by $\sigma(u) = C\epsilon(u)$. The force term is obtained by:

$$F(v - u) = \int_\Omega f_V \cdot (v - u) dx + \int_{\Gamma_N} f_N \cdot (v - u) d\gamma$$

A detailed explanation can be found in [1]. Applying a discretization with the finite element space V_N we obtain a finite dimensional variational inequality. The constraints will be replaced by constraints on the grid points:

Find x^* such that

$$\begin{aligned}(M\ddot{x} + Kx - b)^T(x^* - x) &\geq 0 \\ \text{s.t. } Bx &\leq c\end{aligned} \tag{1}$$

where K is the stiffness matrix, M is the mass matrix

$$(K)_{ij} = \int\limits_{\Omega} C\epsilon(\phi_j) : \epsilon(\phi_i) dx, \qquad (M)_{ij} = \int\limits_{\Omega} \rho\phi_j\phi_i dx$$

and

$$b_i = \int\limits_{\Omega} F_V \cdot \phi_i ds + \int\limits_{\Gamma_N} F_N \cdot \phi_i ds$$

$$(B)_{ij} = n \cdot \phi_j(x_i), \quad c_i = g(x_i) \text{ with } x_i \in \Gamma_N$$

Applying a discretization in time:

$$\ddot{x}_k = \frac{x_k - 2x_{k-1} + x_{k-2}}{(\Delta t_k)^2}$$

Using this in (1) we obtain:

$$\begin{aligned}\left(\bar{A}_k x_k - \bar{b}_k\right)^T (x^* - x_k) &\geq 0 \\ \text{s.t. } Bx &\leq c\end{aligned} \tag{2}$$

with

$$\bar{A}_k = \frac{1}{(\Delta t_k)^2} M + K \quad \text{and} \quad \bar{b}_k = b + M\frac{2x_{k-1} - x_{k-2}}{(\Delta t_k)^2}$$

This linear variational inequality (2) is equivalent to the quadratic program:

$$\begin{aligned}\min\ & \frac{1}{2} x_k^T \bar{A} x_k - \bar{b}^T x_k \\ \text{s.t. } & Bx_k \leq c\end{aligned} \tag{3}$$

3 Proper Orthogonal Decomposition

Proper orthogonal decomposition is a method to determine an optimal subspace basis, similar to the concepts of Karhunen-Loève expansion and principal component analysis. Applied as a method of model reduction, the data that is approximated in an optimal least square sense is given in the form of solutions of the full systems or can even be obtained from experimental measurements. Since the solution of the full system usually is given as a result of a numerical approximation, it is only given at certain time instances t_i. These samples are called Snapshots and are stored in the snapshot matrix

$$Y = [y_1 \dots y_m] = [y(t_1) \dots y(t_m)] \in \mathbb{R}^{n \times m} \tag{4}$$

Then an optimal basis with basis vectors φ_i has to be determined that minimizes the projection error:

$$\min_{\varphi_j} \sum_{i=1}^{m} \| y_i - \sum_{j=1}^{d} \langle y_i, \varphi_j \rangle \varphi_j \|^2$$

The solution to this minimization problem can be obtained by means of singular value decomposition (SVD) of the snapshot matrix Y:

$$Y = U \Sigma V^*$$

where U is a unitary $n \times n$ matrix, V is a unitary $m \times m$ matrix and Σ is a $n \times m$ diagonal matrix containing the singular values:

$$U = [\varphi_1 \dots \varphi_n] \qquad \Sigma = \begin{pmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{pmatrix}$$

The minimal projection error can then be expressed in terms of the singular values corresponding to the omitted singular vectors:

$$\min_{\varphi_j} \sum_{i=1}^{n} \| y_i - \sum_{j=1}^{d} \langle y_i, \varphi_j \rangle \varphi_j \|^2 = \sum_{i=d+1}^{n} \sigma_i^2 \tag{5}$$

A derivation can be found in [8].

To apply a projection to the POD subspace defined by $\phi \in \mathbb{R}^{n \times d}$ we substitute x with the reduced approach $x = \phi \tilde{x}$. This leads to a reduced quadratic program resp. linear variational inequality of the same form as (2) and (3) with

$$\tilde{A}_k = \phi^T \bar{A}_k \phi$$

$$\bar{b}_k = \phi^T b + \phi^T M \phi \frac{2x_{k-1} - x_{k-2}}{(\Delta t_k)^2}$$

$$\tilde{B} = B\phi$$

Note that the POD modes in general do not satisfy the constraints while their linear combination in the solution of the reduced system does. It has to be ensured that the initial value satisfies the constraints in the reduced space as well, hence orthogonal projection might not be applicable or has to be corrected to match the constraints.

If we combine local finite elements on the n_c grid points subject to the contact condition the reduced matrix $\tilde{B}$ for our toy problem is in $\{0, -1, 1\}^{n_c \times d}$. This reduces the cost of incorporating the inequality constraint.

The quadratic program can be solved e.g. using an active set algorithm. It is beneficial to the performance to use the solution of the previous time step as a starting solution for the quadratic program of the next time step.

4 Numerical Results

As a toy problem we consider a two-dimensional beam occupying the initial domain $\Omega = [0, 10] \times [0, 1]\,\mathrm{m}^2$, that is fixed on the left, has scope on the right and a force f_N is applied at the middle of the beam (see Fig. 1).

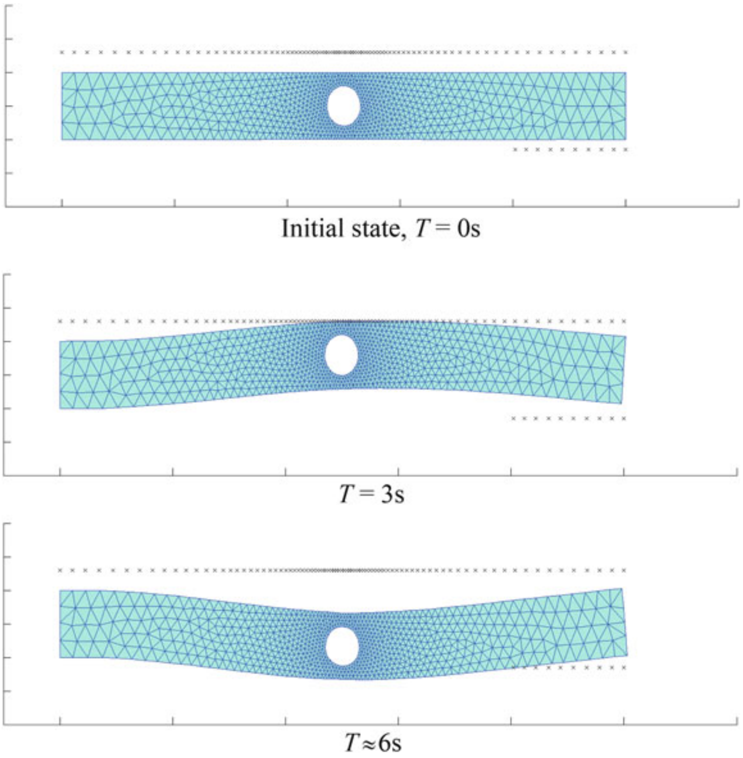
The used material properties are:

- Shear modulus $G = 5{,}000\,\mathrm{N/m}^2$
- Mass density $\rho = 500\,\mathrm{kg/m}^3$
- Poisson's ratio $\nu = 0.3$

The force f_N is applied in the middle of the beam, acts in normal direction with an absolute value of 400 N and the sign is changed with a period of 3 s.

For the spatial discretization we use linear finite elements on a triangular grid. In this test case 1788 degrees of freedom and 1,000 time steps to discretize the time interval $[0, 40]$ s. The computation time of the full model was 1,028 s on our test machine. From (5) we can deduce that the exponentially decreasing singular values of the snapshot matrix (Fig. 2) promise applicability of model reduction. The calculation of the POD modes from the snapshot matrix, that is the principle part of the offline phase, took 2 s. In Fig. 3 we can observe a likewise exponentially decreasing relative error and a drastic reduction of computation time as can be seen in Fig. 4.

In the case of nonlinear material laws a nonlinear variational inequality has to be solved. This can be done using the Josephy-Newton-Method or sequential quadratic programming (SQP), which involves a large number of linear variational inequalities that have to be solved. Here the dimension reduction using POD can give significant improvements in performance.


Initial state, $T = 0$s

$T = 3$s

$T \approx 6$s

Fig. 1 Encastre beam (fixed on the *left*) with scope on the *right*, periodic force applied in normal direction at the *middle* of the beam (deformation scaled for visualization)

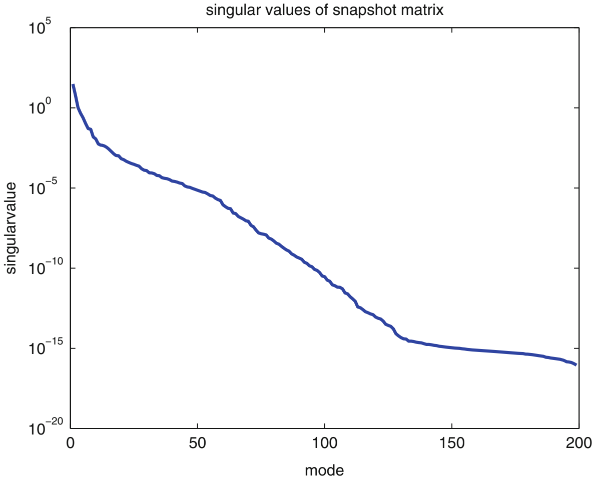


Fig. 2 Decay of singular values of snapshot matrix

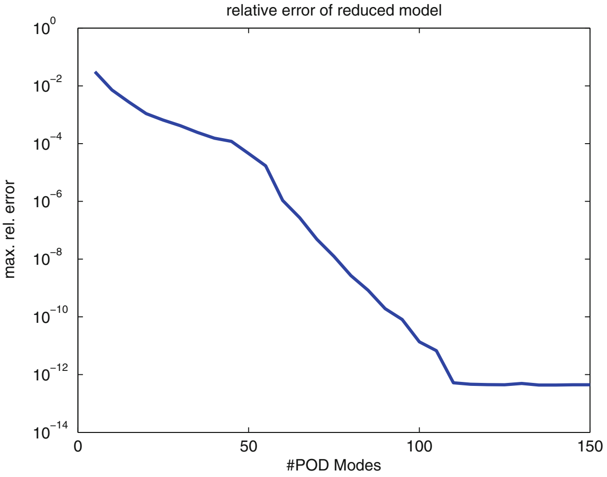


Fig. 3 Maximum relative error of deformation of POD reduced system

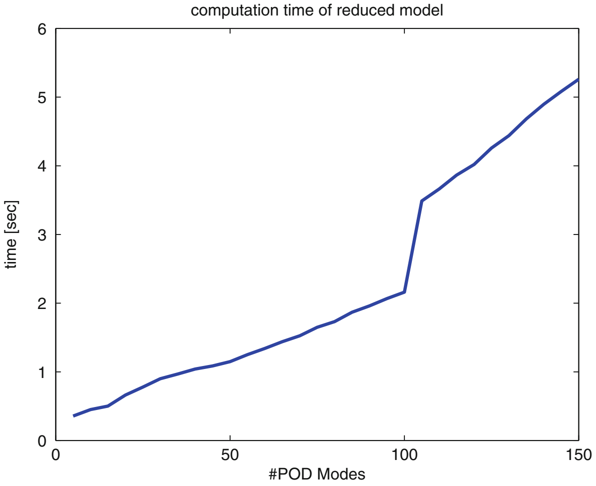


Fig. 4 Computation time of POD reduced system (compared to 1,028 s with full system)

5 Conclusion

Contact problems are inherently nonlinear, therefore model reduction is a challenging task. We demonstrated the application of proper orthogonal decomposition to a Signorini contact problem formulated as a variational inequality. The reduced model with a dimension reduction of more than 90 % (∼120 modes), yielding a relative approximation error of less than 10^{-12}, takes less than 1 % of the computation time of the full model including time used in the offline phase. We intend to investigate the model reduction of contact problems with nonlinear material laws and the trajectory piecewise linear (TPWL) approach in combination with POD under the presence of contact constraints.

References

1. Cao-Rial, M.T., Quintela, P., Moreno, C.: Numerical solution of a time-dependent signorini contact problem. Discrete Continuous Dyn. Syst., 201–211 (2007)
2. Geiger, C., Kanzow, C.: Theorie und Numerik restringierter Optimierungsaufgaben. Springer, Berlin (2002)
3. Haasdonk, B., Salomon, J., Wohlmuth, B.: A reduced basis method for parametrized variational inequalities. SIAM J. Numer. Anal. **50**, 2656–2676 (2012)
4. Hinze, M., Volkwein, S.: Proper orthogonal decomposition surrogate models for nonlinear dynamical systems: error estimates and suboptimal control. In: Benner, P., Sorensen, D.C., Mehrmann, V. (eds.) Dimension Reduction of Large-Scale Systems. Lecture Notes in Computational Science and Engineering, vol. 45, pp. 261–306. Springer, Berlin (2005)
5. Kikuchi, N., Oden, J.: Contact Problems in Elasticity: A Study of Variational Inequalities and Finite Element Methods. SIAM Studies in Applied Mathematics. Society for Industrial and Applied Mathematics, Philadelphia (1988)
6. Kinderlehrer, D., Stampacchia, G.: An Introduction to Variational Inequalities and Their Applications. Academic, New York (1980)
7. Kunisch, K., Volkwein, S.: Galerkin proper orthogonal decomposition methods for parabolic problems. Numerische Mathematik **148**, 117–148 (2001)
8. Kunisch, K., Volkwein, S.: Galerkin proper orthogonal decomposition methods for a general equation in fluid dynamics. SIAM J. Numer. Anal. **40**(2), 492–515 (2010)

Novel Updating Mechanisms for Stochastic Lattice-Free Traffic Dynamics

Alexandros Sopasakis

Abstract We present a novel lattice-free microscopic stochastic process in order to model vehicular traffic. Vehicles advance freely in a multi-lane environment without lattice cells limitations. As a result vehicles perform their moves based on a modified stochastic spin-flip and spin-exchange Arrhenius dynamic potential. Furthermore we put forward a modified kinetic Monte Carlo algorithm which produces the solution for these dynamics in real-time even for the case of a large traffic streams. An up to now unknown discrepancy is revealed between models using classical lattice-based methods versus those implementing this new lattice-free approach. The solution proposed by Renyi as well as the Palasti conjecture help in clarifying this discrepancy by showing that indeed the new proposed lattice-free process is correct in predicting traffic densities while avoiding the overestimates produced by classic Cellular Automata type, lattice-based, approaches.

1 Introduction

We begin by presenting the mechanism behind the lattice-free stochastic process. We refer to [5] for statistical mechanics related details.

We define a two dimensional domain D representing the roadway. The set D consists of the set O comprised of the disjoint union of all space occupied by vehicles and the set E which is the disjoint union of all empty space. Thus $D = O \cup E = V_1 \cup V_2 \cup \cdots \cup V_k \cup E_{k+1} \cup \cdots \cup E_{k+l}$. assuming k vehicles and l empty sets on the roadway. Figure 1 illustrates such a set topology as it would be applied in a single lane vehicular roadway. We now define the microscopic stochastic process $\{\sigma_t\}_{t \geq 0}$ on each of the subsets of the set D representing the roadway. Clearly, based

A. Sopasakis (✉)
Mathematical Sciences, Lund University, Lund, Sweden
e-mail: sopasak@maths.lth.se

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_39,

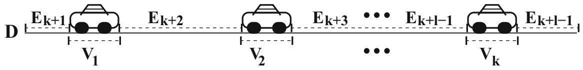


Fig. 1 Schematic of roadway subdivided into respective occupied sets V_i, and empty sets E_i

on our definition of sets E_i and V_i above, there will always be a finite number of occupied and empty sets in D. We can therefore define a spin-like variable $\sigma_t(i) \equiv \sigma(i)$ on each of those sets as follows

$$\sigma(i) = \begin{cases} 1 \text{ if at } V_i, \text{ i.e. vehicle exist at } i \in \{1, \dots, k\}, \\ 0 \text{ if at } E_i, \text{ i.e. there is no particle at } i \in \{k+1, \dots, k+l\}. \end{cases}$$

where $1 \le i \le k + l < M$ assuming k vehicles and l empty sets. Note that for any given number of vehicles the upper bound M will always exists. The stochastic process $\{\sigma_t\}_{t \ge 0}$ will change values in time, signifying vehicle motion, according to specific interaction rules which we provide below.

We follow ideas from classical [1, 4, 5] lattice-based stochastic processes in defining a new lattice-free interaction potential J describing how vehicles interact locally with each other. Using our lattice-free set infrastructure, interactions between vehicles at and are described from,

$$J(i-j) = \frac{1}{L} W\left(\frac{i-j}{L}\right), \text{if } \ i, j \in I_o \text{ and } W(r) = \begin{cases} J_* & \text{for } 0 \le r \le 1 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where I_o is the index set for set O and J_* is a free parameter to be calibrated from actual data as shown in [4,5].

2 Lattice-Free Stochastic Interactions and Arrhenius Dynamics

Following ideas from [4, 5] we choose to implement Arrhenius rates in order to model how vehicles physically interact. For the reasons behind choosing Arrhenius instead of perhaps Metropolis dynamics we refer to [4, 5]. As a result to model vehicles entering the roadway we define the spin-flip mechanism using the following lattice-free Arrhenius type rate

$$c(i,\sigma) = \begin{cases} c_o \exp(-\beta U(i,\sigma)) & \text{if } \sigma(i) = 1 \\ c_o w(i) & \text{if } \sigma(i) = 0. \end{cases} \text{ with } w(i) = \begin{cases} |E_i| - |V| & \text{if } |E_i| > |V| \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where $|V|$ denotes the area occupied by a vehicle. Thus the condition $|E_i| > |V|$ for $w(i)$ in (2) simply denotes the fact that for a vehicle, which occupies space $|V|$, the empty space $|E_i|$ at that location of the roadway must be sufficiently large. Similarly, using the same definition for $w(j)$ as in (2), we define the corresponding lattice-free Arrhenius-type spin-exchange rate as

$$c(i,j,\sigma) = \begin{cases} d_o w(j)\exp(-\beta U(i,\sigma)) & \text{if } \sigma(i) = 1 \text{ and } \sigma(j) = 0 \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

As previously explained, this rule describes how a vehicle moves from location i to location j on the roadway. The parameter d_o in (2) and (3) is a constant representing the characteristic time of the stochastic process. This constant represents driver reaction times and therefore the values chosen for this parameter affect vehicle velocities. Such constants are calibrated directly from actual data as shown in [4] and once calibrated do not need to be changed throughout the simulation of the roadway. The potential function U appearing in (2) and (3) above is defined to be

$$U(i,\sigma) = \sum_{j=1}^{k+l} J(i-j)\sigma(j) \tag{4}$$

with the local interaction potential J as defined previously in (1). Further details about the rates (2) and (3) as well as numerical implementation issues are also treated in [5].

3 Lattice-Free Versus Lattice-Based Dynamics

In this section we present results of traditional LB simulations such as Cellular Automaton versus the new lattice-free stochastic process solutions under the influence of spin-flip dynamics in order to reveal an interesting difference between the two. We compare three cases of vehicle densities (light, medium and heavy) versus time in Fig. 2a.

The results in Fig. 2a above show complete path-wise and long-range agreement between Cellular Automaton and the new lattice-free dynamics only for very light vehicle densities. Even at such light densities however the Cellular Automaton simulations seem to, on the average, always produce solutions which are slightly greater than the lattice-free solutions. Clear differences in solutions start to appear for medium vehicle densities. In fact the larger the vehicle densities the worst the discrepancies are as can be seen in the case of heavy vehicle densities in Fig. 2a. We found that all cases tested (not shown here) with increasingly higher vehicle densities, also produced increasingly higher discrepancies between Cellular Automaton and lattice-free dynamics. The natural question therefore is which is

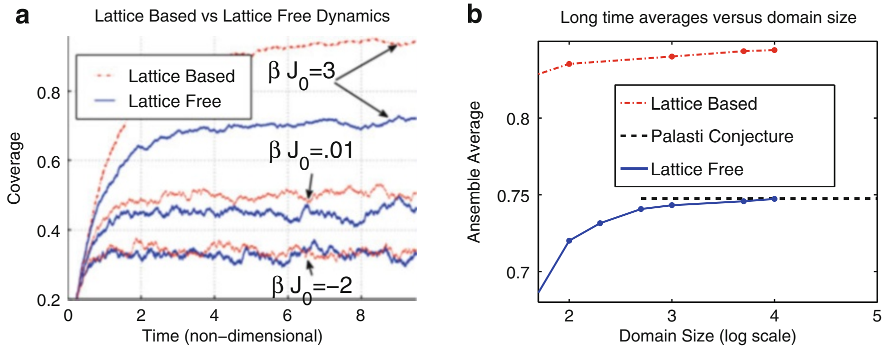


Fig. 2 (**a**) Comparisons of vehicle densities for lattice-free (*red*) versus lattice-based (*black*) cellular Automaton type simulations. Three different (light, medium, heavy) vehicle densities are shown. Clear differences in solutions appear as vehicle densities become larger. (**b**) The heavy density example revisited and analyzed. We present the equilibrated solution versus domain size for classical lattice-based (*blue*) and lattice-free (*red*) dynamics. Dynamics are clearly different. However only the lattice-free solution approaches the correct [2,3] asymptotic limit 0.7476

correct the classic lattice-based, Cellular Automaton type, solution or the proposed here lattice-free solution?

However the Palasti conjecture [2] (and well-known solutions, e.g. [3]) can resolve this question. The assumption behind the conjecture is that you monitor a process randomly placing objects of the same size in a given region. The Palasti conjecture and subsequent Renye result [3] provide an estimate for the maximum resulting coverage for that region. The Palasti conjecture states that the maximum density of such a process in a given region will be 0.7476 on the average. This further validates the solutions produced by the proposed LF stochastic process when compared to the lattice-based, classical, Cellular Automaton solutions (Fig. 2b). As pointed out in that figure such discrepancies are not a numerical artifact of finite vehicle sizes or domain size (e.g. the findings do not change as $D \to \infty$). Those differences are a direct result of whether lattice-based or lattice-free dynamics were used.

Furthermore the discrepancy in solutions does not only occur at equilibration but almost from the onset (see Fig. 2a). In other words in and around locally dense vehicle accumulations lattice-based solutions already start to display deviations when compared with lattice-free solutions under similar conditions. In general this shows that serious numerical overestimates can occur for Cellular Automaton type solutions especially for cases where higher vehicle densities come into play. In such cases the relative error is higher than 20 % (compared to the equivalent lattice-free solution). It is important to note that this is a local result. In other words the error in LB dynamics will occur as soon as the local density increases (even if the overall traffic stream density is low). As a result discrepancies in solutions can occur even for light traffic streams as soon as vehicles start to come close to each other during their travel.

References

1. Liggett, T.M.: Interacting Particle Systems. Springer, New York (1985)
2. Palasti, I.: On some random space filling problems. Publ. Math. Inst. Hung. Acad. Sci. **5**, 353–360 (1960)
3. Renyi, A.: On a one-dimensional problem concerning random spacefilling. Publ. Math. Inst. Hung. Acad. Sci. **3**, 109–127 (1958)
4. Sopasakis, A.: Stochastic noise approach to traffic flow modeling. Physica A **342**(3–4), 741–754 (2004)
5. Sopasakis, A., Katsoulakis, M.A.: Stochastic modeling and simulation of traffic flow: Asep with Arrhenius look-ahead dynamics. SIAM J. Appl. Math. **66**, 921–944 (2005)

Part VII
Further Applications

Overview

This section on mathematics for *Further Applications* contains nine contributions that range from recrystallization kinetics, molecular dynamics over eye protective welding devices and haptic touchscreens to satellite-to-satellite laser tracking.

In *Modelling Some Recrystallization Processes with Random Growth Velocity of the Grains* Elena Villa and Paulo R. Rios consider birth-and-growth processes and study the effect of a random grain's growth velocity on recrystallization kinetics. The modeling framework can also be applied to nucleation and growth reactions.

In *A Mathematical Model for the Melting of Spherical Nanoparticles* Francesc Font et al. deal with the melting process of gold nanoparticles and present asymptotic and numerical results for the melting front of spherical particles. Based on a scale analysis they show that previously neglected terms in the Gibbs–Thomson equation describing the melt temperature as a function of particle size can have a significant effect.

In *Local Quantum-Like Updates in Classical Molecular Simulation Realized within an Uncoupling-Coupling Approach* Konstantin Fackeldey and Alexander Bujotzek present a novel technique to improve the precision of the classical molecular dynamics force field by solving an approximation problem with scattered quantum mechanical data. The performance of the method is studied for the alanine tripeptide.

In *Design of Automatic Eye Protective Welding Devices* Matej Bazec et al. investigate the optimal configuration of LCD light shutters by help of two different numerical approaches. The first approach aims a very accurate, but computationally expensive result by minimizing the Frank elastic energy of a particular liquid crystal layer and solving the Maxwell equations for the complete stack of optical elements. The second approach is dedicated to a better understanding (analysis) of the light polarization and propagation and uses a simplified model.

The occurrence of Color-over-Angle (CoA) variations in the light output of white phosphor-converted LEDs is an undesired effect. In *A Three-Segment*

Inverse Method for the Design of CoA Correcting TIR Collimators Corien Prins et al. propose an inverse method to reduce these CoA variations using a special collimator.

In *Mathematical Modelling of Haptic Touchscreens* William T. Lee et al. report on a feasibility study for implementing haptic keyboards for touchscreen mobile devices. They consider driving transverse waves of the touchscreen using piezoelectric transducers mounted at the edges.

Transformation acoustics focuses on the design of advanced acoustic devices by employing sophisticated mathematical transformation techniques for engineering acoustic meta-materials. These are materials that are artificially fabricated with extraordinary acoustic properties beyond those encountered in nature. In *A Covariant Spacetime Approach to Transformation Acoustics* Michael M. Tung and Jesús Peinado present differential-geometric methods in combination with a variational principle that form the basis for a framework to control acoustic waves in industrial applications.

In *Location and Management of a New Industrial Plant* Miguel E. Vázquez-Méndez et al. discuss the optimal location problem of a new industrial plant taking into account economic and ecological aspects. Embedded in the framework of multi-objective optimization and control they analyze the problem, state Pareto-optimal solutions and use the Pareto-frontier as tool in the decision-making process.

In *A Satellite-to-Satellite Laser Tracking Solution within the Post-Newtonian Model of the Earth Outer Space* Jose M. Gambi and Maria Luisa Garcia del Pino derive two second order post-Newtonian formulae for the two-way frequency shift and the two-way laser ranging by means of Synge's world-function. The formulae can be used to increase the accuracy in tracking passive targets by APT systems on board of Earth satellites.

Nicole Marheineke

Modelling Some Recrystallization Processes with Random Growth Velocity of the Grains

Elena Villa and Paulo R. Rios

Abstract Heterogeneous transformations (or reactions) may be defined as those transformations in which there is a sharp moving boundary between the transformed and untransformed region. Such transformations may be modelled by the so-called birth-and-growth processes. We focus here on the effect that a random velocity of the moving boundaries of the grains has in the overall kinetics. One example of a practical situation in which such a model may be useful is that of recrystallization; a recent review of 3-D experimental results on recrystallization kinetics concluded that there is compelling evidence that every grain has its own distinct growth rate. Motivated by this practical application we present general kinetics expressions for various situations of practical interest, in which a random distribution of growth velocities is assumed. Previously known results follow here as particular cases. Although the motivation was recrystallization, the expressions presented here may be applied to nucleation and growth reactions in general.

1 Basics and Notations

Heterogeneous transformations (or reactions) may be defined as those transformations in which there is a sharp moving boundary between the transformed and untransformed region. This definition aims at chemical reactions in general; specifically it is applied to *nucleation and growth* transformations in Materials Science, but the geometrical idea pertaining to the definition finds a wide range

E. Villa (✉)
Department of Mathematics, University of Milan, Via Saldini 50, 20133 Milano, Italy
e-mail: elena.villa@unimi.it

P.R. Rios
Universidade Federal Fluminense, Escola de Engenharia Industrial Metalúrgica de Volta Redonda, Av. dos Trabalhadores 420, 27255-125 Volta Redonda, RJ, Brazil
e-mail: prrios@id.uff.br

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_40,

of application in diverse fields of knowledge [16], such as, the phase separations in multicomponent alloys [14], the film growth on solid substrates [6], the kinetics of Ising lattice-gas model [12], and the DNA replication [10]. To these we may add a recent extensive work by Aquilano et al. [1] on crystallization processes.

These transformations, or more in general any practical situation in which *nuclei* (germs) are born in time and are located in space randomly, and each nucleus generates a *grain* evolving in time according with a given growth law, may be mathematically modelled by dynamic germ-grain models [15] by means of the so called *birth-and-growth (stochastic) processes*. Specifically, by denoting $\Theta^t_{T_j}(X_j)$ the grain obtained as the evolution up to time $t \geq T_j$ of the nucleus born at time T_j in X_j, then the transformed region Θ^t at time $t > 0$ is given by

$$\Theta^t = \bigcup_{T_j \leq t} \Theta^t_{T_j}(X_j).$$

Of course a site saturated process (i.e., all possible nucleation sites are exhausted at the very beginning of the reaction) may be seen as a particular case of the time-dependent one by assuming $T_j \equiv 0$ for any j. Time-dependent nucleation processes and site-saturated processes may be modelled by marked point processes and by point processes, respectively. In order to define a birth-and-growth process we need to introduce also a growth model. Models of volume growth have been studied extensively, since the pioneering work by Kolmogorov [11]. We consider here a simple case of the so-called *normal growth model* (see also, e.g., [5, 18] and reference therein); namely, we shall consider the case in which all the grains develop with random velocity G constant in time or time dependent, so that for any time t all the grains have spherical shape (this is due to the fact that G is not space-dependent). The family of random sets $\{\Theta^t\}_t$ is called birth-and-growth process.

Since Θ^t is a random set, it gives rise to a random measure $\nu^d(\Theta^t \cap \cdot\,)$ in $\mathbb{R}^d$ for all $t > 0$, having denoted by ν^d the d-dimensional Lebesgue measure in $\mathbb{R}^d$. In particular, it is of interest to consider the *expected volume measure* $\mathbb{E}[\nu^d(\Theta^t \cap \cdot\,)]$ and its density (i.e., its Radon–Nikodym derivative), called *mean volume density of* Θ^t and denoted by V_V, provided it exists:

$$\mathbb{E}[\nu^d(\Theta^t \cap A)] = \int_A V_V(t,x)\mathrm{d}x \qquad \forall A \in \mathscr{B}_{\mathbb{R}^d}.$$

Whenever V_V is independent of x (e.g., under assumptions of homogeneous nucleation and growth), it is also called *volume fraction*. We mention that other quantities of interest in real applications are the so-called *mean extended volume density* at time t, denoted by $V_E(t,\cdot)$, defined as the density of the *mean extended volume measure* at time t, $\mathbb{E}[\mu^{\mathrm{ex}}_{\Theta^t}](\,\cdot\,) := \mathbb{E}[\sum_{j:T_j \leq t} \nu^d(\Theta^t_{T_j}(X_j) \cap \,\cdot\,)]$ on $\mathbb{R}^d$, that is

$$\mathbb{E}[\mu^{\mathrm{ex}}_{\Theta^t}](A) = \int_A V_E(t,x)\mathrm{d}x, \qquad \forall A \in \mathscr{B}_{\mathbb{R}^d},$$

and the *mean surface density* $S_V(t,\cdot)$ and the *mean extended surface density* $S_E(t,\cdot)$ at time t, defined as the density of the *mean surface measure* at time t, $\mathbb{E}[\mu_{\partial\Theta^t}](\,\cdot\,) := \mathbb{E}[\mathscr{H}^{d-1}(\partial\Theta \cap \,\cdot\,)]$, and the density of the *mean extended surface measure* at time t, $\mathbb{E}[\mu^{ex}_{\partial\Theta^t}](\,\cdot\,) := \mathbb{E}[\sum_{j:T_j\leq t}\mathscr{H}^{d-1}(\partial\Theta^t_{T_j}(X_j)\cap \,\cdot\,)]$, respectively, where $\mathscr{H}^{d-1}$ is the $(d-1)$-dimensional Hausdorff measure. It is clear that to find out formulas for the mean volume density V_V (and so for the other quantities we mentioned above, as a consequence) is of particular interest in real applications.

Birth-and-growth processes constitute the basis of a methodology to analyze transformation kinetics, which is often called "formal kinetic". Formal kinetics had its inception in the early work by [11], [9] and [2–4]. These papers were originally motivated by phase transformations, and considered that nucleation sites were located in space according to a homogeneous Poisson point process. They also considered that the velocity of the moving boundaries was constant in time and was the same at every point of the moving boundaries. Namely, in the site-saturated case, if the number of nuclei per unit of volume is N_V, then the volume fraction transformed, V_V, is given by

$$V_V(t) = 1 - e^{-\frac{4\pi}{3}N_V G^3 t^3}$$

whereas in the case of constant nucleation rate per unit of volume, I_V,

$$V_V(t) = 1 - e^{-\frac{\pi}{3}I_V G^3 t^4}.$$

Subsequent works generalized both the distribution of the nuclei in space and the time-dependence of the growth velocity; more general growth models admitting different velocities for different boundary points have been obtained by assuming space-and-time dependent velocity. In particular, by denoting Λ the intensity measure of the nucleation process, and by $\mathscr{C}(t,x)$ the *causal cone* of a point x at time t (i.e., the subset in which at least one nucleation event has to take place in order to cover the point x at time t [11]), we recall that (e.g., see [17]),

$$V_E(t,x) = \Lambda(\mathscr{C}(t,x)), \tag{1}$$

and that

$$G(t) = \frac{1}{S_V(t,x)}\frac{\partial V_V(t,x)}{\partial t} = \frac{1}{S_E(t,x)}\frac{\partial V_E(t,x)}{\partial t}; \tag{2}$$

finally, under Poissonian assumption on the nucleation process, it holds

$$V_V(t,x) = 1 - e^{-V_E(t,x)} \tag{3}$$

and

$$S_V(t,x) = (1 - V_V(t,x))S_E(t,x). \tag{4}$$

2 Random Growth Velocity of the Grains

In all the mentioned models the growth velocity field is assumed to be deterministic; such an assumption is possibly a good approximation for certain practical cases, whereas for others the boundary velocity may not reasonably be thought to be neither deterministic nor to be the same for each grain. We focus here on the effect that a random velocity of the moving boundaries of the grains has in the overall kinetics. One example of a practical situation in which such a model may be useful is that of recrystallization (e.g, a concrete case would be the nucleation and growth of ferrite from austenite in an iron-carbon alloy). A recent review of 3-D experimental results on recrystallization kinetics concluded that there is compelling evidence that "every single grain has its own kinetics different from the other grains"[8]. Nonetheless, in spite of this experimental evidence very few papers deal with this problem theoretically.

Motivated by this practical application we present here general expressions for the mean volume and surface densities of birth-and-growth models where a probability distribution of growth velocities of the grains is assumed, both in the case of site-saturation and in the case of time dependent nucleation [13,19]. Namely, we consider three different cases of interest:

1. the velocity G_i associated to the grain with nucleus located in X_i is constant during the reaction, but random;
2. the velocity G_i associated to the grain with nucleus located in X_i is random and time dependent, of the type

$$G(t) = G_0 g(t, \alpha), \tag{5}$$

   where G_0 is a non-negative random variable and g is a non-negative function depending on time and on a random vector parameter α in $\mathbb{R}^n$;
3. the velocity G_i associated to the grain with nucleus located in X_i is constant during the reaction, but random with probability distribution dependent on the specific location of the associated nucleus.

Note that the case 1 can be seen as a particular case of 3.

In all of the above mentioned cases, we assume that the nucleation process is an inhomogeneous Poisson point process. General results for the mean densities of the transformed region Θ^t are provided in [19], reobtaining the known above mentioned results (1)–(4) when the velocity is not random, as particular case.

2.1 Case G Random and Constant During the Reaction

The basic idea is to consider G_i, the velocity associated to the i-th nucleus with random location $X_i \in \mathbb{R}^d$, as a further mark associated to such a nucleus. With reference to the cases 1 and 3 above, let G_i a random variable with probability distribution Q, and with position-dependent probability distribution $Q(x,\cdot)$, respectively. Then:

- in the *site-saturated case*, the nucleation process $N = \{X_i, G_i\}$ has intensity measure Λ on $\mathbb{R}^d \times \mathbb{R}_+$ of the type

  $$\Lambda(\mathrm{d}(y,\xi)) = \begin{cases} \lambda(y)\mathrm{d}yQ(\mathrm{d}\xi) & \text{in case 1} \\ \lambda(y)\mathrm{d}yQ(y,\mathrm{d}\xi) & \text{in case 3} \end{cases},$$

  and so

  $$\Lambda(\mathscr{C}(t,x)) = \begin{cases} \int_{\mathbb{R}_+}\int_{B_{\xi t}(x)} \lambda(y)\mathrm{d}yQ(\mathrm{d}\xi) & \text{in case 1} \\ \int_{\mathbb{R}^d}\Big(\int_{\mathrm{dist}(y,x)/t}^{\infty} Q(y,\mathrm{d}\xi)\Big)\lambda(y)\mathrm{d}y & \text{in case 3} \end{cases}.$$

  Note that $\Lambda(\mathscr{C}(t,x)) = \lambda b_d t^d \mathbb{E}[G^d]$ in the case 1, whenever the nucleation process of the locations $\{X_i\}_i$ is stationary.
- In the *time-dependent nucleation case*, the nucleation process $N = \{T_i,(X_i,G_i)\}$ (where T_i is the birth-time of the nucleus located in X_i) has intensity measure Λ on $\mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}_+$ of the type

  $$\Lambda(\mathrm{d}s,\mathrm{d}(y,\xi)) = \begin{cases} \lambda(s)\mathrm{d}sQ_X(\mathrm{d}y)Q(\mathrm{d}\xi) & \text{in case 1} \\ \lambda(s)\mathrm{d}sQ(\mathrm{d}y,\mathrm{d}\xi) & \text{in case 3} \end{cases},$$

  where Q_X is the probability distribution of the random location X of the nuclei. It follows

  $$\Lambda(\mathscr{C}(t,x)) = \int_0^t \lambda(s)\Big(\int_{\mathbb{R}_+}\int_{B_{\xi(t-s)}(x)} Q(\mathrm{d}(y,\xi))\Big)\mathrm{d}s.$$

  Note that explicit expressions in particular cases of interests are easy to handle. In particular, we mention that in some applications it is of interest to evaluate the mean volume density in the centre of the specimen; in the particular case in which the nucleation is homogenous in time (i.e. $\lambda(s) \equiv \lambda > 0$), the nuclei are uniformly located in a compact window $[-M,M]^d$ and G is bounded, say $G \le K \in \mathbb{R}_+$, then

  $$\Lambda(\mathscr{C}(t,0)) = \frac{\lambda b_d t^{d+1}}{2^d M^d (d+1)}\mathbb{E}[G^d] \qquad \forall t \in [0, M/K].$$

  For a discussion of further examples, see [19].

2.2 Case G Random and Time-Dependent

We assume that each grain develops with random time-dependent velocity during the reaction, of the type given in (5). We further assume that G_0 and α are independent on the spatial location of the nucleus of the associated grain, with joint probability distribution $Q(\mathrm{d}(\xi, a))$ on $\mathbb{R}_+ \times \mathbb{R}^n$.
Even in this case, different grains may have different velocity, and we may model such a birth-and-growth process by a suitable marked Poisson point process N.

- in the *site-saturated case*, $N = \{X_i, (G_i, \alpha_i)\}_i$ is a marked point process in $\mathbb{R}^d$ with independent marking in $\mathbb{R}_+ \times \mathbb{R}^n$, with mark distribution Q. Then, its intensity measure Λ is of the type

$$\Lambda(\mathrm{d}(y, \xi, a)) = \lambda(y)\mathrm{d}yQ(\mathrm{d}(\xi, a)) \tag{6}$$

  while the transformed region Θ^t at time t is given by

$$\Theta^t = \bigcup_{(X_i,(G_i,\alpha_i))\in N} B_{R_i(t)}(X_i),$$

  with $R_i(t) := G_i \displaystyle\int_0^t g(\tau, \alpha_i)\mathrm{d}\tau.$
  It follows then

$$\Lambda(\mathscr{C}(t, x)) = \int_{\mathbb{R}_+\times\mathbb{R}^n} \int_{B_{R(t)}(x)} \lambda(y)\mathrm{d}yQ(\mathrm{d}(\xi, a)),$$

  Note that if λ is a non-negative harmonic function in the spatial region where the nucleation takes place, and if G_0 and α are independent with probability distribution Q_1 and Q_2, respectively, then the above equation simplifies as follows:

$$\Lambda(\mathscr{C}(t, x)) = \lambda(x)b_d\mathbb{E}[G_0^d]\mathbb{E}\Big[\Big(\int_0^t g(\tau, \alpha)\mathrm{d}\tau\Big)^d\Big].$$

  For further examples and particular cases see [19].
- In the *time-dependent nucleation case*, $N = \{(T_i, (X_i, G_i, \alpha_i))\}$ is a marked point process in $\mathbb{R}_+$ with marks in $\mathbb{R}^d \times \mathbb{R}_+ \times \mathbb{R}^n$, with intensity measure

$$\Lambda(\mathrm{d}(s, y, \xi, a)) = \lambda(s, y)\mathrm{d}s\mathrm{d}yQ(\mathrm{d}(\xi, a),$$

  while the transformed region Θ^t at time t is given by

$$\Theta^t = \bigcup_{(T_i,(X_i,G_i,\alpha_i))\in N:T_i\leq t} B_{R(T_i,t)}(X_i),$$

  with $R_i(s, T_i) := G_i \int_{T_i}^t g(\tau, \alpha_i)\mathrm{d}\tau.$

It follows then

$$\Lambda(\mathscr{C}(t,x)) = \int_0^t \Big(\int_{B_{R(s,t)}(x)\times\mathbb{R}_+\times\mathbb{R}^n} \lambda(s,y) Q(\mathrm{d}(y,\xi,a)) \Big) \mathrm{d}s.$$

If moreover $\lambda(s,\,\cdot\,)$ is harmonic for any $s \in \mathbb{R}_+$, then,

$$\Lambda(\mathscr{C}(t,x)) = \int_0^t \lambda(s,x) b_d \Big(\int_{\mathbb{R}_+\times\mathbb{R}^n} (R(s,t))^d \, Q(\mathrm{d}(\xi,a)) \Big) \mathrm{d}s.$$

We refer to [19] for further examples and particular cases (for instance, with $g(t,\alpha) = (1-\alpha)t^{-\alpha}$ and α having a Beta distribution as studied by Godiksen et al. in [7]).

2.3 Generalization of Eqs. (1)–(4)

In all the three mentioned cases 1–3, we can prove (see [19]) that:

- Equation (1) still holds, by a simple application of Campbell's formula in the definition of V_E.
- Equation (3) still holds under the assumption that the nucleation process is Poissonian.
- Equation (4) still holds under the assumption that the nucleation process is Poissonian, with intensity λ bounded and continuous.
- Equation (2) has to be regarded now in terms of an overall velocity $\mathscr{G}(t)$ defined as

$$\mathscr{G}(t) := \frac{1}{S_V(t,x)} \frac{\partial V_V(t,x)}{\partial t}.$$

In particular, in the above mentioned case 1, with $\mathbb{E}[G^d] < \infty$, if the process is site-saturated such that the intensity λ of the Poisson nucleation process is a harmonic function in the spatial region where the nucleation takes place (i.e., twice continuously differentiable and it satisfies the Laplace's equation $\sum_{i=1}^d \partial^2\lambda(x)/\partial x_i^2 = 0$), then

$$\frac{\mathbb{E}[G^d]}{\mathbb{E}[G^{d-1}]} = \frac{1}{S_V(t,x)} \frac{\partial V_V(t,x)}{\partial t} = \frac{1}{S_E(t,x)} \frac{\partial V_E(t,x)}{\partial t},$$

which generalizes Eq. (2).

(Explicit expressions for $\mathscr{G}(t)$ in particular cases of interest in applications are discussed in [13, 19].)

References

1. Aquilano, D., Capasso, V., Micheletti, A., Patti, S., Pizzocchero, L., Rubbo, M.A.: Birth and growth model for kinetic-driven crystallization processes. I. Modeling. Nonlinear Anal. Real World Appl. **10**, 71–92 (2009)
2. Avrami, M.J.: Kinetics of phase change I. General theory. J. Chem. Phys. **7**, 1103–1112 (1939)
3. Avrami, M.J.: Kinetics of phase change. II. Transformation-time relations for random distribution of nuclei. J. Chem. Phys. **8**, 212–224 (1940)
4. Avrami, M.J.: Granulation phase change, and microstructure kinetics of phase change. III. J. Chem. Phys. **9**, 177–184 (1941)
5. Capasso, V., Villa, E.: On mean densities of inhomogeneous geometric processes arising in material science and medicine. Image Anal. Stereol. **26**, 23–36 (2007)
6. Fanfoni, M., Tomellini, M.: Film growth viewed as stochastic dot processes. Phys.: Condens. Matter **17**, 571–605 (2005)
7. Godiksen, R.B., Schmidt, S., Jensen, D.J.: Effects of distributions of growth rates on recrystallization kinetics and microstructure. Scr. Mater. **57**, 345–348 (2007)
8. Jensen, D.J., Godiksen, R.: Neutron and synchrotron X-ray studies of recrystallization kinetics. Metall. Mater. Trans. A **39**, 3065–3069 (2008)
9. Johnson, W.A, Mehl, R.F.: Reaction kinetics in process of nucleation and growth. Trans. AIME **135**, 416–442 (1939)
10. Jun, S., Bechhoefer, J.: Nucleation and growth in one dimension. II. Application to DNA replication kinetics. Phys. Rev. E **71**, 11909 (2005)
11. Kolmogorov, A.N.: On the statistical theory of the crystallization of metals. Bull. Acad. Sci. USSR Math. Ser. **1**, 355–359 (1937)
12. Ramos, R.A., Rikvold, P.A., Novotny, M.A.: Test of the Kolmogorov-Johnson-Mehl-Avrami picture of metastable decay in a model with microscopic dynamics. Phys. Rev. B **59**, 9053 (1999)
13. Rios, P.R., Villa, E.: An analytical approach to the effect of a distribution of growth velocities on recrystallization kinetics. Scr. Mater. **65**, 938–941 (2011)
14. Starink, M.J.: Analysis of aluminium based alloys by calorimetry: quantitative analysis of reactions and reaction kinetics. Int. Mater. Rev. **49**, 191–226 (2004)
15. Stoyan, D., Kendall, W.S., Mecke, J.: Stochastic Geometry and Its Application. Wiley, New York (1995)
16. Tomellini, M., Fanfoni, M.: Impingement factor in the case of phase transformations governed by spatially correlated nucleation. Phys. Rev. B **78**, 14206 (2008)
17. Villa, E.: A note on mean volume and surface densities for a class of birth-and-growth stochastic processes. Int. J. Contemp. Math. Sci. **3**, 1141–1155 (2008)
18. Villa, E.: On the specific area of inhomogeneous Boolean models. Existence results and applications. Image Anal. Stereol. **29**, 111–119 (2010)
19. Villa, E., Rios, P.R.: On modelling recrystallization processes with random growth velocities of the grains in materials science. Image Anal. Stereol. **31**, 149–162 (2012)

A Mathematical Model for the Melting of Spherical Nanoparticles

Francesc Font, Tim G. Myers, and Michelle MacDevette

Abstract This paper will specifically deal with the melting process of gold nanoparticles. Based on scale analysis we first show that retaining previously neglected terms in the Gibbs–Thomson equation (describing the melt temperature as a function of size) can have a significant effect on results. Asymptotic and numerical results for the position of the melting front are presented for spherical nanoparticles. They appear to match well down to the final stages of melting.

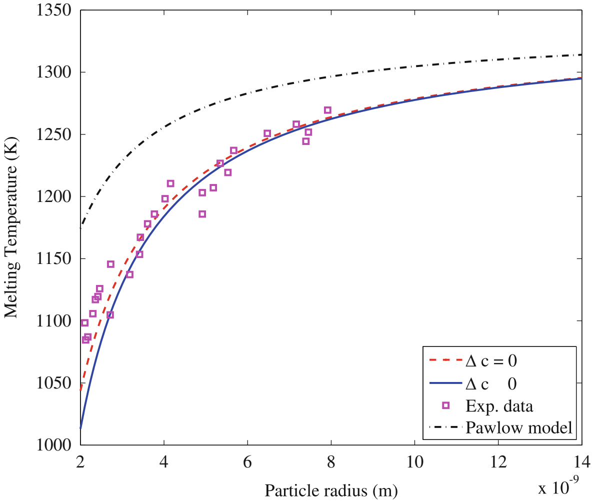
1 Introduction

The classical one-phase Stefan problem involves solving a single heat equation subject to constant temperature boundary conditions over a time-dependent domain whose extent is unknown "a priori". At the phase change boundary, $x = s(t)$, the temperature is fixed at the constant bulk phase change temperature $T(s(t), t) = T_m^*$. However, there are situations where the phase change temperature is also a variable. This is the case with the melting of nanoparticles. Nanoparticles are made up of bulk and surface atoms: the surface atoms are more weakly bound to the cluster than the bulk atoms and melting proceeds by the surface atoms separating from the bulk. Obviously this separation is paid for with energy (the latent heat). With a sufficiently large cluster the energy required is relatively constant since each surface molecule is affected by the same quantity of bulk molecules. However, as the cluster decreases in size the surface molecules feel less attraction to the bulk, consequently less energy is required for separation. This energy drop translates in a depression of the melting temperature that depends on the particle radius [2].

F. Font (✉) • T.G. Myers • M. MacDevette
Centre de Recerca Matemàtica, Campus de Bellaterra Edifici C, 08193 Barcelona, Spain
e-mail: ffont@crm.cat; tmyers@crm.cat; mmacdevette@crm.cat

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_41,

Fig. 1 Melting temperature as a function of particle size



Assuming that the density and specific heat remain constant in each phase the melt temperature, T_m, may be estimated from the following generalized Gibbs–Thomson relation

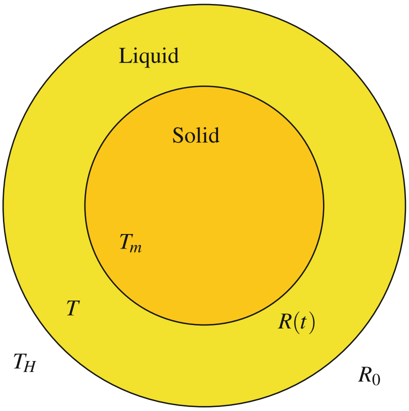
$$\left(\frac{1}{\rho_l} - \frac{1}{\rho_s}\right)(p_l - p_a) = L_m\left(\frac{T_m}{T_m^*} - 1\right) + \Delta c\left[T_m \ln\left(\frac{T_m}{T_m^*}\right) + T_m^* - T_m\right] + \frac{2\sigma_{sl}\kappa}{\rho_s} \tag{1}$$

where T_m^* is the bulk phase change temperature, $\Delta c = c_l - c_s$ the difference between specific heats, p the pressure (and p_a the ambient pressure), σ the surface tension and κ the mean curvature and s and l indicate solid and liquid components. A complete derivation of this equation from thermodynamical principles can be found in [1]. For a gold nanoparticle with radius 6 nm it has been found that $T_m \approx T_m^* - 100\,\mathrm{K}$. Taking $p_l - p_a = 10^5$ the term in the LHS of (1) is $\mathcal{O}(0.6)$ while the rest of the terms are $\mathcal{O}(10^2)$. Hence, we assume that the pressure term is negligible. In Fig. 1 we show experimental results for the melting temperature of gold nanoparticles. For this situation the term on the LHS of (1) is small. The solid line in Fig. 1 is the solution of (1) with the LHS set to zero, the dashed line represents the same but setting $\Delta c = 0$ and the dashed-dotted line represents the Pawlow model [1, 3].

2 Two-Phase Mathematical Model

The practical situation motivating the present study is the melting of gold nanoparticles, consequently the mathematical model is formulated as spherically symmetric. A typical configuration for the appropriate Stefan problem is shown in Fig. 2. This

Fig. 2 Picture of the model



depicts an initially solid, spherical nanoparticle which is heated at the boundary to a temperature $T_H > T_m^*$. The governing equations for the two-phase problem may be written as

$$c_l \rho_l \frac{\partial T}{\partial t} = k_l \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial T}{\partial r} \right), \qquad R < r < R_0 \tag{2}$$

$$c_s \rho_s \frac{\partial \theta}{\partial t} = k_s \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial \theta}{\partial r} \right), \qquad 0 < r < R \tag{3}$$

where T represents the temperature in the liquid, θ the temperature in the solid, $R = R(t)$ the moving boundary, R_0 the initial radius of the particle and k the thermal conductivity. These equations are subject to the following boundary conditions

$$T(R_0, t) = T_H \qquad T(R, t) = \theta(R, t) = T_m \qquad \theta_r(0, t) = 0 \tag{4}$$

and the Stefan condition

$$\rho_l \left[L_m + \Delta c (T_m - T_m^*) \right] \frac{dR}{dt} = k_s \frac{\partial \theta}{\partial r} - k_l \frac{\partial T}{\partial r} \bigg|_{r=R} \tag{5}$$

where T_m is solution of (1).

Introducing the dimensionless variables

$$\hat{T} = \frac{T - T_m^*}{T_H - T_m^*} \qquad \hat{\theta} = \frac{\theta - T_m^*}{T_H - T_m^*} \qquad \hat{r} = \frac{r}{R_0} \qquad \hat{R} = \frac{R}{R_0} \qquad \hat{t} = \frac{\alpha_l}{R_0^2} t \tag{6}$$

in (2)–(5) and dropping the hats the following nondimensional formulation is obtained

$$\frac{\partial T}{\partial t} = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial T}{\partial r}\right) \qquad \frac{\partial \theta}{\partial t} = \frac{k}{c}\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial \theta}{\partial r}\right) \tag{7}$$

with boundary conditions $T(R_0, t) = 1$, $T(R, t) = \theta(R, t) = T_m$, $\theta_r(0, t) = 0$ and the Stefan condition

$$[\beta + (1 - c)T_m]\, R_t = k\frac{\partial \theta}{\partial r} - \frac{\partial T}{\partial r}, \qquad r = R. \tag{8}$$

The nondimensional melting temperature T_m is determined from

$$0 = \beta\left(T_m + \frac{\Gamma}{R}\right) + \frac{(1-c)}{\delta T}\left[\left(T_m + \frac{1}{\delta T}\right)\ln\left(T_m\,\delta T + 1\right) - T_m\right]. \tag{9}$$

The dimensionless parameters above are defined by $\alpha_l = k_l/\rho_l c_l$, $c = c_s/c_l$, $k = k_s/k_l$, $\beta = L_m/c_l\Delta T$, $\delta T = \Delta T/T_m^*$ and $\Gamma = 2\sigma_{sl}T_m^*/R_0\rho L\Delta T$.

2.1 One-Phase Reduction

In order to reduce the complexity of (7)–(9) we reduce the problem to a one-phase system. To do so, we assume the solid to be initially at the melting temperature $T_m = T_m(0)$ given by the Gibbs–Thomson equation (9). If we assume $k/c \gg 1$ in (7b) at leading order we simply find $\theta = T_m(t)$. This permits us to eliminate the term $k\theta_r$ from (8) and the problem (7)–(8) reduces to

$$\frac{\partial T}{\partial t} = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial T}{\partial r}\right), \quad T(1, t) = 1, \quad T(R, t) = T_m \tag{10}$$

with

$$[\beta + (1 - c)T]\, R_t = -\frac{\partial T}{\partial r}, \qquad r = R. \tag{11}$$

So as to apply asymptotic techniques to the system (10)–(11) two further transformations are needed. First, by introducing a new function $T(r, t) = u(r, t)/r$ we transform (10) into a planar heat equation. Second, a boundary fixing transformation is employed

$$\xi = \frac{r - R}{1 - R} \qquad \tau = 1 - R \tag{12}$$

where ξ represents the new space variable and τ the new time variable, so $u = u(\xi, \tau)$. In this new framework the variable region occupied by the liquid is fixed at the unit domain $0 < \xi < 1$. Therefore, the one-phase Stefan problem that we shall be concerned with is stated as follows,

$$u_{\xi\xi} = -\tau_t \tau \left[(\xi - 1)u_\xi - \tau u_\tau\right], \qquad 0 < \xi < 1 \tag{13}$$

$$u(1, \tau) = 1 \tag{14}$$

$$-\tau\tau_t = \frac{1}{\beta}\left[\frac{\tau u}{(1-\tau)} - u_\xi\right]\left[(1-\tau) + \frac{(1-c)u}{\beta}\right]^{-1}, \qquad \xi = 0 \tag{15}$$

and

$$\beta\,(u + \Gamma) + \frac{(1-c)}{\delta T}\left[\left(u + \frac{1-\tau}{\delta T}\right)\ln\left(\frac{u}{1-\tau}\,\delta T + 1\right) - u\right] = 0, \qquad \xi = 0. \tag{16}$$

2.2 Asymptotic Analysis for Large Stefan Number

In this section we will seek series solutions for large Stefan number of the form $u \approx u_0 + \epsilon u_1 \ldots$, where $\epsilon = 1/\beta \ll 1$. Then, the leading order problem is

$$u_{0\xi\xi} = 0 \qquad u_0(1, \tau) = 1 \qquad u_0(0, \tau) = 1 - A \tag{17}$$

with solution

$$u_0 = 1 + A(\xi - 1) \tag{18}$$

where $A = A(\tau)$ is the solution of

$$\beta\,(1 - A + \Gamma) + \frac{(1-c)}{\delta T}\left[\left(1 - A + \frac{1-\tau}{\delta T}\right)\ln\left(\frac{1-A}{1-\tau}\,\delta T + 1\right) - 1 + A\right] = 0. \tag{19}$$

Substituting (18) and (15) into (13) we obtain the $O(\epsilon)$ problem

$$u_{1\xi\xi} = f(A, \tau)(\xi - 1) \qquad u_1(0, \tau) = u_1(1, \tau) = 0 \tag{20}$$

with

$$f = \frac{\beta(\tau - A)(A - \tau A_\tau)}{(1-\tau)\left[\beta(1-\tau) + (1-c)(1-A)\right]}, \tag{21}$$

$$A_\tau = \frac{(1-c)\left[\delta T(1-A) - (1-\tau)\ln\left(\frac{1-A}{1-\tau}\delta T + 1\right)\right]}{\delta T(1-\tau)\left[\delta T\beta + (1-c)\ln\left(\frac{1-A}{1-\tau}\delta T + 1\right)\right]} \tag{22}$$

where A_τ has been found by taking the τ derivative of (19). So, the solution of (20) is

$$u_1 = f(A,\tau)\left(\frac{\xi^3}{6} - \frac{\xi^2}{2} + \frac{\xi}{3}\right) \tag{23}$$

and finally

$$u = 1 + A(\xi - 1) + \epsilon f(A,\tau)\left(\frac{\xi^3}{6} - \frac{\xi^2}{2} + \frac{\xi}{3}\right) + \mathscr{O}(\epsilon^2). \tag{24}$$

Then, replacing $u \approx u_0 + \epsilon u_1$ in (15) leads to the following system of ODEs

$$\tau_t = \epsilon \frac{-3(\tau - A) + \epsilon(1-\tau)f}{3\tau(1-\tau)\left[(1-\tau) + \epsilon(1-c)(1-A)\right]} \tag{25}$$

$$A_t = A_\tau \tau_t \tag{26}$$

that can be solved by means of the Matlab routine `ode15s`.

3 Discussion

The plots in Fig. 3a, b show the evolution of the melting front $R(t)$ ($R = 1 - \tau$) for two different values of the Stefan number $\beta = 147$ and $\beta = 12$. Curve (i) corresponds to the solution of Eqs. (13)–(16), curve (ii) corresponds to the result of the system if we assume $c_l = c_s$ (hence $c = 1$) and curve (iii) is the result of considering $c = 1$ and neglecting the surface effects on (16) by setting $\Gamma = 0$ (hence $T_m^* = T_m$ is constant). The solid lines represent the asymptotic solutions and the dashed lines the numerical results by finite differences.

For large Stefan number it is clear that the asymptotics and numerics agree well. However, in Fig. 3 we see that as β decreases the asymptotics lose accuracy as $R \to 0$. The solution with the full expression (1) to determine T_m shows that as $R \to 0$, $R_t \to \infty$. This abrupt melting has been observed experimentally [4]. Neglecting the variation in specific heat leads to slightly slower melting whereas the standard Stefan formulation is clearly inappropriate for melting at the nanoscale.

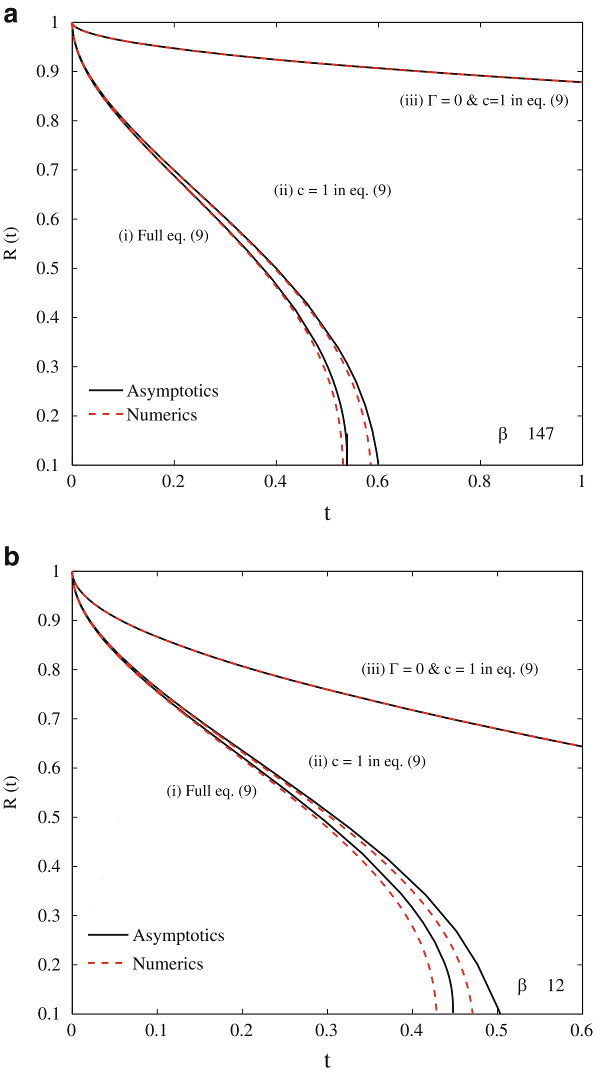


Fig. 3 Solution of the system (13)–(16) at three levels of approximation of Eq. (1) for different β. (**a**) $\beta \approx 147$. (**b**) $\beta \approx 12$

References

1. Alexiades, V., Solomon, A.D.: Mathematical Modelling of Freezing and Melting Processes. Hemisphere Publishing Corporation, Washington (1993)
2. Buffat, P., Borel, J.P.: Size effect on the melting temperature of gold particles. Phys. Rev. A **13**(6), 2287–2298 (1976)
3. Kofman, R., et al.: Surface melting enhanced by curvature effects. Surf. Sci. **303**, 231–246 (1994)
4. Kofman, R., Cheyssac, P., Lereah, Y., Stella, A.: Melting of clusters approaching 0d. Eur. Phys. J. D **106**, 441–444 (1999)

Local Quantum-Like Updates in Classical Molecular Simulation Realized Within an Uncoupling-Coupling Approach

Konstantin Fackeldey and Alexander Bujotzek

Abstract In this article a method to improve the precision of the classical molecular dynamics force field by solving an approximation problem with scattered quantum mechanical data is presented. This novel technique is based on two steps. In the first step a partition of unity scheme is used for partitioning the state space by meshfree basis functions. As a consequence the potential can be localized for each basis function. In a second step, for one state in each meshfree basis function, the precise QM-based charges are computed. These local QM-based charges are then used, to optimize the local potential function. The performance of this method is shown for the alanine tripeptide.

1 Introduction

When simulating molecular systems, we are interested in statistical ensembles of conformational states. In order to obtain observables from molecular simulation one has to compute high-dimensional integrals, i.e., expectation values over these ensembles. Closely related to the complexity of the computation of these expectation values is the choice of the molecular model, which represents the interactions between the atoms.

In quantum mechanics the dynamics of the particles is described by the Schrödinger equation, which provides probabilistic information about the position and impulses of the particles. Unfortunately, the Schrödinger equation is very complex, which permits long simulations of a large number of particles. Thus the Schrödinger equation is simplified by exploiting that the mass of an electron is much smaller than the mass of a nucleus. This allows to consider two coupled equations—one

K. Fackeldey (✉) • A. Bujotzek
Zuse Institute Berlin (ZIB), Takustr. 7, 14109 Berlin, Germany
e-mail: fackeldey@zib.de

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*, Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_42,

for the electron and one for the nuclei, instead of the original Schrödinger equation, which describes both states (electrons and nuclei). However, the resulting energy landscapes are far beyond the capabilities of quantum mechanical calculations.

Classical molecular simulations, in turn, are more tractable by computationally methods since there, the potential energy, which results from the position of configuration of the molecule, is averaged. In particular, the charge interactions are averaged over the whole conformation space, i.e. one average charge distribution for all possible conformations of the protein instead of computing them for each nucleus. As an example we consider ethane (C_2H_6) which has $N = 2 + 8 = 10$ nuclei and $K = 2 \cdot 6 + 1 \cdot 6 = 18$. Of course the smaller dimension of the configuration space has to be paid with less accuracy.

More precisely, force fields are in general empirical and thus one obtains different results when using different force fields for even the same molecule in the same setting [7].

Summing up, in the modeling of molecular systems we have a hierarchy, which leads to a trade off between computational complexity and precision.

2 Quantum-Like Charge Refinement

We now proceed a further step towards coupling quantum mechanical precision with more efficient classical simulations by using a local, "quantum-like" refinement of the partial charges in a classical molecular simulation. Typically, when initiating a classical molecular simulation, a partial charge is assigned to every particle in the system under observation. Partial charges per se are a rough approximation of quantum-mechanical electron density distributions. Due to the fact that the initial assignment of partial charges is assumed to remain invariant during the course of the simulation, the results are bound to become increasingly inaccurate: The more the molecular system departs from the initial configuration for which the charges have been calculated (typically a local energy minimum), the more inaccurate the results will become. As a consequence, now that modern parallel computing facilities enable us to calculate classical trajectories of unprecedented length (a fact that in parts lessens the sampling problem associated with molecular simulation), we are running the risk of producing dubious results due to an increasing error in the force field. The magnitude of this charge-induced error is depending on the (chemical) nature of the system under observation, and will affect some systems more than others. In order to address this problem, we determine accurate partial charges (i.e. partial charges calculated from QM-based methods such as AM1-BCC [4, 5]) for multiple configurations of the molecule, and, accordingly, perform a local update of the classical force field prior to simulation. The actual simulation remains a purely classical one, i.e. it is not a true hybrid simulation scheme such as QMMD (see [6] for well written overview). In order to ensure the validity of the "locally refined" charge, the simulation has to be restrained within the region of conformational space for which the charge has been calculated. This notion is best realized within in uncoupling-coupling sampling scheme such as ZIBgridfree [2] (https://github.com/CMD-at-ZIB/ZIBMolPy). The uncoupling-coupling sampling approach has been developed in order to address the trapping problem inherent to molecular systems: A conventional (single) simulation trajectory is prone to become "trapped" in a single energy minimum for a long time, a phenomenon that is likely to render the sampling process inefficient. In ZIBgridfree, the conformational space Ω is partitioned into fuzzy sets by using a meshless approach based on basis functions defined according to Shepard's method [8]. The partitioning of Ω is also denoted as "decoupling" step. The decoupled partial densities are sampled separately. Due to the fact that in each separate partition of Ω the sampling is confined to a comparatively small region, convergence according to a given criterion (e.g. the Gelman-Rubin convergence criterion [3]) can be achieved within a maintainable time. The converged partial densities obtained from the decoupled samplings are weighted and rejoined in order to yield the overall Boltzmann distribution. This (final) step is denoted as "coupling". The ZIBgridfree algorithm is outlined in the following: The final step identifies the metastable states of the system, which is important for interpreting its chemical properties, e.g. calculating the transition rates between different molecular conformations, or determining binding paths in a ligand-receptor complex [1]. In the context of local charge refinement, we now have the following advantage of a modified potential. Let us denote the nodes of the ith basis function by k_i, then, associated with each basis function $\varphi_i(q)$, $q \in \Omega$, i.e.

$$\varphi_i(q) = \frac{\exp(-\alpha \|q - k_i\|)}{\sum_{j=1}^{n} \exp(-\alpha \|q - k_j\|)},$$

where α is a parameter used to adjust the softness of the partitioning, comes a modified potential function U_i:

$$U_i(q) = \underbrace{U(q)}_{\text{global potential}} - \underbrace{\beta^{-1} \ln(\varphi_i(q))}_{\text{local restraint}}. \qquad (1)$$

In practice, the modified potential U_i is used to sample the partial density ρ_i associated with each φ_i. The potential modification $-\beta^{-1} \ln(\varphi_i(q))$ (softly) restricts the sampling to the region of Ω that is encompassed by basis function φ_i. The more the sampling departs from its node k_i, the more the restraining potential will "push" it back to its original support. This not only assures thorough sampling of the partial density ρ_i, but also opens up the possibility for local optimization of the force field: For each node k_i, one can now calculate precise QM-based charges that enter into an optimized local potential function U_i^{opt}. In summary, the optimized potential $\tilde{U}_i$ used to sample each basis function φ is the following:

$$\tilde{U}_i(q) = U_i^{\text{opt}}(q) - \beta^{-1} \ln(\varphi_i(q)). \qquad (2)$$

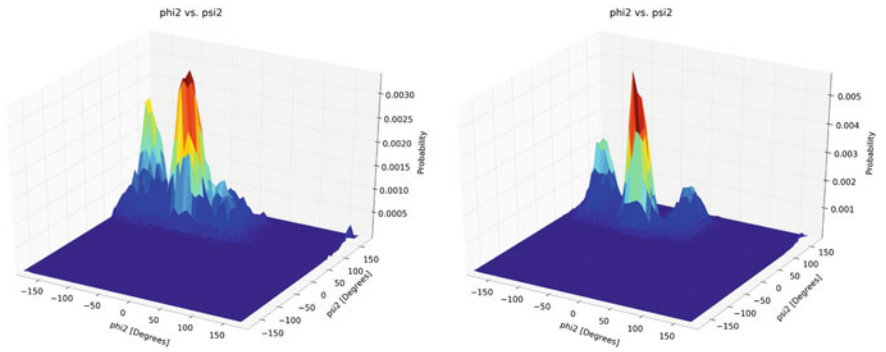


Fig. 1 Histogram of alanine tripeptide in vacuum. *Left*: Applying classical charges, by using the same charge for each basis function. *Right*: Charges were computed for each basis function individually by the local quantum update method

This notable increase in precision comes at the relatively low computational cost of one QM-based charge calculation per discretization node. The number of discretization nodes, in turn, is dependent on the size of Ω.

3 Illustrative Example

As an example we simulated alanine tripeptide in vacuum and used the Amber99sb force field. For the presampling, we calculated a trajectory for 100 ns at 300 K. The conformation space was partitioned into 12 nodes, where to each node a meshfree basis functions is attached. Within each basis function, we started a local trajectory (300 K) for 100 ps. In Fig. 1, the histogram is given by counting the absolute number of the dihedral angles in a certain range. One can clearly see, that the two histograms differ. On the left hand side, the classical scheme, by using one charge calculation for all basis functions, whereas on the left hand side the locally updated charges have been used. The local update of the charges leads to a stronger separation of the clusters. A comparison with quantum mechanical methods will be undertaken in future.

References

1. Bujotzek, A., Weber, M.: Efficient simulation of ligand-receptor binding processes using the conformation dynamics approach. J. Bioinform. Comput. Biol. **7**(5), 811 (2009)
2. Bujotzek, A., Schütt, O., Nielsen, A., Fackeldey, K., Weber, M.: Zibgridfree: efficient conformational analysis by partition-of-unity coupling. J. Math. Chem. **52**, 781–804 (2014)
3. Gelman, A., Rubin, D.: Inference from iterative simulation using multiple sequences. Stat. Sci. **7**(4), 457–472 (1992)

4. Jakalian, A., Bush, B.L., Jack, D.B., Bayly, C.I.: Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. J. Comput. Chem. **21**, 132–146 (2000)
5. Jakalian, A., Jack, D.B., Bayly, C.I.: Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. J. Comput. Chem. **23**, 1623–1641 (2002)
6. Mo, Y., Alhambra, C., Gao, J.: Recent development and applications of combined qm/mm methods. Acta Chim. Sin. **58**, 1504–1510 (2000)
7. Mu, Y., Kosov, D.S., Stock, G.: Conformational dynamics of trialanine in water. 2. Comparison of amber, charmm, gromos, and opls force fields to nmr and infrared experiments. J. Phys. Chem. B **107**(21), 5064–5073 (2003)
8. Shepard, D.: A two-dimensional interpolation function for irregularly spaced data. In: Proceeding of the 1968 23rd ACM National Conference (1968)

Design of Automatic Eye Protective Welding Devices

Matej Bazec, Bernarda Urankar, and Janez Pirs

Abstract LCD light shutters used as eye protective devices in welding environments have different requirements from LCD display devices typically used in consumer electronics. Their contrast must be many scales greater, while in the open state the shutter should be brighter. The light scattering should almost vanish and the switching time should be much shorter. This implies a different approach as typical solutions used in LCD displays don't meet the required criteria. Although there are many solutions and concepts that are shared between both types of devices, the light shutters have a much different design. In order to find an optimal configuration many cells should be built and tested. However, this is a very time consuming task as there are many parameters that should be taken in consideration and each cell may take few days to build and test. This is where the computer simulation steps into. It takes only a few minutes to build an appropriate setup for a particular cell and to simulate it, leaving to the experimental tests only some fine tuning. Furthermore, the simulations give a deeper insight in what is really happening with the light polarization within the cell. Such a way a better understanding can be achieved. The simulator works with two different approaches. In the first approach it tries to give the best possible and exact numerical solution. It does so by minimizing the Frank elastic energy of a particular LC layer and by solving the Maxwell equations for the complete stack of optical elements. For the latter the Berreman method is used reducing a system of partial differential equations to 4 × 4 matrices manipulation

M. Bazec (✉)
Maritime and Transportation Department, University of Ljubljana, Pot pomorscakov 4,
6320 Portoroz, Slovenia
e-mail: matej.bazec@fpp.uni-lj.si

B. Urankar • J. Pirs
Jozef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia
e-mail: bernarda.urankar@ijs.si; janez.pirs@ijs.si

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_43,

(multiplication, inversion and eigenvalue problem). Although such an approach is very accurate and mimics the reality quite well it doesn't give a deeper insight in the cell functionality. In such cases it is better to reduce the LC layer to a few simple uniaxial layers and follow the light polarization change by means of the transformations on the Poincar sphere. This makes it a very efficient tool in shutter design.

1 Background

LCD light shutters used as an eye protection device (welding shield) or in a stereovision application should have much different properties from typical LC display devices due to different technical requirements [1] in their field of use. As such they need to have:

- many scales larger and angularly independent light attenuations,
- higher switching speeds,
- low light distortion,
- low scattering and
- bright open state.

Such requirements lead solutions with much simpler configuration (TN an LTN LCD light shutters [2], Pi-cell [3], STN [4], etc.) in comparison with those popular used for the display devices [5–7].

This inevitably imposes the production of many new cells in the process of development of better shutters in order to test new configurations. However such an experimental approach is very time consuming. A rule of thumb is that a production of a single cell takes a day to produce and test. As such the development of new light shutter concepts would be practical impossible without the aid of computational tools.

It should be stressed out that despite the simplicity of the cell configuration both the liquid crystal director and the light propagation are complex enough they cannot be evaluated analytically. This implies two different but complementary approaches.

The first approach aims at finding a mathematical model that would reproduce the results of an equivalent experimental test as close as possible. In order to solve such a model various numerical tools are used as the equations involved are typically nonlinear.

On the other hand sometimes a better understanding of the ongoing process is wanted. In this case we don't care about the precision of the result. This means some simplifications can be introduced as long as the result is qualitatively similar. In such a way we can eliminate the unnecessary effects and focus our attention on the relevant ones.

2 Model Description

2.1 Light Shutter Simulation

The simulation of the light shutter can be roughly divided in two separate steps: nematic director and light propagation calculation. Sometimes the second step can be omitted (e.g. when only switching times are important).

2.1.1 Director Calculation

The liquid crystals used in light shutters are typically in the nematic phase put between two alignment layers that are displaced few µm apart and mixed with low concentrations of some chiral dopant. As low scattering is required the liquid crystal is defectless. This is usually not difficult to achieve as the LC layer is very thin and the alignment layers are uniform. In these circumstances the liquid crystal configuration can be described only with the nematic director **n** that is a unitary vector instead of the second order tensor order parameter.

The free energy of the nematic can be described in terms of the Frank-Oseen energy [8]:

$$f = \frac{1}{2}K_{11}(\nabla \cdot \mathbf{n})^2 + \frac{1}{2}K_{22}((\nabla \times \mathbf{n})\,\mathbf{n} + q)^2 + \frac{1}{2}K_{33}((\nabla \times \mathbf{n}) \times \mathbf{n})^2 - \frac{1}{2}\epsilon_0 \Delta\epsilon \frac{\epsilon_{\|}}{\epsilon_{\perp}}(\mathbf{nE})^2, \tag{1}$$

where K are the nematic elastic constants for splay, twist and bend, q is the pitch of the chirality, $\epsilon_{\|}$ and $\epsilon_{\perp}$ are the eigenvalues of the dielectric tensor of the liquid crystal and **E** is the applied electrical field. It should be stressed out that such a free energy is used when the constant voltage is applied. In other circumstances a different thermodynamic potential may be needed.

The director will always choose such a configuration that will minimize the free energy. In some cases (e.g. when optimal angular compensation is needed) knowing the minimum suffices. On the other hand, in some cases the time evolution from one state to some another is needed. In some way the time should be introduced. This can be done with the equation of nematodynamics [8]:

$$\left(\frac{\mathrm{d}}{\mathrm{d}z}\frac{\partial f}{\partial \mathbf{n}'} - \frac{\partial f}{\partial \mathbf{n}}\right)_{\perp} = -\gamma \frac{\partial \mathbf{n}}{\partial t}, \tag{2}$$

where γ is the nematic viscosity.

This equation seems numerically unstable if used with an explicit (forward in time) integration scheme. Indeed if used directly is quite unstable. However if each integration discretization step is followed by the normalization of the director, which should stay normalized anyway due to the time derivative being perpendicular to the director, the procedure surprisingly gets very stable. This both simplifies the algorithm structure and decreases computation time in comparison with implicit schemes.

2.1.2 Light Propagation Calculation

The director configuration by itself is useful to optimize for the fast switching time and similar things. However for most problems also the underlying optics should be calculated. For this task a procedure first used by Berreman [9] is used.

It takes the advantage of the planar geometry of the cell. The normal of the cell is put on the z axis so that the material parameters depend on z only. Further we can get rid of the dependency on the variable y if the x axis is defined such as the light comes in the $x - z$ plane. Such a way the Maxwell's equations could be simplified by the following ansatz:

$$\mathbf{E}(x, y, z, t) = \mathbf{E}(z)e^{i(\xi x - \omega t)}. \tag{3}$$

The vector $\mathbf{E}$ could be replaced with $\mathbf{D}$, $\mathbf{B}$ and $\mathbf{H}$.

This reduces the Maxwell's equations to a system of four ordinary differential equations of four independent quantities, the other eight being related with the former four only through the linear algebraic equations. A wise choice would be to choose H_x, H_y, E_x and E_y as independent quantities as their values do not change at the boundaries of the two layers in contact (in the absence of the superficial charges and currents). Those quantities define the vector $\boldsymbol{\psi} = (E_x, \mu_0 c_0 H_y, E_y, -\mu_0 c_0 H_x)$ (in SI units). In this compact form the four equations can be written as

$$\frac{\mathrm{d}\boldsymbol{\psi}}{\mathrm{d}z} = i\frac{\omega}{c_0}\Delta\boldsymbol{\psi}, \tag{4}$$

where

$$\Delta = \begin{bmatrix} -S\frac{\epsilon_{xz}}{\epsilon_{zz}} & 1 - S^2\frac{1}{\epsilon_{zz}} & -S\frac{\epsilon_{yz}}{\epsilon_{zz}} & 0 \\ \epsilon_{xx} - \frac{\epsilon_{xz}^2}{\epsilon_{zz}} & -S\frac{\epsilon_{xz}}{\epsilon_{zz}} & \epsilon_{xy} - \frac{\epsilon_{xz}\epsilon_{yz}}{\epsilon_{zz}} & 0 \\ 0 & 0 & 0 & 1 \\ \epsilon_{xy} - \frac{\epsilon_{xz}\epsilon_{yz}}{\epsilon_{zz}} & -S\frac{\epsilon_{yz}}{\epsilon_{zz}} & -S^2 + \epsilon_{yy} - \frac{\epsilon_{yz}^2}{\epsilon_{zz}} & 0 \end{bmatrix} \tag{5}$$

and S is the Snell's coefficient (the sine of the incident angle in vacuum).

If the layer is homogeneous, then Δ is constant and can be diagonalized. Then the vector at the incoming side $\boldsymbol{\psi}_I$ and on the outgoing side $\boldsymbol{\psi}_F$ are related through a simple linear transformation

$$\boldsymbol{\psi}_F = P\boldsymbol{\psi}_I, \tag{6}$$

where P is the transition matrix and can be calculated from Δ and the layer thickness.

If the layer is not homogeneous then it can be cut in thin slices each of them being so thin that can be considered homogeneous. As $\boldsymbol{\psi}$ does not change on the boundaries of the neighbor layers, the total transition matrix P can be calculated as a product of transition matrices of all the thin slices.

Finally, such a $\boldsymbol{\psi}_I$ should be found, that $\boldsymbol{\psi}_F$ does not have any component that represents the two polarization coming in the cell from the out coming side. This could be done by few simple steps involving only basic linear algebra manipulation. Such a way both the transmitted and reflected light can be calculated.

Although the procedure is very simple it can be numerical unstable when high attenuations are involved. This introduces some extra tricks that reduces the numerical error to a useful level. This is mostly done by filtering out the noise either by the closest angles or by the closest frequencies.

2.2 Light Propagation Analysis

Although the simulations yield results that are very close to the experimental measurements, they do not provide a deeper insight in the mechanism of compensation itself. This is where analytical approximation step in. A very useful tool is the representation of the light polarization with the four Stokes parameters. Further it can be assumed that:

- The light is fully polarized once passed the incoming polarizer.
- The reflected light is low compared to the transmitted.
- The direction of the light propagation does not depend on the polarization of the light.
- Between the polarizers the light is not absorbed (the intensity) ($n_e \approx n_o$).

In such a case the Stokes parameter representing intensity is constant and the other three lay on a sphere also called Poincar sphere. The linear polarizations lay on the equator while the circular polarizations lay on the poles. The propagation of the light through a homogeneous layer produces the rotation of the polarization point on the sphere where the angle depends linearly on the layer thickness and the axis lays on the equator.

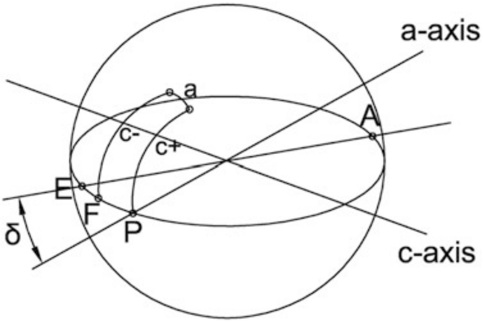


Fig. 1 Compensation layer improving the angular dependence. P is the polarization of the polarize and is the initial polarization. A is the polarization of the analyzer and E is the vanishing point laying at the opposite side. F is the final polarization

Although the transformations involved in a single step can be simple it can get complicated when many layers are involved. This is why the liquid crystal layer should be simplified with a consistent model.

Since the voltages involved are typically far above the Fredericks transition, the director is mainly homeotropically aligned. Only a small fraction close to the alignment layers can be considered nonhomeotropic. This leads us to introduce the three layer model, where the central part is replaced by a thick homogeneous and homeotropic layer and two thin homogeneous and planarly oriented layers [10]. More sophisticated models include five instead of three layers.

A very useful application of this method was used to describe the optimal position of the negative retardation layer in 180° twisted cell [10]. When put on the incoming polarize side it moves the polarization far from the vanishing point, while on the outgoing side it moves close to it improving the angular dependence (see Fig. 1).

3 Results

As mentioned in the previous section using Poincar sphere analysis for the cell design and numerical simulations for the parameters definition were used to achieve a very uniform light shutter using only a single 180° twisted cell with a negative retardation layer [4, 10]. Unfortunately the parameters could only be optimized for a single light attenuation and the cell should be driven at slightly higher voltages than those typically used in such devices.

Recently an improved variable element was developed using only a stack of two STN cells with an additional negative compensation layer [11]. It significantly improves the optical properties of the similar cells (e.g. [2]). It has an almost linear dependence of the shade on the applied voltage and meets the requirements imposed by the standard [1]. As can be seen from Fig. 2 the shade variation is far less than 1 in each direction at the incident angle 15° in a very broad range of shades (from shade 9 to shade 13).

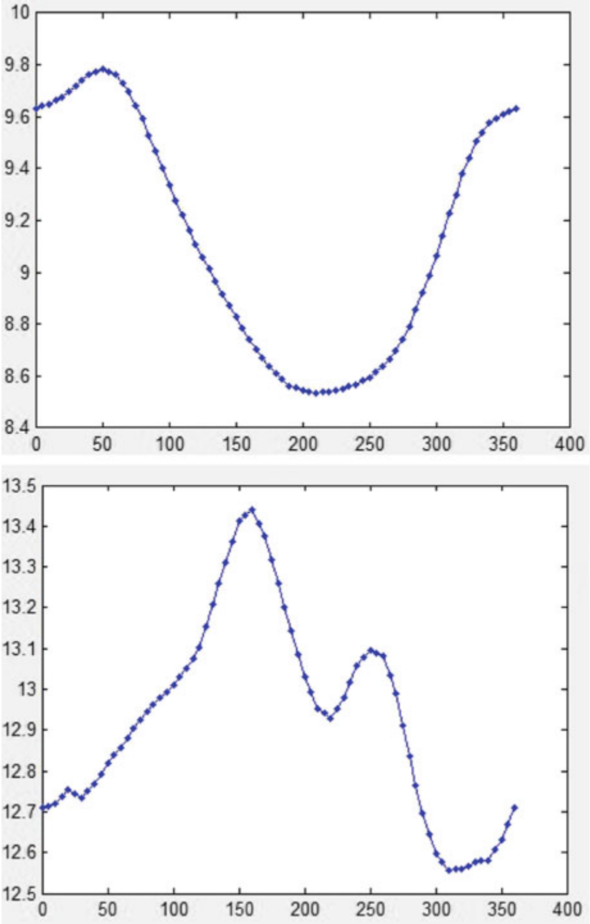


Fig. 2 Azimuthal angular dependence at polar angle 15° and two different driving voltages. At the lower voltage the light attenuation corresponds to shade 9 and at the higher voltage to 13. The variation is less than 1 in each direction

Acknowledgements The authors would like to thank Phil Bos for contributing the source code on which the simulator was built.

References

1. Standard EN 379 (2003)
2. Palmer, S.: Liquid-crystal cell with a wide viewing angle and high cell contrast. Appl. Opt. **36**, 2094 (1997)
3. Bos, P.J.: High contrast light shutter system. US patent US5187603, 1993
4. Pirs, J., Bazec, M., Pirs, S., Marin, B., Vrecko, A.: High contrast, wide viewing angle lcd light-switching element. European patent EP1625445, 2006
5. Mori, H.: Liquid crystal display with optical compensatory sheet having discotic molecules varyingly inclined. US patent US5583679, 1996

6. Clerc, J.F.: Vertically aligned lcd. In: SID Digest, pp. 758–761 (1991)
7. Oh-e, M., Kondo, K.: Response mechanism of nematic liquid crystals using the in-plane switching mode. Appl. Phys. Lett. **69**, 623 (1996)
8. de Gennes, P.G., Prost, J.: The Physics of Liquid Crystals. Oxford University Press, Oxford (1995)
9. Berreman, D.W.: Optics in stratified and anisotropic media: 4 × 4-matrix formulation. Appl. Opt. **36**, 2094 (1997)
10. Vrecko, A., Pirs, J., Bazec, M., Ponikvar, D.: Wide view stn liquid crystal light shutter. Appl. Opt. **47**, 2623–2629 (2008)
11. Pirs, J., Bazec, M., Pirs, S., Marin, B., Urankar, B., Ponikvar, D.: Variable contrast, wide viewing angle lcd light-switching filter. US Patent US8542334, 2013

A Three-Segment Inverse Method for the Design of CoA Correcting TIR Collimators

Corien Prins, Jan ten Thije Boonkkamp, Teus Tukker, and Wilbert IJzerman

Abstract Color-over-Angle (CoA) variation in the light output of white phosphor-converted LEDs is a common and unsolved problem. Recently, the same authors introduced a new inverse method to reduce CoA variation using a special collimator. This short paper introduces a variant of the method with two important advantages compared to the original method.

1 Introduction

White LED technology becomes increasingly important in lighting. White LEDs are starting to replace traditional light sources such as compact fluorescent lamps and halogen spots. LED-based spotlights are already widely available in retailer shops. These spotlights usually contain a highly efficient TIR collimator to direct the light into a compact beam.

Unfortunately, it is difficult to create an LED that emits light with a uniform white color. The color of the emitted light varies with the angle between the light ray and surface normal of the LED. This phenomenon is called Color-over-Angle (CoA) variation. When the light of an LED with a large CoA variation is collimated using a TIR collimator, this color variation appears in the beam. Various methods have been

C. Prins (✉) • J. ten Thije Boonkkamp
Centre for Analysis, Scientific Computing and Applications (CASA), TU/e, Eindhoven, The Netherlands
e-mail: c.r.prins@tue.nl; tenthije@win.tue.nl

T. Tukker
Philips Research, Eindhoven, The Netherlands
e-mail: teus.tukker@philips.com

W. IJzerman
Philips Lighting, Eindhoven, The Netherlands
e-mail: wilbert.ijzerman@philips.com

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_44,

applied to reduce this CoA variation with different advantages and disadvantages. A commonly applied technique is the introduction of bubbles in the phosphor layer of the LED [6]. Another technique is applying a dichroic coating [1]. Both methods reduce the efficiency and increase the production costs of the LED. If the LED is used in combination with a collimating optic, CoA variation can be reduced by using microstructures on top of the collimator. However, microstructures introduce extra costs in the production process of the collimator and make the collimator look unattractive and broaden the light beam. Wang et al. [3] study the reduction of CoA variation using domes which mix light from two different angles. They note that it is theoretically possible to completely remove the CoA variation, but they do not show a proof. In [2] we showed it is indeed possible to completely remove the CoA variation by mixing light from only two different angles. The current paper introduces a variant of the method introduced in [2]. It has two advantages compared to the original method. First, it can be used to design standard type TIR collimators with three transfer functions. Second, it gives the optical designer more design freedom, as the angular width of the refractive part can be chosen freely.

2 A Color Weighted TIR Collimator

A TIR (Total Internal Reflection) collimator is a rotationally symmetric lens, that is used to collimate the light of an LED into a compact beam. TIR collimators are usually made of a transparent plastic like polycarbonate (PC) or polymethyl methacrylate (PMMA). A profile of a TIR collimator can be seen in Fig. 1. A TIR collimator for a point light source can be designed using inverse methods. The design procedure consists of two steps: first we find the relation between the angles t and θ, where t is the angle between the z-axis and a ray leaving the light source and θ is the angle between the z-axis and a ray leaving the collimator. This relation is described by the so-called transfer functions. Subsequently we use these transfer functions to calculate the free surfaces of the collimator. The second step is described extensively in [2] and will not be covered in this article. This article is concerned with the first step of the design process.

A transfer function $\eta : \Theta \to T$ is a monotone function that describes the relation between the angle $t \in T \subset [0, \pi/2]$ of the light emitted from the light source and the angle $\theta \in \Theta \subset [0, \theta_{\text{max}}]$ of the light emitted from the TIR collimator. Here θ_{max} is the maximum angle of emission. It can be seen that there are three different rays width different angles t for every angle $\theta \in [0, \theta_{\text{max}}]$: one ray is refracted at surface A, one is reflected at surface B and crosses the z-axis, and one is refracted at surface C and does not cross the z-axis. This implies that three different transfer functions are needed to design this type of collimator.

In [2] a system of ordinary differential equations for the transfer functions was derived. Given an effective intensity $\mathscr{I}(t)$ [lm/rad] of the LED, the y-chromaticity of the LED $y(t)$, a required effective target distribution $\mathscr{G}(\theta)$ [lm/rad], and the requirement that the chromaticity of the emitted light is constant, we have the

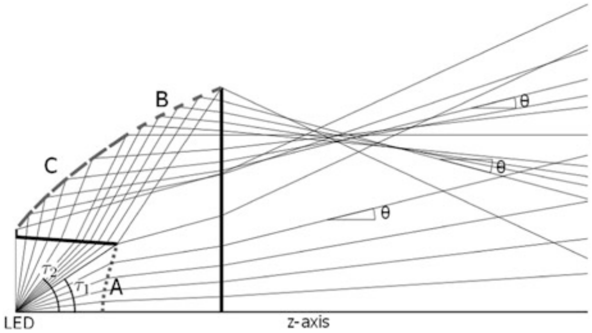


Fig. 1 Profile of a TIR collimator. A full TIR collimator can be obtained by rotating the profile around the z-axis. The free surfaces A, B and C are denoted by *dotted lines*, the *solid lines* show the fixed surfaces. The LED is located at the origin with the surface normal parallel to the z-axis, the *thin lines* are light rays traced from the LED through the collimator

following equations:

$$\sum_{i=1}^{N} \sigma_i \, \eta_i'(\theta) \, \mathscr{I}_i(\theta) = \mathscr{G}(\theta), \qquad \sum_{i=1}^{N} \sigma_i \, \eta_i'(\theta) \, \mathscr{I}_i(\theta)/y_i(\theta) = \mathscr{G}(\theta)/y_{\mathrm{T}}. \tag{1}$$

Here $\eta_i(\theta)$ are the transfer functions, N is the number of segments or transfer functions, $\sigma_i = -1$ for monotonically decreasing transfer functions and $\sigma_i = 1$ for monotonically increasing transfer functions, and y_T is the weighted average y-chromaticity coordinate of the LED. The x-chromaticity coordinate does not appear in these equations because $x(t)$ it is linearly dependent on $y(t)$. We use the following convention: $\mathscr{I}_i(\theta) = \mathscr{I}(\eta_i(\theta))$ is the intensity of the light at the source (in segment i) that is directed to the angle θ. Similarly we write $y_i(\theta) = y(\eta_i(\theta))$ for $i = 1, 2, \ldots, N$. In [2] we chose $N = 2$ so that the system would not be underdetermined. In this article we construct a solution with $N = 3$, so that the solution is better suited to the design of a TIR collimator.

Now we have the transfer functions $\eta_1 : [0, \theta_{\max}] \to [0, \tau_1]$, $\eta_2 : [0, \theta_{\max}] \to [\tau_1, \tau_2]$ and $\eta_3 : [0, \theta_{\max}] \to [\tau_2, \pi/2]$. The angle τ_1 is the angle of the ray that leaves the collimator at angle $\theta_{\max}$ and is refracted at the edge of surface A. The angle τ_2 is the angle of the ray that leaves the collimator at angle $\theta = 0$, this ray marks the separation between surface B and C. To construct a TIR collimator as shown in Fig. 1, we must have η_1 and η_3 monotonically increasing and η_2 monotonically decreasing. Thus, as initial values for the ODE we have $\eta_1(0) = 0$ and $\eta_2(0) = \eta_3(0) = \tau_2$.

The system (1) is underdetermined, we solve this by adding an extra equation. A possible extra equation can be obtained if we require that the intensity resulting from one of the transfer functions contributes a fraction $q \in (0, 1)$ to the total target intensity:

$$\sigma_i \mathscr{I}_i(\theta) \eta_i'(\theta) = q \, \mathscr{G}(\theta). \tag{2}$$

Choosing the extra restriction on the first or third transfer function gives a singular matrix at $\theta = 0$ or $\theta = \theta_{\max}$, therefore we apply the extra restriction to the second transfer function. Choosing the angles τ_1 and τ_2, we find by integration of (2) that $q = \int_{\tau_1}^{\tau_2} \mathscr{I}(t)\,\mathrm{d}t \Big/ \int_0^{\theta_{\max}} \mathscr{G}(\theta)\,\mathrm{d}\theta$. We can write the system of ODEs resulting from (1) and (2) as follows:

$$\eta_1'(\theta) = \frac{\mathscr{G}(\theta)}{\mathscr{I}_1(\theta)} \frac{y_1(\theta)}{y_1(\theta) - y_3(\theta)} \left(1 - q - \frac{y_3(\theta)}{y_T} + q\frac{y_3(\theta)}{y_2(\theta)}\right), \tag{3a}$$

$$\eta_2'(\theta) = -q\frac{\mathscr{G}(\theta)}{\mathscr{I}_2(\theta)}, \tag{3b}$$

$$\eta_3'(\theta) = \frac{\mathscr{G}(\theta)}{\mathscr{I}_3(\theta)} \frac{y_3(\theta)}{y_3(\theta) - y_1(\theta)} \left(1 - q - \frac{y_1(\theta)}{y_T} + q\frac{y_1(\theta)}{y_2(\theta)}\right). \tag{3c}$$

Equation (3a) has a removable singularity at $\theta = 0$, because we have $\mathscr{G}(0) = 0$ and the initial conditions imply that $\eta_1(0) = 0$ and thus $\mathscr{I}_1(0) = 0$ because $\mathscr{I}(0) = 0$ by assumption. Therefore we calculate $\eta_1'(0)$ using l'Hôpital's rule:

$$\eta_1'(0) = \sqrt{\frac{\mathscr{G}_+'(0)}{\mathscr{I}_+'(0)} \frac{y_1(0)}{y_1(0) - y_3(0)} \left(1 - q - \frac{y_3(0)}{y_T} + q\frac{y_3(0)}{y_2(0)}\right)}. \tag{4}$$

Here $\mathscr{G}_+'(0)$ and $\mathscr{I}_+'(0)$ are the right derivatives of $\mathscr{G}(\theta)$ at $\theta = 0$ and of $\mathscr{I}(t)$ at $t = 0$ respectively.

3 Numerical Results

The ODE system (3) is solved using the ODE-solver `ode45` in matlab, substituting (4) for small values of θ. The functions $\mathscr{I}(t)$ and $y(t)$ are least squares fits to the measured data of an LED with a high CoA variation. The target intensity was chosen to be a Gaussian profile [5] with Full Width at Half Maximum (FWHM) [4] at $\pi/9$. This yields an effective target intensity

$$\mathscr{G}(\theta) = C \sin(\theta) \exp\left(-4\ln(2)\left(\frac{\theta}{\theta_{\mathrm{H}}}\right)^2\right), \tag{5}$$

with $0 \le \theta \le 5/4\,\theta_{\mathrm{H}} = \theta_{\max}$, $\theta_{\mathrm{H}} = \pi/9$ and C chosen such that

$$\int_0^{5/4\,\theta_{\mathrm{H}}} \mathscr{G}(\theta)\,\mathrm{d}\theta = \int_0^{\pi/2} \mathscr{I}(t)\,\mathrm{d}t. \tag{6}$$

The chosen angles are $\tau_1 = 0.16\,\pi$ and $\tau_2 = 0.3\,\pi$. The calculated transfer functions can be seen in Fig. 2. Subsequently a TIR-collimator was designed, and

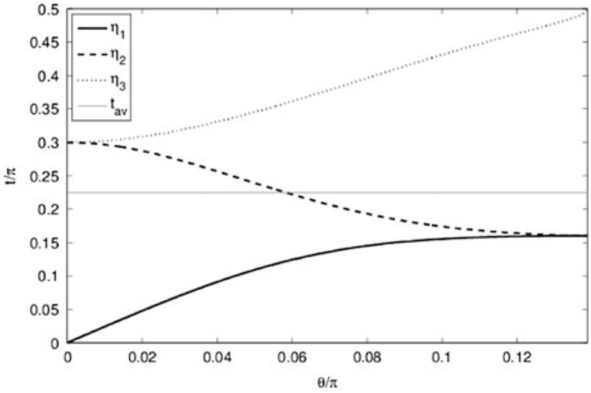


Fig. 2 Transfer functions

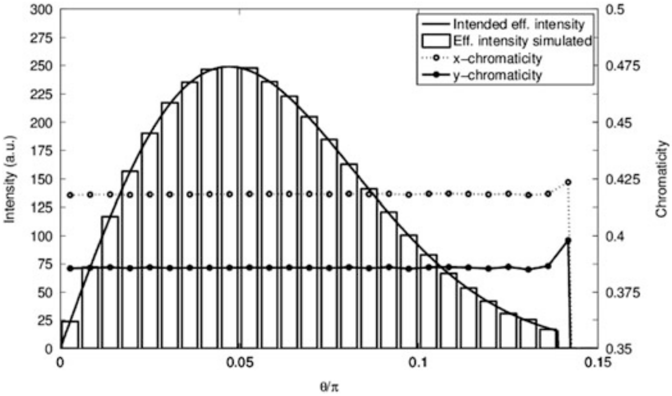


Fig. 3 Intensity and chromaticity of the Monte-Carlo simulations of the LED with the collimator

evaluated using the raytracing software LightTools. Results of the simulation can be seen in Fig. 3. The effective intensity shows the expected profile of a sine times a Gaussian, and the chromaticity is constant over the beam. An irregularity in the chromaticity is visible around $\theta = 25\pi/180$, because there is no light at this angle to properly determine the chromaticity. The collimator achieved the goal of eliminating the CoA variation.

4 Conclusions

A variant of the method in [2] has been introduced for reducing CoA variation in LED lighting systems. The improvement allows the design of regular TIR collimators using inverse methods with a given rotationally symmetric intensity pattern and a uniform color. Additionally, the method provides the optical designer with extra design freedom to choose the angular width of the refractive part of the TIR collimator.

References

1. Mueller, G.O.: Luminescent ceramic for a light emitting device. United States patent US7361938 B2, 2004. Assignee: Philips Lumileds Lighting Company LLC
2. Prins, C., ten Thije Boonkkamp, J., Tukker, T., IJzerman, W.: An inverse method for the design of TIR collimators to achieve a uniform color light beam. J. Eng. Math. **81**, 177–190 (2013). doi:10.1007/s10665-012-9584-7
3. Wang, K., Wu, D., Chen, F., Liu, Z., Luo, X., Liu, S.: Angular color uniformity enhancement of white light-emitting diodes integrated with freeform lenses. Opt. Lett. **35**, 1860–1862 (2010)
4. Weisstein, E.: Full width at half maximum. http://mathworld.wolfram.com/Full\discretionary-Width\discretionary-at\discretionary-HalfMaximum.html (2012)
5. Weisstein, E.: Gaussian function. http://mathworld.wolfram.com/GaussianFunction.html (2012)
6. Wu, H., Narendran, N., Gu, Y., Bierman, A.: Improving the performance of mixed-color white LED systems by using scattered photon extraction technique In: 7th Int. Conf. on Solid State Lighting. Proc. SPIE, vol. 6669, p. 666905 (2007)

Mathematical Modelling of Haptic Touchscreens

William T. Lee, Eoin English, and Mark Murphy

Abstract Haptic keyboards for touchscreen mobile devices would increase the accuracy of users typing, allow touchtyping and increase the satisfaction of users interacting with the devices. We report the results of a feasibility study of one method of implementing such haptic keyboards: driving transverse waves of the touchscreen using piezoelectric transducers mounted at the edges. Our results, while very preliminary, do suggest that this approach is feasible, and that a more detailed investigation is worthwhile.

1 Introduction

Touchscreens are a very popular form of interface to mobile phone and tablet devices. Their advantages are that nearly the whole surface area of the phone can be used as a screen and that differently configured, context sensitive, keyboards can be displayed as needed. The disadvantage of touchscreen keyboards is that the keys have no tactile characteristics. This makes it impossible to touchtype on a touchscreen keyboard, and even while watching the keyboard typists make more mistakes on a touchscreen keyboard than with a physical keyboard.

The disadvantages of touchscreen keyboards could be overcome using haptics. Localised vibrations of the surface could be used to create the tactile sensation of a key, enabling all the benefits of a physical keyboard to be realised. Haptic keyboards would enable touchtyping and reduce the rate of typing errors [2]. As well as practical benefits, adding a tactile dimension to phones would increase users

W.T. Lee (✉)
MACSI, Department of Mathematics and Statistics, University of Limerick, Limerick, Ireland
e-mail: william.lee@ul.ie

E. English • M. Murphy
Analog Devices Limerick, Limerick, Ireland

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_45,

feeling of attachment to their phones. The key barrier to implementation is that the waveforms needed to drive the transducers are not known.

There are physiological and technical constraints on the types of waves that can be generated and used for this application [1]. In order to be detectable displacements must be approximately $u_0 = 30\,\mu\text{m}$ in amplitude. The lowest frequency component of the vibration at the fingertip must be in the range 20–500 Hz. The delay between the action (key press) and the response (vibration) must be less than $T = 100\,\text{ms}$. Transducers are opaque and can only be placed at the edge of the touchscreen. The maximum voltage that can be applied to a transducer is $V_0 = 5\,\text{V}$.

2 Model

There are two simplifications we can make to the problem, both of which do require the assumption that the touchscreen is linear. Firstly if the material is linear then if we know a voltage signal $V_0(t; x_0, y_0, t_0)$ that results in a spatially and temporally localised response on the touch screen, in the ideal case composed of Dirac delta functions,

$$u(x, y, t) = \delta(x - x_0)\,\delta(y - y_0)\,\delta(t - t_0)\,, \tag{1}$$

then we can construct a waveform with any desired spatial and temporal structure from this localised solution. This follows immediately from the definition of the Dirac delta function. For instance to generate the profile $f(x, y, t)$ the required voltage signal is

$$V(t) = \int f(x_0, y_0, t_0)\, V(t, x_0, y_0, t_0)\ \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}t \tag{2}$$

In practice, of course, it would be impossible to generate a Dirac delta function and instead some localised function $w(x - x_0, y - y_0, t - t_0)$ would be produced. In that case the resulting waveforms resulting from the above construction would be the convolution of f and w. The second simplification is illustrated in Fig. 1. If we can produce a localised solution in one dimension, then this can be used to construct localised solutions in two dimensions. Therefore we focus on producing spatially and temporally localised solutions in one dimension.

We model the touchscreen in one dimension using the Euler beam equation [3] as illustrated in Fig. 2

$$\rho\frac{\partial^2 u}{\partial t^2} + \frac{Eh^2}{3\,(1 - \nu^2)}\frac{\partial^4 u}{\partial x^4} = \frac{p_{\text{piezo}}}{2h}, \tag{3}$$

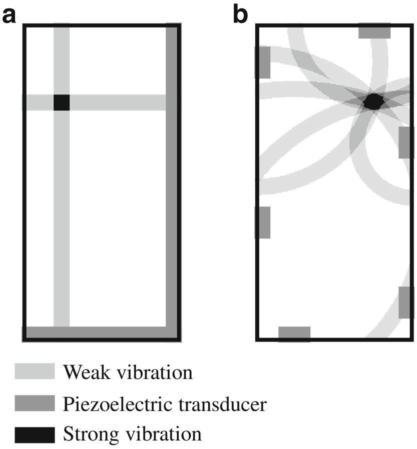


Fig. 1 If localised solutions can be constructed in one dimension, these can be used to construct localised solutions in two dimensions [1]

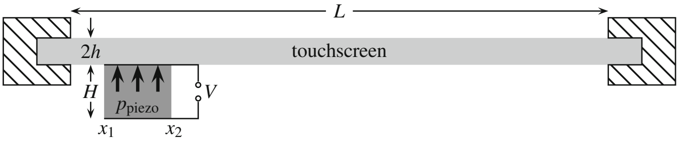


Fig. 2 Schematic of the thin plate model used. The touchscreen is modelled in one dimension as a thin plate, clamped at each end, in contact with a piezoelectric transducer to which a voltage signal is applied

where $\rho = 2{,}500\,\text{kg m}^{-3}$ is the (volume) density of the material making up the plate, $E = 70\,\text{GPa}$ is the Young's modulus, $\nu = 0.25$ is the Poisson's ratio, $h = 0.5\,\text{mm}$ is the half thickness of the plate, and p_{piezo} is the pressure exerted by the piezoelectric transducer. As shown in Fig. 2 the edges of the touchscreen are clamped so the boundary conditions of this equation are $u = \frac{\partial u}{\partial x} = 0$. The model of the transducer we use is

$$p_{\text{piezo}} = \frac{Y}{H}(u - n d_{33} V) \qquad x_1 < x < x_2 \tag{4}$$

where $Y = 50\,\text{GPa}$ is the elastic constant of the transducer, $d_{33} = -150\,\text{pm V}$ is the piezoelectric coefficient, $H = 1\,\text{mm}$ is the height of transducer, $n = 4$ the number of layers, the transducer is placed between x_1 and x_2.

To scale the equations we scale x with L, and use this scale and Eq. (3) to define a timescale. We scale displacements with u_0, and voltages with V_0. The dimensionless equations are

$$\frac{\partial^2 u}{\partial t^2} + \frac{\partial^4 u}{\partial x^4} = p_{\text{piezo}}, \tag{5}$$

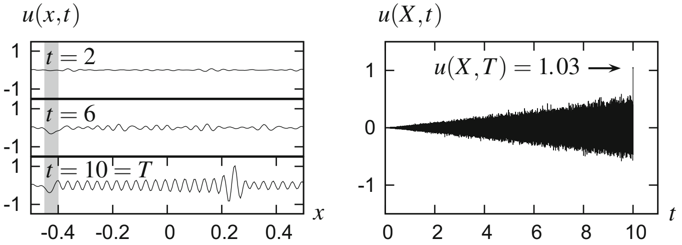


Fig. 3 Simulation results. *Left*: Displacement, u, as a function of x at times $t = 2$, $t = 6$ and $t = 10 = T$. The location of the transducers is shown in *gray*. *Right*: Displacement, u, at point $x = 0.25 = X$ as a function of time

$$p_{\text{piezo}} = \alpha V + \beta u, \qquad x_1 < x < x_2 \tag{6}$$

where $\alpha = 8 \times 10^4$, $\beta = 8 \times 10^8$, $x_1 = -0.45$ and $x_2 = -0.40$. In these dimensionless units $T = 15$.

3 Results and Discussion

A numerical simulation was carried out using the method of lines: 200 gridpoints were used with a 5 point stencil to represent spatial derivatives. The resulting system of ordinary differential equations was integrated with an explicit fourth order Runge-Kutta integrator and a timestep of $\delta t = 10^{-5}$. In order to find the input $V(t)$ needed to generate a localised pulse at a specific point $X = 0.25$ and time $T = 10$ we used the following procedure:

- Input a short $(11\delta t)$ voltage pulse and read off the resulting displacement $u(X,t)$, $0 < t < T$ at the designated point.
- Extract the sign of u: $f(t) = 1$ if $u(X,t) > 0$, $f(t) = -1$ if $u(X,t) < 0$
- Construct the voltage waveform by running time backwards $V(t) = f(T-t)$.

In other words set up the input so that the contribution it makes to the displacement at $u(X,T)$ is always positive. The results from this simulation are shown in Fig. 3. As can be seen the height of the resulting pulse is above the detection threshold.

The results suggests that this approach to generating Haptic keyboards could succeed, and that investigating this approach more carefully would be worthwhile. The three key improvements needed to make the model more realistic are to include: (1) damping, within the transducer and the screen and from the air the screen is in contact with; (2) two dimensions, the effect of the clamping of the screen at the side needs to be included; (3) robustness, the procedure used to construct the localised pulse must be robust to small changes in the properties of the system.

Acknowledgements WTL acknowledges the support of the Mathematics Applications Consortium for Science and Industry (http://www.macsi.ul.ie) funded by the Science Foundation Ireland Mathematics Initiative Grant 06/MI/005.

References

1. Cooker, M., Cribbin, L., Dellar, P., Fitt, A., Gaburro, R., Gibb, T., Kennedy, J., King, J., Kubat, I., Lapin, V., Lee, W., Murphy, E., Nolan, C., Parker, J., Power, O., Timoney, C.: Haptic touchscreens. Technical report, Smith Institute (2012)
2. Koskinen, E.: Optimising tactile feedback for virtual buttons in mobile devices. Master's thesis, Helsinki University of Technology (2008)
3. Landau, L.D., Lifshitz, E.M., Kosevich, A.M., Pitaevskii, L.P.: Theory of Elasticity. Course of Theoretical Physics, 3rd edn. Butterworth, Oxford (1986). Translated from the Russian by Sykes, J.B., Reid, W.H.

A Covariant Spacetime Approach to Transformation Acoustics

Michael M. Tung and Jesús Peinado

Abstract Transformation acoustics focuses on the design of advanced acoustic devices by employing sophisticated mathematical transformation techniques for engineering acoustic metamaterials—materials artificially fabricated with extraordinary acoustic properties beyond those encountered in nature. We present differential-geometric methods together with a variational principle and show how they form the basis for a powerful framework to control acoustic waves in industrial applications. We conclude with a practical example and implement the acoustic wave equation within a uniform accelerating rigid frame (UAF). As expected, an acoustic event horizon emerges, i.e., a boundary in spacetime beyond which events cannot acoustically affect any outside observer.

1 Background

The theoretical design and subsequent industrial engineering of artificial materials with formidable properties, which may go beyond what is found in Nature, is one of the ongoing tasks to improve the standard of living. Acoustic metamaterials [8] may contribute in this endeavour.

By making use of the mathematics similar to the differential-geometric framework of general relativity, transformation acoustics centres on the design of advanced acoustic devices by employing sophisticated coordinate transformation

M.M. Tung (✉)
Instituto de Matemática Multidisciplinar, Universitat Politècnica de València, Camino de Vera, s/n, 46022 Valencia, Spain
e-mail: mtung@mat.upv.es

J. Peinado
Instituto de Instrumentación para Imagen Molecular, Universitat Politècnica de València, Camino de Vera, s/n, 46022 Valencia, Spain
e-mail: jpeinado@dsic.upv.es

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_46,

techniques for the conception of acoustic metamaterials. Industrial applications in this field have wide repercussions and range from acoustically improving concert halls to constructing ships and submarines invisible to sonar detection. Recent applications to acoustic cloaking can be found in [3, 4, 10].

Previously, we have elaborated a Lagrangian framework to describe macroscopic electrodynamics for optical metamaterials [7] and diffusion on curved manifolds [9, 11, 12]. Here, we use a similar approach to derive the equations of motion for non-dissipative acoustic phenomena from a fundamental Lagrangian density [10]. In electrodynamics Maxwell's equations are already inherently covariant, acoustics however does not possess this advantage. Nevertheless, we succeed in reformulating Hamilton's principle for acoustics in a fully covariant fashion for spacetime.

In the following discussion, we will first introduce a fully spacetime covariant formalism most suitable to tackle the coordinate transformations of transformation acoustics. Hamilton's principle for acoustics in spacetime will provide the fundamental starting point and permit to derive the general relations between the constitutive parameters of the virtual and physical acoustic metafluid, κ (bulk modulus) and ρ_{ij} (mass-density tensor), and their spacetime metrics.

Finally, we conclude with a practical example and show how the design of an acoustic devices depends on the tuning of these material parameters of the physical and virtual spaces. For an example spacetime geometry we have chosen to implement the acoustic wave equation within a uniform accelerating rigid frame (UAF), a metric framework introduced by Desloge [1, 5, 6]. We will illustrate how in this example an acoustic event horizon emerges, i.e., a boundary in spacetime beyond which events cannot acoustically affect any outside observer.

2 Results and Discussion

2.1 Hamilton's Principle for Spacetime Acoustics

Hamilton's principle states that the behaviour of a deterministic physical system is completely described by a variational principle. Its solutions are the equations of motion governing the dynamics of the system and are found as the extremal solution of the corresponding action functional. In the case of first-order classical field theories, partial derivatives of configuration space with respect to all spacetime coordinates are required, and the Lagrangian density function will therefore be a mapping

$$\mathscr{L} : J^1 N \to \mathbb{R}, \qquad (1)$$

where $J^1 N = M \times TP$ is the jet bundle of the associated configuration space $N = M \times P$, and P is the ambient space defined by the acoustic potential $\phi : M \to \mathbb{R}$. Here, as usual, M denotes a smooth four-dimensional manifold endowed with a

Lorentzian metric $\mathbf{g}$ having a mixed signature $(-,+,+,+)$. The acoustic potential is the scalar velocity potential of the acoustic metafluid such that

$$v_\mu = -\phi_{,\mu} = \begin{pmatrix} -p/c\rho_0 \\ \mathbf{v} \end{pmatrix}, \tag{2}$$

where p are the acoustic pressure and density, $\mathbf{v}$ is the local fluid velocity, and c the acoustic wave speed (which is assumed to be time-independent, i.e. $\partial c/\partial t = 0$).

For transformation acoustics the variation of the following action integral must then vanish

$$\delta\mathscr{A} = \delta \int_\Omega d^4x \; \mathscr{L}(\phi_{,\mu}) = 0, \tag{3}$$

where the invariant volume element is $d^4x\sqrt{-g} = dx^0dx^1dx^2dx^3\sqrt{-g}$ with $g = \det\mathbf{g}$, and integration occurs over a bounded, closed set of spacetime $\Omega \subset M$ (see [2] and references therein). Notice that Greek indices will be used for the full spacetime values of four-tensors, whereas Latin indices run over the spatial values only. We also use the standard comma and semicolon notation for partial and covariant derivatives.

The simplest possible choice for the acoustic Lagrangian density consists of only a covariant kinetic term:

$$\mathscr{L}(\phi_{,\mu}) = \tfrac{1}{2}\sqrt{-g}\, g^{\mu\nu}\phi_{,\mu}\phi_{,\nu}. \tag{4}$$

A straightforward calculations then yields for the action (3) in combination with (4) the corresponding Euler-Lagrange equations

$$\Delta_M\phi = g^{\mu\nu}\phi_{;\mu\nu} = \frac{1}{\sqrt{-g}}\left(\sqrt{-g}\, g^{\mu\nu}\phi_{,\mu}\right)_{,\nu} = 0, \tag{5}$$

where Δ_M is the Laplace-Beltrami operator on the Riemannian manifold $(M, \mathbf{g})$.

To arrive at the fundamental relations of the constitutive parameters in transformation acoustics, we require the description of the same acoustic phenomena in either anisotropic *physical space* or flat *virtual space*, both represented by their Lagrangian densities, $\mathscr{L}_{\text{phys}}$ and $\mathscr{L}_{\text{virt}}$, respectively. The constitutive relations, which establish the correspondence of the curvilinear coordinate transformations between physical and virtual acoustic space and their material properties, are then given by [10]:

$$\kappa = \frac{\sqrt{-g}}{\sqrt{-\bar{g}}}\,\bar{\kappa}, \qquad \rho_0\rho^{ij} = \frac{\sqrt{-\bar{g}}}{\sqrt{-g}}\,\bar{g}^{ij}, \tag{6}$$

where ρ^{ij} are the contravariant components of the mass density expressed as a $(0, 2)$-tensor field $\boldsymbol{\rho} : T_pM \times T_pM \to \mathbb{R}$, defined at any point p on the smooth manifold M within the metafluid region.

2.2 Acoustic Waves in a Uniformly Accelerating Reference Frame

In order to investigate the gravitational redshift in a uniform field, Desloge [5] proposed the following line element

$$ds^2 = -\left(1 + \frac{g_0}{c^2}y\right)^2 c^2dt^2 + d\mathbf{r}^2, \tag{7}$$

where $g_0 > 0$ is the proper acceleration in y-direction for an observer located at the origin, and $d\mathbf{r}$ represents the infinitesimal spatial displacement. Suppressing the third space component, the corresponding metric for this uniformly accelerating rigid frame (UAF) in field-free space is

$$g_{\mu\nu} = \begin{pmatrix} -\left(1 + g_0y/c^2\right)^2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{8}$$

It is not difficult to show that the underlying physical spacetime of the UAF metric is flat.

In the next step, we associate with virtual space the flat Minkowski metric with two spatial components

$$\bar{g}_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{9}$$

Substituting the metrics (8) and (9) into (6) readily yields the constitutive relations

$$\kappa = 1 + \frac{g_0}{c^2}y, \quad \rho_0\rho^{ij} = \left(1 + \frac{g_0}{c^2}y\right)^{-1} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \tag{10}$$

which exactly describes how the parameters of the acoustic metamaterial have to be fine-tuned to implement an acoustic media within a rigid, uniformly accelerating reference frame.

This fully establishes the design of the desired acoustic device. Furthermore, a detailed analysis of the corresponding solutions of the Euler-Lagrange equations, (5), shows that the y-dependence of the acoustic potential $\phi(t, x, y) = \phi_0(t)\phi_1(x)\phi_2(y)$ for is dictated by the differential equation

$$\phi_2'' + \frac{g_0}{c^2 + g_0 y}\,\phi_2' + \left[\frac{\omega^2}{(c^2 + g_0 y)^2} - k_x^2\right]\phi_2 = 0, \tag{11}$$

where ω and k_x are the angular frequency and wavenumber of the plane-wave solution for a wave travelling in x-direction, respectively.

It is easy to see that in the asymptotic limits $y \to \pm\infty$, Eq. (11) gives for the y-amplitude the dependence $\phi_2 \sim e^{\pm k_x y}$, where the physically relevant solution is chosen. On the other hand, a coordinate singularity occurs at $y_0 = -c^2/g_0$ where the metric tensor (8) has a vanishing determinant, and thus corresponds to a *Rindler event horizon*.

Since the functions $\phi_0(t)$ and $\phi_1(x)$ describe an oscillatory harmonic motion which is bounded, the full acoustic potential $\phi(t, x, y)$ will display the same asymptotic behavior for sufficiently large y as $\phi_2(y)$ and also possess an event horizon at y_0. This means that any acoustic wave will be trapped either within region $y < y_0$ or region $y > y_0$, and no communication to the outside of each domain will be possible. In practice, g_0 can be fine-tuned such that this effect may be detected for desirable values of y_0, The boundary $y = y_0$ demarcates the region at which the accelerational pull becomes so great as to make the escape of an acoustic signal impossible, that is, it constitutes an *acoustic event horizon*. For a full numerical discussion of this effect consult [13], which also compares the UAF model with systems containing a uniform gravitational field (UGF).

3 Conclusions

We have outlined a novel differential-geometric approach to transformation acoustics based on Hamilton's principle for the acoustic potential in a fully covariant spacetime setting [10]. This enabled us to immediately establish the general relations between the constitutive parameters κ (bulk modulus) and ρ_{ij} (mass-density tensor) linking the physical and virtual spaces of the acoustic metafluid and their respective spacetime metrics.

We hope that the proposed approach to transformation acoustics will aid in the efficient design and analysis of acoustic metamaterials and open up hitherto unknown possibilities in this area of research.

Apart from allowing for a thorough examination of new acoustic devices with much more complicated spacetime geometries, it may also serve to implement and investigate many intriguing features of general relativity in a laboratory environment, especially by constructing analogue horizons [13, 14].

Acknowledgements The authors wish to thank for financial support by the Universidad Politécnica de Valencia under grant PAID-06-11-2020.

References

1. Acedo, L., Tung, M.M.: Electromagnetic waves in a uniform gravitational field and Planck's postulate. Eur. J. Phys. **33**, 1073–1082 (2012)
2. Calin, O., Chang, D.C.: Geometric Mechanics on Riemannian Manifolds: Applications to Partial Differential Equations. Applied and Numerical Harmonic Analysis. Birkhäuser, Boston (2005)
3. Chen, H.Y., Chan, C.T.: Acoustic cloaking and transformation acoustics. J. Phys. D **43**(11), 113001 (2010)
4. Cummer, S.A., Schurig, D.: One path to acoustic cloaking. New J. Phys. **9**(3), 45–52 (2007)
5. Desloge, E.A.: Nonequivalence of a uniformly accelerating reference frame and a frame at rest in a uniform gravitational field. Am. J. Phys. **57**(12), 1121–1125 (1989)
6. Desloge, E.A.: The gravitational red shift in a uniform field. Am. J. Phys. **58**(9), 856–858 (1989)
7. García-Meca, C., Tung, M.M.: The variational principle in transformation optics engineering and some applications. Math. Comput. Model. **57**(7–8), 1773–1779 (2013)
8. Norris, A.N.: Acoustic metafluids. J. Acoust. Soc. Am. **125**(2), 839–849 (2009)
9. Tung, M.M.: Basics of a differential-geometric approach to diffusion: uniting Lagrangian and Eulerian models on a manifold. In: Bonilla, L.L., Moscoso, M.A., Platero, G., Vega, J.M. (eds.) Progress in Industrial Mathematics at ECMI 2006. Mathematics in Industry, vol. 12, pp. 897–901. Springer, Berlin (2007)
10. Tung, M.M.: A fundamental Lagrangian approach to transformation acoustics and spherical spacetime cloaking. Europhys. Lett. **98**, 34002–34006 (2012)
11. Tung, M.M.: Diffusion on surfaces of revolution. In: Günther, M., Bartel, A., Brunk, M., Schöps, S., Striebel, M. (eds.) Progress in Industrial Mathematics at ECMI 2010. Mathematics in Industry, vol. 17, pp. 643–650. Springer, Berlin (2012)
12. Tung, M.M., Hervás, A.: A differential-geometric approach to model isotropic diffusion on circular conic surfaces in uniform rotation. In: Fitt, A.D., Norbury, J., Ockendon, H., Wilson, E. (eds.) Progress in Industrial Mathematics at ECMI 2008. Mathematics in Industry, vol. 15, pp. 1053–1058. Springer, Berlin (2010)
13. Tung, M.M., Weinmüller, E.B.: Gravitational frequency shifts in transformation acoustics. Europhys. Lett. **101**, 54006–54011 (2013)
14. Visser, M., Barcelo, C., Liberati, S.: Analogue models of and for gravity. Gen. Relat. Gravit. **34**, 1719–1734 (2002)

Location and Management of a New Industrial Plant

Miguel E. Vázquez-Méndez, Lino J. Alvarez-Vázquez, Néstor García-Chan, and Aurea Martínez

Abstract Within the framework of numerical simulation and multi-objective control of partial differential equations (PDE), in this work we deal with the problem of determining the optimal location of a new industrial plant. We begin presenting a mathematical model (a system of nonlinear parabolic PDE) for the numerical simulation of air pollution. Based on this model, and taking into account economic and ecological objectives, we formulate the problem in the field of multi-objective optimal control. We analyze the problem from a cooperative viewpoint, recalling the standard concept of Pareto-optimal solution, and pointing out the Pareto-frontier as a very useful tool in the decision-making process. Finally, some preliminary results for a hypothetical situation in the region of Galicia (NW Spain) are also presented.

1 Mathematical Modeling

Let $\Omega \subset \mathbb{R}^2$ be a bounded domain where several already existing industrial plants, located at points $p_i \in \Omega,\ i = 0, \ldots, N$, discharge pollutants into the atmosphere. We suppose that these pollutants are transported through the atmosphere by air

M.E. Vázquez-Méndez (✉)
Departamento de Matemática Aplicada, E.P.S., Universidad de Santiago de Compostela, 27002 Lugo, Spain
e-mail: miguelernesto.vazquez@usc.es

L.J. Alvarez-Vázquez • A. Martínez
Departamento de Matemática Aplicada II, E.I. Telecomunicación, Universidad de Vigo, 36310 Vigo, Spain
e-mail: lino@dma.uvigo.es; aurea@dma.uvigo.es

N. García-Chan
Departamento de Física, C.U. Ciencias Exactas e Ingeniería, Universidad de Guadalajara, 44420 Guadalajara, Mexico
e-mail: netog_g@hotmail.com

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_47,

masses and turbulent diffusion, and that the kinetics of all reactions taking place in the process can be expressed quantitatively by a rate law (the reaction rate is proportional to the concentration of reactants). So, we consider N_S substances (pollutants) to control and, for $j = 1, \dots, N_S$, we denote by $\phi_j(x,t)$ the concentration of j-th pollutant in point $x \in \Omega$ and at time $t \geq 0$. Under previous hypotheses, functions $\phi_j(x,t)$ should satisfy for a given time interval $(0,T)$ the following coupled system of partial differential equations, for $j = 1, \dots, N_S$:

$$\begin{aligned} \tfrac{\partial \phi_j}{\partial t} + \mathbf{u} \cdot \nabla \phi_j - \nabla \cdot (\mu_j \nabla \phi_j) + f_j(\phi_1, \dots, \phi_{N_S}) = \\ \sum_{i=0}^{N} Q_i^j(t)\delta(x - p_i) \qquad \text{in } \Omega \times (0,T), \end{aligned} \tag{1}$$

where $\mathbf{u}(x,t)$ is the wind velocity (which we assume experimentally known, and verifying the continuity equation $\nabla . \mathbf{u} = 0$), $\mu_j(x,t)$ is the horizontal turbulent diffusion coefficient, $Q_i^j(t)$ is the mass flow rate of j-th pollutant discharged at point $p_i \in \Omega$, $\delta(x - p_i)$ represents the Dirac delta function at p_i, and, finally,

$$f_j(\phi_1, \dots, \phi_{N_S}) = \kappa_j {\phi_1}^{\alpha_j^1} \dots {\phi_{N_S}}^{\alpha_j^{N_S}}$$

denotes the reaction term for j-th pollutant, where κ_j is a temperature-dependent rate and powers α_j^i are the reaction orders (see, for instance, [1]).

We suppose that initial pollutant concentrations are given by known functions $\phi_1^0(x), \dots, \phi_{N_S}^0(x)$ in such a way that, for $j = 1, \dots, N_S$:

$$\phi_j(x,0) = \phi_j^0(x) \ \text{ in } \ \Omega. \tag{2}$$

Finally, for each $x \in \partial\Omega$, we denote by $\mathbf{n}(x)$ the unit outward normal vector to the boundary $\partial\Omega$, and we write $\partial\Omega \times (0,T) = S^- \cup S^+$, where $S^- = \{(x,t) \in \partial\Omega \times (0,T) \ \text{ such that } \ \mathbf{u}.\mathbf{n} < 0\}$ represents the inflow boundary, and $S^+ = \{(x,t) \in \partial\Omega \times (0,T) \ \text{ such that } \ \mathbf{u}.\mathbf{n}(x) \geq 0\}$ the outflow boundary. No pollution sources outside Ω are considered and, consequently, the combined (diffusive plus advective) pollution flow is assumed to be zero on S^-. On the other hand, on S^+ we assume that the diffusive pollution flow can be negligible as compared to the advective pollution outflow from Ω. Thus, for $j = 1, \dots, N_S$, we consider the following boundary conditions (see [4]):

$$\mu_j \frac{\partial \phi_j}{\partial \mathbf{n}} - \phi_j \, \mathbf{u}.\mathbf{n} = 0 \ \text{ on } \ S^-, \qquad \mu_j \frac{\partial \phi_j}{\partial \mathbf{n}} = 0 \ \text{ on } \ S^+. \tag{3}$$

2 Multi-objective Optimal Control Problem

Our main objective consists of determining the optimal location and management of a new industrial plant. We suppose that at the present moment there exist N plants working in the domain Ω (i.e., points $p_1, \ldots, p_N \in \Omega$ and functions $Q_1^j(t), \ldots, Q_N^j(t)$, for $j = 1, \ldots, N_S$, are known), and that a new industrial plant is to be built in a point $p_0 \in \Omega$, which has to be determined. This new plant will be working for a time $(0, T)$, and during this period of time its emission flow rates will be given by a vector function $\mathbf{Q}_0(t) = (Q_0^1(t), \ldots, Q_0^{N_S}(t))$, which has to be also determined. The new plant should be as cost-effective as possible, but also as *green* (harmless from an ecological perspective) as possible.

From an economic viewpoint, the cost-effectiveness of the plant depends on two aspects. First, more emission rates correspond with a higher production and, consequently, with a higher cost-effectiveness. Therefore, we suppose that there exists a known function F giving the cost-effectiveness of the plant in terms of the emission flow rates, in such a way that the gross profit (to be maximized) for the time interval $(0, T)$ is given by $\int_0^T F(\mathbf{Q}_0(t))\, dt$. On the other hand, the building and managing costs depend on the plant location $p_0 \in \Omega$. For building costs we assume that they are given by a known function $G(p_0)$. For managing costs, we can suppose that there exist an *ideal point* $p_0^I \in \Omega$ (representing, for instance, the source of raw material) and a known function (estimated, for example, from fuel costs) in such a way that managing costs (to be minimized) for the time interval $(0, T)$ are given by $\int_0^T s(t)\, ||p_0 - p_0^I||^2\, dt$. Thus, from an economic point of view, the objective consists of minimizing the economic cost function:

$$J_E(p_0, \mathbf{Q}_0) = -\int_0^T F\,(\mathbf{Q}_0(t))\, dt + \int_0^T s(t)\, ||p_0 - p_0^I||^2\, dt + G(p_0). \tag{4}$$

From an ecological viewpoint, N_Z sensitive environmental areas $A_k \subset \Omega$, for $k = 1, \ldots, N_Z$, are considered, and the environmental impact of the plant is measured in terms of the mean pollutant concentrations in each of these areas. Thus, the ecological objectives consist of minimizing the following functions, for $k = 1, \ldots, N_Z$, and for $j = 1, \ldots, N_S$:

$$J_k^j(p_0, \mathbf{Q}_0) = \frac{1}{|A_k|\, T} \int_0^T \int_{A_k} \phi_j(x, t)\, dx\, dt, \tag{5}$$

where $|A_k|$ denotes the area of A_k, and functions $\phi_1(x, t), \ldots, \phi_{N_S}(x, t)$ are the solutions of the state system (1)–(3).

Finally, we have to take into account some technological constraints limiting both plant location and emission rates. If $X_{ad} \subset \Omega$ denotes the admissible plant locations, and $Q_{ad} \subset (L^\infty(0, T))^{N_S}$ is the set of admissible emission flow rates, then the problem of finding the optimal location and management of the new industrial plant can be formulated as the following Multi-objective Optimal Control problem (MOC): For $k = 1, \dots, N_Z$, and for $j = 1, \dots, N_S$, find the point $p_0 \in X_{ad}$ and the function vector $\mathbf{Q}_0 \in Q_{ad}$ minimizing the economic cost J_E, given by (4), and the ecological costs J_k^j, given by (5), in the admissible set $X_{ad} \times Q_{ad}$.

3 Pareto-Optimal Solutions

Obviously, economic and ecological objectives are contradictory and, consequently, it is not possible to find an element $(p_0, \mathbf{Q}_0) \in X_{ad} \times Q_{ad}$ minimizing J_E and J_k^j, for $k = 1, \dots, N_Z$, $j = 1, \dots, N_S$, simultaneously. In this sense, the problem (MOC) (as it usual in many multi-objective optimization problems) are ill-defined. Anyway, some elements of the admissible set can be extracted for examination. Such vectors are those where none of the components can be improved without a deterioration of, at least, one of the other components. These vectors are usually called Pareto-optimal solutions. A more formal definition is the following (see, for instance, [5]):

Definition 1. $(p_0^*, \mathbf{Q}_0^*) \in X_{ad} \times Q_{ad}$ is a Pareto-optimal solution of problem (MOC) if there does not exist any $(p_0, \mathbf{Q}_0) \in X_{ad} \times Q_{ad}$ satisfying the following conditions:

1. $J_E(p_0, \mathbf{Q}_0) \leq J_E(p_0^*, \mathbf{Q}_0^*)$ and $J_k^j(p_0, \mathbf{Q}_0) \leq J_k^j(p_0^*, \mathbf{Q}_0^*)$ for all $k = 1, \dots, N_Z$, $j = 1, \dots, N_S$.
2. $J_E(p_0, \mathbf{Q}_0) < J_E(p_0^*, \mathbf{Q}_0^*)$ or $J_k^j(p_0, \mathbf{Q}_0) < J_k^j(p_0^*, \mathbf{Q}_0^*)$ for at least one $k = 1, \dots, N_Z$, or $j = 1, \dots, N_S$.

If $(p_0^*, \mathbf{Q}_0^*) \in X_{ad} \times Q_{ad}$ is a Pareto-optimal solution, the corresponding objective vector $(J_E(p_0^*, \mathbf{Q}_0^*), J_1^1(p_0^*, \mathbf{Q}_0^*), \dots, J_1^{N_S}(p_0^*, \mathbf{Q}_0^*), \dots, J_{N_Z}^1(p_0^*, \mathbf{Q}_0^*), \dots, J_{N_Z}^{N_S}$ $(p_0^*, \mathbf{Q}_0^*))$ is also known as Pareto-optimal. The set of Pareto-optimal solutions is called Pareto-optimal set, and the set of Pareto-optimal objective vectors is known as the Pareto-optimal frontier.

Figure 1 shows the geometrical interpretation for two objectives, J_E and J_1^1. Bearing in mind that the Pareto-optimal frontier is very important (crucial) for decision makers, several techniques of multi-objective optimization have been developed during last decades in order to be applied in the computation of the Pareto-optimal frontier (see, for example, a brief historical review in [5] or [2]). To apply any of these techniques, an alternative formulation of problem (MOC), in terms of adjoint state [3], can be very useful.

4 Numerical Results

In this section, we present some preliminary results for a hypothetical situation in the region of Galicia (northwest of Spain). The domain Ω is a rectangular area of 57,600 km^2, covering the surface of Galicia, where we have considered

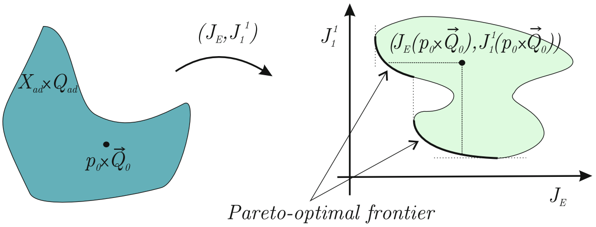


Fig. 1 Geometrical interpretation of Pareto-optimal frontier for two objectives J_E and J_1^1

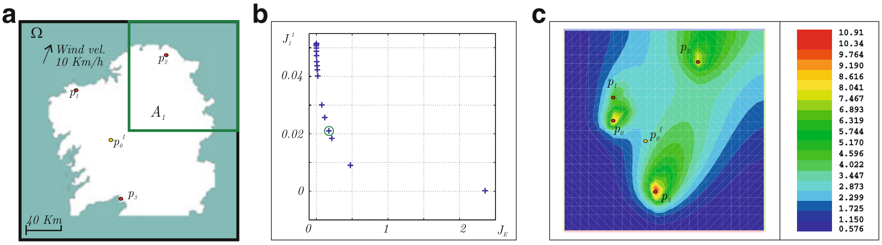


Fig. 2 Numerical results: (**a**) domain Ω for numerical simulation. (**b**) Pareto-optimal frontier. (**c**) Pollutant concentration at final time $\phi_1(x, T)$

$N_Z = 1$ sensible area, and $N = 3$ industrial plants located at points $p_1 = (62{,}165)$, $p_2 = (161{,}203)$ and $p_3 = (11{,}145)$ (see Fig. 2a). Only one pollutant is controlled ($N_S = 1$) with unitary reaction order ($\alpha_1^1 = 1$). The time period for controlling is 1 year ($T = 8{,}760\,$h), and we assume that, for this period of time, the emission rates of the three plants are constant: $Q_1 = 1{,}000\,$kg/h, $Q_2 = 2{,}000\,$kg/h and $Q_3 = 4{,}000\,$kg/h. For the new plant, we take $p_0^I = (100, 110)$ as the ideal point, and we assume that its emission flow rate $Q_0(t)$, that can be no constant, has to be greater than $Q_0^{min} = 100\,$kg/h and lower than $Q_0^{max} = 4{,}000\,$kg/h for all time t. For these first numerical experiences we take a constant wind velocity $\mathbf{u}(x, t) = (4, 9)$ km/h, a constant diffusion coefficient $\nu(x, t) = 100\,\text{km}^2$/h and a reaction rate $\kappa_1 = 10^{-5}\,\text{h}^{-1}$. Finally, null pollutant initial concentration is considered ($\phi_1^0(x) = 0\,\text{kg/m}^2$), building costs are neglected ($G(x) = 0$) and, for cost-effectiveness and managing costs, functions $F(Q) = 10^{-15} Q^2$ and $s(t) = 10^{-8}$ are defined.

In this situation, we obtain the Pareto-optimal frontier which is shown in Fig. 2b. For the Pareto-optimal solution pointed out with a circle in Fig. 2b, the pollutant concentration at the final time of simulation ($\phi_1(x, T)$) is shown in Fig. 2c.

Acknowledgements This work was supported by Project MTM2012-30842 of M.E.C. (Spain). The third author thanks the support from Sistema Nacional de Investigadores SNI-52768 (Mexico).

References

1. Chapra, S.C.: Surface Water-Quality Modeling. McGraw-Hill, New York (1997)
2. Coello, C.A., Van Veldhuizen, D.A., Lamont, G.B.: Evolutionary algorithms for solving multi-objective problems. Kluwer Academic, New York (2002)
3. García-Chan, N., Alvarez-Vázquez, L.J., Martínez, A., Vázquez-Méndez, M.E.: On optimal location and management of a new industrial plant: numerical simulation and control. J. Franklin Inst. (2013). http://dx.doi.org/10.1016/j.jfranklin.2013.11.005
4. Marchuck, G.I.: Mathematical Models in Environmental Problems. Elsevier, New York (1986)
5. Miettinen, K.M.: Nonlinear Multiobjective Optimization. Kluwer Academic, Boston (1999)

A Satellite-to-Satellite Laser Tracking Solution Within the Post-Newtonian Model of the Earth Outer Space

Jose M. Gambi and Maria Luisa Garcia del Pino

Abstract Two second order post-Newtonian formulae, one for the two-way frequency shift and the other for the two-way Laser ranging, are derived by means of Synge's world-function. The formulae can be used to increase the Classical accuracy in tracking passive targets by means of APT systems on board Earth satellites.

1 Introduction

The emerging importance of space-based systems for communications and surveillance is making the implementation of accurate space-based acquisition, pointing, and tracking (APT) systems a relevant issue. In particular, the Satellite-to-Satellite (SST) Laser tracking problem is attracting more and more attention due to the fact that Laser technology has matured substantially in the recent years (see e.g. [1, 2]).

The two-way frequency shift and Laser ranging formulae introduced below correspond to the post-Newtonian model of the exterior of an spherical earth. Therefore, they can also be used to derive TDOA and FDOA equations for them to meet the present needs in locating passive radio transmitters placed on the Earth surface or in space [3–5]. In fact, the formulae can be used for a great variety of tracker-target configurations, since the target need not be active and the tracker may be given a discrete number of orbital impulses. In addition, the tool used, Synge's world-function [6], allows us to follow a procedure with which the Classical and post-Newtonian approaches can be compared at each step.

J.M. Gambi (✉) • M.L.G. del Pino
Gregorio Millan Institute, Universidad Carlos III de Madrid, Avda. de la Universidad, 30, 28911-Leganes, Madrid, Spain
e-mail: gambi@math.uc3m.es; lgarciadelpino@educa.madrid.org

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_48,

2 The World-Function for the Earth Surrounding Space

Synge's world function is an efficient tool so plenty of physical content that many practical results have been obtained with it (see e.g. [7–10]).

As was mentioned above, to derive the two-way frequency shift and Laser ranging formulae we consider an spherical earth. Therefore, we adopt as model of space-time about the Earth the post-Newtonian approximation of the exterior Schwarzschild field. The metric form that characterizes this field written in terms of ECI coordinates, $x^i \equiv (x^\alpha, t)$, is [6, 11] ($c = G = 1$)

$$ds^2 = \epsilon g_{ij} dx^i dx^j = \epsilon[(\delta_{\alpha\beta} + \gamma_{\alpha\beta}) dx^\alpha dx^\beta + (-1 + \gamma_{44}) dt^2] + \mathscr{O}(\varepsilon^3), \tag{1}$$

where $\gamma_{\alpha\beta} = 2mx^\alpha x^\beta / r^3$, $\gamma_{44} = 2m/r$, $r^2 = x^\alpha x^\alpha$, m is the mass of the earth, and $\epsilon = \{-1, 0, 1\}$ for timelike, null and spacelike vectors, dx^i, respectively. $\eta_{ij} = diag(1, 1, 1, -1)$ and ε is a small dimensionless parameter such that ε^2 is of the order of m/r and v^2, where v is the characteristic Classical 3-speed of the objects in orbit about the Earth with respect to the Earth. (Latin indices range from 1 to 4, and Greek from 1 to 3.)

For any space-time characterized by the pseudo-Riemann metric tensor $g_{ij}(x^k)$ (signature $+, +, +, -$) and for any two events, $P_1(x^{k_1})$, $P_2(x^{k_2})$, for which there is a unique geodesic, $\Gamma_{P_1 P_2}$, joining them, Synge's world-function, $\Omega(P_1, P_2)$, is defined by the line integral

$$\Omega(P_1, P_2) = \frac{1}{2} \int_0^1 g_{ij} U^i U^j \, d\omega \tag{2}$$

taken along $\Gamma_{P_1 P_2}$, where $\Gamma_{P_1 P_2}$ is given by $x^i = x^i(\omega)$, ω being an affine parameter satisfying $0 \leq \omega \leq 1$, so that $P_1 \equiv x^i(0)$, $P_2 \equiv x^i(1)$ and $U^i = dx^i / d\omega$.

There are two expressions for Ω for the space-time in (1), one for events $P_1(x^{\alpha_1}, t_1)$, $P_2(x^{\alpha_2}, t_2)$ whose spots $x^{\alpha_1}, x^{\alpha_2}$, at t_1 and t_2 respectively, are not aligned with the ECI center, and the other for events whose spots at those instants are aligned with the ECI center. In fact, from (1) and (2) we have [7]

$$\Omega(P_1, P_2) = \frac{1}{2}\Big[\Delta x^\delta \Delta x^\delta - (\Delta t)^2\Big]$$

$$+ m|\Delta x^\delta|\Big[\log \frac{r_2 + d_2}{r_1 + d_1} + \frac{d_1}{r_1} - \frac{d_2}{r_2}\Big] + \frac{m\,(\Delta t)^2}{|\Delta x^\delta|} \log \frac{r_2 + d_2}{r_1 + d_1} + \mathscr{O}(\varepsilon^3), \tag{3}$$

$$\Omega(P_1, P_2) = \frac{1}{2}\Big[\Delta x^\delta \Delta x^\delta - (\Delta t)^2\Big] + m\Big[|\Delta x^\delta| + \frac{(\Delta t)^2}{|\Delta x^\delta|}\Big] \log \frac{r_2}{r_1} + \mathscr{O}(\varepsilon^3), \tag{4}$$

for the first and second case respectively. In (3) and (4) $r_1^2 = |x^{\delta_1}|^2 = x^{\delta_1} x^{\delta_1}$, $r_2^2 = |x^{\delta_2}|^2 = x^{\delta_2} x^{\delta_2}$, $\Delta x^\delta = x^{\delta_2} - x^{\delta_1}$, $|\Delta x^\delta|^2 = \Delta x^\delta \Delta x^\delta$, $\Delta t = t_2 - t_1$,

$(d_1)^2 = r_1^2 - d^2$ and $(d_2)^2 = r_2^2 - d^2$, where d is the Euclidean distance from the ECI center to the straight line joining x^{δ_1} and x^{δ_2}.

3 Covariant Derivatives of $\boldsymbol{\Omega}$ and Inertial Local Reference Frames

The first covariant derivatives of Ω with respect to P_1 and P_2, Ω_{i_1}, Ω_{i_2}, are needed to derive the frequency shift formula. There are eight derivatives for each case. Thus, for (3) we have at P_1, up to $\mathscr{O}(\varepsilon^3)$,

$$\Omega_{\alpha_1} = -\triangle x^\alpha - m\frac{\triangle x^\alpha}{|\triangle x^\alpha|}\Big[\log\frac{r_2+d_2}{r_1+d_1} + \frac{d_1}{r_1} - \frac{d_2}{r_2}\Big]$$

$$-m|\triangle x^\alpha|\ \frac{d_1}{r_1^3}\ x^{\alpha_1} + m(\triangle t)^2\frac{\triangle x^\alpha}{|\triangle x^\alpha|^3}\log\frac{r_2+d_2}{r_1+d_1} - \frac{m(\triangle t)^2}{|\triangle x^\alpha|}\frac{x^{\alpha_1}}{r_1 d_1},$$

$$\Omega_{4_1} = \triangle t - 2m\frac{\triangle t}{|\triangle x^\alpha|}\log\frac{r_2+d_2}{r_1+d_1} \tag{5}$$

and similar expressions for their relatives at P_2. For (4) we have at P_2, again up to $\mathscr{O}(\varepsilon^3)$,

$$\Omega_{\alpha_2} = \triangle x^\alpha + m\Big[\frac{\triangle x^\alpha}{|\triangle x^\alpha|} - \frac{\triangle x^\alpha}{|\triangle x^\alpha|^3}(\triangle t)^2\Big]\log\frac{r_2}{r_1} + m\Big[|\triangle x^\alpha| + \frac{(\triangle t)^2}{|\triangle x^\alpha|}\Big]\frac{x^{\alpha_2}}{r_2^2},$$

$$\Omega_{4_2} = -\triangle t + 2m\frac{\triangle t}{|\triangle x^\alpha|}\log\frac{r_2}{r_1} \tag{6}$$

and similar expressions for their relatives at P_1. (Note that these derivatives are indicated with simple subscripts, that is to say, without the usual stroke.)

Let us now assume that the world line of the tracking satellite, S, is $L_1 \equiv (x^{\alpha_1}(s_1), t(s_1))$ where s_1 is the proper time of S. Let us also assume that $P_1 \in L_1$. Then, according to (1), the unit tangent vector to L_1 at P_1, A^{i_1}, is given by

$$A^{\alpha_1} = v^{\alpha_1} + \mathscr{O}(\varepsilon^3), \quad A^{4_1} = \Big(\frac{ds_1}{dt}\Big)^{-1} = 1 + \frac{m}{r_1} + \frac{1}{2}(v_1)^2 + \mathscr{O}(\varepsilon^3), \tag{7}$$

where v^{α_1} is the velocity of S at t_1 and $(v_1)^2 = v^{\alpha_1}v^{\alpha_1}$. The importance of A^{i_1} is that it characterizes the reference frames, $\lambda^{i_1}_{(\alpha)}$, co-moving with S at P_1 ($\alpha = 1, 2, 3$). In particular, for $\lambda^{i_1}_{(\alpha)}$ to be inertial the following must be satisfied [12]

$$\lambda^{4_1}_{(\alpha)} = v^{\alpha_1} + \mathscr{O}(\varepsilon^3), \quad \lambda^{\beta_1}_{(\alpha)} = \delta^\beta_\alpha - m\frac{x^{\alpha_1}x^{\beta_1}}{r_1^3} + \frac{1}{2}v^{\alpha_1}v^{\beta_1} + \mathscr{O}(\varepsilon^3). \tag{8}$$

4 The Two-Way Frequency Shift and Ranging Formulae

Let us assume that P_1 is the emission event of a Laser beam that reaches $P_2 \in L_2$, where L_2 is the world line of the target, T. Since the geodesic joining P_1 and P_2 is null, we have $\varOmega((x^{\alpha_1}, t_1), (x^{\alpha_2}, t_2)) = 0$. Then, from (3) it can be deduced that the time taken by the beam to travel from P_1 to P_2 when S (at t_1) and T (at t_2) are not aligned with the ECI center is

$$\varDelta t = |\varDelta x^{\delta}| \left\{ 1 + \frac{m}{|\varDelta x^{\delta}|} \left[2 \log \frac{r_2 + d_2}{r_1 + d_1} + \frac{d_1}{r_1} - \frac{d_2}{r_2} \right] \right\} + \mathscr{O}(\varepsilon^3), \tag{9}$$

and from (4) we have that the time is

$$\Delta t = |\Delta x^{\delta}| \left\{ 1 + \frac{2m}{|\Delta x^{\delta}|} \log \frac{r_2}{r_1} \right\} + \mathscr{O}(\varepsilon^3) \tag{10}$$

when S and T are aligned at those instants with the ECI center.

Hence, substitutions of $\varDelta t$ from (9) into (5) and from (10) into (6) give

$$\varOmega_{\alpha_1} = -\Delta x^{\alpha} - m \frac{\Delta x^{\alpha}}{|\Delta x^{\alpha}|} \left[\log \frac{r_2 + d_2}{r_1 + d_1} + \frac{d_1}{r_1} - \frac{d_2}{r_2} \right]$$

$$- m \, \frac{|\Delta x^{\alpha}|}{r_1^2} \frac{d_1}{r_1} \, x^{\alpha_1} + m \frac{\Delta x^{\alpha}}{|\Delta x^{\alpha}|} \log \frac{r_2 + d_2}{r_1 + d_1} - m |\varDelta x^{\alpha}| \frac{x^{\alpha_1}}{r_1 d_1} + \mathscr{O}(\varepsilon^3), \tag{11}$$

$$\varOmega_{\alpha_2} = \Delta x^{\alpha} + 2m |\varDelta x^{\alpha}| \frac{x^{\alpha_2}}{r_2^2} + \mathscr{O}(\varepsilon^3), \tag{12}$$

respectively, and similar expressions for their relatives.

Now, from (7) and (11) and their relatives we have

$$\varOmega_{i_1} A^{i_1} + \varOmega_{i_2} A^{i_2} = \varDelta x^{\alpha} (v^{\alpha_2} - v^{\alpha_1}) + |\varDelta x^{\alpha}| \left[\frac{m}{r_1} - \frac{m}{r_2} + \frac{1}{2} \left[(v_1)^2 - (v_2)^2 \right] \right] + \mathscr{O}(\varepsilon^3),$$

$$\varOmega_{j_1} A^{j_1} = |\varDelta x^{\alpha}| \left\{ 1 - \frac{\varDelta x^{\alpha} v^{\alpha_1}}{|\varDelta x^{\alpha}|} \right\} + \mathscr{O}(\varepsilon^2), \tag{13}$$

and from (7) and (12) and their relatives we have the same expressions.

Therefore, there is one single one-way formula for the frequency shift. In fact, taking into account (13) we have according to Synge [6] that the formula is

$$f_2 = f_1 \left\{ 1 - \frac{\varDelta x^{\alpha}}{|\varDelta x^{\alpha}|} (v^{\alpha_2} - v^{\alpha_1}) + \frac{m}{r_2} - \frac{m}{r_1} + \frac{1}{2} \left[(v_2)^2 - (v_1)^2 \right] \right.$$

$$\left. - \frac{\varDelta x^{\alpha}}{|\varDelta x^{\alpha}|} (v^{\alpha_2} - v^{\alpha_1}) \frac{\varDelta x^{\beta}}{|\varDelta x^{\beta}|} v^{\beta_1} \right\} + \mathscr{O}(\varepsilon^3), \tag{14}$$

where f_1 is the emission frequency of the beam at P_1 and f_2 the reception frequency at P_2.

Now, if we assume that the beam is reflected by T at P_2, and if we also assume that the reflected beam reaches S at $\bar{P}_1(\bar{x}^{\alpha_1}, \bar{t}_1)$, then it is straightforward to deduce from (14), and from the behavior of $\lambda^{i_1}_{(\alpha)}$ detailed in (8), that the two-way frequency shift formula is

$$\bar{f}_1 = f_1\Big\{1 - \frac{\Delta\bar{x}^\alpha}{|\Delta\bar{x}^\alpha|}(\bar{v}^{\alpha_1} - v^{\alpha_1}) + \frac{m}{\bar{r}_1} - \frac{m}{r_1} + \frac{1}{2}\big[(\bar{v}_1)^2 - (v_1)^2\big]$$

$$- \frac{\Delta\bar{x}^\alpha}{|\Delta\bar{x}^\alpha|}(\bar{v}^{\alpha_1} - v^{\alpha_1})\frac{\Delta\bar{x}^\beta}{|\Delta\bar{x}^\beta|}v^{\beta_1}\Big\} + \mathscr{O}(\varepsilon^3), \tag{15}$$

where $\bar{f}_1$ is the frequency of reception at $\bar{P}_1$; $(\bar{r}_1)^2 = \bar{x}^{\alpha_1}\bar{x}^{\alpha_1}$, $\bar{v}^{\alpha_1}$ is the velocity of S at $\bar{t}_1$, and $\Delta\bar{x}^\beta = \bar{x}^{\alpha_1} - x^{\alpha_2}$, so that $\Delta\bar{x}^\alpha/|\Delta\bar{x}^\alpha|$ is the direction of the line of sight (LOS) of T from S at $\bar{P}_1$.

So far we have not assumed that L_1 is a geodesic in space-time. Let us now assume that S is in free motion, i.e. orbiting the Earth, between two consecutive impulses. In that case L_1 is a geodesic between the events corresponding to those impulses. Let us also assume that $\hat{P}_1(\hat{x}^{\alpha_1}, \hat{t}_1)$ is the foot at L_1 of the geodesic, $\Gamma_{\hat{P}_1 P_2}$ drawn from P_2 to cut orthogonally L_1. Then $\hat{P}_1$ is an event with unknown location along L_1, which occurs between P_1 and $\bar{P}_1$, so that P_2 is in the instantaneous local space of S at $\hat{t}_1$.

In terms of the world function the post-Newtonian relative position of T with respect to S at $\hat{t}_1$, $(\hat{r}_{12})_\beta$, is given by $(\hat{r}_{12})_\beta = -\Omega_{i_1}\lambda^{i_1}_{(\beta)} + \mathscr{O}(\varepsilon^3)$, with Ω_{i_1} as in (5) or (6), and $\lambda^{i_1}_{(\beta)}$ as in (8), evaluated at $\hat{P}_1$ [9].

Since the orthogonality condition between $\Gamma_{\hat{P}_1 P_2}$ and L_1 at $\hat{P}_1$ is given by $\Omega_{i_1}A^{i_1} = \mathscr{O}(\varepsilon^3)$, then solving this condition we have $\Delta\hat{t} = \Delta\hat{x}^\alpha\hat{v}^{\alpha_1} + \mathscr{O}(\varepsilon^3)$, where $\Delta\hat{t} = t_2 - \hat{t}_1$, $\Delta\hat{x}^\alpha = x^{\alpha_2} - \hat{x}^{\alpha_1}$, and $\hat{v}^{\alpha_1}$ is the velocity of S at $\hat{t}_1$. On the other hand, it is clear from (2) that the length of $(\hat{r}_{12})_\beta$, $\hat{r}_{12}$, is $[2\Omega(\hat{P}_1, P_2)]^{1/2}$.

Therefore, for the expressions of the world function in (3) and (4) we have

$$\hat{r}_{12} = |\Delta\hat{x}^\delta| - \frac{1}{2}\frac{(\Delta\hat{x}^\delta\hat{v}^{\delta_1})^2}{|\Delta\hat{x}^\delta|} + m\Big[1 + \frac{(\Delta\hat{x}^\delta\hat{v}^{\delta_1})^2}{|\Delta\hat{x}^\delta|^2}\Big]\log\frac{r_2 + d_2}{\hat{r}_1 + \hat{d}_1} + m\Big[\frac{\hat{d}_1}{\hat{r}_1} - \frac{d_2}{r_2}\Big] + \mathscr{O}(\varepsilon^3), \tag{16}$$

$$\hat{r}_{12} = |\Delta\hat{x}^\delta| - \frac{1}{2}\frac{(\Delta\hat{x}^\delta\hat{v}^{\delta_1})^2}{|\Delta\hat{x}^\delta|} + m\Big[1 + \frac{(\Delta\hat{x}^\delta\hat{v}^{\delta_1})^2}{|\Delta\hat{x}^\delta|^2}\Big]\log\frac{r_2}{\hat{r}_1} + \mathscr{O}(\varepsilon^3), \tag{17}$$

respectively, and for both cases we have that the two-way ranging for S is

$$\hat{r}_{12} = \frac{\bar{s}_1 - s_1}{2}\Big[1 + \frac{m}{6}\Big(\frac{\delta_{\alpha\beta}}{\hat{r}_1^3} - 3\frac{\hat{x}^{\alpha_1}\hat{x}^{\beta_1}}{\hat{r}_1^5}\Big)\Delta\hat{x}^\alpha\Delta\hat{x}^\beta\Big] + \mathscr{O}(\varepsilon^3), \tag{18}$$

where s_1 is the proper time of S at P_1, i.e. the emission time of the laser as measured by S, and $\bar{s}_1$ is the proper time of S at $\bar{P}_1$, i.e. the reception time as measured by S.

5 Conclusions

Once more the world-function has revealed an efficient tool, on this occasion to derive two-way frequency shift and Laser ranging post-Newtonian formulae for SST. These formulae are suitable to increase the accuracy not only in tracking object in space, but also in the location of passive radio transmitters placed on the Earth surface or in the vicinity of the Earth by using the respective TDOA and FDOA equations.

References

1. Guelma, M., Kogan, A., Kazarian, A., Livne, A., Orenstain, M., Michalik, H., Arnold, S.: Acquisition and pointing control for inter-satellite laser communications. IEEE Trans. Aerosp. Electron. Syst. **40**(4), 1239–1248 (2004)
2. Norton, T., Conner, K., Covington, R., Ngo, H., Rink, C.: Development of reprogrammable high frame-rate detector devises for laser communication pointing, acquisition and tracking. In: Aerospace Conference 2008 IEEE, pp. 1–7 (2008)
3. Gambi, J., Rodriguez-Teijeiro, M., Garcia del Pino, M.: The post-newtonian geolocation problem by tdoa. In: Günther, M., Bartel, A., Brunk M., Schöps, S., Striebel, M. (eds.) Progress in Industrial Mathematics at ECMI 2010. Mathematics in Industry, vol. 17, pp. 489–495. Springer, Berlin (2012)
4. Gambi, J., Rodriguez-Teijeiro, M., Garcia del Pino, M., Salas, M.: Shapiro time-delay within the geolocation problem by tdoa. IEEE Trans. Aerosp. Electron. Syst. **47**(3), 1948–1962 (2011)
5. Ho, K.C., Chan, Y.T.: Solution and performance analysis of geolocation by tdoa. IEEE Trans. Aerosp. Electron. Syst. **29**(4), 1311–1322 (1993)
6. Synge, J.L.: Relativity: The General Theory. North-Holland, New York (1960)
7. Bahder, T.B.: Navigation in curved space-time. Am. J. Phys. **69**, 315–321 (2001)
8. Bahder, T.B.: Relativity of gps measurement. Phys. Rev. D **68**(6), 063005 (2013)
9. Lathrop, J.: Covariant description of motion in general relativity. Ann. Phys. **79**, 580–595 (1973)
10. Teyssandier, P., Le Poncin-Lafitte, C., Linet, B.: A universal tool for determining the time delay and the frequency shift of light: Synge's world function. In: Dittus, H., Lammerzahl, C., Turyshev, S.G. (eds.) Lasers, Clocks and Drag-Free Control. Astrophysics and Space Science Library, vol. 349, pp. 153–180. Springer, Berlin (2008)
11. Tapley, B.D., Schutz, B.E., Born, G.H.: Statistical Orbit Determination. Academic, Burlington (2004)
12. Soffel, M.H.: Relativity in Astrometry, Celestial Mechanics and Geodesy. Springer, Berlin (1989)

Part VIII
Methods

Overview

This section contains an interesting collection of ten papers with method development and modeling presented at the ECMI Conference 2012. Let me comment on two important fields with contributions in this chapter:

(a) Statistical methods for simulation. Uncertainty quantification receives a still growing interest. It is of major importance to industry, since uncertainty enters the design and production of products at many levels. In the numerical simulation, random processes have to be considered.
(b) Modeling, coupling, adaptivity. Also due to the complexity of problems, adaptivity and coupling are indispensable, nowadays. A current trend goes towards isogeometric finite elements.

The contribution by Roland Pulch (on *Polynomial-Chaos Based Methods for Differential Algebraic Equations with Random Parameters*) illustrates uncertainty quantification using an example from electric circuit simulation. The Schmitt trigger circuit is considered, where two resistances in the DAE-model are treated stochastically and may vary by 20 % (from the respective mean value). The results and statistics from stochastic collocation are discussed.

Additionally, the efficiency of determination of uncertainty is of high interest. ter Maten et al. (in *Efficient Calculation of Uncertainty Quantification*) consider the stochastic Galerkin approach with a large sequence of deterministic simulations. They use a kind of binning technique also coupled with a parameterized model reduction to enable the use of cheap approximate models. Strategies for the subspace extensions are discussed.

The authors Giacomo Aletti et al. propose and investigate a geometrical approach to represent a birth-and-growth process (*A Stochastic Geometric Framework for Dynamical Birth-and-Growth Processes. Related Statistical Analysis*). This is important for technological applications such as semiconductor crystal growth or DNA replication. Using suitable combinations of set-valued processes, they are able to avoid model problems. Furthermore, the authors are enabled to do a statistical investigation and derive related estimators.

Maria A. Churilova and Maxim E. Frolov consider a stationary reaction-diffusion equation with discontinuous coefficients. In their work (*MATLAB Implementation of Functional Type a Posteriori Error Estimates with Raviart-Thomas Approximation*), they compare adaptive algorithms for different error indicators, also with respect to efficiency.

Anh-Vu Vuong considers isogeometric finite elements with local refinement (in *Finite Element Concepts and Bezier Extraction in Hierarchical Isogeometric Analysis*). The local refinement is addressed by a hierarchical approach and a finite element concept is derived. Using Bezier extraction the relation to standard finite elements is discussed.

A steady state elasticity problem can be used to model volcano activities. This is numerically investigated by Armando Coco et al. (*A Second Order Finite-Difference Ghost-Cell Method for the Steady-State Solution of Elasticity Problems*). In their work, they derive an elasticity model and apply a Cartesian grid with a level set method to take care of the complex geometry. They present numerical results on an Etna profile using a second order stencil with a special ghost-cell treatment.

Mike A. Botchev proposes a two stage Krylov method to solve large systems of inhomogeneous ODEs (*Time-Exact Solution of Large Linear ODE Systems by Block Krylov Subspace Projections*). In the first step, a truncated SVD is used to approximate the source term (by a piecewise polynomial function). It is followed by a residual-based block Krylov method. Numerical experiments are given, which demonstrate the efficiency of the derived method.

In microwave heating processes and optics, hyperbolic matrix functions [e.g. $\cosh(A)$] are needed. The work (*Computing Hyperbolic Matrix Functions Using Orthogonal Matrix Polynomials*) by Emilio Defez et al. discusses and illustrates the approximate calculation of this matrices by the means of truncated Hermite matrix series.

The paper by Jesús Angulo considers morphology as application. This image processing technique is based on computation of the supremum and infimum operator of positive definite matrices. In his work (*Counter-Harmonic Mean of Symmetric Positive Denite Matrices: Application to Filtering Tensor-Valued Images*), he approximates the operators by a nonlinear averaging technique. Properties of this method are discussed and an example is given.

Modeling and simulation for intensive steel quenching is the topic of Sanda Blomkalna and Andris Buikis' paper (*Heat Conduction Problem for Double-Layered Ball*). For their setting, they derive a heat conduction model based on parabolic and hyperbolic PDEs. To reduces numerical difficulties, they apply a conservative averaging technique and present corresponding numerical results.

Andreas Bartel

Polynomial-Chaos Based Methods for Differential Algebraic Equations with Random Parameters

Roland Pulch

Abstract Mathematical modelling of technical applications often yields systems of differential algebraic equations (DAEs), for example, in the simulation of electric circuits or mechanical multibody problems. Imperfections of a manufacturing procedure cause undesired variations in the produced devices. These variations can be taken as uncertainties of physical parameters in a DAE model. We replace the varying parameters by random variables to achieve an uncertainty quantification. The time-dependent solution of the DAEs becomes a random process, which is expanded into a series of the polynomial chaos. We can use either a stochastic Galerkin method or a stochastic collocation technique to determine the unknown coefficient functions. The Galerkin method yields a larger coupled system to be solved once, whereas the collocation approach requires to solve the original systems many times. We present numerical simulations of an illustrative example from electrical engineering.

1 Introduction

The design and production of electronic circuits is based on numerical simulation of mathematical models. Network approaches typically yield systems of differential algebraic equations (DAEs), see [3]. Miniaturisation causes significant imperfections in the industrial production. Thus numerical simulations have to quantify these uncertainties. A common approach consists in the substitution of uncertain parameters by random variables. The solution of the DAEs becomes a random process, which can be expanded in a series of the so-called polynomial chaos.

R. Pulch (✉)
Institut für Mathematik und Informatik, Ernst-Moritz-Arndt-Universität Greifswald,
Walther-Rathenau-Str. 47, 17487 Greifswald, Germany
e-mail: pulchr@uni-greifswald.de

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_49,

Now numerical methods for the stochastic model are based on the polynomial chaos. This strategy has been applied successfully to elliptic partial differential equations in [2], where stochastic finite element methods are introduced. Ordinary differential equations as well as partial differential equations are considered in [1], for example.

2 DAE Models with Random Parameters

We focus on initial value problems of systems of DAEs

$$A(\mathbf{p})\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{x}(t,\mathbf{p}) = \mathbf{f}(t,\mathbf{x}(t,\mathbf{p}),\mathbf{p}), \qquad \mathbf{x}(t_0) = \mathbf{x}_0(\mathbf{p}). \tag{1}$$

The right-hand side $\mathbf{f}$ as well as the mass matrix A include physical parameters $\mathbf{p} = (p_1, \ldots, p_Q)^\top \in \Pi \subseteq \;\; ^{Q}$. Thus the solution depends on time and the parameters, i.e., $\mathbf{x} : [t_0, t_1] \times \Pi \to \;\; ^{N}$. We assume that A is singular for all $\mathbf{p} \in \Pi$.

The properties of a system of DAEs (1) are characterised by its index, where different concepts for the definition of the index exist, see [4]. We consider the differential index. The index is often determined by the topology and not by the values of physical parameters for models of electric circuits. However, the index can become parameter-dependent in some special cases, cf. [3].

Now let the chosen parameters exhibit uncertainties. For an uncertainty quantification, we replace the parameters by random variables $\mathbf{p} : \Omega \to \Pi$ on some probability space. We apply independent random variables with traditional distributions like Gaussian, uniform, beta, etc. For a function $u : \Pi \to \;\;$ depending on the random parameters, the expected value is denoted as $\langle u(\mathbf{p})\rangle$ if it exists. The expected value implies an inner product $\langle u(\mathbf{p})v(\mathbf{p})\rangle$ for two functions $u, v \in L^2(\Omega)$. We apply this notation also to vector-valued and matrix-valued functions by components. Now we are interested in key data of the stochastic process solving (1) like the expected value, the variance or more sophisticated quantities. In industrial problems, often failure probabilities have to be determined approximately, see [6] for an example.

3 Methods Based on Polynomial Chaos

Assuming that $x_j(t, \cdot) \in L^2(\Omega)$ for each $t \in [t_0, t_1]$ and each component $j = 1, \ldots, N$, the solution of the dynamical system (1) can be expanded into the polynomial chaos, i.e.,

$$\mathbf{x}(t,\mathbf{p}) = \sum_{i=0}^{\infty} \mathbf{v}_i(t)\Phi_i(\mathbf{p}). \tag{2}$$

Therein, the basis functions $(\Phi_i)_{i\in}$ are orthogonal polynomials with respect to the inner product of $L^2(\Omega)$. The polynomials are known from the selected probability distributions. The time-dependent coefficient functions satisfy

$$\mathbf{v}_i(t) = \langle \mathbf{x}(t,\mathbf{p})\Phi_i(\mathbf{p})\rangle \qquad \text{for each } \; i \tag{3}$$

and thus they are unknown a priori. The two main classes of numerical methods to determine the coefficient functions are: stochastic collocation techniques and stochastic Galerkin methods. A general overview can be found in [10, 11], for example.

On the one hand, the strategy of stochastic collocation applies Eq. (3), where a coefficient function is given by an expected value, i.e., a probabilistic integral. A quadrature scheme yields an approximation of the integrals. It follows that a system of DAEs (1) has to be solved for each node of the quadrature. The choice of numerical methods for initial value problems may be critical if the index of the systems depends on the parameters $\mathbf{p} \in \Pi$. A special case of collocation methods represent (quasi) Monte-Carlo simulations.

On the other hand, the stochastic Galerkin method is based on a truncation of the series (2). Inserting a truncated series in the system (1) yields a residual. The Galerkin approach determines the coefficient functions by assuming the orthogonality of the residual to the space of involved basis polynomials. It follows the larger coupled system

$$\sum_{i=0}^{M} \Big\langle A(\mathbf{p})\Phi_i(\mathbf{p})\Phi_l(\mathbf{p}) \Big\rangle \frac{\mathrm{d}}{\mathrm{d}t}\tilde{\mathbf{v}}_i(t) = \left\langle \mathbf{f}\left(t, \sum_{i=0}^{M} \tilde{\mathbf{v}}_i(t)\Phi_i(\mathbf{p}), \mathbf{p}\right) \Phi_l(\mathbf{p}) \right\rangle \tag{4}$$

for $l = 0,1,\ldots,M$ including approximations $\tilde{\mathbf{v}}_0, \tilde{\mathbf{v}}_1, \ldots, \tilde{\mathbf{v}}_M$ of the exact coefficient functions in (2). The required initial values are identified at t_0 via (3). The coupled system (4) has to be solved just once to obtain the numerical approximation.

The coupled system (4) typically represents a system of DAEs again. In some rare cases, an implicit system of ordinary differential equations appears. If the index of the coupled system coincides with the index of the original systems (1), then often the same numerical methods can be reused. However, the index of the coupled system (4) can increase or decrease in comparison to the original systems (1). An increase of the index makes the problem more complicated. Sufficient conditions for an identical index are proven for certain types of DAEs in [7, 8]. Results on the spectrum and the numerical range of involved matrices given in [9] are useful in this analysis.

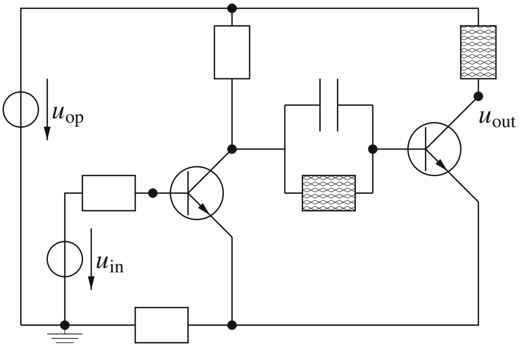


Fig. 1 Electric circuit of a Schmitt trigger with two random resistances (shown as *shaded boxes*)

4 Illustrative Example

The electric circuit of a Schmitt trigger, see Fig. 1, transforms a sinusoidal input voltage into a digital output voltage. We apply a model of this circuit given in [5], where a nonlinear system of five DAEs appears with differential index 1. This example was simulated with a random capacitance in [6]. Now we choose two resistances as random parameters with independent uniform distributions, which vary 20 % around the respective mean value.

The polynomial chaos expansion (2) of the solution involves the Legendre polynomials. We include all polynomials up to degree 3, i.e., 10 basis functions are considered. The coefficient functions are determined by a stochastic collocation using Gauss-Legendre quadrature on a grid of size 4×4 in the domain of the random parameters. The backward differentiation formula of second order resolves the initial value problems of the DAEs (1) in time.

Figure 2 shows the expected value and the standard deviation of the output voltage, which are reconstructed from the computed coefficient functions. We recognise that the uncertainties of the resistances influence only the lower value of the digital output signal.

Furthermore, Fig. 3 depicts the coefficient functions of the different polynomial degrees. Note that the expected value corresponds to the constant polynomial of degree zero. Although the relative amount of variation coincides in the two random resistances (20 %), we observe that the impact on the output voltage is much smaller for the second resistance. The magnitude of the coefficient functions decreases significantly for degree two and three, which reflects the convergence of the polynomial chaos expansion. The coefficient functions of these higher degrees exhibit an overshooting behaviour at the transitions from lower to upper values and vice versa. However, zooming indicates that the computed solutions are smooth and thus resolved correctly.

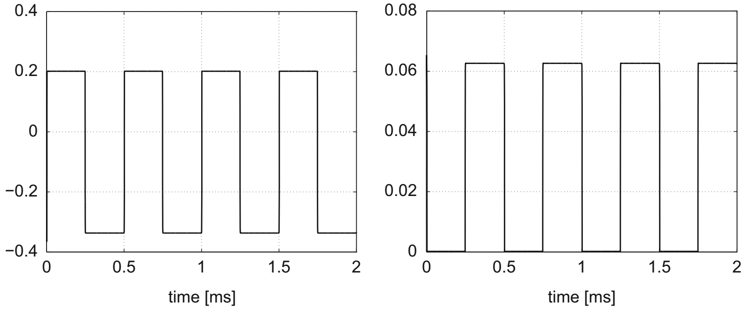


Fig. 2 Expected value (*left*) and standard deviation (*right*) of output voltage

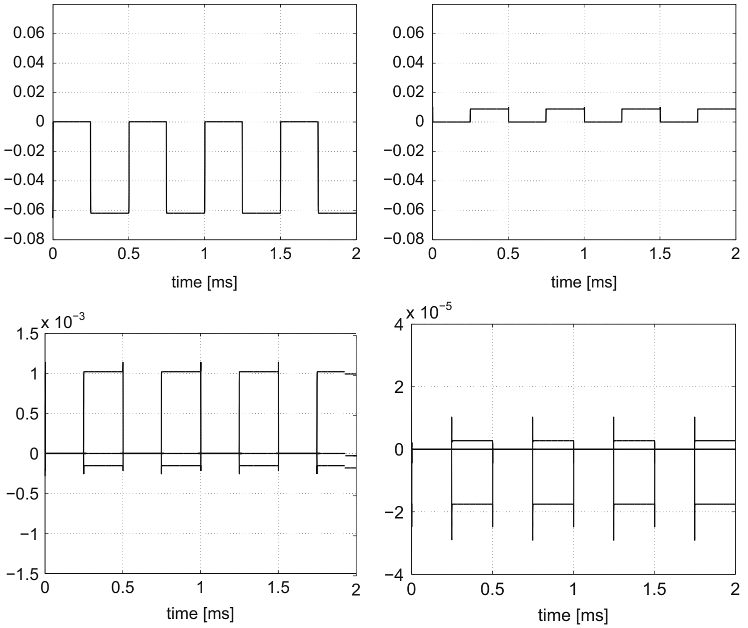


Fig. 3 Coefficient functions of output voltage. (**a**) Degree 1 for first resistance. (**b**) Degree 1 for second resistance. (**c**) Degree 2 (all functions). (**d**) Degree 3 (all functions)

References

1. Augustin, F., Gilg, A., Paffrath, M., Rentrop, P., Wever, U.: Polynomial chaos for the approximation of uncertainties: chances and limits. Eur. J. Appl. Math. **19**, 149–190 (2008)
2. Ghanem, R.G., Spanos, P.: Stochastic Finite Elements: A Spectral Approach. Springer, New York (1991)
3. Günther, M., Feldmann, U., ter Maten, E.J.W.: Modelling and discretization of circuit problems. In: Ciarlet, P.G., Schilders, W.H.A., ter Maten, E.J.W. (eds.) Numerical Methods in Electromagnetics. Handbook of Numerical Analysis, vol. XIII, pp. 523–650. Elsevier B.V., Amsterdam (2005)

4. Hairer, E., Wanner, G.: Solving Ordinary Differential Equations. Vol. 2: Stiff and Differential-Algebraic Equations, 2nd edn. Springer, Berlin (1996)
5. Kampowsky, W., Rentrop, P., Schmitt, W.: Classification and numerical simulation of electric circuits. Surv. Math. Ind. **2**, 23–65 (1992)
6. Pulch, R.: Polynomial chaos for the computation of failure probabilities in periodic problems. In: Roos, J., Costa, L. (eds.) Scientific Computing in Electrical Engineering SCEE 2008. Mathematics in Industry, vol. 14, pp. 191–198. Springer, Berlin (2010)
7. Pulch, R.: Polynomial chaos for linear differential algebraic equations with random parameters. Int. J. Uncertain. Quantif. **1**(3), 223–240 (2011)
8. Pulch, R.: Polynomial chaos for semi-explicit differential algebraic equations of index 1. Int. J. Uncertain. Quantif. **3**(1), 1–23 (2013)
9. Sonday, B., Berry, R., Debusschere, B., Najm, H.: Eigenvalues of the Jacobian of a Galerkin-projected uncertain ODE system. SIAM J. Sci. Comput. **33**(3), 1212–1233 (2011)
10. Xiu, D.: Numerical Methods for Stochastic Computations: A Spectral Method Approach. Princeton University Press, Princeton (2010)
11. Xiu, D., Hesthaven, J.S.: High order collocation methods for differential equations with random inputs. SIAM J. Sci. Comput. **27**(3), 1118–1139 (2005)

Efficient Calculation of Uncertainty Quantification

E. Jan W. ter Maten, Roland Pulch, Wil H.A. Schilders, and H.H.J.M. Janssen

Abstract We consider Uncertainty Quantification (UQ) by expanding the solution in so-called generalized Polynomial Chaos expansions. In these expansions the solution is decomposed into a series with orthogonal polynomials in which the parameter dependency becomes an argument of the orthogonal polynomial basis functions. The time and space dependency remains in the coefficients. In UQ two main approaches are in use: Stochastic Collocation (SC) and Stochastic Galerkin (SG). Practice shows that in many cases SC is more efficient for similar accuracy as obtained by SG. In SC the coefficients in the expansion are approximated by quadrature and thus lead to a large series of deterministic simulations for several parameters. We consider strategies to efficiently perform this sequence of deterministic simulations within SC.

E.J.W. ter Maten (✉)
Bergische Universität Wuppertal, FB C, AMNA, Gaußstraße 20, 42119 Wuppertal, Germany

Department of Mathematics and Computer Science (CASA), Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands
e-mail: Jan.ter.Maten@math.uni-wuppertal.de; E.J.W.ter.Maten@tue.nl

R. Pulch
Department of Mathematics and Computer Science, Ernst-Moritz-Arndt Universität Greifswald, Walther-Rathenau-Straße 47, 17487 Greifswald, Germany
e-mail: Roland.Pulch@uni-greifswald.de

W.H.A. Schilders
Department of Mathematics and Computer Science (CASA), Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands
e-mail: W.H.A.Schilders@tue.nl

H.H.J.M. Janssen
NXP Semiconductors, High Tech Campus 46, 5656 AE Eindhoven, The Netherlands
e-mail: Rick.Janssen@nxp.com

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_50,

1 Polynomial Chaos for Dynamical Systems with Random Parameters

We will denote parameters by $\mathbf{p} = (p_1, \ldots, p_q)^T$ and assume a probability space $(\Omega, \mathscr{A}, \mathscr{P})$ given where $\mathscr{A}$ represents a σ-algebra, $\mathscr{P} : \mathscr{A} \rightarrow \mathbb{R}$ is a measure and $\mathbf{p} = \mathbf{p}(\omega) : \Omega \rightarrow Q \subseteq \mathbb{R}^q$. Here we will assume that the p_i are independent.

For a function $f : Q \rightarrow \mathbb{R}$, the mean or expected value is defined by

$$\mathbf{E}_p[f(\mathbf{p})] =< f >= \int_\Omega f(\mathbf{p}(\omega))\mathrm{d}\mathscr{P}(\omega) = \int_Q f(\mathbf{p})\ \rho(\mathbf{p})\mathrm{d}\mathbf{p}. \tag{1}$$

The specific probability distribution density is given by the function $\rho(\mathbf{p})$. Because $\mathscr{P}(\Omega) = 1$, we have $< 1 >= 1$. A bilinear form (with associated norm L^2_ρ) is defined by

$$< f, g >= \int_Q f(\mathbf{p})\ g(\mathbf{p})\ \rho(\mathbf{p})\mathrm{d}\mathbf{p} =< f\ \ g > . \tag{2}$$

The last form is convenient when products of more functions are involved. Similar definitions hold for vector- or matrix-valued functions $\mathbf{f} : Q \rightarrow \mathbb{R}^{m\times n}$.

We assume a complete orthonormal basis of polynomials $(\phi_i)_{i\in\mathbb{N}}, \phi_i : \mathbb{R}^q \rightarrow \mathbb{R}$, given with $< \phi_i, \phi_j >= \delta_{ij}$ $(i, j, \geq 0)$. When $q = 1$, ϕ_i has degree i. To treat a uniform distribution (i.e., for studying effects caused by robust variations) Legendre polynomials are optimal in some sense; for a Gaussian distribution one can use Hermite polynomials [17, 28]. A polynomial ϕ_i on $\mathbb{R}^q$ can be defined from one-dimensional polynomials: $\phi_i(\mathbf{p}) = \prod_{d=1}^q \phi_{i_d}(p_d)$. Actually i orders a vector $\mathbf{i} = (i_1, \ldots, i_q)^T$.

We will denote a dynamical system by

$$\mathbf{F}(\mathbf{x}(t, \mathbf{p}), t, \mathbf{p}) = \mathbf{0}, \quad \text{for } t \in [t_0, t_1]. \tag{3}$$

Here $\mathbf{F}$ may contain differential operators. The solution $\mathbf{x} \in \mathbb{R}^n$ depends on t and on $\mathbf{p}$. In addition initial and boundary values are assumed. In general these may depend on $\mathbf{p}$ as well.

A solution $\mathbf{x}(t, \mathbf{p}) = (x_1(t, \mathbf{p}), \ldots, x_n(t, \mathbf{p}))^T$ of the dynamical system becomes a random process. We assume that second moments are finite: $< x_j^2(t, \mathbf{p}) > < \infty$, for all $t \in [t_0, t_1]$ and $j = 1, \ldots, n$. We express $\mathbf{x}(t, \mathbf{p})$ in a Polynomial Chaos expansion

$$\mathbf{x}(t, \mathbf{p}) = \sum_{i=0}^{\infty} \mathbf{v}_i(t)\ \phi_i(\mathbf{p}), \tag{4}$$

where the coefficient functions $\mathbf{v}_i(t)$ are defined by

$$\mathbf{v}_i(t) =< \mathbf{x}(t,\mathbf{p}), \phi_i(\mathbf{p}) > . \tag{5}$$

A finite approximation $\mathbf{x}^m(t,\mathbf{p})$ to $\mathbf{x}(t,\mathbf{p})$ is defined by

$$\mathbf{x}^m(t,\mathbf{p}) = \sum_{i=0}^{m} \mathbf{v}_i(t)\ \phi_i(\mathbf{p}). \tag{6}$$

For traditional random distributions $\rho(.)$ convergence rates for $||\mathbf{x}(t,.) - \mathbf{x}^m(t,.)||$ for functions $\mathbf{x}(t,\mathbf{p})$, that depend smoothly on $\mathbf{p}$, are known (see [2] and [28] for an expansion in Hermite or in Legendre polynomials, respectively). For more general distributions $\rho(.)$ convergence may not be true. For instance, polynomials in a lognormal variable are not dense in L^2_ρ. For convergence one requires that the probability measure is uniquely determined by its moments [8]. One at least needs that the expected value of each polynomial has to exist.

The integrals (5) can be computed by (quasi) Monte Carlo, or by multi-dimensional quadrature. We assume quadrature grid points $\mathbf{p}^k$ and quadrature weights w_k, with $0 \le k \le K$, such that

$$\mathbf{v}_i(t) =< \mathbf{x}(t,\mathbf{p}), \phi_i(\mathbf{p}) > \approx \sum_{k=0}^{K} w_k\ \mathbf{x}(t,\mathbf{p}^k)\ \phi_i(\mathbf{p}^k). \tag{7}$$

Typically, Gaussian quadrature is used with corresponding weights. We solve (3) for $\mathbf{x}(t,\mathbf{p}^k)$, $k = 0,\ldots,K$ ($K+1$ deterministic simulations). Here any suitable numerical solver for (3) can be used. By post-processing we determine the $\mathbf{v}_i(t)$ in (7).

As alternative approach, Stochastic Galerkin can be used. Then the sum (6) is put into Eq. (3) and the residues are made orthogonal to the basis functions. This results into one big system for the coefficient functions $\mathbf{v}_i(t)$ [17,22,28]. Due to averaging, this system does not depend on particular parameter values anymore.

2 Statistical Information and Sensitivity

We note that the expansion $\mathbf{x}^m(t,\mathbf{p})$, see (6), gives full detailed information when varying $\mathbf{p}$; it serves as a response surface model. From this the actual (and probably biased) range of solutions can be determined. These can be different from envelope approximations based on mean and variances.

Let ϕ_0 be the polynomial that is constant c; orthonormality implies that $c = 1$. By further use of the orthogonality, the mean of $\mathbf{x}(t,\mathbf{p})$ is given by

$$\mathbf{E}_p[\mathbf{x}(t,\mathbf{p})] \approx \int_Q \mathbf{x}^m(t,\mathbf{p})\rho(\mathbf{p})\,\mathrm{d}\mathbf{p} =< \mathbf{x}^m(t,\mathbf{p})\,1 >=< \mathbf{x}^m(t,\mathbf{p})\,\phi_0 >= \mathbf{v}_0(t) \tag{8}$$

(for the finite expansion with exact coefficients the equality sign holds). This involves all p_k together. One may want to consider effects of p_i and p_j separately. This restricts the parameter space $Q \subseteq \mathbb{R}^q$ to a one-dimensional subset with individual distribution densities $\rho_i(p)$ and $\rho_j(p)$. A covariance function of $\mathbf{x}(t,\mathbf{p})$ can also be easily expressed

$$\mathbf{E}_p[(\mathbf{x}(t_1,\mathbf{p}) - \mathbf{E}_p[\mathbf{x}(t_1,\mathbf{p})])^T\,(\mathbf{x}(t_2,\mathbf{p}) - \mathbf{E}_p[\mathbf{x}(t_2,\mathbf{p})])] \approx \sum_{i=1}^{m} \mathbf{v}_i^T(t_1)\mathbf{v}_i(t_2). \tag{9}$$

Having a gPC expansion also the sensitivity (matrix) w.r.t. $\mathbf{p}$ is easily obtained

$$\mathbf{S}_p(t,\mathbf{p}) = \left[\frac{\partial \mathbf{x}(t,\mathbf{p})}{\partial \mathbf{p}}\right] \approx \sum_{i=0}^{m} \mathbf{v}_i(t)\,\frac{\partial \phi_i(\mathbf{p})}{\partial \mathbf{p}}. \tag{10}$$

From this a relative sensitivity can be defined by

$$\mathbf{S}_p^r(t,\mathbf{p}) = \left[\left(\frac{\partial x_i(t,\mathbf{p})}{\partial p_j}\cdot\frac{p_j}{x_i(t,\mathbf{p})}\right)_{ij}\right] = \mathbf{S}_p(t,\mathbf{p}) \circ \left[\left(\frac{p_j}{x_i(t,\mathbf{p})}\right)_{ij}\right]. \tag{11}$$

It describes the amplification of a relative error in $\mathbf{p}_j$ to the relative error in $\mathbf{x}_i(t,\mathbf{p})$ (here $\circ$ denotes the Hadamard product of two matrices).

The sensitivity matrix also is subject to stochastic variations. With a gPC expansion it is possible to determine a mean global sensitivity matrix by

$$\mathbf{S}_p(t) = \mathbf{E}_p\left[\frac{\partial \mathbf{x}(t,\mathbf{p})}{\partial \mathbf{p}}\right] \approx \sum_{i=0}^{m} \mathbf{v}_i(t) \int_Q \frac{\partial \phi_i(\mathbf{p})}{\partial \mathbf{p}}\,\rho(\mathbf{p})\,\mathrm{d}\mathbf{p}. \tag{12}$$

Note that the integrals at the right-hand side can be determined in advance and stored in tables.

3 Failure and Tolerance Analysis

Failure may be defined after introducing a criterion function $g(t,\mathbf{x}(t,\mathbf{p}))$, e.g., $g(t,\mathbf{x}(t,\mathbf{p})) \equiv \mathbf{x}(t,\mathbf{p}) - \theta$, with a threshold θ. Then failure is measured by a function χ

$$\chi(g(t,\mathbf{x}(t,\mathbf{p}))) = \begin{cases} 0 \text{ for } & g > 0 \\ 1 \text{ for } & g \leq 0 \end{cases}. \tag{13}$$

The Failure Probability is then

$$P_{\mathrm{F}}(t) = \int \chi(g(t, \mathbf{x}(t, \mathbf{p})))\, \rho(\mathbf{p})\, \mathrm{d}\mathbf{p} \approx \int \chi(g(t, \mathbf{x}^m(t, \mathbf{p})))\, \rho(\mathbf{p})\, \mathrm{d}\mathbf{p}. \qquad (14)$$

In (14) the expression at the left of the approximation symbol may be obtained using Monte Carlo methods for the original problems, probably speeded up by methods like Importance Sampling [7,26]. In [26], after applying results from Large Deviations Theory, also realistic, but sharp, upper bounds were derived involving the number of samples that have to be drawn.

Alternatively, after having spent the effort in determining $\mathbf{x}^m(t, \mathbf{p})$ in (6) the evaluation for different $\mathbf{p}$ is surprisingly cheap. Monte Carlo, Quasi Monte Carlo, Importance Sampling can be used again for statistics, but at a much lower price [21]. Determination of Failure Probability, however, deserves additional attention, because the expansion $\mathbf{x}^m(t, \mathbf{p})$ in (6) may be less accurate in areas of interest for this kind of statistics. The software tool RODEO of Siemens AG seems to be the only industrial implementation of failure probability calculation that fits within the polynomial chaos framework [20].

A hybrid method to compute small failure probabilities that exploits surrogate models has been introduced by [18]. Their method can be slightly generalized as follows. By this we can determine the effect of approximation on the Failure Probability. To each sample $\mathbf{z}^i$ we assume a numerically obtained approximation $\tilde{\mathbf{z}}^i$. In addition g is approximated by $\tilde{g}$. The probabilities one checks are

$$\tilde{P}_\varepsilon(t) = \int \chi\big(\tilde{g}(t, \tilde{\mathbf{z}}(t, \mathbf{p})) + \varepsilon\big)\, \rho(\mathbf{p})\mathrm{d}\mathbf{p},$$

$$\tilde{Q}_\varepsilon(t) = \int \chi\big(-\tilde{g}(t, \tilde{\mathbf{z}}(t, \mathbf{p})) - \varepsilon\big)\, \chi\big(\tilde{g}(t, \tilde{\mathbf{z}}(t, \mathbf{p})) - \varepsilon\big)\, \chi\big(g(t, \mathbf{z}(t, \mathbf{p}))\big)\, \rho(\mathbf{p})\mathrm{d}\mathbf{p}.$$

Note that in $\tilde{P}_\varepsilon(t)$ one deals with $\tilde{g}(t, \tilde{\mathbf{z}}(t, \mathbf{p})) \leq -\varepsilon$. In $\tilde{Q}_\varepsilon$ the first two factors involve $|\tilde{g}(t, \tilde{\mathbf{z}}(t, \mathbf{p}))| \leq \varepsilon$. The two quantities result in a Failure Probability $\tilde{P}_F(t) = \tilde{P}_\varepsilon(t) + \tilde{Q}_\varepsilon(t)$. The impact of the last factor in $\tilde{Q}_\varepsilon$ is that one additionally evaluates the exact $g(t, \mathbf{z}(t, \mathbf{p}))$ (or one approximates it more accurately) when its approximation $\tilde{g}(t, \tilde{\mathbf{z}}(t, \mathbf{p}))$ is small.

Now let

$$\tilde{D}_\varepsilon(t) = \int_{|\tilde{g}(\tilde{\mathbf{z}}(t,\mathbf{p})) - g(\mathbf{z}(t,\mathbf{p}))| > \varepsilon} \rho(\mathbf{p})\mathrm{d}\mathbf{p}$$

be the combined quality of both approximations. One should be able to make this small. Note that $|\tilde{g}(\tilde{\mathbf{z}}(t, \mathbf{p})) - g(\mathbf{z}(t, \mathbf{p}))| < |\tilde{g}(\tilde{\mathbf{z}}(t, \mathbf{p})) - \tilde{g}(\mathbf{z}(t, \mathbf{p}))| + |\tilde{g}(\mathbf{z}(t, \mathbf{p})) - g(\mathbf{z}(t, \mathbf{p}))|$. The first term needs Lipschitz continuity for $\tilde{g}$ to deal with $\tilde{\mathbf{z}}(t, \mathbf{p}) - \mathbf{z}(t, \mathbf{p})$, the second one deals with $|\tilde{g} - g|$. By this and exploiting the finite probability measure one may assume, f.i., that $\tilde{D}_\varepsilon(t) < \delta P_F(t)$, for $0 < \delta < 1$.

One can proof (similar to [18], Theorem 4.1)

$$|\tilde{P}_F(t) - P_F(t)| < \tilde{D}_\varepsilon(t) < \delta P_F(t). \tag{15}$$

One may order the (remaining) approximative samples $\tilde{g}^{(i)}(t) = \tilde{g}(t, \tilde{\mathbf{z}}(t, \mathbf{p}^i))$ according to $|\tilde{g}^{(i)}(t)|$ and replace the smallest ones by $g^{(i)}(t) = g(t, \tilde{\mathbf{z}}(t, \mathbf{p}^i))$ and reduce the set of (remaining) approximative samples accordingly. One can stop if the Failure Probability does not change that much anymore [18]. This procedure resembles algorithmic steps in [20].

4 Strategies for Efficient Stochastic Collocation

Stochastic Collocation implies that the problem has to be solved for a sequence (or sweep) of parameter settings $\mathbf{p}^0, \ldots, \mathbf{p}^K$. One can obtain some benefit by exploiting knowledge derived before.

In [16], the parameters $\mathbf{p}^k$ are grouped in blocks and in each block one simulation is made, say for $\mathbf{p}^{k_0}$. At the subset of the $\mathbf{p}^{k_0}$ the solution $\mathbf{x}(t, \mathbf{p}^{k_0})$ is calculated at some higher accuracy (f.i., with a smaller stepsize h_0). The solution is used to estimate the truncation error of the time integration for $\mathbf{x}(t, \mathbf{p}^k)$. One determines the residue $\mathbf{r}(t, \mathbf{x}(t, \mathbf{p}^{k_0}))$ for $\mathbf{x}(t, \mathbf{p}^{k_0})$ using the same integration method as intended to be used for $\mathbf{x}(t, \mathbf{p}^k)$, with stepsize h, but using $\mathbf{p}^{k_0}$ in all expressions. By this the discretization error for $\mathbf{x}(t, \mathbf{p}^k)$ is estimated automatically when $\mathbf{p}^{k_0}$ is close to $\mathbf{p}^k$. By subtracting $\mathbf{r}(t, \mathbf{x}(t, \mathbf{p}^{k_0}))$ from the equations for $\mathbf{x}(t, \mathbf{p}^k)$, one may expect a larger stepsize h to be used then without this modification. Note that

$$\begin{aligned}
&\mathbf{r}\big(t, \mathbf{x}(t, \mathbf{p}^k)\big) - \mathbf{r}\big(t, \mathbf{x}(t, \mathbf{p}^{k_0})\big) \\
&\qquad = \frac{\partial \mathbf{r}}{\partial \mathbf{x}}\big(t, \mathbf{x}(t, \mathbf{p}^k)\big) \cdot \big(\mathbf{x}(t, \mathbf{p}^k) - \mathbf{x}(t, \mathbf{p}^{k_0})\big) + \mathscr{O}(||\mathbf{p}^k - \mathbf{p}^{k_0}||^2) \\
&\qquad = \frac{\partial \mathbf{r}}{\partial \mathbf{x}}\big(t, \mathbf{x}(t, \mathbf{p}^k)\big) \cdot \frac{\partial \mathbf{x}}{\partial \mathbf{p}} \cdot (\mathbf{p}^k - \mathbf{p}^{k_0}) + \mathscr{O}(||\mathbf{p}^k - \mathbf{p}^{k_0}||^2).
\end{aligned} \tag{16}$$

Here the first factor equals the last Jacobian. The second factor is the sensitivity matrix of the solution with respect to the parameter variation [13, 14]; it can be estimated from its value at $\mathbf{p}^{k_0}$. When the usual error control is too pessimistic, this approach may be an alternative.

In [25] also first the solution for $\mathbf{p}^{k_0}$ is calculated for the next time discretization point and used as predictor for the time step integration of the problems for other $\mathbf{p}^k$. Here as well the prediction can be improved by additional sensitivity estimates. If parameters are values for capacitors, inductors or resistors they are model bound. Then hierarchy techniques [11] can be exploited to by-pass certain parts of the circuit during the Newton iteration. Of course, the time step integration for the other $\mathbf{p}^k$ can be solved in parallel.

In [2,20] one builds an estimator by a moderately-sized gPC approximation

$$\tilde{\mathbf{x}}^{m'} = \sum_{i=0}^{m'} \tilde{\mathbf{v}}_i(t)\phi_i(\mathbf{p}). \tag{17}$$

As before the best $\tilde{\mathbf{v}}_i(t)$ has $\tilde{\mathbf{v}}_i(t) = \int \mathbf{x}(t,\mathbf{p})\rho(\mathbf{p})\mathrm{d}\mathbf{p}$. We can approximate them by a Least Squares procedure at each time t

$$\begin{aligned}\min_{\tilde{\mathbf{v}}_i(t)} \int (\mathbf{x}(t,\mathbf{p}) - \tilde{\mathbf{x}}^{m'})^2\rho(\mathbf{p})\mathrm{d}\mathbf{p} &\approx \min_{\tilde{\mathbf{v}}_i(t)} \sum_{k=0}^{K} w_k \left(\mathbf{x}(t,\mathbf{p}^k) - \sum_{i=0}^{m'} \tilde{\mathbf{v}}_i(t)\phi_i(\mathbf{p}^k)\right)^2 \\ &= \min_{\mathbf{y}} ||\mathbf{M}\mathbf{y} - \mathbf{b}||_2^2, \quad \text{where} \end{aligned} \tag{18}$$

$$\mathbf{M} = \left(\begin{pmatrix} \sqrt{w_0} & & \\ & \ddots & \\ & & \sqrt{w_K} \end{pmatrix} \begin{pmatrix} \phi_0(\mathbf{p}^0) & \dots & \phi_{m'}(\mathbf{p}^0) \\ \vdots & & \vdots \\ \phi_0(\mathbf{p}^K) & \dots & \phi_{m'}(\mathbf{p}^K) \end{pmatrix}\right) \otimes \mathbf{I}_n,$$

$$\mathbf{b} = (\sqrt{w_0}\mathbf{x}^T(t,\mathbf{p}^0), \dots, \sqrt{w_K}\mathbf{x}^T(t,\mathbf{p}^K))^T,$$

$$\mathbf{y} = (\tilde{\mathbf{v}}_0^T(t), \dots, \tilde{\mathbf{v}}_{m'}^T(t))^T.$$

In [2,20] one applies a Least Squares procedure (18) not for the final solution values $\mathbf{x}(t,\mathbf{p}^0)$, ..., $\mathbf{x}(t,\mathbf{p}^K)$, but after splitting the sequence in already determined values $\mathbf{x}(t,\mathbf{p}^0), \dots, \mathbf{x}(t,\mathbf{p}^{\tilde{K}})$, and approximated values $\tilde{\mathbf{x}}(t,\mathbf{p}^{\tilde{K}+1}), \dots, \tilde{\mathbf{x}}(t,\mathbf{p}^K)$. Clearly the error $\Delta\mathbf{y}$ is determined by $\Delta\mathbf{y} = \mathbf{M}^+\Delta\mathbf{b}$, where the $\Delta\mathbf{b}$ comes from the errors in the $\mathbf{z}_k \equiv \sqrt{w_k}\tilde{\mathbf{x}}(t,\mathbf{p}^k)$, $k = \tilde{K}+1, \dots, K$. One can sort the $\mathbf{z}_k$ and update the $\tilde{\mathbf{x}}(t,\mathbf{p}^k)$ to final solution values $\mathbf{x}(t,\mathbf{p}^k)$ for the $\Delta\tilde{K}$ largest $\mathbf{z}_k$. This allows to update $\tilde{\mathbf{x}}^{m'}$ iteratively and the approximation values $\tilde{\mathbf{x}}(t,\mathbf{p}^{\tilde{K}+1})$, ..., $\tilde{\mathbf{x}}(t,\mathbf{p}^K)$ may come from the previous $\tilde{\mathbf{x}}^{m'}$. Interpreting the values $\mathbf{x}(t,\mathbf{p}^0)$, ..., $\mathbf{x}(t,\mathbf{p}^{\tilde{K}})$, $\tilde{\mathbf{x}}(t,\mathbf{p}^{\tilde{K}+1})$, ..., $\tilde{\mathbf{x}}(t,\mathbf{p}^K)$ as coming from a function $\hat{\mathbf{x}}(t,\mathbf{p})$. Then for $\hat{\mathbf{x}}(t,\mathbf{p})$ the mean, variance and sensitivity simply follow from the gPC expansion. The mean and variance can be used to check their change after an update. Note that here one can exploit the average sensitivity as well, which also simply follows from the gPC expansion. In this way one can assure that one includes dominant parameters first. We finally note that the approximations may come from (parameterized) Model Order Reduction.

5 Parameterized Model Order Reduction

Model Order Reduction (MOR) techniques can be applied to reduce the size of the deterministic problems that have to be simulated using SC. For good general introductions we refer to [1,5,23]. For parameterized MOR we refer to [3,9,10,24].

We consider a linear system for circuit equations with capacitance matrix $\mathbf{C} = \mathbf{C}(\mathbf{p})$, conductance matrix $\mathbf{G} = \mathbf{G}(\mathbf{p})$ and source $\mathbf{u}(t) = \mathbf{u}(t, \mathbf{p})$ that involve parameters $\mathbf{p}$,

$$\mathbf{C}(\mathbf{p})\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} + \mathbf{G}(\mathbf{p})\mathbf{x}(t,\mathbf{p}) = \mathbf{B}\mathbf{u}(t,\mathbf{p}), \tag{19}$$
$$\mathbf{y}(t,\mathbf{p}) = \mathbf{B}^T\mathbf{x}(t,\mathbf{p}).$$

Here $\mathbf{y}(t,\mathbf{p})$ is some output result. This separation of $\mathbf{p}$ and $\mathbf{x}$ in the expressions in each equation of (19) is quite common in circuit simulation (capacitors, inductors and resistors depend on $\mathbf{p}$), but for more general expressions (like when using controlled sources) this may require some organization in the evaluation tree of the expression handler. In [10] a parameterized system in the frequency domain is considered in which the coefficient matrices have been expanded. We consider, however, the nonexpanded form. Let s be the (angular) frequency. It is assumed that a set $\mathbf{p}^1, \mathbf{p}^2, \ldots, \mathbf{p}^K$ is given in advance, together with frequencies $s_1, s_2, \ldots, s_K$. In our case the $\mathbf{p}^1, \mathbf{p}^2, \ldots, \mathbf{p}^K$ can come from quadrature points in SC. Let $\Psi^k = (s_k, \mathbf{p}^k)$. Furthermore, let $\mathbf{A} = s\mathbf{C}(\mathbf{p}) + \mathbf{G}(\mathbf{p})$ and $\mathbf{A}\mathbf{X} = \mathbf{B}$, where $\mathbf{X}$ is the Laplace Transform of $\mathbf{x}$. Similarly, let $\mathbf{A}_k = \mathbf{A}(\Psi^k) = s_k\mathbf{C}(\mathbf{p}^k) + \mathbf{G}(\mathbf{p}^k)$ and $\mathbf{A}_k\mathbf{X}_k = \mathbf{B}$.

A projection matrix $\mathbf{V}$ (with orthonormal columns $\mathbf{v}_i$) is searched for such that $\mathbf{X}(s,\mathbf{p}) \approx \bar{\mathbf{X}}(s,\mathbf{p}) \equiv \mathbf{V}\hat{\mathbf{X}}(s,\mathbf{p}) \equiv \sum_{i=1}^{K'} \alpha_i(s,\mathbf{p})\mathbf{v}_i$.

We assume that we have already found some part of the (orthonormal) basis, $\mathbf{V} = (\mathbf{v}_1, \ldots, \mathbf{v}_k)$. Then for any Ψ^j that was not selected before to extend the basis the actual error is formally given by $\mathbf{E}^j = \mathbf{X}(\Psi^j) - \sum_{i=1}^{k} \alpha_i(\Psi^j)\mathbf{v}_i$ and thus for the residue we have $\mathbf{R}^j = \mathbf{A}_j\mathbf{E}^j = \mathbf{B} - \sum_{i=1}^{k} \alpha_i(\Psi^j)\mathbf{A}_j\mathbf{v}_i$. Note that the residues deal with $\mathbf{B}$ and with $\mathbf{x}$ and not with the effect in $\mathbf{y}$. For UQ one may consider a two-sided projection here, which will bring in the effect due to the quadrature weights. The method of [10] was used in [6] (using expansions of the matrices in moments of $\mathbf{p}$). In [6] the parameter variation in $\mathbf{C}$ and $\mathbf{G}$ did come from parameterized layout extraction of RC circuits. In the extraction it was assumed that $\mathbf{B}$, as well as the fill-in patterns of $\mathbf{C}(\mathbf{p})$ and of $\mathbf{G}(\mathbf{p})$, did not depend on $\mathbf{p}$. When $\mathbf{B}$ also becomes dependent on $\mathbf{p}$ one should determine a basis for the range of $\mathbf{B}(\mathbf{p})$. In fact one needs MOR for multi-input, multi-output [4, 15, 27].

The selection of the next parameter introduces a notion of "dominancy" from an algorithmic point of view: this parameter most significantly needs extension of the Krylov subspace. To invest for this parameter will automatically reduce work for other parameters (several may even drop out of the list because of zero residues).

We finally describe two ideas to include sensitivity in parameterized MOR. One can calculate the sensitivities of the solution of the reduced system by adjoint techniques as described by [13, 14]. Alternatively one can exploit the sensitivity indication based on the gPC expansion of the combined list of exact evaluations and outcomes of approximations as mentioned in Sect. 4.

If first order sensitivity matrices are available for $\mathbf{C}(\mathbf{p}) = \mathbf{C}_0(\mathbf{p}_0) + \mathbf{C}'(\mathbf{p}_0)\mathbf{p}$ and for $\mathbf{G}(\mathbf{p}) = \mathbf{G}_0(\mathbf{p}_0) + \mathbf{G}'(\mathbf{p}_0)\mathbf{p}$ one can apply a Generalized Singular Value

Decomposition [12] to both pairs $(\mathbf{C}_0^T(\mathbf{p}_0), [\mathbf{C}']^T(\mathbf{p}_0))$ and $(\mathbf{G}_0^T(\mathbf{p}_0), [\mathbf{G}']^T(\mathbf{p}_0))$. In [19] this was applied in MOR for linear coupled systems. The low-rank approximations for $\mathbf{C}'(\mathbf{p}_0)$ and $\mathbf{G}'(\mathbf{p}_0)$ give way to increase the basis for the columns of $\mathbf{B}$ of the source function. Note that by this one automatically will need MOR methods that can deal with many terminals [4, 15, 27].

6 Conclusion

We have derived strategies to efficiently determine the coefficients in generalized polynomial chaos expansions. When determined by Stochastic Collocation and numerical quadrature this leads to a large number of deterministic simulations. Parameterized Model Order Reduction is a natural choice to reduce sizes. In selecting a next parameter for the subspace extension different options have been described: residue size and options for sensitivity. For UQ however, one should involve the influence of the quadrature weights and one may check the contribution to global statistical quantities. A related algorithm can be used for Failure Probabilities.

Acknowledgements The first and last author did part of the work within the project ARTEMOS (Ref. 270683-2), http://www.artmeos.eu/ (ENIAC Joint Undertaking).

References

1. Antoulas, A.C.: Approximation of Large-Scale Dynamical Systems. SIAM Publications, Philadelphia (2005)
2. Augustin, F., Gilg, A., Paffrath, M., Rentrop, P., Wever, U.: Polynomial chaos for the approximation of uncertainties: chances and limits. Eur. J. Appl. Math. **19**, 149–190 (2008)
3. Baur, U., Beattie, C., Benner, P., Gugercin, S.: Interpolatory projection methods for parameterized model reduction. SIAM J. Comput. **33**, 2489–2518 (2011)
4. Benner, P., Schneider, A.: Model reduction for linear descriptor systems with many ports. In: Günther, M., Bartel, A., Brunk, M., Schöps, S., Striebel, M. (eds.) Progress in Industrial Mathematics at ECMI 2010. Mathematics in Industry, vol. 17, pp. 137–143. Springer, Berlin (2012)
5. Benner, P., Hinze, M., ter Maten, E.J.W. (eds.): Model Reduction for Circuit Simulation. Lecture Notes in Electrical Engineering, vol. 74. Springer, Berlin (2011)
6. Bi, Y., van der Kolk, K.-J., Fernández Villena, J., Silveira, L.M., van der Meijs, N.: Fast statistical analysis of RC nets subject to manufacturing variabilities. In: Proceedings of DATE 2011, Grenoble, 14–18 March 2011
7. Doorn, T.S., Croon, J.A., ter Maten, E.J.W., Di Bucchianico, A.: A yield centric statistical design method for optimization of the SRAM active column. In: Proceedings of the IEEE ESSCIRC'09, 35th European Solid-State Circuits Conference, Athens, pp. 352–355 (2009)
8. Ernst, O.G., Mugler, A., Starkloff, H.-J., Ullmann, E.: On the convergence of generalized polynomial chaos expansions. ESAIM Math. Model. Numer. Anal. **46**, 317–339 (2012)

9. Feng, L., Benner, P.: A robust algorithm for parametric model order reduction. PAMM Proc. Appl. Math. Mech. **7**, 1021501–1021502 (2007)
10. Fernández Villena, J., Silveira, L.M.: Multi-dimensional automatic sampling schemes for multi-point modeling methodologies. IEEE Trans. Comput. Aided Des. Integr. Circuits Syst. **30**(8), 1141–1151 (2011)
11. Fijnvandraat, J.G., Houben, S.H.M.J., ter Maten, E.J.W., Peters, J.M.F.: Time domain analog circuit simulation. J. Comput. Appl. Math. **185**, 441–459 (2006)
12. Golub, G.H., Van Loan, C.F.: Matrix Computations, 3rd edn. Johns Hopkins University Press, Baltimore (1996)
13. Ilievski, Z.: Model order reduction and sensitivity analysis. Ph.D. thesis, TU Eindhoven (2010). http://alexandria.tue.nl/extra2/201010770.pdf
14. Ilievski, Z., Xu, H., Verhoeven, A., ter Maten, E.J.W., Schilders, W.H.A., Mattheij, R.M.M.: Adjoint transient sensitivity analysis in circuit simulation. In: Ciuprina, G., Ioan, D. (eds.) Scientific Computing in Electrical Engineering SCEE 2006. Mathematics in Industry, vol. 11, pp. 183–189. Springer, Berlin (2007)
15. Ionutiu, R.: Model order reduction for multi-terminal systems - with applications to circuit simulation. Ph.D. thesis, TU Eindhoven (2011). http://alexandria.tue.nl/extra2/716352.pdf
16. Jansen, L., Tischendorf, C.: Effective numerical computation of parameter dependent problems. In: Michielsen, B.L., Poirier, J.-R. (eds.) Scientific Computing in Electrical Engineering SCEE 2010. Mathematics in Industry, vol. 16, pp. 49–57. Springer, Berlin (2012)
17. Le Maître, O.P., Knio, O.M.: Spectral Methods for Uncertainty Quantification, with Applications to Computational Fluid Dynamics. Springer, Science+Business Media B.V., Dordrecht (2010)
18. Li, J., Xiu, D.: Evaluation of failure probability via surrogate models. J. Comput. Phys. **229**, 8966–8980 (2010)
19. Lutowska, A.: Model order reduction for coupled systems using low-rank approximations. Ph.D. thesis, TU Eindhoven (2012). http://alexandria.tue.nl/extra2/729804.pdf
20. Paffrath, M., Wever, U.: Adapted polynomial chaos expansion for failure detection. J. Comput. Phys. **226**, 263–281 (2007)
21. Pulch, R.: Polynomial chaos for the computation of failure probabilities in periodic problems. In: Roos, J., Costa, L.R.J. (eds.) Scientific Computing in Electrical Engineering SCEE 2008. Mathematics in Industry, vol. 14, pp. 191–198. Springer, Berlin (2010)
22. Pulch, R.: Polynomial chaos for linear differential algebraic equations with random parameters. Int. J. Uncertain. Quantif. **1**(3), 223–240 (2011)
23. Schilders, W.H.A., van der Vorst, H.A., Rommes, J. (eds.): Model Order Reduction: Theory, Research Aspects and Applications. Mathematics in Industry, vol. 13. Springer, Berlin (2008)
24. Stavrakakis, K., Wittig, T., Ackermann, W., Weiland, T.: Parametric model order reduction by neighbouring subspaces. In: Michielsen, B., Poirier, J.-R. (eds.) Scientific Computing in Electrical Engineering SCEE 2010. Series Mathematics in Industry, vol. 16, pp. 443–451. Springer, Berlin (2012)
25. Tasić, B., Dohmen, J.J., ter Maten, E.J.W., Beelen, T.G.J., Schilders, W.H.A., de Vries, A., van Beurden, M.: Robust DC and efficient time-domain fast fault simulation. COMPEL (2014, accepted)
26. ter Maten, E.J.W., Wittich, O., Di Bucchianico, A., Doorn, T.S., Beelen, T.G.J.: Importance sampling for determining SRAM yield and optimization with statistical constraint. In: Michielsen, B.L., Poirier, J.-R. (eds.) Scientific Computing in Electrical Engineering SCEE 2010. Mathematics in Industry, vol. 16, pp. 39–48. Springer, Berlin (2012)
27. Ugryumova, M.V.: Applications of model order reduction for IC modeling. Ph.D. thesis, TU Eindhoven (2011). http://alexandria.tue.nl/extra2/711015.pdf
28. Xiu, D.: Numerical Methods for Stochastic Computations - A Spectral Method Approach. Princeton University Press, Princeton (2010)

A Stochastic Geometric Framework for Dynamical Birth-and-Growth Processes: Related Statistical Analysis

Giacomo Aletti, Enea G. Bongiorno, and Vincenzo Capasso

Abstract A birth-and-growth model is rigorously defined as a suitable combination, involving the Minkowski sum and the Aumann integral, of two very general set-valued processes representing nucleation and growth respectively. The simplicity of the proposed geometrical approach let us avoid problems arising from an analytical definition of the front growth such as boundary regularities. In this framework, growth is generally anisotropic and, according to a mesoscale point of view, is not local, i.e. for a fixed time instant, growth is the same at each point space. The proposed setting allows us to investigate nucleation and growth processes also from a statistical point of view. Different consistent set-valued estimators for growth processes and for the nucleation hitting function are derived.

1 Introduction

The importance of nucleation and growth processes is well known. They arise in several natural and technological applications (e.g. [7, 8]) such as, for example, solidification and phase-transition of materials, semiconductor crystal growth, biomineralization, and DNA replication, e.g. [14]. During the years, several authors studied stochastic spatial processes (e.g. [12, 16, 21]), nevertheless they essentially

G. Aletti (✉) • V. Capasso
ADAMSS (Interdisciplinary Centre for Advanced Applied Mathematical and Statistical Sciences) and Department of Mathematics, Università degli Studi di Milano, Milano, Italy
e-mail: giacomo.aletti@unimi.it; vincenzo.capasso@unimi.it

E.G. Bongiorno
Department of Studies in Economics and Business, Università degli Studi del Piemonte Orientale, Vercelli, Italy
e-mail: enea.bongiorno@unipmn.it

V. Capasso
Gregorio Millan Institute, Universidad Carlos III de Madrid, Madrid, Spain

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_51,

consider static approaches modeling real phenomena. For what concerns the dynamical point of view, a parametric *birth-and-growth process* was studied in [17, 18]. A birth-and-growth process is a family of random closed sets (RaCS) given by $\Theta_t = \bigcup_{n:T_n \leq t} \Theta^t_{T_n}(X_n)$, for $t \geq 0$, where $\Theta^t_{T_n}(X_n)$ is the RaCS obtained as the evolution up to time $t > T_n$ of the germ born at (random) time T_n in (random) location X_n, according to some growth model. An analytical approach is often used to model birth-and-growth process, in particular it is assumed that the growth of a spherical nucleus of infinitesimal radius is driven according to a non negative normal velocity, i.e. for every instant t, a border point of the crystal $x \in \partial\Theta_t$ "grows" along the outwards normal unit, e.g. [4–6, 10, 13]. In view of the chosen framework, different parametric and non parametric estimations have been proposed over the years, e.g. [1, 7, 9, 11, 19]. Note that the existence of an outwards normal vector imposes a regularity condition on $\partial\Theta_t$ (and also on the nucleation process; it cannot be a point process of Hausdorff dimension zero).

In this paper, we offer an outline of recent results obtained by the authors [2, 3]. In order to avoid regularity assumptions describing birth-and-growth processes, the authors have offered an original approach based on a purely stochastic geometric point of view that leads to novel and significant statistical results. In [3], they derive a computationally tractable mathematical model (based on Minkowski sum and Aumann integral) rigorously defined as a suitable combination of two very general set-valued processes representing nucleation $\{B_t\}_{t\in[t_0,T]}$ and growth $\{G_t\}_{t\in[t_0,T]}$ respectively. In [2], different set-valued parametric estimators of the rate of growth of the process are introduced. These are consistent as the observation window expands to the whole space. Moreover, keeping in mind that distributions of random closed sets are determined by their hitting functions and that the nucleation process cannot be observed directly, an estimation procedure of the hitting function of the nucleation process is provided.

2 Preliminary Results

Let $\mathbb{F}$ be the family of all closed subsets of $\mathbb{R}^d$ and $\mathbb{F}' = \mathbb{F} \setminus \{\emptyset\}$. The subscripts b, k and c denote boundedness, compactness and convexity properties respectively (e.g. $\mathbb{F}_{kc}$ denotes the family of all compact convex subsets of $\mathbb{R}^d$). For all $A, B \subseteq \mathbb{R}^d$ and $\alpha \geq 0$, let us consider

$$A+B = \{a+b : a \in A,\ b \in B\}, \qquad A \ominus B = \left(A^C + B\right)^C, \qquad \check{A} = \{-a : a \in A\},$$

where $A^C = \mathbb{R}^d \setminus A$. In what follows, we shall use: if $A \in \mathbb{F}$ and $B \in \mathbb{F}_k$ then $A + B \in \mathbb{F}$ [20]. Let $(\Omega, \mathfrak{F}, \mu)$ be a finite measure space, $X : \Omega \to \mathbb{F}$ is a measurable map if $\{\omega \in \Omega : X(\omega) \cap K \neq \emptyset\} \in \mathfrak{F}$ is measurable for each compact set K in $\mathbb{R}^d$. If μ is a probability measure, then X is a random closed set (RaCS). Let X be a RaCS, then $\{T_X(K) = \mathbb{P}(X \cap K \neq \emptyset), K \in \mathbb{F}_k\}$, is its *hitting function*. The Matheron theorem states that, the probability law $\mathbb{P}_X$ of any RaCS X is uniquely determined by its hitting function [15] and hence by $Q_X(K) = 1-T_X(K)$. Let $(\Omega, \mathfrak{F}, \mu)$ be a finite measure space. The *Aumann integral* of a non empty measurable closed set-valued map X is defined by $\int_\Omega X d\mu = \left\{\int_\Omega x d\mu : x \in L^1[\Omega; \mathbb{R}^d] \text{ and } x \in X \ \mu\text{–a.e.}\right\}$, where $\int_\Omega x d\mu$ is the usual Bochner integral in $L^1[\Omega; \mathbb{R}^d]$.

3 Geometric Random Process

Here, we refer to [3]. Let $[t_0, T] \subset \mathbb{R}$ be the *time interval*, and $(\Omega, \mathfrak{F}, \{\mathfrak{F}_t\}_{t\in[t_0,T]}, \mathbb{P})$ be a filtered probability space, where the filtration $\{\mathfrak{F}_t\}_{t\in[t_0,T]}$ is assumed to have the usual properties. Let B and G be two processes, *Nucleation* and *Growth* processes respectively, defined on $\Omega \times [t_0, T]$ with non empty closed set values, for which the following assumptions hold.

(A-1) For every $t, s \in [t_0, T]$ with $s < t$, $B_t = B(\cdot, t)$ is an $\mathfrak{F}_t$-measurable RaCS and $B_s \subseteq B_t$.

(A-2) For every $\omega \in \Omega$ and $t \in [t_0, T]$, $G(\omega, t)$ is convex and, there exists $K \in \mathbb{F}'_b$ such that $0 \in G(\omega, t) \subseteq K$.

Let $\mathscr{P}$ denote the *previsible* (or *predictable*) σ-algebra on $\Omega \times [t_0, T]$ generated by the processes $\{X_t\}_{t\in[t_0,T]}$ adapted, w.r.t. $\{\mathfrak{F}_t\}_{t\in[t_0,T]}$, with left Hausdorff-continuous trajectories on $[t_0, T]$. Thus, let us assume the following fact,

(A-3) G is $\mathscr{P}$-measurable.

It can be proven that, for any $a, b \in [t_0, T]$, $G_{a,b} = \int_a^b G(\omega, \tau) d\tau$ is a non empty bounded (compact) convex RaCS. For every $t \in [t_0, T] \subset \mathbb{R}$, $n \in \mathbb{N}$, and $\Pi = (t_i)_{i=0}^n$ partition of $[t_0, t]$, let us define

$$s_\Pi = s_\Pi(t) = \big(B_{t_0} + \int_{t_0}^t G(\tau) d\tau\big) \cup \bigcup_{i=1}^n \big(\Delta B_{t_i} + \int_{t_i}^t G(\tau) d\tau\big) \tag{1}$$

$$S_\Pi = S_\Pi(t) = \big(B_{t_0} + \int_{t_0}^t G(\tau) d\tau\big) \cup \bigcup_{i=1}^n \big(\Delta B_{t_i} + \int_{t_{i-1}}^t G(\tau) d\tau\big) \tag{2}$$

where $\Delta B_{t_i} = B_{t_i} \setminus B^o_{t_{i-1}}$ ($B^o_{t_{i-1}}$ denotes the interior set of $B_{t_{i-1}}$) and where the integral is in the Aumann sense w.r.t. the Lebesgue measure $d\tau = d\mu_\lambda$.

Clearly, both s_Π and S_Π are well defined RaCS, with $s_\Pi \subseteq S_\Pi$ (as a consequence of different time intervals integration). Moreover, it can be proven that $\{s_\Pi\}$ ($\{S_\Pi\}$) does not decrease (does not increase) whenever a refinement of Π is considered and, s_Π and S_Π are closer to each other (in the Hausdorff distance sense) as the partition Π is finer. Finally, their "limit" is independent of the choice of the refinement. In other words, s_{Π_j} and S_{Π_j} play the same role as lower sums and upper sums play in classical analysis when we define the Riemann integral. In fact, if Θ_t denotes their limit value (cf. Definition 1), s_{Π_j} and S_{Π_j} are a lower and an upper approximation of Θ_t respectively. This argument prevents problems that may arise considering uncountable unions in (1), (2) instead of countable unions and, allows the definition of a set-valued, continuous time, stochastic process.

Definition 1. For every $t \in [t_0, T]$, let $\left\{\Pi_j\right\}_{j\in\mathbb{N}}$ be a refinement sequence of the time interval $[t_0, t]$ and let Θ_t be the RaCS defined by

$$\overline{\bigcup_{j\in\mathbb{N}} s_{\Pi_j}(t)} = \overline{(\lim_{j\to\infty} s_{\Pi_j}(t))} = \Theta_t = \lim_{j\to\infty} S_{\Pi_j}(t) = \bigcap_{j\in\mathbb{N}} S_{\Pi_j}(t),$$

then, $\Theta = \{\Theta_t : t \in [t_0, T]\}$ is called *geometric random process G-RaP* (on $[t_0, T]$).

As a consequence, Θ is an a.s. non decreasing process, i.e.

$$\mathbb{P}\left(\Theta_s \subseteq \Theta_t,\ \forall t_0 \le s < t \le T\right) = 1.$$

Further, Θ is adapted w.r.t. $\{\mathfrak{F}_t\}_{t\in[t_0,T]}$. Thus, we justify the following integral and differential formulations. Let $t \in [t_0, T]$,

$$\begin{gathered}\Theta_t = \left(B_{t_0} + \int_{t_0}^t G(\tau)d\tau\right) \cup \bigcup_{s=t_0}^t \left(dB_s + \int_s^t G(\tau)d\tau\right),\\ \Theta_{t+dt} = (\Theta_t + G_t dt) \cup dB_t.\end{gathered}$$

Roughly speaking, the increment of the set Θ_t, during an infinitesimal time interval dt, is an enlargement due to an infinitesimal addend $G_t dt$ followed by the union with the infinitesimal nucleation dB_t. Note that, as a consequence of the definition of $+$, at any instant t, each point $x \in \Theta_t$ grows up by $G_t dt$ and no regularity assumptions on the boundaries are required. In particular it is sufficient to consider points $x \in \partial\Theta_t$ to describe the set evolution. Then we deal with *non local* growth; i.e. growth is the same addend for every $x \in \Theta_t$. Nevertheless, under mesoscale hypotheses we may only consider constant growth region as described, for example, in [5]. On the other hand, growth is anisotropic whenever G_t is not a ball.

4 Statistical Aspects

Clearly, one may derive the following discrete time formulation of above model

$$\Theta_n = \begin{cases} (\Theta_{n-1} + G_n) \cup B_n, & n \ge 1, \\ B_0, & n = 0. \end{cases}$$

In view of applications, note that a sample of a birth-and-growth process is usually a time sequence of pictures that represent process Θ at different temporal step; namely Θ_{n-1}, Θ_n that, for the sake of simplicity, we shall also denote by X and Y

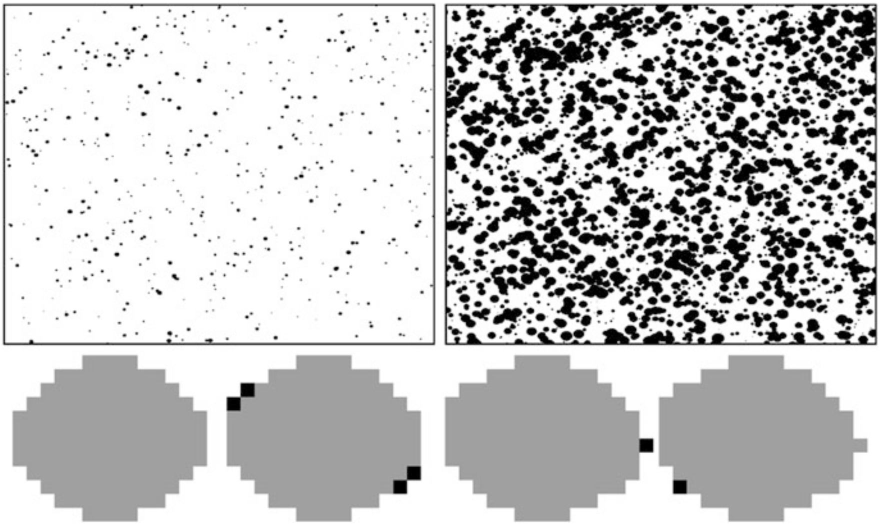


Fig. 1 Two different time instants (X and Y) pictures of a simulated birth-and-growth process. The magnified pictures of the true growth used for the simulation, the computed $\hat{G}^2_W$, $\hat{G}^1_W$ and $\hat{G}^1_{W\ominus\check{K}}$

respectively. In [2], the rate growth of Θ and the hitting function of B_n are estimated. In fact, G_n is not identified univocally, while the RaCS $Y \ominus \check{X}$ (denoted, from now on, by G) is unique, since it is the greatest RaCS, w.r.t. set inclusion, for which $(X + G) \subseteq Y$. Let us assume the following facts.

(A-4) There exists $K \in \mathbb{F}'_b$ such that $G \subseteq K$.

(A-5) For every $n \geq 1$, $\left(B_n \ominus \check{\Theta}_{n-1}\right) = \emptyset$ a.s.

Roughly speaking, Assumption (A-4) means that process Θ does not grow too "fast", whilst Assumption (A-5) means that it cannot be born something that, up to a translation, is larger than (or equal to) what there already exists.

In practical cases, data are bounded by some observation window and edge effects may cause problems in estimating G. As the standard statistical scheme for spatial processes suggests [16], we wonder if there exists a consistent estimator of G as $W_i \uparrow \mathbb{R}^d$. Thus, let $W \in \{W_i\}$ and let us set $Y_W = Y \cap W$. Edge effects are reduced by considering the following estimators of G

$$\hat{G}^1_W = \left(Y_W \ominus \check{X}_{W\ominus\check{K}}\right) \cap K, \qquad \hat{G}^2_W = \left(\left[Y_W \cup \left(\partial^{+K}_W X_W\right)\right] \ominus \check{X}_W\right) \cap K;$$

where K is given in Assumption (A-4) and where $\left(\partial^{+K}_W X_W\right) = \overline{(X_W + K) \setminus W}$. The following results hold (Fig. 1 shows how Proposition 1 works).

Proposition 1. *Let Y, X be RaCS, let $0 \in G = Y \ominus \check{X} \subseteq K$. Thus, for any $W_2 \supseteq W_1$, $G \subseteq \hat{G}^1_{W_2} \subseteq \hat{G}^1_{W_1}$. In particular, $\bigcap_{i\in\mathbb{N}} \hat{G}^1_{W_i} = G$ and $\lim_{i\to\infty} \delta_H(\hat{G}^1_{W_i}, G) = 0$.*

Moreover, for every $W \in \mathbb{F}'$, $G \subseteq \hat{G}_W^2 \subseteq \hat{G}_W^1$. Thus, $\hat{G}_W^2$ is consistent too (i.e. if $W \uparrow \mathbb{R}^d$ $\hat{G}_W^2 \downarrow G$).

It is also interesting to test whenever the nucleation process $B = \{B_n\}_{n \in \mathbb{N}}$ is a specific RaCS (e.g. Boolean model vs. point process). Although, we cannot directly observe the n-th nucleation B_n (it can be overlapped by other nuclei or by their evolutions), we shall infer on the hitting function associated to the nucleation process $T_{B_n}(\cdot)$. In particular, for any $K \in \mathbb{F}_k$, let $\tilde{Q}_{B,W}(K) = \hat{Q}_{Y,W}(K)/\hat{Q}_{X+\hat{G}_W,W}(K)$, where $\hat{Q}_{(\cdot)} = 1 - \hat{T}_{(\cdot)}$ is defined in [16] and $\hat{G}_W$ is one between $\hat{G}_W^2$ and $\hat{G}_W^1$.

Theorem 1. *Let X, Y be a.s. regular closed (i.e. $G = \overline{IntG}$). Let G, B be two RaCS such that $Y = (X + G) \cup B$, with B a stationary ergodic RaCS independent on G and X, and with G a.s. regular closed. Then, for any $K \in \mathbb{F}_k$, $|\tilde{Q}_{B,W}(K) - Q_B(K)| \to 0$ as $W \uparrow \mathbb{R}^d$ almost surely.*

Acknowledgements The second author was partially supported by PRIN 2009RNH97Z-002-2009.

References

1. Aletti, G., Saada, D.: Survival analysis in Johnson–Mehl tessellation. Stat. Inference Stoch. Process. **11**, 55–76 (2008)
2. Aletti, G., Bongiorno, E.G., Capasso, V.: Statistical aspects of fuzzy monotone set-valued stochastic processes. Application to birth-and-growth processes. Fuzzy Set. Syst. **160**, 3140–3151 (2009)
3. Aletti, G., Bongiorno, E.G., Capasso, V.: Integration in a dynamical stochastic geometric framework. ESAIM: Probab. Stat. **15**, 402–416 (2011)
4. Aquilano, D., Capasso, V., Micheletti, A., Patti, S., Pizzocchero, L., Rubbo, M.: A birth and growth model for kinetic-driven crystallization processes, Part I: Modeling. Nonlinear Anal. Real World Appl. **10**, 71–92 (2009)
5. Burger, M., Capasso, V., Pizzocchero, L.: Mesoscale averaging of nucleation and growth models. Multiscale Model. Simul. **5**, 564–592 (2006)
6. Burger, M., Capasso, V., Micheletti, A.: An extension of the Kolmogorov–Avrami formula to inhomogeneous birth-and-growth processes. In: Aletti, G., et al. (eds.) Math Everywhere, pp. 63–76. Springer, Berlin (2007)
7. Capasso, V. (ed.): Mathematical Modelling for Polymer Processing: Polymerization, Crystallization, Manufacturing. Mathematics in Industry, vol. 2. Springer, Berlin (2003)
8. Capasso, V.: On the stochastic geometry of growth. In: Sekimura, T., et al. (eds.) Morphogenesis and Pattern Formation in Biological Systems, pp. 45–58. Springer, Tokyo (2003)
9. Capasso, V., Villa, E.: Survival functions and contact distribution functions for inhomogeneous, stochastic geometric marked point processes. Stoch. Anal. Appl. **23**, 79–96 (2005)
10. Chiu, S.N.: Johnson–Mehl tessellations: asymptotics and inferences. In: Probability, Finance and Insurance, pp. 136–149. World Scientific Publication, River Edge (2004)
11. Chiu, S.N., Molchanov, I.S., Quine, M.P.: Maximum likelihood estimation for germination-growth processes with application to neurotransmitters data. J. Stat. Comput. Simul. **73**, 725–732 (2003)

12. Cressie, N.: Modeling growth with random sets. In: Spatial Statistics and Imaging. IMS Lecture Notes Monograph Series, vol. 20, pp. 31–45. Institute of Mathematical Statistics, Hayward (1991)
13. Frost, H.-J., Thompson, C.V.: The effect of nucleation conditions on the topology and geometry of two-dimensional grain structures. Acta Metall. **35**, 529–540 (1987)
14. Herrick, J., Jun, S., Bechhoefer, J., Bensimon, A.: Kinetic model of DNA replication in eukaryotic organisms. J. Mol. Biol. **320**, 741–750 (2002)
15. Matheron, G.: Random Sets and Integral Geometry. Wiley, New York (1975)
16. Molchanov, I.S.: Statistics of the Boolean Model for Practitioners and Mathematicians. Wiley, Chichester (1997)
17. Møller, J.: Random Johnson–Mehl tessellations. Adv. Appl. Probab. **24**, 814–844 (1992)
18. Møller, J.: Generation of Johnson–Mehl crystals and comparative analysis of models for random nucleation. Adv. Appl. Probab. **27**, 367–383 (1995)
19. Møller, J., Sørensen, M.: Statistical analysis of a spatial birth-and-death process model with a view to modelling linear dune fields. Scand. J. Stat. **21**, 1–19 (1994)
20. Serra, J.: Image Analysis and Mathematical Morphology. Academic, London (1984)
21. Stoyan, D., Kendall, W.S., Mecke, J.: Stochastic Geometry and its Applications. Wiley, Chichester (1995)

MATLAB Implementation of Functional Type A Posteriori Error Estimates with Raviart-Thomas Approximation

Maria A. Churilova and Maxim E. Frolov

Abstract Work is devoted to comparison of adaptive algorithms based on the functional approach to a posteriori error estimation. Classical elliptic boundary value problems with discontinuities of the first kind in coefficients are considered. Adaptive algorithms are implemented in MATLAB. Both, a standard finite element with continuous piecewise linear approximations and the simplest Raviart-Thomas finite element are used. For mesh adaptations different error indicators are applied. Sequences of finite-element meshes, effectivity indexes for estimates, relative errors of approximate solutions are compared for different error indicators. The results demonstrate that the usage of the Raviart-Thomas approximation considerably improves the efficiency of the corresponding adaptive algorithm.

1 Introduction

The problem of error control arises in the numerical analysis of various types of boundary value problems due to the necessity to guarantee the reliability of computed results. Various techniques for error estimation were developed for this purpose. There are hundreds of publications concerning approaches to the construction of a posteriori error estimates in the finite element method. An overview of them can be found, for example, in [3, 4] and many publications referenced therein. Here we examine the so called "functional approach" that is based purely on variational and functional methods. The approach is general and reliable, the results are directly applicable to any approximate solution from the corresponding energy space for a problem under consideration. It means that the error estimate remains valid regardless of the approach used to compute an approximation.

M.A. Churilova (✉) • M.E. Frolov
Saint Petersburg State Polytechnical University, Saint Petersburg, Russian Federation
e-mail: m_churilova@mail.ru; frolov_me@mail.ru

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_52,

2 Stationary Reaction-Diffusion Problem

2.1 Functional A Posteriori Error Estimate

As the model problem we consider the stationary reaction-diffusion equation with Dirichlet type boundary conditions, which is as follows:

$$\begin{cases} -\mathrm{div}(A\nabla u) + \rho^2 u = f & \text{in } \Omega \\ u = 0 & \text{on } \partial\Omega \end{cases} \tag{1}$$

where Ω is a bounded connected domain in $\mathbb{R}^2$ with a Lipschitz continuous boundary $\partial\Omega$, $f \in L^2(\Omega)$, ρ^2 is the reaction coefficient and A is a symmetric, positive definite matrix, which possess the property

$$\alpha_1|\eta|^2 \leqslant A\eta\cdot\eta \leqslant \alpha_2|\eta|^2 \quad \forall\eta\in\mathbb{R}^2,$$

where α_1 and α_2 are some positive constants. It is well known that the problem has the following weak formulation: find $u \in V_0$ satisfying the integral identity

$$\int\limits_\Omega \big(A\nabla u\cdot\nabla w + \rho^2 uw\big)\,dx = \int\limits_\Omega fw\,dx \quad \forall w\in V_0 = \mathrm{H}^1_0(\Omega).$$

The norm of the deviation of any approximation u_h from the exact solution u is defined as

$$|[u-u_h]| = \left(\int\limits_\Omega A\nabla(u-u_h)\cdot\nabla(u-u_h)\,dx + \int\limits_\Omega \rho^2(u-u_h)^2\,dx\right)^{1/2}.$$

For problem (1), a reliable upper estimate of this norm can be obtained in several ways (see [4]), dependent on the value of the coefficient ρ^2. We use the following majorant, which is known from the theory as suitable to wide range of ρ^2:

$$\begin{aligned} |[u-u_h]|^2 \leqslant M^2 = (1+\beta)\int\limits_\Omega (A\nabla u_h - y^*)\cdot(\nabla u_h - A^{-1}y^*)\,dx \\ + \frac{\mathbb{C}^2(1+\beta)}{\beta\alpha_1^2 + \rho^2\mathbb{C}^2(1+\beta)}\int\limits_\Omega (\mathrm{div}y^* + f - \rho^2 u_h)^2\,dx, \end{aligned} \tag{2}$$

with arbitrary element $y^* \in \mathrm{H}(\Omega,\mathrm{div})$,

$$\mathrm{H}(\Omega,\mathrm{div}) = \Big\{q\in L^2(\Omega,\mathbb{R}^2)\,\Big|\,\mathrm{div}q\in L^2(\Omega)\Big\},$$

arbitrary positive number β and the constant $\mathbb{C}$, which comes from the Friedrichs inequality.

2.2 Approximations of Free Variable

For the construction of y^* two approaches are used: the classical piecewise linear continuous approximation for both components and zero order Raviart-Thomas approximation. In both cases, for given β it is necessary to solve the minimization problem

$$\min_{y^*} M^2(u_h, \beta, y^*).$$

Necessary condition of minimum yields a system of linear algebraic equations with a positive definite, symmetric and sparse matrix. For the stationary diffusion problem numerical examples and a detailed description for the continuous approximation can be found in [2]. Results obtained in present research agree with ones from [2] and [1].

For mesh adaptations the following error indicators are used:

$$\eta_T = \left(\int_T (A\nabla u_h - y^*) \cdot (\nabla u_h - A^{-1} y^*)\, dx \right)^{1/2}$$

denoted as η^{con} in the case of the continuous approximation or η^{RT} in the case of the Raviart-Thomas approximation. Also an indicator based on the reference solution was used for verification, namely

$$\eta^{ref} = \left(\int_T A\nabla \tilde{e} \cdot \nabla \tilde{e}\, dx \right)^{1/2},$$

where $\tilde{e} = u_{ref} - u_h$ and the reference solution u_{ref} is obtained on a refined mesh (here $u_{ref} = u_{h/4}$). The reference solution is also used to calculate the relative error $e\,\% = |[u_h - u_{ref}]|/|[u_{ref}]| * 100\,\%$ and the effectivity index $I_{eff} = M/|[u_h - u_{ref}]|$.

2.3 Numerical Example

Let consider one example of mesh adaptations. Domain geometry and the initial mesh are depicted in Fig. 1. Matrix $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ in the subdomains I and IV, and

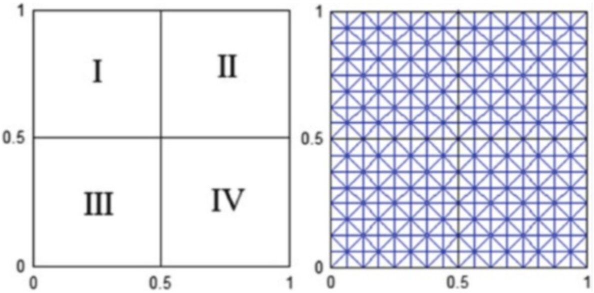


Fig. 1 Numerical example: domain geometry and the initial mesh

Table 1 Mesh adaptation steps

η^{ref}		η^{con}			η^{RT}		
Nodes	*e %*	*Nodes*	*e %*	I_{eff}	*Nodes*	*e %*	I_{eff}
289	13.97	289	13.97	1.64	289	13.97	1.28
380	10.32	382	10.79	1.91	377	10.70	1.36
928	6.25	1,032	6.84	1.94	1,014	6.38	1.32
2,209	3.95	3,970	3.85	1.87	2,576	3.92	1.33
4,006	2.94	7,910	2.86	1.81	4,661	2.94	1.31

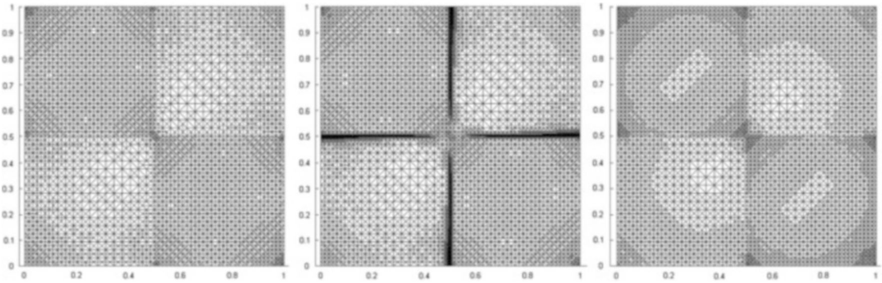


Fig. 2 Numerical example: final meshes for indicators η^{ref}, η^{con} and η^{RT}

$A = \begin{pmatrix} 10 & 0 \\ 0 & 10 \end{pmatrix}$ in the subdomains II and III. Right hand side f and the reaction coefficient ρ^2 are equal to 1 in the whole domain. In Table 1 several adaptation steps are presented. First two columns refer to the reference error indicator—we compare these results with ones obtained for η^{con} and η^{RT}. From the results one can conclude that meshes obtained with the indicator based on the Raviart-Thomas approximation are closer to the reference ones. It can also be seen from Fig. 2 where final meshes are depicted from left to right: for η^{ref}, η^{con}, η^{RT}, respectively. The corresponding number of nodes, relative errors and effectivity indexes are collected in the last row of Table 1.

Finally, we note that close results are obtained for other stationary reaction-diffusion problems. Also for some stationary diffusion problems with mixed boundary conditions a similar behavior is observed.

3 Conclusion

For vector fields from $\mathrm{H}(\Omega, \mathrm{div})$ the tangential component can be discontinuous. Such a discontinuity can not be well represented by any continuous approximation. For functional approach, this drawback has a large influence on the quality of local error indication near the discontinuity zone. Raviart-Thomas approximation helps to overcome difficulties arising in the case of discontinuity in coefficients of the reaction-diffusion equation.

Acknowledgements The reported study was partially supported by RFBR, research project No. 11-01-00531-a.

References

1. Frolov, M.E., Churilova, M.A.: Mesh adaptation based on functional a posteriori estimates with Raviart-Thomas approximation. Comput. Math. Math. Phys. **52**(7), 1044–054 (2012)
2. Frolov, M., Neittaanmäki, P., Repin, S.: On computational properties of a posteriori error estimates based upon the method of duality error majorants. In: Numerical Mathematics and Advanced Applications (ENUMATH, 2003), pp. 346–357. Springer, Berlin (2004)
3. Neittaanmäki, P., Repin, S.: Reliable Methods for Computer Simulation—Error Control and A Posteriori Estimates. Elsevier, Amsterdam (2004)
4. Repin, S.: A Posteriori Estimates for Partial Differential Equations. Walter de Gruyter, Berlin (2008)

Finite Element Concepts and Bezier Extraction in Hierarchical Isogeometric Analysis

Anh-Vu Vuong

Abstract Isogeometric analysis is an emerging approach combining computer aided geometric design and numerical analysis. Still local refinement techniques for isogeometric analysis are a major issue. One solution is proposed in Vuong et al. (Comput. Methods Appl. Mech. Eng. 200:3554–3567, 2011) and employs a hierarchical concept. This paper is an extension of this work and will discuss the corresponding element concept and apply the Bézier extraction to illustrate the connection to standard finite elements.

1 Introduction

In the classical FEM the finite dimensional subspace $V_h \subset V$ for the Galerkin projection typically consists of piecewise polynomials defined over a subdivision with global C^0 continuity. Isogeometric analysis [5], in contrast, makes use of the spline space that allows higher continuity and the initial geometric description from a CAGD program is already formulated with respect to this function space.

Therefore, our point of departure is a spline parameterization $\boldsymbol{G} : \Omega_0 \to \Omega$, $\boldsymbol{G}(\boldsymbol{u}) = \sum_i N_i(\boldsymbol{u})\boldsymbol{P}_i$ with control points $\boldsymbol{P}_i$ and with respect to a basis of B-Splines or NURBS N_i, which maps from the parametric space Ω_0 onto the computational domain Ω. The basic idea is to formulate the finite dimensional variational formulation

$$a(\varphi_h, \psi) = (l, \psi) \qquad \forall \psi \in V_h \tag{1}$$

A.-V. Vuong (✉)
Felix-Klein Centre for Mathematics, University of Kaiserslautern, Paul-Ehrlich Straße 31, 67663 Kaiserslautern, Germany
e-mail: vuong@mathematik.uni-kl.de

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_53,

with respect to basis functions defined on the parameter domain Ω_0 and to use the geometry mapping $\boldsymbol{G}$ as a global push-forward operator to map these functions to the physical domain Ω. Therefore we get the ansatz space $V_h = \operatorname{span}\{N_i \circ \boldsymbol{G}^{-1}\}$. The fact that we employ the same basis functions that describe the geometry and are used for the Galerkin projection is the reason why this method was named "iso-geometric".

2 Isogeometric Element Concept

Starting from the basics of isogeometric analysis it is straightforward to set up an element structure given by the knots of the definition of splines. We call $\hat{T}_{i,j} := [u_i, u_{i+1}] \times [v_i, v_{i+1}]$ a knot domain and the set of all knot domains *isogeometric mesh*

$$\hat{\mathscr{T}} := \left\{ \hat{T}_{i,j} \right\}_{i=0\ldots n+p, j=0\ldots m+q} . \tag{2}$$

Lemma 1. *Let $\mathscr{T}$ be the set of non-empty knot domains, an isogeometric subdivision. There are exactly $(p_u+1)(p_v+1)$ nonzero basis function within each element $T \in \mathscr{T}$ with p_u and p_v the degrees of the spline space in first and second parameter direction, respectively.*

An example is shown for two dimensions in Fig. 1 for $p_u = p_v = 2$. The support spreads over nine knot domains as visualized in Fig. 1a, b shows the support extension of the element in the middle. The reference element for the two-dimensional case with $p_u = p_v = 2$ is shown in Fig. 1c where we have a ordered grid of 3×3 degrees of freedom. We can conclude that not only the number of nonzero basis function over an element is constant as shown in Lemma 1, but also how these function are placed related to the element.

Corollary 1. *Let $\mathscr{T}$ be an isogeometric subdivision and $\hat{\mathscr{T}}$ the isogeometric mesh. For any element T the nonzero basis functions over T are B_{ij} with $i = k - p_u, \ldots, k$, $j = \ell - p_v, \ldots, \ell$, whereas k and ℓ are chosen so that $T = \hat{T}_{k,\ell}$.*

Note that this is valid for any element independent of its exact shape or position. We want to stress here that this also shows the usefulness of the isogeometric elements and the knot domains. On elements we can make use of the information about the basis functions that are non-zero but the support of these basis functions spreads over knot domains.

This is for example used in a FEM framework by using "general elements" in the commercial software LS-DYNA (see [1,4]).

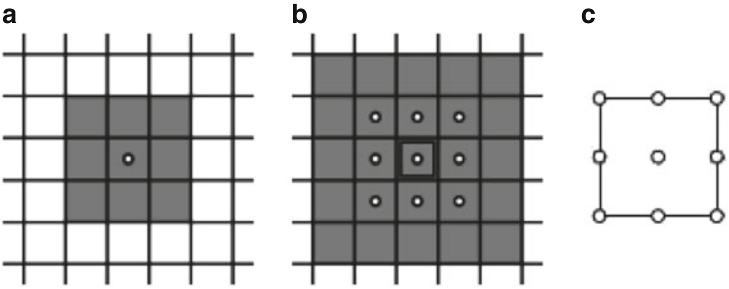


Fig. 1 Bivariate B-splines over knot domains. (**a**) Support of a function of degree two. (**b**) Influences on an element for degree two. (**c**) Reference element

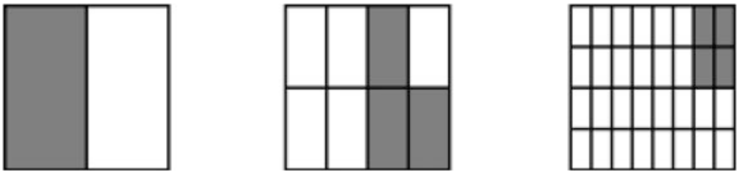


Fig. 2 Sequence of h-refined parameter spaces with the hierarchical subdivision marked in *grey*

2.1 Hierarchical Mesh Structure and Refinement

Before extending these concept to the hierarchical refinement approach, we will shortly revisit its basics. All details and numerical examples can be found in [7]. Point of departure is a hierarchy of subdivisions $\mathscr{T}^\ell$, which are created by uniform h-refinement and define a hierarchy of bases $\mathscr{B}^\ell$. The corresponding spline spaces $\mathscr{S}^\ell$ therefore form a chain of inclusion

$$\mathscr{S}^1 \subset \cdots \subset \mathscr{S}^\ell \subset \mathscr{S}^{\ell+1} \subset \cdots \subset \mathscr{S}^L. \tag{3}$$

Definition 1. We call a selection $\mathscr{M} \subset \bigcup_{\ell=1}^{L} \mathscr{T}^\ell$ a *hierarchical subdivision* if the following conditions hold

$$\operatorname{int} T_i \cap \operatorname{int} T_j = \emptyset \quad \forall T_i, T_j \in \mathscr{M}, T_i \neq T_j. \tag{4}$$

$$\bigcup_{T \in \mathscr{M}} T = \overline{\Omega}_0. \tag{5}$$

An element $T \in \mathscr{M}$ is called an *active element*. The set of active elements with level ℓ is denoted by $M^\ell := \mathscr{M} \cap \mathscr{T}^\ell$.

An example for a hierarchical subdivision is shown in Fig. 2 where eight elements are selected to be active out of three levels.

In order to establish the relation to the support of the basis functions we have to study the region filled out by particular elements in a set. For ease of notation we define the domain $U_X \subseteq \Omega_0$ to be

$$U_X := \bigcup_{C \in X} C \tag{6}$$

for a subset X of the power set $\mathfrak{P}(\Omega_0)$.

Definition 2. The set of active basis functions $\mathscr{A}$ is defined as follows: a function $\varphi \in \mathscr{B}^k$ of level k is an element of $\mathscr{A}$ if

$$\operatorname{supp}\varphi \subseteq \bigcup_{\ell=k}^{L} U_{M^\ell} \text{ and } \operatorname{supp}\varphi \nsubseteq \bigcup_{\ell=k+1}^{L} U_{M^\ell}. \tag{7}$$

We now want to investigate how these active basis function distribute over an element.

2.2 Hierarchical Element

In order to cope with the complexity added by hierarchical local refined meshes we have to extend the isogeometric reference element that only holds information about one level. Following observation can be made: in a hierarchically refined grid and a given element of level k, we can find an element of level r that contains the given element for each lower level r. Based on this, we define the hierarchical isogeometric reference element of level k as a sequence of reference elements from level $\ell = 1, \dots, k$, where the active basis functions from level $\ell = 1, \dots, k$ can be positioned.

It should be remarked that in the hierarchical subdivision these elements have different size, but this has no influence on the hierarchical reference element. It only indicates for all level, which basis functions are non-zero on this element, just like for the non-hierarchical case.

We illustrate the active basis functions over the reference element by an example. For the sake of simplicity we look at a hierarchical basis of degree one. In Fig. 3a the same the hierarchical subdivision like in Fig. 2 is shown. We want to describe the configuration of the element marked in grey. The six active basis function are symbolized by dots, whereas those that are nonzero on the grey element are highlighted with their level number. Finally, the corresponding reference elements and the position of the active functions are shown in Fig. 3b.

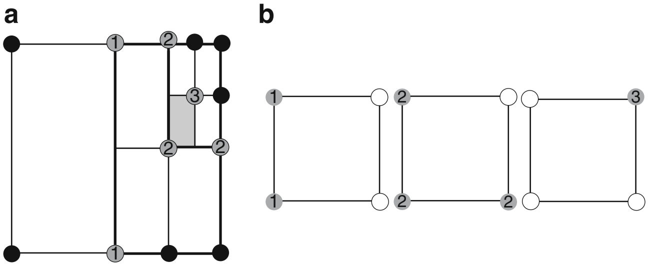


Fig. 3 Active basis functions on hierarchical reference element. (**a**) Hierarchical subdivision with active basis functions. (**b**) Configuration of the element

2.3 *Bézier Extraction*

Bézier extraction introduced in [2] is a point of view that is mainly used to illustrate and use the connection between the isogeometric approach and existing FEM codes. The main idea is to choose a representation of the basis functions that is more local than B-splines or NURBS and enforces less continuity. The well-known Bernstein polynomials $b_{i,p}$ are a suitable choice for this. For all B-splines over an element the representation in Bernstein polynomials is computed by knot insertion and we get the representation

$$B_{i,p} = \sum \lambda_j b_{j,p}. \tag{8}$$

As these polynomials can now be defined element-wise we have returned to the classical finite element setting. Bézier extraction was for example used to transfer refined T-spline meshes [6] to a FEM framework. As we have seen in the previous section the active functions of a hierarchical refined basis on one element can be expressed in a hierarchy of reference elements. The basis change to Bernstein polynomials is applicable on one level, but also from all previous basis level to the level of the element, because of the chain of inclusion of the hierarchy in Eq. (3). Therefore it is in the same manner possible to apply these techniques to hierarchically refined meshes.

3 Conclusions

We have discussed several finite elements point of views to isogeometric analysis and its hierarchical refinement. The usage of an element concept allows to employ generalized finite elements like for example in the commercial FEM software LS DYNA to implement isogeometric analysis. Furthermore, we discussed that Bézier extraction, which created the connection to finite elements also on the implementation level, is also extendable for the hierarchical refinement approach. Future work includes the investigation of element concepts for more advanced techniques like truncated hierarchical spline spaces [3].

Acknowledgements Financial supported within the 7th Framework Programme of the European Union by the project TERRIFIC is greatly acknowledged.

References

1. Benson, D.J., Bazilevs, Y., De Luycker, E., Hsu, M.C., Scott, M., Hughes, T.J.R., Belytschko, T.: A generalized finite element formulation for arbitrary basis functions: from isogeometric analysis to xfem. Int. J. Numer. Methods Eng. **83**, 765–785 (2010)
2. Borden, M.J., Scott, M.A., Evans, J., Hughes, T.J.R.: Isogeometric finite element data structures based on Bézier extraction of NURBS. Int. J. Numer. Methods Eng. **87**, 15–47 (2010)
3. Giannelli, C., Jüttler, B., Speleers, H.: THB-splines: the truncated basis for hierarchical splines. Comput. Aided Geom. Des. **29**, 485–498 (2012)
4. Hartmann, S., Benson, D.J., Lorenz, D.: About isogeometric analysis and the new NURBS-based finite elements in LS-DYNA. In: 8th European LS-DYNA Users Conference (2011)
5. Hughes, T.J.R., Cottrell, J.A., Bazilevs, Y.: Isogeometric analysis: CAD, finite elements, NURBS, exact geometry and mesh refinement. Comput. Methods Appl. Mech. Eng. **194**, 4135–4195 (2005)
6. Scott, M.A., Borden, M.J., Verhoosel, C.V., Sederberg, T.W., Hughes, T.J.R.: Isogeometric finite element data structures based on Bézier extraction of T-splines. Int. J. Numer. Methods Eng. **88**, 126–156 (2011)
7. Vuong, A.V., Giannelli, C., Jüttler, B., Simeon, B.: A hierarchical approach to local refinement in isogeometric analysis. Comput. Methods Appl. Mech. Eng. **200**, 3554–3567 (2011)

A Second Order Finite-Difference Ghost-Cell Method for the Steady-State Solution of Elasticity Problems

Armando Coco, Gilda Currenti, and Giovanni Russo

Abstract This work presents a second order finite-difference ghost cell method for the steady-state solution of elasticity problems. Numerical results are shown for the application of underground volcano activities.

1 Introduction

Underground volcano activity has several observable effects besides earthquakes and seismic activity. For example, a sudden increase in pressure in a magmatic chamber produces deformations in the surrounding, which can be observed and are indeed accurately monitored by satellite observations; front sliding in a fault causes horizontal and vertical displacements on the earth surface, with a configuration which depends on the geometry and strength of the slide. One of the objective of the present research is to infer the underground activity from the measured ground displacement. This inverse problem requires the solution of several direct problems: given an underground source of stress/strain, compute the displacement field (in particular on the surface). The starting point to model the physical system is based on static linear elastic problem.

Accurate solution of the static problem in complex geometry can be computed with several commercial packages, such as, for example, COMSOL multiphysics. Such software uses Finite Element discretization with tetrahedral elements, which can be adapted to the geometry. Such methods, however, are not straightforward to

A. Coco (✉) • G. Russo
University of Catania, Viale Andrea Doria 6, 95125 Catania, Italy
e-mail: coco@dmi.unict.it; russo@dmi.unict.it

G. Currenti
INGV - Sezione di Catania, Piazza Roma 2, 95125 Catania, Italy
e-mail: gilda.currenti@ct.ingv.it

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_54,

implement and difficult to use especially in the case of moving domain, because of the computationally expensive meshing procedures needed for each domain.

Here we adopt a different strategy. We solve the equation on a regular Cartesian grid, and use level set to define the geometry. This approach presents several advantages over the use of tetrahedral grids. For example it is automatically second order accurate with a very compact stencil, it requires a simpler data structure and it allows the construction of a geometric multigrid solver.

2 Model

We start with two space dimensions, so that we can perform a more careful comparison with some analytical solution and with other solvers. The problem is described by the linearized steady-state equations of elasticity:

$$\nabla \cdot \sigma = 0 \Longrightarrow G\, \Delta \mathbf{u} + (\lambda + G)\, \nabla \cdot (\nabla \mathbf{u}) = 0 \tag{1}$$

where $\sigma := \sigma(\mathbf{u})$ is the stress tensor, determined by the Hooke's law:

$$\sigma_{ij} = \lambda\, e_{kk}\, \delta_{ij} + 2\, G\, e_{ij},$$

e_{ij} is the linearized Almansi strain tensor:

$$e_{ij} = \frac{1}{2}\left(\nabla \mathbf{u} + \nabla \mathbf{u}^T\right)_{ij}$$

$\mathbf{u} = (u, v, w)$ is the displacement, λ is Lamé's first parameter, G is the rigidity (see [2]). To obtain the two dimensional problem, we consider the plane strain model, that is we suppose that the z-component of displacement w vanishes everywhere, and the displacements u, v are functions of x, y only, and not of z. The basic equation (1) becomes in two dimensions:

$$G\, \Delta \mathbf{u} + \frac{1}{1 - 2\nu}\, G\, \nabla \cdot (\nabla \mathbf{u}) = 0 \tag{2}$$

where $\mathbf{u} = (u, v)$.

The geometry of the problem is represented in Fig. 1, where Ω_p is the source in pressurization, Γ_s is the free surface, while $\Gamma_{l,r,b}$ are boundaries taken far enough from Ω_p in such a way they do not influence the results. Let Ω be the domain below the surface Γ_s. The domains Ω and Ω_p are implicitly described by two level-set functions, i.e. [4]:

$$\Omega = \{\phi(x, y) < 0\}, \quad \Omega_p = \left\{\phi_p(x, y) < 0\right\}.$$

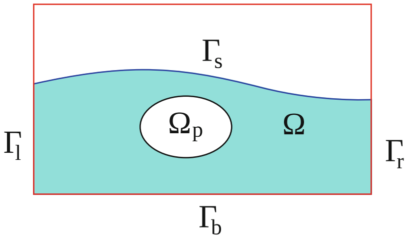


Fig. 1 Geometry of the model

We suppose that we know the signed distance function [3], which is a special case of a level-set function:

$$\phi(x,y) = \begin{cases} -d\left((x,y),\Gamma_s\right) & \text{if } (x,y) \in \Omega \\ \ \ d\left((x,y),\Gamma_s\right) & \text{if } (x,y) \notin \Omega \end{cases},$$

$$\phi_p(x,y) = \begin{cases} -d\left((x,y),\Gamma_p\right) & \text{if } (x,y) \in \Omega_p \\ \ \ d\left((x,y),\Gamma_p\right) & \text{if } (x,y) \notin \Omega_p \end{cases}.$$

We solve Eq. (2) in $\Omega\backslash\Omega_p$, with a free-stressed boundary condition $\sigma\cdot\mathbf{n} = 0$ on Γ_s, where $\mathbf{n} = \nabla\phi$ is the normal to Γ_s, while on $\partial\Omega_p$ we impose $\sigma\cdot\mathbf{n} = -p\,\mathbf{n}$, where $\mathbf{n} = \nabla\phi_p$ is the normal to $\partial\Omega_p$ and p is the pressure. On $\Gamma_{l,r,b}$ we impose homogeneous Dirichlet or Neumann boundary conditions, i.e. $\mathbf{u} = 0$ or $\nabla\mathbf{u}\cdot\mathbf{n} = 0$, where $\mathbf{n}$ is the normal to $\Gamma_{l,r,b}$.

3 Numerical Scheme

Let us discretize the domain by a regular Cartesian grid with spatial step $\Delta x = \Delta y = h$ and let us call D_h the set of grid points. The linear system coming from the discretization of the problem is composed as following. For each grid point of $\Omega\backslash\Omega_p$ we discretize Eq. (2) separately for u and w by central differences. For instance, the discretization of the derivatives of u reads:

$$\frac{\partial^2 u}{\partial x^2} \approx \frac{1}{h^2}\begin{bmatrix} 0 & 0 & 0 \\ 1 & -2 & 1 \\ 0 & 0 & 0 \end{bmatrix} u_{i,j} = \frac{u_{i-1,j} - 2u_{i,j} + u_{i+1,j}}{h^2}.$$

$$\frac{\partial^2 u}{\partial y^2} \approx \frac{1}{h^2}\begin{bmatrix} 0 & 1 & 0 \\ 0 & -2 & 0 \\ 0 & 1 & 0 \end{bmatrix} u_{i,j} = \frac{u_{i,j-1} - 2u_{i,j} + u_{i,j+1}}{h^2}.$$

$$\frac{\partial^2 u}{\partial x\partial y} \approx \frac{1}{4\,h^2}\begin{bmatrix} -1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & -1 \end{bmatrix} u_{i,j} = \frac{u_{i+1,j+1} + u_{i-1,j-1} - u_{i+1,j-1} - u_{i-1,j+1}}{4\,h^2}.$$

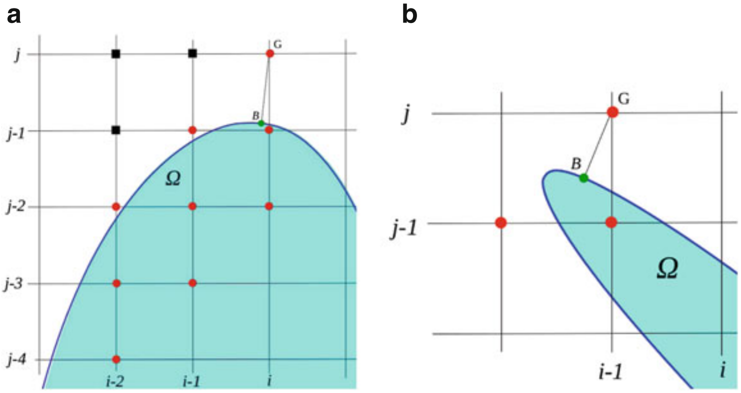


Fig. 2 Stencils for the ghost point G. (**a**) Nine-point stencil. (**b**) Reduction of the nine-point stencil to a three-point stencil

The whole stencil results in a nine-point stencil. For grid points of $\Omega \backslash \Omega_p$ which are close to the boundary, some of the points of the stencil may lie outside $\Omega \backslash \Omega_p$ (i.e. outside Ω or inside Ω_p). Such grid points are called *ghost points* and a suitable value should be defined for them to close the linear system.

To this aim, we write an equation for each ghost point. Let G be a ghost point outside Ω (if G is inside Ω_p the discretization is analogous). We compute the outward unit normal in G, that is $\mathbf{n}_G = \left(n_G^x, n_G^y\right) = \nabla\phi$, using a second-order accurate discretization for $\nabla\phi$, such as central difference in G. Now we can compute the closest boundary point to G, that we call B, by the signed distance function:

$$B = G - \phi(G)\mathbf{n}_G. \tag{3}$$

Therefore, the equation of the linear system for the ghost point G is:

$$\sigma(\tilde{u}) \cdot \tilde{\mathbf{n}}(B) = 0 \Longrightarrow \sigma(\tilde{u}) \cdot (\nabla\tilde{\phi})(B) = 0 \tag{4}$$

where $\tilde{u}$ and $\tilde{\phi}$ are the biquadratic interpolants of u and ϕ respectively on a suitable upwind nine-point stencil [1]. We choose the upwind nine-point stencil in the following manner (see Fig. 2a for the case $n_x, n_y > 0$, the other cases are analogous). If $|x_B - x_G| < |y_B - y_G|$ (as in Fig. 2a, b), the nine-point stencil will be composed by three points of the column i, three points of the column $i-1$, three points of the column $i-2$; while if $|x_B - x_G| \geq |y_B - y_G|$ it will be composed by three points of the row j, three points of the row $j-1$, three points of the row $j-2$. When possible we prefer the 3×3 squared stencil. If it is not possible to build the nine-point stencil, we revert to a more robust (less accurate) three-point stencil (Fig. 2b).

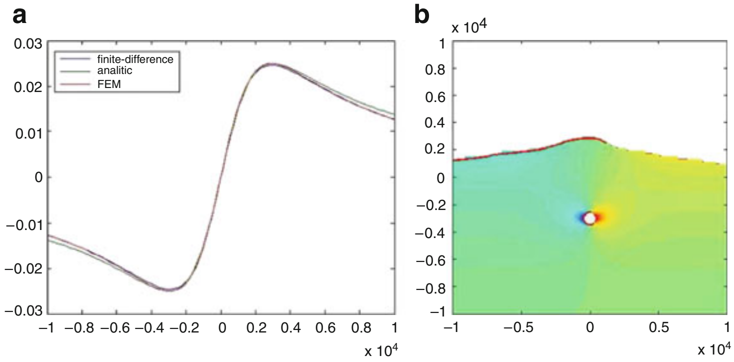


Fig. 3 Numerical results. (**a**) Comparison between the method proposed in this work, the FEM, and the analytic solution. (**b**) Displacement *u* with the real Etna profile

4 Results and Outlook

Some preliminary tests are the following. In Fig. 3a we compare the method with FEM and an analytic solution. We plot the displacement *u* along Γ_s. In Fig. 3b we performed a test using the real Etna profile, plotting the displacement *u* on all the domain.

Some works in progress concern the extension of the method to the case of variable coefficients λ and G (heterogeneous medium), multigrid technique (using a recent approach adopted in elliptic problems), grid adaptation (since only small portions of the computational domain require fine resolution) and three dimensional extension.

References

1. Coco, A., Russo, G.: Finite-difference ghost-point multigrid methods on Cartesian grids for elliptic problems in arbitrary domains. J. Comput. Phys. **241**, 464–501 (2013)
2. Fung, Y.: Foundations of Solid Mechanics. Prentice-Hall, Englewood Cliffs (1965)
3. Russo, G., Smereka, P.: A remark on computing distance functions. J. Comput. Phys. **163**, 51–67 (2000)
4. Sussman, M., Smereka, P., Osher, S.: A level set approach for computing solutions to incompressible 2-phase flow. J. Comput. Phys. **114**, 146–159 (1994)

Time-Exact Solution of Large Linear ODE Systems by Block Krylov Subspace Projections

Mike A. Botchev

Abstract We propose a time-exact Krylov-subspace-based method for solving large linear inhomogeneous systems of ODE (ordinary differential equations). The method consists of two stages. The first stage is an accurate piecewise polynomial approximation of the inhomogeneous source term, constructed with the help of the truncated SVD (singular value decomposition). The second stage is a special residual-based block Krylov subspace method for the matrix exponential. The accuracy of the method is only restricted by the accuracy of the piecewise polynomial approximation and by the error of the block Krylov process. Since both errors can, in principle, be made arbitrarily small, this yields, at some costs, a time-exact method. Numerical experiments are presented to demonstrate efficiency of the new method, as compared to an exponential time integrator with Krylov subspace matrix function evaluations. This conference paper is based on the preprint (Botchev, A block Krylov subspace time-exact solution method for linear ODE systems, Memorandum 1973, Department of Applied Mathematics, University of Twente, Enschede, 2012, http://eprints.eemcs.utwente.nl/21277/).

1 Introduction and Problem Formulation

Consider initial-value problem (IVP)

$$y' = -Ay + g(t), \qquad y(0) = v = 0, \qquad t \in [0, T], \tag{1}$$

M.A. Botchev (✉)
Applied Mathematics Department and MESA+ Institute for Nanotechnology, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands
e-mail: m.a.botchev@utwente.nl

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_55,

where $y(t)$ is the unknown vector function, $y : \mathbb{R} \to \mathbb{R}^n$, and the matrix $A \in \mathbb{R}^{n\times n}$, vector function $g : \mathbb{R} \to \mathbb{R}^n$, and vector $v \in \mathbb{R}^n$ are given. We assume, without loss of generality, that $v = 0$ (otherwise a change of variables $\tilde{y}(t) \equiv y(t) - v$ transforms (1) to an equivalent IVP with homogeneous initial value). Problems of type (1) appear in numerous applications, in particular, in the context of numerical solution of partial differential equations (PDEs) by the method of lines. This means that a discretization of a PDE in space is followed by a time integration of the resulting ODE system (1). We are thus interested in problems (1) where A is a large, typically sparse matrix.

The time step size in explicit time integration methods can often be unacceptably small, for instance, due to the stiffness of the ODE system or due to a locally refined spatial mesh. In this case implicit time integration is of interest. Since recently, a lot of research has been carried out on the so-called exponential time integration schemes, see a recent comprehensive survey [10]. These are time integration schemes involving the matrix exponential and related matrix functions. The interest in exponential time integration is due to the new, challenging applications [11,12] as well as to the recent progress in techniques to compute actions of matrix functions for large matrices (see e.g. [3–6, 8, 9, 15–18]).

The first stage of our method is a truncated SVD approximation of the source term $g(t) \approx Up(t)$, with $U \in \mathbb{R}^{n\times m}$ and $p : \mathbb{R} \to \mathbb{R}^m$. It is described in [1] and leads to initial-value problem

$$y' = -Ay + Up(t), \qquad y(0) = 0, \qquad t \in [0, T]. \tag{2}$$

1.1 EBK: Exponential Block Krylov Method

Define residual $r_k(t)$ of an approximate solution $y_k(t)$ of (2) as

$$r_k(t) \equiv -Ay_k(t) - y_k'(t) + Up(t).$$

This residual concept (well known in the ODE literature [7, 12, 14]) can be used as a stopping criterion and for restarting in Krylov subspace methods for matrix exponential [2]. The methods presented here are based on this residual-based restarting approach.

Choosing the initial guess $y_0(t)$ to be a zero vector function, we see that the corresponding initial residual is

$$r_0(t) = Up(t). \tag{3}$$

The approximate solution $y_k(t)$ at Krylov iteration k is then obtained as $y_k(t) = y_0(t) + \xi_k(t)$. Here the vector function $\xi_k(t)$ is the Krylov subspace approximate solution of the correction problem

$$\xi' = -A\xi + r_0(t), \qquad \xi(0) = 0, \qquad t \in [0, T], \tag{4}$$

Note that if $\xi_k(t)$ solves (4) exactly then $y_k(t)$ is the sought-after exact solution of (2). It is natural to solve (4) by projecting it onto a block Krylov subspace defined as $\mathscr{K}_k(A, U) \equiv \operatorname{span}\{U, AU, A^2U, \dots, A^{k-1}U\}$, with dimension at most $k \cdot m$. An orthonormal basis for this subspace can be generated by the block Arnoldi or Lanczos process (see e.g. [13, 19]). The process produces, after k block steps, matrices $V_{[k+1]} = \begin{bmatrix} V_1 \ V_2 \ \dots \ V_{k+1} \end{bmatrix} \in \mathbb{R}^{n\times(k+1)m}$, $H_{[k+1,k]} \in \mathbb{R}^{(k+1)m\times km}$. Here $V_i \in \mathbb{R}^{n\times m}$, V_1 is the matrix U from $g(t) \approx Up(t)$ and $V_{[k+1]}$ has orthonormal columns spanning the Krylov subspace, namely, $\operatorname{colspan}(V_{[k]}) = \mathscr{K}_k(A, U)$. The matrix $H_{[k+1,k]}$ is block upper Hessenberg, with $m\times m$ blocks H_{ij}, $i = 1, \dots, k+1$, $j = 1, \dots, k$. The matrices $V_{[k+1]}$ and $H_{[k+1,k]}$ satisfy the block Arnoldi (Lanczos) decomposition [13, 19],

$$AV_{[k]} = V_{[k+1]}H_{[k+1,k]} = V_{[k]}H_{[k,k]} + V_{k+1}H_{k+1,k}E_k^T, \tag{5}$$

where $H_{k+1,k}$ is the only nonzero block in the last $k+1$ block row of $H_{[k+1,k]}$ and $E_k \in \mathbb{R}^{n\times k}$ is formed by the last m columns of the $km \times km$ identity matrix.

Once the Krylov basis matrix $V_{[k]}$ is built, the Krylov subspace solution $\xi_k(t)$ of (4) can be computed as $\xi_k(t) = V_{[k]}u(t)$, where $u(t)$ solves the projected IVP

$$u'(t) = -H_{[k,k]}u(t) + E_1 p(t), \qquad u(0) = 0, \qquad t \in [0, T], \tag{6}$$

where $E_1 \in \mathbb{R}^{km\times m}$ is formed by the first m columns of the $km \times km$ identity matrix. Note that $E_1 p(t) = V_{[k]}^T r_0(t) = V_{[k]}^T V_1 p(t)$. Using (3), (5) and (6), we can show [1] that for the exponential residual $r_k(t)$ holds

$$r_k(t) = -Ay_k - y_k' + Up(t) = -V_{k+1}H_{k+1,k}E_k^T u(t). \tag{7}$$

There are two important messages relation (7) provides. First, the residual can be computed efficiently during the iteration process because the matrices V_{k+1} and $H_{k+1,k}$ are readily available in the Arnoldi or Lanczos process. Second, the residual after k block steps has the same form as the initial residual (3), namely it is a matrix of m orthonormal columns times a time dependent vector function. This allows for a restart in the block Krylov method: set $y_0(t) := y_k(t)$, then relation (3) holds with $U := V_{k+1}$ and $p(t) := -H_{k+1,k}E_k^T u(t)$. The just described correction with k block Krylov iterations can then be repeated, which results in a restarted block Krylov subspace method for solving (2).

We will refer to the just described scheme as EBK, exponential block Krylov method.

2 Implementation of the EBK Methods

We now sketch an algorithm for the EBK method.

```
1. Approximate $g(t) \approx Up(t)$.
2. Set $y_0(t) := 0$ and $r_0(t) := Up(t)$. Stop if $\|r_0(t)\|$ is small enough.
   Otherwise set $V_1 := U$.
3. main Krylov subspace loop:
   for $k = 1, \dots, \texttt{restart}$

     a. Perform step $k$ of the block Arnoldi/Lanczos process (5):
        compute $V_{k+1}$ and the block column $k$ of $H_{[k+1,k]}$,
        $AV_{[k]} = V_{[k+1]}H_{[k+1,k]} = V_{[k]}H_{[k,k]} + V_{k+1}H_{k+1,k}E_k^T$.
     b. Find solution $u(t)$ of the projected IVP (6) approximately,
        compute residual with (7): $r_k(t) := -V_{k+1}H_{k+1,k}E_k^T u(t)$.
     c. if $k = \texttt{restart}$ or $\|r_k(t)\|$ is small enough
              solve the projected IVP (6) accurately,
              update solution $y_k(t) := y_0(t) + V_{[k]}u(t)$
              if $\|r_k(t)\|$ is small enough
                   stop
              endif
              if $k = \texttt{restart}$
                   $y_0(t) := y_k(t)$, $U := V_{k+1}$, $p(t) := -H_{k+1,k}E_k^T u(t)$
                   return to step 2.
              endif

   endfor
```

It is important to stop only if $\|r_k(t)\|$ is small enough for *several* values $t \in [0, T]$, checking only $\|r_k(T)\|$ is not enough. Ideally, one should check the $L_2[0, T]$ integral norm of $\|r_k(t)\|$. Furthermore, note that the projected problem is not solved to a full accuracy most of the time. This is only necessary when the solution is updated due to a restart or satisfied stopping criterion. In EBK the projected IVP is solved with the `ode15s` MATLAB ODE solver. For numerical experiments with EBK see [1].

Acknowledgements This work was supported by the Russian Federal Program "Scientific and scientific-pedagogical personnel of innovative Russia", Grant 8500.

References

1. Botchev, M.A.: A block Krylov subspace time-exact solution method for linear ODE systems. Memorandum 1973, Department of Applied Mathematics, University of Twente, Enschede (2012). http://eprints.eemcs.utwente.nl/21277/
2. Botchev, M.A., Grimm, V., Hochbruck, M.: Residual, restarting and Richardson iteration for the matrix exponential. SIAM J. Sci. Comput. **35**(3), A1376–A1397 (2013)

3. Druskin, V.L., Knizhnerman, L.A.: Two polynomial methods of calculating functions of symmetric matrices. U.S.S.R. Comput. Math. Math. Phys. **29**(6), 112–121 (1989)
4. Druskin, V.L., Knizhnerman, L.A.: Krylov subspace approximations of eigenpairs and matrix functions in exact and computer arithmetic. Numer. Linear Algebra Appl. **2**, 205–217 (1995)
5. Druskin, V.L., Greenbaum, A., Knizhnerman, L.A.: Using nonorthogonal Lanczos vectors in the computation of matrix functions. SIAM J. Sci. Comput. **19**(1), 38–54 (1998). doi:10.1137/S1064827596303661
6. Eiermann, M., Ernst, O.G., Güttel, S.: Deflated restarting for matrix functions. SIAM J. Matrix Anal. Appl. **32**(2), 621–641 (2011). http://dx.doi.org/10.1137/090774665
7. Enright, W.H.: Continuous numerical methods for ODEs with defect control. J. Comput. Appl. Math. **125**(1–2), 159–170 (2000). doi:10.1016/S0377-0427(00)00466-0. Numerical analysis 2000, vol. VI, Ordinary differential equations and integral equations
8. Hochbruck, M., Lubich, C.: On Krylov subspace approximations to the matrix exponential operator. SIAM J. Numer. Anal. **34**(5), 1911–1925 (1997)
9. Hochbruck, M., Niehoff, J.: Approximation of matrix operators applied to multiple vectors. Math. Comput. Simul. **79**(4), 1270–1283 (2008)
10. Hochbruck, M., Ostermann, A.: Exponential integrators. Acta Numer. **19**, 209–286 (2010). doi:10.1017/S0962492910000048
11. in 't Hout, K.J., Weideman, J.A.C.: A contour integral method for the Black-Scholes and Heston equations. SIAM J. Sci. Comput. **33**(2), 763–785 (2011). doi:10.1137/090776081. http://dx.doi.org/10.1137/090776081
12. Lubich, C.: From Quantum to Classical Molecular Dynamics: Reduced Models and Numerical Analysis. Zurich Lectures in Advanced Mathematics. European Mathematical Society (EMS), Zürich (2008). doi:10.4171/067. http://dx.doi.org/10.4171/067
13. Saad, Y.: Iterative Methods for Sparse Linear Systems. Book out of print (2000). www-users.cs.umn.edu/~saad/books.html
14. Shampine, L.F.: Solving ODEs and DDEs with residual control. Appl. Numer. Math. **52**(1), 113–127 (2005)
15. Tal-Ezer, H.: Spectral methods in time for parabolic problems. SIAM J. Numer. Anal. **26**(1), 1–11 (1989)
16. Tal-Ezer, H.: On restart and error estimation for Krylov approximation of $w = f(A)v$. SIAM J. Sci. Comput. **29**(6), 2426–2441 (2007). doi:10.1137/040617868. http://dx.doi.org/10.1137/040617868
17. van den Eshof, J., Hochbruck, M.: Preconditioning Lanczos approximations to the matrix exponential. SIAM J. Sci. Comput. **27**(4), 1438–1457 (2006)
18. van der Vorst, H.A.: An iterative solution method for solving $f(A)x = b$, using Krylov subspace information obtained for the symmetric positive definite matrix A. J. Comput. Appl. Math. **18**, 249–263 (1987)
19. van der Vorst, H.A.: Iterative Krylov Methods for Large Linear Systems. Cambridge University Press, Cambridge (2003)

Computing Hyperbolic Matrix Functions Using Orthogonal Matrix Polynomials

Emilio Defez, Jorge Sastre, Javier Ibáñez, and Pedro A. Ruiz

Abstract Hyperbolic matrix functions play a fundamental role in the exact solution of coupled partial differential systems of hyperbolic type. For the numerical solution of these problems, analytic-numerical approximations are most suitable obtained by using the hyperbolic matrix functions $\sinh(A)$ and $\cosh(A)$. It is well known that the computation of both functions can be reduced to the cosine of a matrix $\cos(A)$, which can be effectively calculated, with the disadvantage, however, to require complex arithmetic even though the matrix A is real. In this work we focus on approximate calculation of the hyperbolic matrix cosine $\cosh(A)$ using the truncation of a Hermite matrix polynomials series for $\cosh(A)$. The proposed approximation allows the efficient computation of this matrix function. An illustrative example is given.

E. Defez (✉)
Instituto de Matemática Multidisciplinar, Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia, Spain
e-mail: edefez@imm.upv.es

J. Sastre
Instituto de Telecomunicaciones y Aplicaciones Multimedia, Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia, Spain
e-mail: jorsasma@iteam.upv.es

J. Ibáñez • P.A. Ruiz
Instituto de Instrumentación para Imagen Molecular, Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia, Spain
e-mail: jjibanez@dsic.upv.es; pruiz@dsic.upv.es

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_56,

1 Introduction

Coupled partial differential systems are frequent in many different situations [1–5] and many other fields. Coupled hyperbolic systems appear in microwave heating processes [6] and optics [7] for instance. The exact solution of a class of this problems, see [8], is given in terms of matrix functions, in particular, of hyperbolic sine and cosine of a matrix, $\sinh(A), \cosh(A)$, defined respectively by

$$\cosh(Ay) = \frac{e^{Ay} + e^{-Ay}}{2}, \ \sinh(Ay) = \frac{e^{Ay} - e^{-Ay}}{2}. \tag{1}$$

For the numerical solution of these problems, analytic-numerical approximations are most suitable obtained by using the hyperbolic matrix functions $\sinh(A)$ and $\cosh(A)$, see [8]. It is well known that the computation of both functions can be reduced to the cosine of a matrix, because $\sinh(A) = i\cos(A - \frac{i\pi}{2}I)$ and $\cosh(A) = cos(iA)$. Thus, the matrix cosine can be effectively calculated, [9, 10], with the disadvantage, however, to require complex arithmetic even though the matrix A is real, which contributes substantially to the computational overhead. Direct calculation through exponential matrix using (1) is costly. In this paper, we apply Hermite matrix polynomials to approximate $\sinh(A)$ and $\cosh(A)$, providing sharper bounds for Hermite matrix polynomials and the approximation error. Throughout this paper, $[x]$ and $Re(z)$ denote the integer part of the real number x and the real part of a complex number z. For a matrix $A \in C^{r\times r}$, $\parallel A \parallel_2$ and $\sigma(A)$ denote the two-norm and the spectrum (the set of all the eigenvalues) of the matrix A, respectively, and I_r denotes the identity matrix of order r.

2 Hermite Matrix Polynomial Series Expansions of Matrix Hyperbolic Cosine

For the sake of clarity in the presentation of the following results we recall some properties of Hermite matrix polynomials which have been established in [9, 11, 12]. From (3.4) of [11], for an arbitrary matrix A in $C^{r\times r}$, the nth Hermite matrix polynomial satisfies

$$H_n\left(x, \frac{1}{2}A^2\right) = n! \sum_{k=0}^{[\frac{n}{2}]} \frac{(-1)^k (xA)^{n-2k}}{k!(n-2k)!}, \tag{2}$$

and from its generating function in (3.1) and (3.2) [11] one gets

$$e^{txA - t^2 I} = \sum_{n\geq 0} H_n\left(x, \frac{1}{2}A^2\right) t^n/n!, \quad x, t \in C, |t| < \infty, \tag{3}$$

Taking $y = tx$ and $\theta = 1/t$ in (3) it follows that

$$e^{Ay} = e^{\frac{1}{\theta^2}} \sum_{n \geq 0} \frac{1}{\theta^n n!} H_n \left(\theta y, \frac{1}{2} A^2 \right), \ (\theta, y) \in C^2, \ \ A \in C^{r \times r} \ . \tag{4}$$

It is important to pay attention to the fact that the matrix A which defines the Hermite matrix polynomial sequence must be *positive definite*, see [12], i.e. $Re(z) > 0$ for all $z \in \sigma(A)$. This positive stable condition was imposed on the matrix A to guarantee the existence of $\sqrt{A}$ and some integral properties of Hermite polynomials, see [11], but it is not necessary to guarantee the expansion (4). Now, we will look for the Hermite matrix polynomials series expansion of the matrix hyperbolic cosine $\cos(Ax)$. To obtain it, given an arbitrary matrix $A \in C^{r \times r}$, by (1) using (4) and taking into account that, from [11], it follows that

$$H_n(-x, A) = (-1)^n H_n(x, A) \, ,$$

one gets the locking for expression:

$$\cosh(Ay) \ = e^{-\frac{1}{\lambda^2}} \sum_{n \geq 0} \frac{1}{\lambda^{2n} (2n)!} H_{2n} \left(y\lambda, \frac{1}{2} A^2 \right) \ . \tag{5}$$

Denoting by $CH_N(\lambda, A^2)$ the Nth partial sum of series (5) for $y = 1$, one gets the approximation

$$CH_N(\lambda, A^2) = e^{-\frac{1}{\lambda^2}} \sum_{n=0}^{N} \frac{1}{\lambda^{2n} (2n)!} H_{2n} \left(\lambda, \frac{1}{2} A^2 \right) \approx \cosh(A), \ \lambda \in C. \tag{6}$$

From [10] we have the following bound $\left\| H_{2n} \left(x, \frac{1}{2} A^2 \right) \right\|$ for Hermite matrix polynomials based on $||A^2||$:

$$\left\| H_{2n} \left(x, \frac{1}{2} A^2 \right) \right\| \leq (2n)! \, e \cosh \left(x \left\| A^2 \right\|^{\frac{1}{2}} \right), \ \forall x \in R, \ n \geq 0, \ \forall A \in C^{r \times r}. \tag{7}$$

Taking into account approximation (6) and bound (7), it follows that

$$\left\| \cosh(A) - CH_N(\lambda, A^2) \right\| \leq \frac{e^{1-\frac{1}{\lambda^2}} \cosh \left(\lambda \left\| A^2 \right\|^{\frac{1}{2}} \right)}{(\lambda^2 - 1)\lambda^{2N}}. \tag{8}$$

A similar approximate expression (6) and error bound (8) can be found for $\sinh(A)$.

3 Example

Let A be the non-diagonalizable matrix defined by

$$A = \begin{pmatrix} 3 & -1 & 1 \\ 2 & 0 & 1 \\ 1 & -1 & 2 \end{pmatrix}.$$

Using the minimal theorem the exact value of $\cosh(A)$ is

$$\cosh(A) = \begin{pmatrix} 7.389056098931 & -3.62686040784702 & 3.62686040784702 \\ 5.8459754641154 & -2.0837797730318 & 3.62686040784702 \\ 2.21911505626839 & -2.21911505626839 & 3.76219569108363 \end{pmatrix}.$$

Using (8), if $\lambda > 1$, for an admissible error $\varepsilon > 0$, we need choose a positive integer N so that the next inequality holds:

$$N \geq \frac{\log\left(\dfrac{e^{\left(1-\frac{1}{\lambda^2}\right)} \cosh\left(\lambda \left\|A^2\right\|^{\frac{1}{2}}\right)}{(\lambda^2-1)\,\epsilon}\right)}{2\log\lambda} \tag{9}$$

For example, if $\lambda = 1.8$ and $\varepsilon = 10^{-5}$ we need $N = 15$ to provide the required accuracy:

$$CH_{15}(1.8, A^2) = \begin{pmatrix} 7.3890560989307 & -3.62686040784702 & 3.62686040784702 \\ 5.8459754641154 & -2.08377977303177 & 3.62686040784702 \\ 2.21911505626839 & -2.21911505626839 & 3.76219569108363 \end{pmatrix},$$

and

$$\left\|\cosh(A) - CH_{15}(1.8, A^2)\right\|_2 = 1.85095 \times 10^{-15}.$$

In practice, the number of terms required to obtain a prefixed accuracy uses to be smaller than the one provided by (9). So for instance, taking the same $\lambda = 1.8$ and $N = 6$ one gets:

$$CH_6(1.8, A^2) = \begin{pmatrix} 7.3890548171477 & -3.6268592817884 & 3.6268592817884 \\ 5.84597418233707 & -2.08377864697777 & 3.6268592817884 \\ 2.21911490054867 & -2.21911490054867 & 3.76219553535930 \end{pmatrix},$$

and

$$\left\|\cosh(A) - CH_6(1.8, A^2)\right\|_2 = 2.90352 \times 10^{-6} \, .$$

The choice of parameter λ can still be refined. For example, taking $\lambda = 5$ and $N = 9$ one gets

$$\left\|\cosh(A) - CH_9(5, A^2)\right\|_2 = 3.07199 \times 10^{-14} \, .$$

Similar results are being obtained for $\sinh(A)$.

Acknowledgements This work has been supported by the *Universitat Politècnica de València* PAID-06-11-2020.

References

1. Sezgin, M.: Magnetohydrodynamics flows in a rectangular duct. Int. J. Numer. Methods Fluids **7**(7), 697–718 (1987)
2. King, A., Chou, C.: Mathematical modeling simulation and experimental testing of biochemical systems crash response. J. Biomech. **9**, 301–317 (1976)
3. Jódar, L., Navarro, E., Martín, J.A.: Exact and analytic-numerical solutions of strongly coupled mixed diffusion problems. Proc. Edinb. Math. Soc. **43**, 269–293 (2000)
4. Winfree, A.: When Times Breaks Down. Princeton University Press, Princeton (1987)
5. Morimoto, H.: Stability in the wave equation coupled with heat flows. Numerische Mathematik **4**(1), 136–145 (1962)
6. Pozar, D.: Microwave Engineering. Addison-Wesley, New York (1991)
7. Das, P.: Optical Signal Processing. Springer, New York (1991)
8. Jódar, L., Navarro, E., Posso, A., Casabán, M.: Constructive solution of strongly coupled continuous hyperbolic mixed problems. Appl. Numer. Math. **47**(3), 477–492 (2003)
9. Defez, E., Sastre, J., Ibáñez, J.J., Ruiz, P.A.: Computing matrix functions solving coupled differential models. Math. Comput. Model. **50**(5–6), 831–839 (2009)
10. Defez, E., Sastre, J., Ibáñez, J.J., Ruiz, P.A.: Computing matrix functions arising in engineering models with orthogonal matrix polynomials. Math. Comput. Model. **57**(7–8), 1738–1743 (2013)
11. Jódar, L., Company, R.: Hermite matrix polynomials and second order matrix differential equations. J. Approx. Theory Appl. **12**(2), 20–30 (1996)
12. Defez, E., Jódar, L.: Some applications of Hermite matrix polynomials series expansions. J. Comput. Appl. Math. **99**, 105–117 (1998)

Counter-Harmonic Mean of Symmetric Positive Definite Matrices: Application to Filtering Tensor-Valued Images

Jesús Angulo

Abstract Mathematical morphology is a nonlinear image processing methodology based on the computation of supremum (dilation operator) and infimum (erosion operator) in local neighborhoods called structuring elements. This paper deals with computation of supremum and infimum operators for symmetric positive definite (SPD) matrices, which are the basic ingredients for the extension mathematical morphology to SPD matrices-valued images. Approximation to the supremum and infimum associated to the Löwner ellipsoids are computed as the asymptotic cases of nonlinear averaging using the original notion of counter-harmonic mean for SPD matrices. Properties of this approach are explored, including also image examples.

1 Context, Aim and State-of-the-Art

Mathematical morphology is a nonlinear image processing methodology originally developed for binary and greyscale images [13]. It is based on the computation of maximum $\bigwedge$ (dilation operator) and minimum $\bigvee$ (erosion operator) in local neighborhoods called structuring elements [14]. That means that the definition of morphological operators needs a partial ordering relationship $\leq$ between the points to be processed. More precisely, for a real valued image $f : E \to \mathbb{R}$, the flat dilation and erosion of image f by structuring element B are defined respectively by $\delta_B(f)(\mathbf{x}) = \left\{f(\mathbf{y}) : f(\mathbf{y}) = \bigwedge_{\mathbf{z}}[f(\mathbf{z})], \mathbf{z} \in B_{\mathbf{x}}\right\}$ and $\varepsilon_B(f)(\mathbf{x}) = \left\{f(\mathbf{y}) : f(\mathbf{y}) = \bigvee_{\mathbf{z}}[f(\mathbf{z})], \mathbf{z} \in \check{B}_{\mathbf{x}}\right\}$, where $B_{\mathbf{x}} \subset E$ is the structuring element centered at point $\mathbf{x} \in E$, and $\check{B}$ is the reflection of structuring element with respect to the origin.

J. Angulo (✉)
CMM-Centre de Morphologie Mathématique, Mathématiques et Systèmes, MINES Paristech, 35, rue Saint-Honor, 77305 Fontainebleau cedex, France
e-mail: jesus.angulo@mines-paristech.fr

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_57,

Theory of morphological operators has been formulated in the general framework of complete lattices [11]: a complete lattice $(\mathscr{L}, \leq)$ is a partially ordered set $\mathscr{L}$ with order relation $\leq$, a supremum written $\bigvee$, and an infimum written $\bigwedge$, such that every subset of $\mathscr{L}$ has a supremum (smallest upper bound) and an infimum (greatest lower bound). Let $\mathscr{L}$ be a complete lattice. A dilation $\delta : \mathscr{L} \rightarrow \mathscr{L}$ is a mapping commuting with suprema, i.e., $\delta\left(\bigvee_i X_i\right) = \bigvee_i \delta\left(X_i\right)$. An erosion $\varepsilon : \mathscr{L} \rightarrow \mathscr{L}$ commutes with infima, i.e., $\delta\left(\bigwedge_i X_i\right) = \bigwedge_i \delta\left(X_i\right)$. Then the pair (ε, δ) is called an adjunction on $\mathscr{L}$ if for very $X, Y \in \mathscr{L}$, it holds: $\delta(X) \leq Y \Leftrightarrow X \leq \varepsilon(Y)$.

Matrix and tensor valued images appear nowadays in various image processing fields and applications [15]: structure tensor images representing the local orientation and edge information [10]; diffusion tensor magnetic resonance imaging (DT-MRI) [5]; covariance matrices in different modalities of radar imaging [4]; etc. In this paper we are interested in matrix-valued images considered as a spatial structured matrix field $f(\mathbf{x})$ such that $f : E \subset \mathbb{Z}^2, \mathbb{Z}^3 \longrightarrow \mathrm{SPD}(n)$, where E is the support space of pixels and, in particular, we focuss on (real) symmetric positive definite $n \times n$ matrices $\mathrm{SPD}(n)$. The reader interested in positive definite matrices is referred to the excellent monograph [6]. More precisely, let $\mathfrak{A} = \{A_i\}_{i=1}^N$ be a finite set of N matrices, where $A_i \in \mathrm{SPD}(n)$, we are aiming at computing the supremum $\sup(\mathfrak{A}) = A_\vee$ and the infimum $\inf(\mathfrak{A}) = A_\wedge$ matrices, such that $A_\vee$, $A_\wedge \in \mathrm{SPD}(n)$. As mentioned above, if the operators $\sup(\mathfrak{A})$ and $\inf(\mathfrak{A})$ are defined, dilation and erosion operators are stated for any image $f \in \mathscr{F}(E, \mathrm{SPD(n)})$ and any structuring element.

Extension of mathematical morphology to matrix-valued images has been previously addressed according to two different approaches. The first one [9] is based on the Löwner partial ordering $\leq^L$: $\forall A, B \in \mathrm{SPD}(n)$, $A \leq^L B \Leftrightarrow B - A \in \mathrm{SPD}(n)$, and where the supremum and infimum of a set of matrices are computed using convex matrix analysis tools (penumbral cones of each matrix, minimal enclosing circle of basis, computation of vertex of associated penumbra matrix). There is a geometrical interpretation viewing the tensors $\mathrm{SPD}(n)$ as ellipsoids: the supremum of a set of tensors is the smallest ellipsoid enclosing the ellipsoids associated to all the tensors; the infimum is the largest ellipsoid which is contained in all the ellipsoids. The second approach [8] corresponds to the generalization of a morphological PDE to matrix data. Finding the unique smallest enclosing ball of a set of points in a particular space (also known as the minimum enclosing ball or the one-center problem) is related to the Löwner ordering in the case of $\mathrm{SPD}(n)$ matrices [1, 3].

We have recently shown in [2] how the counter-harmonic mean [7] can be used to introduce nonlinear operators which asymptotically mimic dilation and erosion. In particular, we have proved in [2] the advantages of the counter-harmonic mean against the classical P-mean to approximate supremum and infimum. The extension of P-mean to $\mathrm{SPD}(n)$ matrices was considered in [12] for diffusion tensor imaging. We introduce in this paper how the extension of counter-harmonic mean to $\mathrm{SPD}(n)$ matrices is very natural and leads to an efficient operator to robustly approximate the supremum/infimum of a set of matrices.

2 Counter-Harmonic Mean for SPD Matrices

The counter-harmonic mean (CHM) belongs to the family of the power means [7]. We propose a straightforward generalization of CHM for SPD(n) matrices.

Definition 1. Given $\mathfrak{A} = \{A_i\}_{i=1}^{N}$, a finite set of N matrices, where $A_i \in \mathrm{SPD}(n)$, the symmetrized counter-harmonic matrix mean (CHMM) of order P, $P \in \mathbb{R}$, is defined by

$$\kappa^P(\mathfrak{A}) = \left(\sum_{i=1}^{N} A_i^P\right)^{-1/2} \left(\sum_{i=1}^{N} A_i^{P+1}\right)\left(\sum_{i=1}^{N} A_i^P\right)^{-1/2} \tag{1}$$

The asymptotic values of the CHMM with $P \to +\infty$ and $P \to -\infty$ can be used to define approximations to the supremum and infimum of a set of matrices.

Definition 2. The supremum and the infimum of a set $\mathfrak{A} = \{A_i\}_{i=1}^{N}$ of SPD(n) matrices are defined respectively as

$$A_\vee = \sup(\mathfrak{A}) = \lim_{P\to+\infty} \kappa^P(\mathfrak{A}), \tag{2}$$

and

$$A_\wedge = \inf(\mathfrak{A}) = \lim_{P\to-\infty} \kappa^P(\mathfrak{A}), \tag{3}$$

Proposition 1. *Given a set $\mathfrak{A}$ of $SPD(n)$ matrices, the following properties hold.*

(i) CHMM of $\mathfrak{A}$ is a rotationally invariant operation for any value of P (including $P \to \pm\infty$).

(ii) CHMM of $\mathfrak{A}$ is for any value of P (including $P \to \pm\infty$) invariant to scaling transformations, i.e., multiplication by a real constant $\alpha \in \mathbb{R}$.

(iii) CHMM of $\mathfrak{A}$ produces a symmetric positive definite matrix for any value of P (including $P \to \pm\infty$).

(iv) Due to the fact that the CHMM is not associative, sup($\mathfrak{A}$) *and* inf($\mathfrak{A}$) *do not yield dilation and erosion operators over $SPD(n)$ (they do not commute with the "union" and the "intersection").*

Proof. (i) Let us consider that the rotation is given by the matrix $O \in SO(n)$. We know from linear algebra that the P-th power A^P of a diagonalized matrix is achieved by taking the P-th power of the eigenvalues:

$$A^P = V\mathrm{diag}\left((\lambda_1(A_i))^P, \cdots, (\lambda_n(A_i))^P\right) V^{\mathrm{T}}.$$

On the other hand, since $\sum_{i=1}^{N} A_i^P$ is positive definite, there exists an orthogonal matrix V_P and a diagonal matrix Λ_P such that $\sum_{i=1}^{N} A_i^P = V_P \Lambda_P V_P^{\mathrm{T}}$. Hence, if we apply the rotation, we have

$$\begin{aligned}
&\left(\sum_{i=1}^{N}(OA_iO^{\mathrm{T}})^P\right)^{-1/2}\left(\sum_{i=1}^{N}(OA_iO^{\mathrm{T}})^{P+1}\right)\left(\sum_{i=1}^{N}(OA_iO^{\mathrm{T}})^P\right)^{-1/2}\\
&=\left(\sum_{i=1}^{N}OA_i^PO^{\mathrm{T}}\right)^{-1/2}\left(\sum_{i=1}^{N}OA_i^{P+1}O^{\mathrm{T}}\right)\left(\sum_{i=1}^{N}OA_i^PO^{\mathrm{T}}\right)^{-1/2}\\
&=\left(O\left(\sum_{i=1}^{N}A_i^P\right)O^{\mathrm{T}}\right)^{-1/2}\left(O\left(\sum_{i=1}^{N}A_i^{P+1}\right)O^{\mathrm{T}}\right)\left(O\left(\sum_{i=1}^{N}\right)A_i^PO^{\mathrm{T}}\right)^{-1/2}\\
&=\left(OV_P\Lambda_PV_P^{\mathrm{T}}O^{\mathrm{T}}\right)^{-1/2}\left(OV_{P+1}\Lambda_{P+1}V_{P+1}^{\mathrm{T}}O^{\mathrm{T}}\right)\left(OV_P\Lambda_PV_P^{\mathrm{T}}O^{\mathrm{T}}\right)^{-1/2}
\end{aligned}$$

Considering the fact that $OO^{\mathrm{T}}=I$ and that $OV_P\in SO(3)$, we can write

$$O\left(\sum_{i=1}^{N}A_i^P\right)^{-1/2}\left(\sum_{i=1}^{N}A_i^{P+1}\right)\left(\sum_{i=1}^{N}A_i^P\right)^{-1/2}O^{\mathrm{T}}$$

and consequently

$$\kappa^P\left(\{OA_iO^{\mathrm{T}\}_{i=1}^{N}}\right)=O\kappa^P\left(\{A_i\}_{i=1}^{N}\right)O^{\mathrm{T}}$$

(ii) By considering scaling by parameter $\alpha\in\mathbb{R}$, $\alpha\neq 0$, we have

$$\begin{aligned}
\kappa^P\left(\{\alpha A_i\}_{i=1}^{N}\right)&=\left(\sum_{i=1}^{N}(\alpha A_i)^P\right)^{-1/2}\left(\sum_{i=1}^{N}(\alpha A_i)^{P+1}\right)\left(\sum_{i=1}^{N}(\alpha A_i)^P\right)^{-1/2}\\
&=\alpha^{-P/2}\left(\sum_{i=1}^{N}A_i^P\right)^{-1/2}\alpha^{P+1}\left(\sum_{i=1}^{N}A_i^{P+1}\right)\alpha^{-P/2}\left(\sum_{i=1}^{N}A_i^P\right)^{-1/2}\\
&=\alpha^{-P/2}\alpha^{P+1}\alpha^{-P/2}\left(\sum_{i=1}^{N}A_i^P\right)^{-1/2}\left(\sum_{i=1}^{N}A_i^{P+1}\right)\left(\sum_{i=1}^{N}A_i^P\right)^{-1/2}\\
&=\alpha\kappa^P\left(\{A_i\}_{i=1}^{N}\right)
\end{aligned}$$

(iii) By construction, the P-th power A^P and the inverse square root $A^{-1/2}$ have positive eigenvalues whenever A has. Similarly, the sum and the product of positive definite matrices preserves also the positiveness.

(iv) Let consider two sets of SPD(n) matrices $\mathfrak{A}=\{A_i\}_{i=1}^{N}$ and $\mathfrak{A}'=\{A_j\}_{j=N+1}^{M}$. Due to the fact that the counter-harmonic matrix mean is not associative, it cannot be ensured that there exist always a value of P such that

$$\lim_{P\to+\infty}\kappa^P\left(\{A_k\}_{k=1}^{M}\right)$$

is equal to

$$\lim_{P\to+\infty} \kappa^P \left(\lim_{P\to+\infty} \kappa^P \left(\{A_i\}_{i=1}^N\right), \lim_{P\to+\infty} \kappa^P \left(\{A_j\}_{j=N+1}^M\right)\right)$$

and consequently the operators $\sup(\mathfrak{A})$ do not commute with "supremum". A similar result is observed for the erosion.

The following result gives a spectral interpretation of asymptotic cases in SPD(2).

Proposition 2. *Given $\mathfrak{A} = \{A_i\}_{i=1}^N$, a finite set of N matrices, where $A_i \in \mathrm{SPD}(2)$. Let $\Lambda(A_i)$ and $\lambda(A_i)$ (with $\Lambda(A_i) \geq \lambda(A_i) \geq 0$) be the two eigenvalues of A_i. Then*

$$A_\vee = \sup(\mathfrak{A}) = \lim_{P\to+\infty} \kappa^P(\mathfrak{A}),$$

is a SPD(2) matrix with eigenvalues $\Lambda(A_\vee)$ and $\lambda(A_\vee)$, where $\Lambda(A_\vee) = \max(\Lambda(A_1), \Lambda(A_2) \cdots \Lambda(A_N))$, and its corresponding eigenvector is the eigenvector of $A_\vee$, and the remaining eigenvalue $\lambda(A_\vee)$ is the second largest eigenvalue from $\{\Lambda(A_i), \lambda(A_i)\}$; the corresponding eigenvector is the orthogonal to the major one.

A spectral characterization of $A_\wedge$ is obtained by replacing largest by smallest eigenvalues. We conjecture that this result may be extended to SPD(n), $n > 2$, but the proof is not straightforward.

Proof. Let us write each SPD(2) matrix in the form $A = V_i \operatorname{diag}(\Lambda_i \lambda_i) V_i^T$ such that $\Lambda_i \geq \lambda_i > 0$ and where the rotation matrix is parameterized by the angle θ_i:

$$V_i = \begin{pmatrix} \cos\theta_i & -\sin\theta_i \\ -\sin\theta_i & \cos\theta_i \end{pmatrix}.$$

Hence we have

$$\sum_{i=1}^N A_i^{P+1} = \begin{pmatrix} \sum_{i=1}^N \Lambda_i^{P+1}\cos^2\theta_i + \lambda_i^{P+1}\sin^2\theta_i & \sum_{i=1}^N(\lambda_i^{P+1} - \Lambda_i^{P+1})\cos\theta_i\sin\theta_i \\ \sum_{i=1}^N(\lambda_i^{P+1} - \Lambda_i^{P+1})\cos\theta_i\sin\theta_i & \sum_{i=1}^N \lambda_i^{P+1}\cos^2\theta_i + \Lambda_i^{P+1}\sin^2\theta_i \end{pmatrix}.$$

The eigenvalues of $\sum_{i=1}^N A_i^{P+1}$ are given by

$$(\Lambda(P+1), \lambda(P+1))$$
$$= \frac{1}{2}\left[\sum_{i=1}^N \Lambda_i^{P+1}\cos^2\theta_i + \lambda_i^{P+1}\sin^2\theta_i + \sum_{i=1}^N \lambda_i^{P+1}\cos^2\theta_i + \Lambda_i^{P+1}\sin^2\theta_i\right.$$

$$\pm \left\{ \left(\sum_{i=1}^{N} \Lambda_i^{P+1} \cos^2\theta_i + \lambda_i^{P+1} \sin^2\theta_i - \sum_{i=1}^{N} \lambda_i^{P+1} \cos^2\theta_i + \Lambda_i^{P+1} \sin^2\theta_i \right)^2 \right.$$
$$\left. + 4 \left(\sum_{i=1}^{N} (\lambda_i^{P+1} - \Lambda_i^{P+1}) \cos\theta_i \sin\theta_i \right)^2 \right\}^{1/2} \Bigg],$$

which can be simplified to

$$(2N)^{-1} \textstyle\sum_{i=1}^{N} \left(\Lambda_i^{P+1} + \lambda_i^{P+1} \right)$$
$$\pm (2N)^{-1} \sqrt{ \left(\textstyle\sum_{i=1}^{N} \left(\Lambda_i^{P+1} - \lambda_i^{P+1} \right) \cos(2\theta_i) \right)^2 + \left(\textstyle\sum_{i=1}^{N} \left(\Lambda_i^{P+1} - \lambda_i^{P+1} \right) \sin(2\theta_i) \right)^2 }.$$

We are interested in the limit case:

$$\Lambda(A_\vee) = \lim_{P \to +\infty} \Lambda(P)^{-1/2} \Lambda(P+1) \Lambda(P)^{-1/2} = \lim_{P \to +\infty} \frac{\Lambda(P+1)}{\Lambda(P)} = \max\{\Lambda_i\}.$$

For the second eigenvalue, we first consider the product

$$\Lambda(P+1)\lambda(P+1) = (4N^2)^{-1} \left(\sum_{i=1}^{N} \left(\Lambda_i^{P+1} + \lambda_i^{P+1} \right) \right)^2$$
$$- (4N^2)^{-1} \left(\sum_{i=1}^{N} \left(\Lambda_i^{P+1} - \lambda_i^{P+1} \right) \cos(2\theta_i) \right)^2$$
$$- (4N^2)^{-1} \left(\sum_{i=1}^{N} \left(\Lambda_i^{P+1} - \lambda_i^{P+1} \right) \sin(2\theta_i) \right)^2 .$$

By the invariance under scaling, we can consider without loss of generality that $\Lambda_1 = \Lambda_2 = \cdots \Lambda_n = 1$ and $\Lambda_i \leq 1$. We also assume $\theta_1 = \theta_2 = \cdots, \theta_n = 0$ and $\theta_j \neq 0$, $j = n+1, n+2, \cdots, N$. As $\alpha(P+1) = \sum_{i=1}^{n} \Lambda_i^{P+1}$, with $1 \leq \alpha(P+1) \leq N$, is the dominant term, and considering defining the element

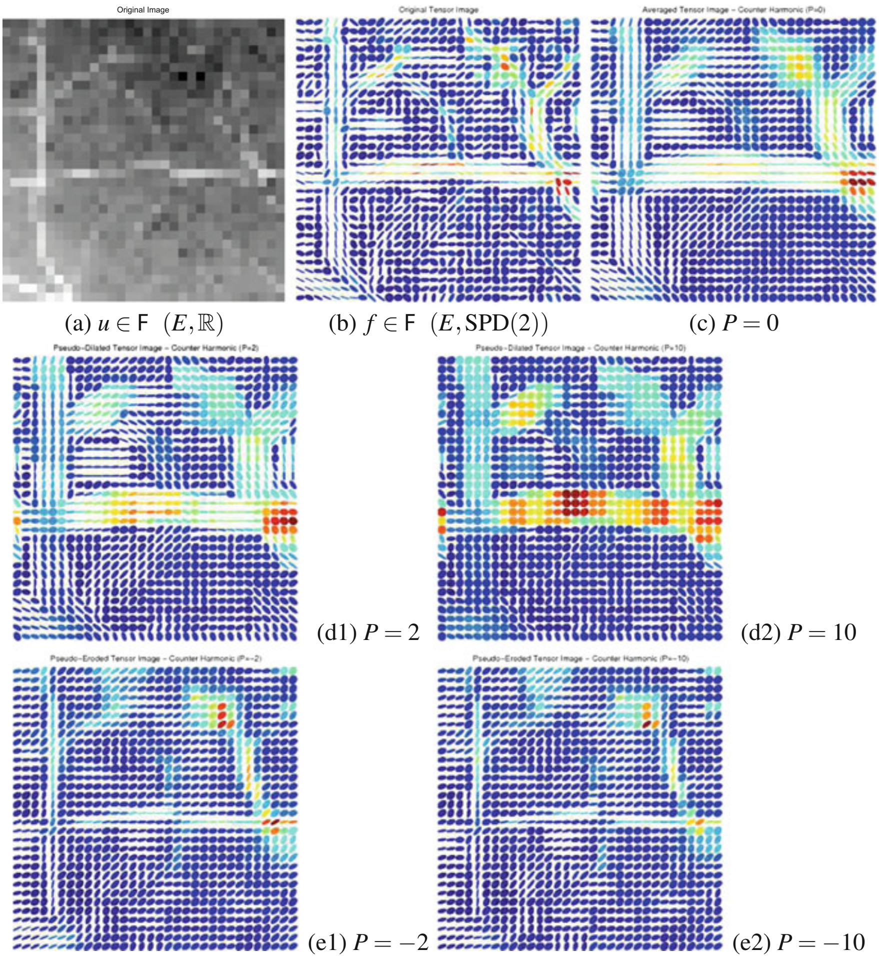
$$M = \max\left(\Lambda_{n+1}, \Lambda_{n+2}, \cdots, \Lambda_N, \lambda_1, \cdots, \lambda_N \right),$$

then we can approximate

$$\Lambda(P+1)\lambda(P+1) = Cte \cdot \alpha(P+1) M^{P+1} + M^{P+1}\,\mathrm{O}(1)$$

Finally, we have

$$\lambda(A_\vee) = \lim_{P \to +\infty} \frac{\alpha(P+1) M^{P+1} + M^{P+1}\,\mathrm{O}(1)}{\alpha(P) M^{P} + M^{P}\,\mathrm{O}(1)} = M.$$



(a) $u \in \mathsf{F}\ (E,\mathbb{R})$ (b) $f \in \mathsf{F}\ (E,\mathrm{SPD}(2))$ (c) $P = 0$

(d1) $P = 2$ (d2) $P = 10$

(e1) $P = -2$ (e2) $P = -10$

Fig. 1 Counter-harmonic matrix mean based processing of SPD(2) matrix-valued image: (**a**) initial *gray-level* image from retina vessels, (**b**) corresponding structure tensor image, (**c**–**e**) tensor filtered image by CHMM κ^P ($\mathfrak{A}$), for different values of order P. The local neighborhood (structuring element B) is a square of 3 × 3 pixels

Figure 1b depicts an example of $SPD(n)$ matrix-valued image. This image corresponds to the structure tensors obtained from the gray-level image Fig. 1a, representing the local orientation and edge information, which is computed by Gaussian smoothing of the dyadic product $\nabla u \nabla u^T$ of an image $u(x, y)$ [10]. Using the symmetrized counter-harmonic matrix mean operator κ^P ($\mathfrak{A}$) computed in local neighborhoods, various values of P are compared. In particular, $P = 0$ in Fig. 1c which corresponds to the arithmetic mean filtered image, $P = 2$ and $P = 10$ in Fig. 1d1, d2 are pseudo-dilations, $P = -2$ and $P = -10$ in Fig. 1e1, e2 can be considered as pseudo-erosions. It is natural to consider that the matrices $A_\vee$ and $A_\wedge$, associated respectively to the limit cases $P = 10$ and $P = -10$, can be interpreted geometrically similarly to the supremum/infimum associated to the Löwner ordering: $A_\vee$ "tends to be" the smallest ellipsoid enclosing the ellipsoids of $\mathfrak{A}$ and $A_\wedge$ "tends to be" the largest ellipsoid which is contained in all the ellipsoids.

References

1. Afsari, B.: Riemannian Lp center of mass: existence, uniqueness, and convexity. Proc. Am. Math. Soc. **139**, 655–674 (2011)
2. Angulo, J.: Pseudo-morphological image diffusion using the counter-harmonic paradigm. In: Proceedings of Acivs'2010 (2010 Advanced Concepts for Intelligent Vision Systems). Lecture Notes in Computer Science, Part I, vol. 6474, pp. 426–437. Springer, Heidelberg (2010)
3. Arnaudon, M., Nielsen, F.: On approximating the Riemannian 1-center. Comput. Geom. **46**(1), 93–104 (2013)
4. Barbaresco, F.: New foundation of radar Doppler signal processing based on advanced differential geometry of symmetric spaces: Doppler matrix CFAR and radar application. In: Proceedings of Radar 09 Conference, Bordeaux (2009)
5. Basser, P.J., Mattiello, J., LeBihan, D.: MR diffusion tensor spectroscopy and imaging. Biophys. J. **66**, 259–267 (1994)
6. Bhatia, R.: Positive Definite Matrices. Princeton University Press, Princeton (2007)
7. Bullen, P.S.: Handbook of Means and Their Inequalities. Springer, New York (1987)
8. Burgeth, B., Bruhn, A., Didas, S., Weickert, J., Welk, M.: Morphology for tensor data: ordering versus PDE-based approach. Image Vis. Comput. **25**(4), 496–511 (2007)
9. Burgeth, B., Papenberg, N., Bruhn, A., Welk, M., Weickert, J.: Mathematical morphology for matrix fields induced by the Loewner ordering in higher dimensions. Signal Process. **87**(2), 277–290 (2007)
10. Förstner, W., Gülch, E.: A fast operator for detection and precise location of distinct points, corners and centres of circular features. In: Proceedings of ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data, pp. 281–304 (1987)
11. Heijmans, H.J.A.M.: Morphological Image Operators. Academic, Boston (1994)
12. Herberthson, M., Brun, A., Knutsson, H.: P-averages of diffusion tensors. In: Swedish Symposium in Image Analysis 2007 (SSBA'07)
13. Serra, J.: Image Analysis and Mathematical Morphology. Academic, London (1982)
14. Soille, P.: Morphological Image Analysis. Springer, Berlin (1999)
15. Weickert, J., Hagen, H. (eds.): Visualization and Processing of Tensor Fields. Springer, Berlin (2006)

Heat Conduction Problem for Double-Layered Ball

Sanda Blomkalna and Andris Buikis

Abstract Heat conduction models for double layered spherical sample are developed. Parabolic (classic, based on Fourier's Law) and hyperbolic (based on Modified Fourier's Law) heat conduction equations are used to describe processes in the sample during Intensive Quenching. Solution and numerical results are obtained for 1D model using Conservative Averaging method and transforming the original problem for a sphere to a new problem for a slab, with non classic boundary condition. Models include boundary conditions of third kind and non-linear BC case. Numerical results are presented for several relaxation time and initial heat flux values.

1 Introduction

Classical heat conduction equation, based on Fourier's Law

$$q(x,t) = -k\nabla T(x,t), \tag{1}$$

where $q(x,t)$—heat flux vector, $\nabla T(x,t)$—temperature gradient, k—thermal conductivity, usually is suitable for describing heat conduction processes. However, for some specific modern problems, modified Fourier's Law is more appropriate [10]:

$$q(x,t+\tau) = -k\nabla T(x,t). \tag{2}$$

S. Blomkalna (✉)
Faculty of Physics and Mathematics, University of Latvia, Zellu iela 8-50, Riga, LV-1002, Latvia
e-mail: sanda.blomkalna@lu.lv

A. Buikis
Institute of Mathematics and Computer Sciences, University of Latvia, Raina bulv. 29, Riga, LV-1459, Latvia
e-mail: buikis@latnet.lv

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*, Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_58,

where τ denotes relaxation time, $\tau > 0$. It is material dependent and represents time lag needed to establish heat flux when temperature gradient is suddenly imposed.

Hyperbolic heat conduction equation allows finite thermal signal speed, wave-like behaviour of heat and is better suited for describing fast transient effects or processes that happen for very short time intervals, have extreme cooling or heating rates, processes with relaxation time, laser heating, processing biological materials, etc. In our case we are interested in the Intensive Steel Quenching [5]—steel parts are rapidly and uniformly cooled in water-polymer solutions. This method is environmentally friendly and cheaper than quenching using oil.

We model quenching process for a layered spherical sample, taking into account some practical limitations and conditions. Our main goal is to model industrial process and obtain approximate solutions for otherwise difficult problems.

This paper is organized as follows. In Sect. 2 we develop models for quenching process. In Sect. 3 we transform original models and use Conservative Averaging method (CAM) to reduce complexity of problems. Results are presented in Sect. 4 for multiple values of parameters.

2 Mathematical Formulation of Models

Our models are formulated for a ball consisting of two spherical layers made of materials with possibly different properties (Fig. 1). The inner layer is much smaller compared to the outer layer (corresponding to the experimental sample and thermocouple nozzle at the very centre of it) [7, 8]. When examining experimental results (Fig. 2) [4], hyperbolic heat conduction equation was proposed as a better mathematical description of processes in the sample.

We develop two 1D models—parabolic heat conduction equation corresponds to the inner part (function $U_0(x,t)$ and matching parameters). For the outer layer (function $U_1(x,t)$ and matching parameters) we use parabolic equation for the first model and hyperbolic for the second one so we can consider parabolic-parabolic problem and parabolic-hyperbolic problem. We also include hyperbolic-hyperbolic model for theoretical point of view, however we note that relaxation time corresponding to the inner part would be significantly smaller compared to the outer steel layer, so parabolic equation generally is sufficient for describing inner layer. It should be noted that we use relatively large relaxation time values. Physically it is connected with martensite forming.

Let $r = 0$ be the symmetry centre of the sample. On the outer surface there is heat exchange with environment (third type boundary condition or non-linear boundary condition

$$\frac{\partial T}{\partial x}\,|_{x=R} = -\frac{1}{k_1}\cdot\beta^m[T - T_B(t)]^m \tag{3}$$

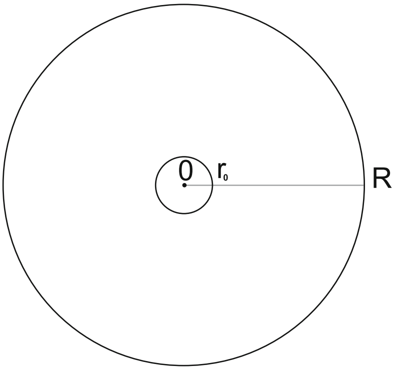


Fig. 1 Geometry of the sample

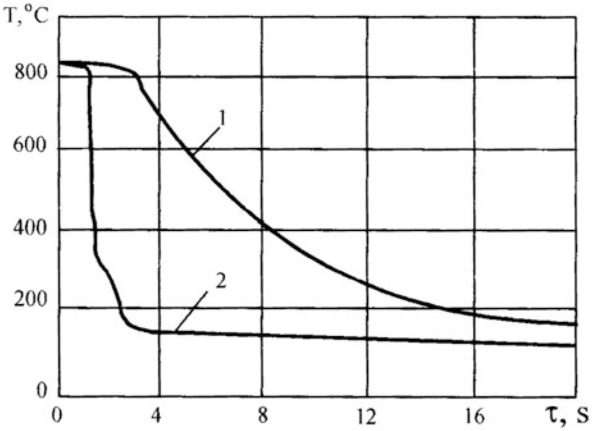


Fig. 2 Experimental results—temperature function of cylinder-shaped sample. 1—centre, 2—surface

that describes water boiling, $m \in [3, 3\frac{1}{3}]$, $T_B(t)$—saturation temperature (boiling point of quenchant), $\beta > 0$—a constant.) The nonlinear condition was proposed in [6] as mathematical description of situation. Here the third kind BC is used in models, but numerical results also cover case (3).

Parabolic-parabolic problem (4):

$$\begin{array}{cc}
\frac{\partial T_0}{\partial t} = a_0^2 \frac{1}{r^2} \frac{\partial}{\partial r}(r^2 \frac{\partial T_0}{\partial r}) + F_0'(r,t), & \frac{\partial T_1}{\partial t} = a_1^2 \frac{1}{r^2} \frac{\partial}{\partial r}(r^2 \frac{\partial T_1}{\partial r}) + F_1'(r,t), \\
r \in (0, r_0) & r \in (r_0, R) \\
r^2 \frac{\partial T_0}{\partial r} \mid_{r=0} = 0, & (k_1 r^2 \frac{\partial T_1}{\partial r} + h_1' T_1) \mid_{r=R} = \varphi_1'(t), \\
T_0 \mid_{t=0} = N_0; & T_1 \mid_{t=0} = N_0.
\end{array} \quad (4)$$

Parabolic-hyperbolic problem (5):

$$\begin{array}{cc}
\frac{\partial T_0}{\partial t}=a_0^2\frac{1}{r^2}\frac{\partial}{\partial r}(r^2\frac{\partial T_0}{\partial r})+F_0'(r,t), & \tau_r\frac{\partial^2 T_1}{\partial t^2}+\frac{\partial T_1}{\partial t}=a_1^2\frac{1}{r^2}\frac{\partial}{\partial r}(r^2\frac{\partial T_1}{\partial r})+F_1'(r,t),\\
r\in(0,r_0) & r\in(r_0,R)\\
r^2\frac{\partial T_0}{\partial r}\mid_{r=0}=0, & (k_1r^2\frac{\partial T_1}{\partial r}+h_1'T_1)\mid_{r=R}=\varphi_1'(t),\\
T_0\mid_{t=0}=N_0; & T_1\mid_{t=0}=N_0,\\
 & \frac{\partial T_1}{\partial t}\mid_{t=0}=M_1,
\end{array} \tag{5}$$

Hyperbolic-hyperbolic problem (6):

$$\begin{array}{cc}
\tau_0\frac{\partial^2 T_0}{\partial t^2}+\frac{\partial T_0}{\partial t}=a_0^2\frac{1}{r^2}\frac{\partial}{\partial r}(r^2\frac{\partial T_0}{\partial r})+F_0'(r,t), & \tau_r\frac{\partial^2 T_1}{\partial t^2}+\frac{\partial T_1}{\partial t}=a_1^2\frac{1}{r^2}\frac{\partial}{\partial r}(r^2\frac{\partial T_1}{\partial r})+F_1'(r,t),\\
r\in(0,r_0) & r\in(r_0,R)\\
r^2\frac{\partial T_0}{\partial r}\mid_{r=0}=0, & (k_1r^2\frac{\partial T_1}{\partial r}+h_1'T_1)\mid_{r=R}=\varphi_1'(t),\\
T_0\mid_{t=0}=N_0, & T_1\mid_{t=0}=N_0,\\
\frac{\partial T_0}{\partial t}\mid_{t=0}=M_1, & \frac{\partial T_1}{\partial t}\mid_{t=0}=M_1,
\end{array} \tag{6}$$

where for $i=0,1$: $a_i^2=\frac{k_i}{c_i\cdot\rho_i}$, k_i—heat conduction coefficient, c_i—specific heat capacity, ρ_i—density, h_i'—heat exchange coefficient, $\varphi_1'(t)$—temperature of environment, F_i'—inner heat generation function.

Conjunction conditions on the surface between both layers:

$$T_0\mid_{r=r_0-0}=T_1\mid_{r=r_0+0}, \tag{7}$$

$$r^2\cdot k_0\frac{\partial T_0}{\partial r}\mid_{r=r_0-0}=r^2\cdot k_1\frac{\partial T_1}{\partial r}\mid_{r=r_0+0}. \tag{8}$$

3 Transformation of the Original Problem

The process for parabolic-hyperbolic model is described (procedure is similar for the parabolic-parabolic and hyperbolic-hyperbolic models).

We want to simplify the problem, so we use CAM [1–3]. Conservative Averaging method is an approximate analytical and numerical method for solving partial differential equations. It reduces the complexity of problem by decreasing the domain where we look for the solution. According to the method we introduce integral average values:

$$u_0(t)=\frac{1}{H}\int_0^{r_0}r^2T_0(r,t)dr,\qquad f_0(t)=\frac{1}{H}\int_0^{r_0}r^2F_0(r,t)dr, \tag{9}$$

$$H=\int_0^{r_0}r^2dr, \tag{10}$$

multiply main Eq. (5) by r^2 and integrate main equation over $[0,r_0]$

$$\int_0^{r_0} r^2 \frac{\partial T_0}{\partial t} dr = \int_0^{r_0} a_0^2 r^2 \frac{1}{r^2} (\frac{\partial}{\partial r} (r^2 \frac{\partial T_0}{\partial r}) dr + \int_0^{r_0} r^2 F_0 dr. \tag{11}$$

$$\frac{\partial u_0}{\partial t} = a_0^2 (r^2 \frac{\partial T_0}{\partial r} |_0^{r_0}) + f_0. \tag{12}$$

The second conjunction condition can be expressed in form

$$r^2 \frac{\partial T_0}{\partial r} \mid_{r=r_0-0} = r^2 \frac{k_1}{k_0} \frac{\partial T_1}{\partial r} |_{r=r_0+0} \qquad \frac{k_1}{k_0} \neq 1. \tag{13}$$

Using conjunction condition and boundary condition, we get fundamental relation:

$$\frac{\partial u_0}{\partial t} = a_0^2 (r^2 \frac{k_1}{k_0} \frac{\partial T_1}{\partial r}) |_{r=r_{0+0}} + f_0. \tag{14}$$

The original problem for two-layered ball can now be transformed into a new one, in a smaller region $r \in (r_0, R)$. Since r_0 is physically small value we can use approximation with a constant in r direction. We assume that function T_0 is constant over $(0, r_0)$ and using first conjunction condition and fundamental relation we get boundary condition on $r = r_0$.

$$T_0(r,t) \approx u_0(t) \approx T_1(r_0, t). \tag{15}$$

Approximation with higher order polynomials or exponential approximation can also be used to describe unknown function T_0.

To emphasize the difference between original and transformed problem, we denote the function we are looking for as $W(r,t)$ instead of $T_1(r,t)$. The fundamental relation is in form

$$\frac{\partial W}{\partial t} = a_0^2 (r^2 \frac{k_1}{k_0} \frac{\partial W}{\partial r}) |_{r=r_{0+0}} + f_0. \tag{16}$$

From the fundamental relation we derive a non classic boundary condition for the new problem:

$$\frac{\partial W}{\partial r} |_{r=r_{0+0}} = \frac{k_0}{a_0^2 \cdot r_0^2 \cdot k_1} (\frac{\partial W}{\partial t} - f_0). \tag{17}$$

Transformed parabolic-hyperbolic problem:

$$\begin{aligned} \tau_r \tfrac{\partial^2 W}{\partial t^2} + \tfrac{\partial W}{\partial t} = a_1^2 \tfrac{1}{r^2} \tfrac{\partial}{\partial r} (r^2 \tfrac{\partial W}{\partial r}) + F_1'(r,t), &\qquad r \in (r_0, R) \\ W |_{t=0} = N_0, & \\ \tfrac{\partial W}{\partial t} |_{t=0} = M_1, & \\ \tfrac{\partial W}{\partial r} |_{r=r_{0+0}} = \tfrac{k_0}{a_0^2 \cdot r_0^2 \cdot k_1} (\tfrac{\partial W}{\partial t} - f_0), & \\ (k_1 r^2 \tfrac{\partial W}{\partial r} + h_1' W) |_{r=R} = \varphi_1'(t). & \end{aligned} \tag{18}$$

We notice that

$$a_1^2 \frac{1}{r^2} \frac{\partial}{\partial r}(r^2 \frac{\partial W}{\partial r}) = a_1^2 \frac{1}{r} \frac{\partial^2 (r \cdot W)}{\partial r^2} \tag{19}$$

and, using standard transformation

$$U(r,t) = r \cdot W(r,t), \tag{20}$$

for governing equations, boundary and initial conditions and conjugation conditions, we can put our 1D problem for a ball to a problem for a slab [9], so we indicate $x = r$ and introduce

$$h_1 = \frac{h_1'}{R^2} - \frac{k_1}{R}, \quad \varphi_1(t) = \frac{\varphi_1'(t)}{R}, \quad F = x \cdot F_1', \quad K = \frac{k_0}{a_0^2 \cdot k_1 \cdot r_0^2}.$$

Parabolic-hyperbolic transformed problem for slab:

$$\begin{array}{c} \tau_r \frac{\partial^2 U}{\partial t^2} + \frac{\partial U}{\partial t} = a_1^2 \frac{\partial^2 U}{\partial x^2} + F, \qquad x \in (r_0, R) \\ \frac{\partial U}{\partial x} \mid_{x=r_0} = K(\frac{\partial U}{\partial t} - f_0 \cdot r_0) + \frac{1}{r_0} U, \\ (k_1 \frac{\partial U}{\partial x} + h_1 U) \mid_{x=R} = \varphi_1(t), \\ U \mid_{t=0} = x \cdot N_1. \\ \frac{\partial U}{\partial t} \mid_{t=0} = x \cdot M_1. \end{array} \tag{21}$$

Likewise we obtain parabolic-parabolic transformed problem for slab:

$$\begin{array}{c} \frac{\partial U}{\partial t} = a_1^2 \frac{\partial^2 U}{\partial x^2} + F, \qquad x \in (r_0, R) \\ \frac{\partial U}{\partial x} \mid_{x=r_0} = K(\frac{\partial U}{\partial t} - f_0 \cdot r_0) + \frac{1}{r_0} U, \\ (k_1 \frac{\partial U}{\partial x} + h_1 U) \mid_{x=R} = \varphi_1(t), \\ U \mid_{t=0} = x \cdot N_1, \end{array} \tag{22}$$

and hyperbolic-hyperbolic transformed problem for slab:

$$\begin{array}{c} \tau_r \frac{\partial^2 U}{\partial t^2} + \frac{\partial U}{\partial t} = a_1^2 \frac{\partial^2 U}{\partial x^2} + F, \qquad x \in (r_0, R) \\ \frac{\partial U}{\partial x} \mid_{x=r_0} = K(\tau_0 \frac{\partial^2 U}{\partial t^2} + \frac{\partial U}{\partial t} - f_0 \cdot r_0) + \frac{1}{r_0} U, \\ (k_1 \frac{\partial U}{\partial x} + h_1 U) \mid_{x=R} = \varphi_1(t), \\ U \mid_{t=0} = x \cdot N_1, \\ \frac{\partial U}{\partial t} \mid_{t=0} = x \cdot M_1. \end{array} \tag{23}$$

Initial condition—temperature at the beginning is known, but it is experimentally impossible to determine the initial heat flux. As an additional condition we can use temperature distribution at the end of process ($t = T$)

$$T_1 \mid_{t=T} = N_T, \tag{24}$$

to determine the heat flux theoretically—we have to solve inverse problem.

4 Solution and Numerical Results

We use CAM to obtain approximate solutions for our transformed models. It leads to ordinary differential equation problems which are relatively easier to solve. We denote $R_0 = R - r_0$. New integral average values:

$$u(t) = \tfrac{1}{R_0}\int\limits_{r_0}^{R} U(x,t)dx, \quad f_1(t) = \tfrac{1}{R_0}\int\limits_{r_0}^{R} F_1(x,t)dx. \tag{25}$$

ODE for parabolic-parabolic problem:

$$\begin{gathered} \tfrac{\partial u}{\partial t}(1 + K \cdot \tfrac{a_1^2}{R_0}) + u(\tfrac{h_1 a_1^2}{k_1 R_0} + \tfrac{a_1^2}{R_0 r_0}) = f_1 + \tfrac{a_1^2}{R_0}\tfrac{\varphi_1}{k_1} + \tfrac{a_1^2}{R_0} r_0 f_0 K, \\ u(0) = \tfrac{N_1}{2}(R + r_0). \end{gathered} \tag{26}$$

ODE for parabolic-hyperbolic problem:

$$\begin{gathered} \tau\tfrac{\partial^2 u}{\partial t^2} + \tfrac{\partial u}{\partial t}(1 + K \cdot \tfrac{a_1^2}{R_0}) + u(\tfrac{h_1 a_1^2}{k_1 R_0} + \tfrac{a_1^2}{R_0 r_0}) = f_1 + \tfrac{a_1^2}{R_0}\tfrac{\varphi_1}{k_1} + \tfrac{a_1^2}{R_0} r_0 f_0 K, \\ u(0) = \tfrac{N_1}{2}(R + r_0), \\ \tfrac{\partial u}{\partial t} \mid_{t=0} = \tfrac{M_1}{2}(R + r_0). \end{gathered} \tag{27}$$

As mentioned before, it is possible to use exponential or higher order polynomial approximations, but calculations show that differences in outcomes are almost negligible.

Parabolic-hyperbolic and hyperbolic-hyperbolic models are split in two sub-problems. For the inverse problem we use (24). ODE for inverse parabolic-hyperbolic problem is in form

$$\begin{gathered} \tau\tfrac{\partial^2 u}{\partial t^2} + \tfrac{\partial u}{\partial t}(1 + K \cdot \tfrac{a_1^2}{R_0}) + u(\tfrac{h_1 a_1^2}{k_1 R_0} + \tfrac{a_1^2}{R_0 r_0}) = f_1 + \tfrac{a_1^2}{R_0}\tfrac{\varphi_1}{k_1} + \tfrac{a_1^2}{R_0} r_0 f_0 K, \\ u(0) = \tfrac{N_1}{2}(R + r_0), \\ u(T) = \tfrac{N_T}{2}(R + r_0). \end{gathered} \tag{28}$$

Sub-problem with non-homogeneous conditions has initial heat flux as one condition, so after we derive solution for ODE, we express it with respect to $\frac{\partial w}{\partial t} \mid_{t=0}$ and compute its value at according time $t = T$. Solution is sensitive to changes in initial heat flux, so more research is needed to obtain precise results. We assume that temperature at the beginning of the process is 800 °C.

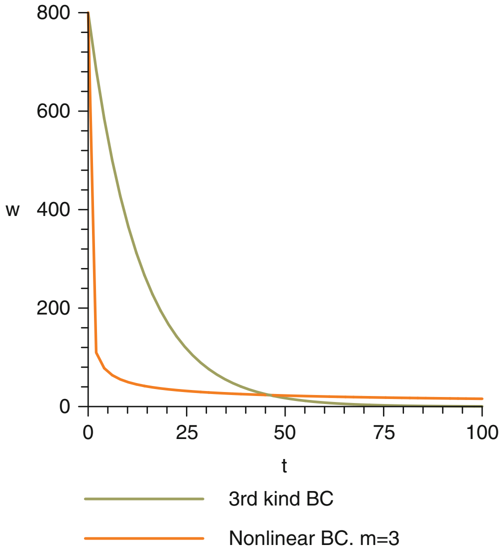


Fig. 3 Parabolic-parabolic problem. Approximate solutions with Conservative Averaging method

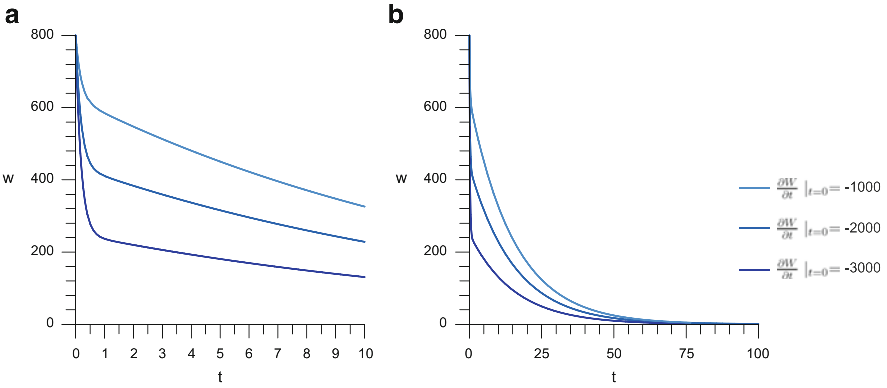


Fig. 4 Parabolic-hyperbolic problem. Approximate solutions with Conservative Averaging method. Solutions' dependence on initial feat flux. (**a**) t=10. (**b**) t=100

Modelling for parabolic-parabolic includes BC of third kind and nonlinear BC case $m = 3$ (Fig. 3). Parabolic-hyperbolic model is the one we are interested in, so results for third kind BC model also show solution's dependence on initial heat flux value (accordingly Figs. 4 and 5). It is clear that hyperbolic part is important at the very beginning of quenching process. When we compare parabolic and hyperbolic models for realistic description of physical process, one can easily see that hyperbolic model corresponds with experimental evidence better and without using nonlinear BC. Since nonlinear BC case was proposed by developers of Intensive Quenching method, modelling was done, however more detailed investigation on

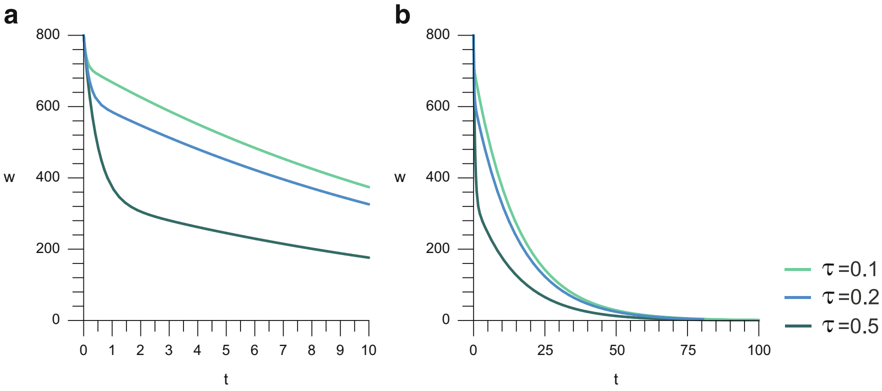


Fig. 5 Parabolic-hyperbolic problem. Approximate solutions with Conservative Averaging method. Solutions' dependence on relaxation time τ. (**a**) t=10. (**b**) t=100

this case is expected. Results of parabolic-hyperbolic model can better describe temperature values on different places (chosen radius) in the sample. These results are in accordance with previously done simulations.

5 Conclusion

We have developed heat conduction models for double layered spherical sample. Conservative Averaging method can be successfully used for reducing problems difficulty and obtaining approximate solutions. Results are in accordance to experimental outcomes. We propose Parabolic-hyperbolic model with BC of third kind as the most realistic one. Numerical experiments show that there are little differences in parabolic/hyperbolic models solutions if we look at longer period of process, but differences at the beginning of process are important because for industrial purposes critical heat fluxes and temperature drops determine quality of parts.

It is extremely important to determine relaxation time and initial heat flux values accurately since solutions are sensitive to small changes in these values.

Acknowledgements This work was supported by ESF project Nr.2009/0223/1DP/1.1.1.2.0/09/-APIA/VIAA/008.

References

1. Buike, M., Buikis, A.: Approximate solutions of heat conduction problems in multi-dimensional cylinder type domain by conservative averaging method, part 1. In: Proceedings of 5th International Conference on Heat Transfer, Thermal Engineering and Environment, Greece, pp. 15–21 (2007)

2. Buike, M., Buikis, A.: Several intensive steel quenching models for rectangular samples. In: Latest Trends on Theoretical and Applied Mechanics, Fluid Mechanics and Heat & Mass Transfer, Greece, pp. 88–94 (2010)
3. Buikis, A.: Conservative averaging as an approximate method for solution of some direct and inverse heat transfer problems. In: Advanced Computational Methods in Heat Transfer, pp. 311–320 (2006)
4. Guseynov, S.E., Kobasko, N.I.: Hyperbolic heat transfer equations for the mathematical modelling of the quenching processes. In: Proceedings of the 8th International Scientific and Technical Congress on the Equipment and Technologies for Metal and Alloy Heat-Treatment in the Machinery (OTTOM-8), pp. 213–219 (2007)
5. Kobasko, N.I.: Intensive Steel Quenching Methods, 2nd edn. CRC Press/Taylor & Francis Group, London (2010)
6. Kobasko, N.I., Krukovskyi, P., Yurchenko, D.: Initial and critical heat flux densities evaluated on the basis of cfd modelling and experiments during intensive quenching. In: Proceedings of the 5th IASME/WSEAS International Conference on Heat Transfer, Thermal Engineering and Environment, pp. 295–298 (2007)
7. Kobasko, N.I., Moskalenko, A.A., Deyneko, L.N., Dobryvechir, V.V.: Electrical and noise control systems for analysing film and transient nucleate boiling processes. In: Proceedings of the 7th IASME/WSEAS International Conference on Heat Transfer, Thermal Engineering and Environment, pp. 101–105 (2009)
8. Moskalenko, A.A., Kobasko, N.I., Protsenko, L.M., Rasumtseva, O.V.: Acoustical system analyzes the cooling characteristics of water and water salt solutions. In: Proceedings of the 7th IASME/WSEAS International Conference on Heat Transfer, Thermal Engineering and Environment, pp. 117–122 (2009)
9. Ozisik, M.N.: Heat Conduction. Wiley, New York (1993)
10. Wang, L., Zhou, X., Wei, X.: Heat Conduction. Springer, Berlin (2008)

Part IX
Education

Overview

Since its early years, the educational programme, started in 1987, was one of the backbones of ECMI's activities. This section covers four contributions regarding ECMI's educational profile.

In a first paper, Matti Heiliö and Allesandra Micheletti, former and current chair of ECMI's educational committee, look back on 25 years of education within ECMI. Their contribution *The ECMI Educational Programme in Mathematics for Industry: A Long Term Success Story* does not only give a brief overview on history and motivation behind ECMI's educational programme, but explains structure and time schedule of its two branches: "technomathematics and economathematics"—information, helpful not only for the readers interested in joining the programme.

The following three papers discuss developments inspired by ECMI's educational programme in France, Spain and Bulgaria. Edwige Godlewski discusses *Recent Evolution Enhancing the Interface between Mathematics and Industry in French Higher Education*, initiated by the creation of AMIES, the Agency for the interaction of Mathematics with the Industry and the Society. Francisco Pena's contribution *Two Examples of Collaboration between Industry and University in Spain* shows that industry and university can cooperate successfully in the field of mathematics in industry, if based on strong educational structures in mathematical engineering. The last contribution *ECMI Master Programmes at the Faculty of Mathematics and Informatics, Sofia University* by Stefka Dimova demonstrates the impact of ECMI's educational programme on Eastern Europe: guided by ECMI's educational programme, both technomathematics and economathematics have been successfully implemented in Sofia.

Michael Günther

The ECMI Educational Programme in Mathematics for Industry: A Long Term Success Story

Matti Heilio and Alessandra Micheletti

Abstract Here a description of the history and the main characteristics of the ECMI Educational Programme in Mathematics for Industry is provided. The Programme started in 1987 and evolved in time, according to the increasing new requirments coming both from the industrial and academic world. It is now running since 25 years and the success and brilliant career, both in Industry and Academy, of many students who followed the Programme in these years are the best recognition of the long term success of this educational activity.

1 History

During the academic year 1986–1987, representatives of universities belonging to the European Consortium for Mathematics in Industry (ECMI) (see Fig. 1) designed a 2 year postgraduate programme and reported the results in accounts dated 20-2-1987 and 20-3-1987. As intended, an educational programme which included exchange of students, exchange of teachers, central international courses and cooperation with industry became operational. This original ECMI Educational program was planned at the time when the Bologna model was not yet established. This was almost like a "pathfinder project" in the line of the emerging Bologna Model. The program was initially called *ECMI postgraduate programme in Math for Industry*. In many countries this was initially understood as a 2 year extension after the first degree—which often was a Master's degree, Engineer with diploma, etc.

M. Heilio
Department of Mathematics and Physics, Lappeenranta University of Technology, P.O. Box 20, 53851 Lappeenranta, Finland
e-mail: matti.heilio@lut.fi

A. Micheletti (✉)
Department of Mathematics, Università degli Studi di Milano, via Saldini 50, 20133 Milano, Italy
e-mail: alessandra.micheletti@unimi.it

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_59,

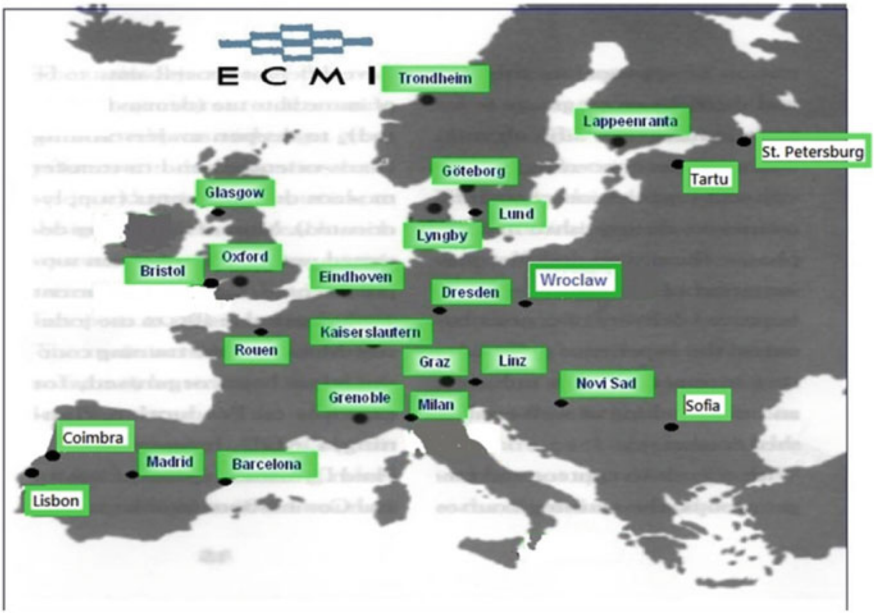


Fig. 1 The established and provisional ECMI Educational Centers in 2014

This variability of study structures in European universities was also in many cases a hindrance to its implementation. The initial structure of ECMI postgraduate program was defined in a rather detailed rigorous manner. A list of recommended textbooks and model syllabi were published to emphasize the spirit and to set a standard of ECMI education. A thorough description of the Introductory Phase (required prerequisites for admittance) was published, to harmonize the entrance qualifications.

Such quality management was part of the initial idealism and determination of the founders. The vision of the time was to provide a standardized European brand. The real scale of variation in European academic life turned out to be a challenge. These matters were primary reasons why the adoption of ECMI model was slow and many member universities were not able to fit into the given frame. That was the reason for the need of various revisions, that we describe in the following.

After the first few years of implementation of the ECMI-educational system the single experiences were discussed in detail by its partners and they led to an agreement for small changes in the philosophy and execution of the original Programme. The resulting description, dated 17-8-1990, has been the guideline for the Programme for a period of about 5 years in which the educational system of ECMI was consolidated and gradually extended.

The programme Mathematics for Industry initially placed emphasis on ODE's, PDE's and numerics and consequently on industrial problems that can be attacked by these mathematical techniques. When it became apparent that staff and students from the fields of Operation Research, Statistics and related areas would like to join the Programme (with an emphasis on these parts of mathematics, and consequently on other types of industrial problems) the Programme was again re-considered. The decision was taken, to reconstruct the programme so as to consist of two branches, closely linked together, of which the existing one was called "Technomathematics" and the new one "Economathematics". Contents of and interaction between the two branches as well as the way to arrange the execution of the international aspects have been discussed in several meetings of the Educational Committee; the final result has been approved in the meeting of the Council of ECMI on July 8, 1995.

The Programme with the two branches was then running for about 10 years, during which the number of ECMI Educational Centers increased, and the programme was also exported in other countries in Europe, also outside the EU, for example to Serbia, University of Novi Sad,via an EU funded Tempus Project, or, more recently, to Bulgaria (Sofia University) and Russia (St. Petersburg State Polytechnical University).

Starting from 2005, with the gradual revision of the educational programmes of European universities, according to the $3 + 2$ Bologna Scheme, a need for a deeper revision of the ECMI Educational Programme emerged, giving rise to new projects, funded by the EU (in particular the Erasmus Mundus ESIM, and the Erasmus Curriculum Development ECMIMIM) to modify the programme in order also to facilitate the establishment of double or joint degree master programmes between the ECMI educational centers. The established Bologna Scheme, providing a standardization of the structure of European graduate programmes, gave a more natural frame to the ECMI Educational Programme as an ECMI Master Program.

2 Motivation

Let us describe the main motivations which led and still push the ECMI Centers to establish, maintain, and update the Educational Programme in Math for Industry.

In modern industry, mathematical methods play an increasingly important role in research and development, production, distribution and management. These methods come not only from classical applied mathematics (mathematical physics, numerical mathematics, probability theory and statistics), but also involve e.g. operations research, control theory, signal processing and cryptography. Furthermore, mathematicians are more and more involved in the formulation, analysis and evaluation of mathematical models. For this development at least three reasons can be given:

1. Industry in Europe is increasingly engaged in knowledge-intensive activities. Research and development are important and a certain sophistication in production is needed to survive (flexible automation, optimization of products and production processes, quality control). Notice, that the word "industry" here and elsewhere in this description has to be interpreted in a broad sense, covering also

e.g. transport, finance, medical science, data-communication and any activity with an economical, technological or societal impact.
2. The possibilities for the use of mathematical models are now superior to and more extensive than those of some years ago. This is due to the rapid development of mathematical methods and to the increased capability of computers and their programming facilities.
3. Mathematics in industry has traditionally been exploited by engineers, chemists and physicists, with occasional support from a mathematician. Nowadays the need for more advanced mathematical methods, not familiar to those scientists, introduces an increased demand for industrial mathematicians.

It should be remarked, however, that it is rare for Mathematics to be used as an independent science for the benefit of an industrial company. The common situation is, that Mathematics is called in to assist with the solution of problems that arise from other fields. For this reason, a mathematician often has to be member of an interdisciplinary team. A consequence is that the training of an industrial mathematician should contain communication techniques, knowledge of other disciplines and experience in teamwork.

3 Structure of the Programme

Here we present the structure of the Educational Programme which was running up to the recent revision in the 3 + 2 scheme. The main ingredients of the Programme are still contained in the new versions which have been developed after the ESIM and ECMIMIM projects.

Originally the programme was studied to fit a 5 years cycle of graduate studies, being concentrated on the last 2 years of the cycle.

Each student had and still has to complete the following components:

- A mandatory common core of course work, designed to give the student a command of basic mathematical tools emphasizing constructive aspects, and with problem solving and modelling as the primary goals. The sections that constitute the common core must be regularly offered at all participating institutions.
- An individual selection of special topics which may vary from center to center, according to the different local expertise.
- Practical training in mathematical modelling, organized in a regular modelling seminar. In addition, ECMI organizes yearly a Modelling Week where students from the participating institutions meet and work in international teams on industrial problems in a simulated "Study Group with Industry" environment.
- A project thesis of at least half a year's work involving a real industrial problem, preferably carried out in an interdisciplinary environment, involving participants from Industry. Ideally, the thesis should demonstrate the candidate's ability to model the problem, to treat it with mathematical tools and to present the results in a way understandable and useful to the client. To be acceptable, the project thesis

must meet the standards of the profession with regard to each of these aspects. The thesis must be written in English and is reviewed by an expert appointed by the ECMI Educational Committee.
- A student exchange programme requiring each ECMI student to spend a period at another participating university or develop the final project abroad.

Students successfully completing the previous requests are awarded of a Certificate by the ECMI Board.

4 The Two Branches

Since “Technomathematics” and “Economathematics” are artificial names, there is some need to describe in more detail what is meant by them. This description will focus on the relation between the two branches and provide some examples of subjects in industrial practice. It is in no way meant to draw a line, distinguishing types of mathematics or even of mathematicians. The description is, on purpose, not a sharp one, since in the Programme it is an advantage rather than a problem that certain subjects can be reached from either branch.

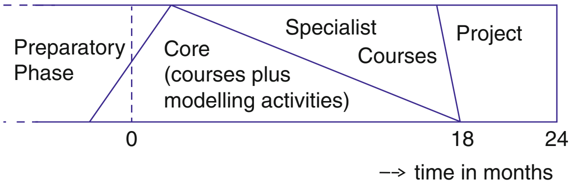
“Technomathematics” has to be considered as the part of the programme “Mathematics for Industry” in which real world physical, technological or biomedical problems are treated, in areas like e.g. heat exchange, fluid dynamics, electromagnetic fields, polymer science, population dynamics. “Economathematics” on the other hand deals with problems like e.g. planning and scheduling, quality control, distribution management, financial decision processes, data communication and data mining.

The general policy is that the two branches have to be closely linked together. In any case, students from the different branches in the Programme must be able to “talk to each other”. In order to reach this, the conditions for admission to the Programme have been made nearly the same for the two branches. The International Modelling Week is organized for both branches together.

5 Time Schedule of the Programme

Since the very beginning, each branch of the Programme consisted of courses and problem-solving activities from its Common Core, courses of a specialist nature, and a project, and was planned to extend over a 2 year period. The Preparatory Phase was the range of knowledge which a student entering the Programme should have. However, it was recognized that the backgrounds of different students may be very varied and that most students would not have covered all the Preparatory Phase topics before commencing the Programme. Thus in each individual case it was expected that some topics in the Preparatory Phase would be studied during the

Fig. 2 Time schedule



Programme, and conversely that exemption of some courses of the Programme can be given when they are proved to be known from the preparatory university study. Further, there was no need for a strict order in time between the core courses and the specialist courses. This induced a time profile as given in Fig. 2.

For further and more updated information on the ECMI Educational Programme please visit the web site of ECMI http://www.ecmi-indmath.org/

Recent Evolution Enhancing the Interface Between Mathematics and Industry in French Higher Education

Edwige Godlewski

Abstract The paper focuses on some recent initiatives in the French higher education system, in particular the creation of an Agency for the interactions of Mathematics with the Industry and the Society, AMIES, and its possible impact on already existing MSc programmes in industrial mathematics in the French university.

1 Recent Initiatives

In the last few years, several initiatives have been carried out in France with the aim of enforcing collaborations between mathematics and industry. The important actors are the French Ministry of Higher Education and Research which launched in 2011 a specific project entitled "Laboratory of Excellence" and INSMI, the National Institute for Mathematical Sciences in CNRS (the National Centre for Scientific Research). Some projects were initiated by SMAI, the French Society of Applied and Industrial Mathematics.

E. Godlewski (✉)
Sorbonne Universités, UPMC Univ Paris 06, UMR 7598, Laboratoire Jacques-Louis Lions, 75005 Paris, France

AMIES, Agence pour les Mathématiques en Interaction avec l'Entreprise et la Société, CNRS UMS 3458, Grenoble, France
e-mail: edwige.godlewski@upmc.fr; edwige.godlewski@agence-maths-entreprises.fr

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_60,

1.1 AMIES

AMIES, an acronym for Agency for the interactions of Mathematics with the Industry and the Society, is a French national distributed Laboratory of Excellence. An international panel of judges appointed by the National Research Agency (ANR) has awarded the proposal AMIES in spring 2011 and consequently the project has been provided with significant funding. It is sponsored by the CNRS in partnership with the University of Grenoble and INRIA (the National Institute for Research in computer science and automatics).

AMIES is based on a network of regional correspondents who have two roles: they promote links between the companies, laboratories and universities of their respective regions; and they monitor technological progress in their respective technical areas and help AMIES stakeholders to understand and benefit from it.

AMIES targets three main areas: Education, Research, and Interaction between mathematics and industry. The agency will chart industry-relevant research activities and training opportunities in universities and laboratories nationwide, highlighting successful industrial collaborations. It will also act as a contact point to the mathematical community and to research funding agencies and coordinate activities with similar programs abroad, particularly in Europe (for instance AMIES is already an ECMI member[1]).

Concerning more specially Education, AMIES aims at raising the awareness of students and instructors to opportunities in industry, notably via joint study weeks on modelling industrial problems. It has already organized study weeks (see the SEME section below); it intends to organize internships in industry and will try to coordinate the activities already taking place in some master programmes in French universities.

The agency also aids the integration of students in industry by supporting exploratory projects between academics and industry (see also Cemracs below). These projects may involve internships and lead to some industrial PhD grant. Note that industrial PhD contracts in France have been existing for 30 years through the CIFRE process, a national research funding agency (ANRT) providing part of the salary while expenses remaining at the charge of the company are eligible for a Research Tax Credit.

1.2 Some Realizations

SEME The creation of AMIES was preceded by that of a CNRS Research Group *Mathématiques & Entreprises*, which was inaugurated in 2010. The M&E group organized the first SEME (acronym for Study week for Mathematics with the

[1] More details in ECMI Newsletter 51, http://www.mafy.lut.fi/EcmiNL/issues.php?issueNumber=51.

Industry, following the idea of European Study Group in Industry ESGI) in April 2011 in Paris. Both structures now work hand in hand and a second SEME took place in Lyon in December 2011 and the third one is to be held in Toulouse in June 2012 (then Paris again, followed by Nancy, Grenoble, Limoges, Orléans). If they are inspired by ESGI, these weeks are dedicated to students undergoing a PhD; they work in teams on selected problems presented by representatives from industry. They may get some help from some academic instructors and be guided by the industrial representative but mostly they work by themselves, trying to innovate and bring new ideas. They report on the problem at the end of the meeting.

Job Forum for Mathematics. Following an idea of some SMAI members, the first French Job Forum for Mathematics (the French acronym is FEM) was held on January 26, 2012 in Paris. The aim was to present job opportunities both in Academia and in companies and services to master students, engineering students, graduate students and young doctors in mathematics. Nearly a thousand people participated, among which students and faculty from the whole country and industrial partners and it was a great success. A second FEM is to be held in January 2013 (and a third one in December 2013).

Cemracs. AMIES can help project with industrial partner at Cemracs (Summer mathematic center for advanced research in scientific computing). The Cemracs is a scientific event of the SMAI, the concept was initiated in 1996 by two French applied mathematicians, Y. Maday and F. Coquel. It consists in two types of events: a 1 week summer school mid July and a 5 week research session (end July–August), the research project and the organizing team change every year. During the first week, a classical summer school is proposed. The remaining 5 weeks are dedicated to working on the research projects, after a daily morning seminar. The Cemracs '12 will be devoted this year to Numerical Methods and Algorithms for High Performance Computing, it is organized by the French Research Group on scientific computing. The goal of this event is to bring together scientists from both the academic and industrial communities and discuss these topics. Each participant will work in a team on a project proposed by an industrial or an academic partner. Each team will be composed of young researchers assisted by one or more senior researchers.

1.3 Other Initiatives

There are other recent local initiatives, most of them linked to the above mentioned French Investments for the Future funding program; to cite a few of them: the Gaspard Monge Program for Optimization and operation research launched by EDF (French Energy Company) and the Jacques Hadamard Mathematical Foundation in Paris South; it aims at organizing the master curricula of this scientific field in the region Île de France; IRMIA, a Laboratory of Excellence funded in University of

Strasbourg, aiming at developing high performance computing and creating a local relay for AMIES, it has also some Education projects, a School of statistics and Master classes in Mathematics with nearby German Universities; MaiMoSiNE in Grenoble University (already an ECMI partner), involving in particular a scientific computing and modelling network; ICS at UPMC-Paris 6 University: this institute for Computation and Simulation will organize in July and August 2 2-week summer schools which in 2012 concerned biology and mathematics or computer science.

2 Evolution of Masters Programmes

2.1 Education Context: Universities and Grandes Écoles d'ingénieurs

The situation in France is specific because of the coexistence of two separate tracks for the training of students: Universities and Schools of Engineering (*Écoles d'ingénieurs*). While in Schools of Engineering, less and less mathematics are being taught (outside Mathematical Finance) since more time is given to management and economy, in French Universities the traditional high level of training in mathematics has been more or less preserved and is appreciated in industry, so that French students following this path still find good job opportunities thanks to their specific skills. Moreover, students from *Grandes Écoles d'ingénieurs* interested in mathematics often get a joint M-level diploma in mathematics in University.

Then many "Schools of Engineering" have been created and are growing inside Universities, thus, the training of engineers may be provided as part of a component of a University such as EPU belonging to the Polytech' network. These schools often share with University curricula part of the faculty teams so that both tracks get more and more interwoven. Even when the schools and university faculty do not mix for teaching, they are often linked in research teams, in particular because research is in general more active in the university laboratories than elsewhere.

Though Universities have been greatly supported in the last few years, the trend consists more in encouraging "excellence" than a clear choice between the two different tracks. In order to favor the creation of "poles of excellence" that are aimed at improving the ranking of French universities, the LRU reform encourages competition between public institutions of education and research. The French mathematical community tries to promote excellence while maintaining contact and some coordination.

2.2 Programmes in Industrial Mathematics

Some History. The first diploma DESS (*Diplôme d'Etudes Supérieures Spécialisées*) in applied mathematics were launched in French Universities in the late

1970s [2]: they correspond to the fifth year of higher education, the graduate program is chosen only in the fifth year, but coherent with the four previous years. The first diploma DESS in applied mathematics was created in *Université Paris 6*-UPMC. It appears that the second one was created in 1979 in *Université de Pau*[5], now *Université de Pau et des Pays de l'Adour* (UPPA), a smaller multidisciplinary University in the southwest of France, this University benefiting from a favourable industrial environment thanks to the presence of important industries. This is illustrative of the fact that mathematical sciences were present all over the country, and CNRS has encouraged this situation and INSMI continue to support high level teams in most regions.

The French organization of higher education, following the Bologna declaration on the European space for higher education in 1998 has undergone a change and passed to the so-called LMD frame (for *Licence—Master—Doctorat*). All courses are organized in compliance with the European Credit Transfer System (ECTS) of credits accumulation. Some French specificity is that the diploma has now a more complex structure than the previous DESS, involving several levels with *domaine, mention, specialité, parcours, filière* which does not help the foreign student in finding its way through the different programmes. However, a national directory tool has been developed aiming at collecting the information in order to promote the mathematics graduate programmes [1].

An Example: The UPMC Programme. The above mentioned *Dess de mathématiques appliquées* in UPMC has become in 2004 a Master programme *Ingénierie mathématique, Mathématiques Pour l'Entreprise* which exactly means Mathematics for Industry in the broad sense given to the word after the OECD report on Mathematics in Industry (2008).

This change of name reflects a greater awareness in the teaching team of the importance of identifying a core curriculum of mathematics for industry or, at least, the teams wanted to make the program more appealing and exciting to students and the professionals in industry [4]. If the requirement of spending one semester in another university is not fulfilled, the programme does fulfill all the other requirements listed in the Forward look model[3], including internships.

Other Programmes. One can find similar Master programmes in industrial mathematics in France: Grenoble, Pau, Orléans, Rennes, Toulouse, etc. besides some particular programmes in Schools of engineering: Insa in Rouen, Matmeca in Bordeaux, Ensimag in Grenoble, etc. and also an Erasmus Mundus MSc Course in MathMods—Mathematical Modelling in Engineering, whose French participant is UNSA (University of Nice-Sophia Antipolis).

2.3 Future Prospects

However, not all universities offer coherent courses with modelling activities, industrial projects and internship in a company, and there is some need to broaden and harmonize such programmes. A possible project is that AMIES could work at defining and delivering a label "Master in industrial mathematics" to French programmes fulfilling a list of requirements similar to those of the Forward look model. In a first time, so as to include the existing successful programmes, international might be encouraged but not yet compulsory. AMIES might also help in defining some industrial projects in the common interest, and organize internships in industry as already mentioned. We hope the numerous initiatives to promote applied mathematics at all levels and in industry will converge in particular to an increasing number of students enrolling.

Appendix

ANR	*Agence Nationale pour la Recherche*
ANRT	*Association Nationale pour la Recherche et la Technologie*
AMIES	*Agence pour les Mathématiques en Interaction avec les Entreprises et la Société*
CEMRACS	*Centre d'Eté Mathématique de Recherche Avancée en Calcul Scientifique*
CIFRE	*Conventions Industrielles de Formation par la REcherche*
CNRS	*Centre National de la Recherche Scientifique*
DESS	*Diplôme d'Etudes Supérieures Spécialisées*
EPU	*Écoles Polytechniques Universitaires*
FEM	*Forum Emploi maths*
GDR	*Groupement de Recherche*
ICS	*Institut du calcul et de la simulation*
IDEFI	*Initiatives d'excellence en formations innovantes*
INRIA	*Institut National de Recherche en Informatique et Automatique*
INSMI	*Institut National des Sciences Mathématiques et de leurs interactions*
IRMIA	*Institut de Recherche en Mathématiques, ses Interactions et Applications*
LMD	*Licence - Master - Doctorat*
LRU	*loi relative aux Libertés et Responsabilités des Universités*
MaiMoSiNE	*Maison de la Modélisation et de la Simulation, Nanosciences et Environnement*
SEME	*Semaine d'Etude Mathématiques et Entreprises*
SMAI	*Société de Mathématiques Appliquées et Industrielles*

References

1. Annuaire des masters de mathématiques. http://masters.emath.fr/main/emath_fr.html; http://www.agence-maths-entreprises.fr/; http://www.maths-entreprises.fr/; http://smai.emath.fr/forum-emploi/; http://smai.emath.fr/cemracs/cemracs12/; http://www.fondation-hadamard.fr/en/PGMOeng; http://www.maimosine.fr/; http://www.ics.upmc.fr/fr/institut/actualites/calsimlab.html; http://www.oecd.org/dataoecd/47/1/41019441.pdf
2. CNE: Les formations supérieures en mathématiques orientées vers les applications. Rapport du Comité National d'Evaluation, France (juillet 2002). ISSN: 0983-8740
3. European Science Foundation, Forward Look "Mathematics and Industry" (November 2010)
4. Godlewski, E.: Enseignement professionnel de mathématiques au niveau Master. In: Proceedings Conférence EMF 2012 (Genève), Espace mathématique Francophone (2012)
5. Godlewski, E., Madaune-Tort, M., Dossou-Gbete, S.: Two examples of a program 'Mathematics for Industry' at the master's level in a French University: Université Pierre et Marie Curie-Paris 6 and Université de Pau et des Pays de l'Adour. In: Proceedings of Educational Interfaces Between Mathematics and Industry (EIMI) 2010 Conference, CIM/Comap Proceedings, Lisbon/Bedford, pp. 211–226 (2010)

Two Examples of Collaboration Between Industry and University in Spain

Francisco Pena

Abstract Modelling of industrial processes is one of the ground lines of the research group Ingeniería Matemática (mat+i), from the University of Santiago de Compostela. Different activities have been developed in order to be in contact with the industry needs. Two examples of this close collaboration are presented here: the first one was proposed by the company FerroAtlántica to simulate the magnetic field and the temperature evolution of an electrode for electric-arc furnaces. The second one was proposed by company Gamelsa to simulate the energy efficiency of a newly designed solar collector. The difficulties arisen in the numerical simulation are summarized, as well as the benefits for both, the industry and the academic community.

1 Background

Collaboration between the research group mat+i and industry has been intense along time. Since the first contact in the 1980s with the energy company Endesa, there have been dozens of projects with companies and public administrations, in a wide range of fields: solid and fluid mechanics, heat transfer, electromagnetism, environmental modelling, finances, etc. Some of them are showed in [1]. We have exploited this experience on shared projects with industry to develop stable partnership formulas, together with two other research groups from universities of A Coruña and Vigo, and the CESGA node of the i-Math project:

- **Forums for Mathematics-Industry Interaction:** Last year the eighth edition of this forum was held in A Coruña. These 1-day meetings serve to present several

F. Pena (✉)
Departamento de Matemática Aplicada, Facultade de Matemáticas, Campus Vida, Santiago de Compostela, Spain
e-mail: fran.pena@usc.es

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_61,

industrial problems where their solution involves some numerical simulation techniques.

- **Master in Mathematical Engineering:** Its first edition was in 2006 and the subject *Industrial problems workshop* has been implemented since then. Every course about one dozen companies propose problems related to their industrial needs. Students must attend most presentations and choose one problem to develop its solution. This work will be their Master dissertation.
- **Mathematical Consultation Sessions:** These 3-day meetings have been financed by the i-Math project throughout Spain in the last years; four of them were held in Santiago de Compostela and 13 more in the rest of Spain. A reduced number of open problems are proposed by companies related to numerical modelling, statistics and operational research. Participants are organized in groups and they try to find a feasible solution for the problem, led by an invited expert.

These activities have been funded in recent years thanks to several projects, the most important of them being the aforementioned i-Math project, ended this year. The research groups involved in the previous enterprises have joined forces with other groups and entities to promote two new initiatives:

- **Technological Institute for Industrial Mathematics (ITMATI):** This institute, supported by the Galician autonomous government and the Galician universities, will try to continue with the i-Math objectives [2].
- **The Spanish Network for Mathematics and Industry (math-in.net):** It is a private association composed by more than 30 research groups in applied mathematics and statistics to improve collaboration between university and industry [3].

2 Results and Discussion

We present here two examples of collaboration between the research group mat+i and the industry.

2.1 Numerical Simulation of Metallurgical Processes in Silicon Production

The collaboration started in 1996, when company FerroAtlántica was interested in modelling a new compound electrode, named ELSA, patented in those years by them.

The research activity was financed under annual contracts. The company invested more than EUR 100,000 to simulate the behaviour of the electrode. The study was carried out in several phases: (a) the thermo-electric and thermo-mechanical study of a single electrode using axisymmetric models; (b) the thermo-electric study of a horizontal cut of the pot using bi-dimensional models and (c) the electromagnetic study of the whole pot, using tri-dimensional models.

For the thermo-electric study of the electrode, a harmonic eddy-currents model for the electric and magnetic fields was considered. This model is obtained from Maxwell equations by assuming alternate low frequency current. The resulting set of equations for the complex electric and magnetic fields, **E** and **H**, is:

$$\operatorname{curl}\mathbf{H} = \mathbf{J}, \qquad \operatorname{curl}\mathbf{E} = -\mathrm{i}\omega\mathbf{B}, \qquad \operatorname{div}\mathbf{B} = 0,$$
$$\mathbf{B} = \mu\mathbf{H}, \qquad \mathbf{J} = \sigma\mathbf{E}.$$

For the thermal model, an enthalpy formulation was considered:

$$\frac{\partial e}{\partial t} + \mathbf{v}\cdot\operatorname{grad} e - \operatorname{div}\left(k\operatorname{grad} T\right) = \frac{|\operatorname{curl}\mathbf{H}|^2}{2\sigma},$$

where enthalpy e is expressed through a multivalued operator depending of temperature T, due to the phase change (see [4]).

For the tri-dimensional case of the eddy-currents model, there are a wide variety of formulations depending on the chosen unknowns (see [5]). The approach considered in [6] tries to use the most usual boundary conditions in the industrial applications; besides, it permits to consider general geometries without complicating the mesh.

The project provided a way to understood how the density current distributes throughout the electrode to produce the electric arc, an aspect that was not completely clear when the electrode was planned. It also served to test the performance of the electrode under different operational conditions, allowing to make recommendations about how to operate it. Besides, the numerical results eased to explain the benefits of the new electrodes to possible clients. To our surprise, mathematical modelling was useful, not only to interpret the physical phenomena and to improve its operational performance, but also to sell the electrode to other companies.

2.2 Numerical Simulation of a Solar Collector

The second work was proposed by company Gamelsa for the Master subject *Industrial problems workshop*. They wanted to model a novel design for a collector with high surface contact to the absorbent surfaces. This problem was chosen by the student Ana Álvarez, under the supervision of M.C. Muñiz. Part of the work described here was completed in her Ph.D. Thesis, presented in 2011.

The work consisted in the numerical calculation of the thermal parameters of a low-temperature solar collector, in order to estimate its energy efficiency. To get a solution with reduced computational cost, the idea was to couple the two-dimensional heat equation at the cross section of the collector with the one-dimensional convective heat equation modelling the behaviour of the fluid temperature.

For the boundary conditions, some additional terms must be added to consider the wind exposition, the glazing and the associated greenhouse effect. In the steady-state, non-local boundary conditions were obtained on the tube-to-fluid boundary, written in terms of the tube-to-fluid boundary temperature and the inlet fluid temperature (see [7]).

Since the tube inside the collector is a serpentine, the previous approximation must be checked prior. The solution of the bi-dimensional problem for a tube of circular section was compared with some well-established analytical solutions, obtaining a relative error smaller than 0.2 %.

The corrugated topology introduced another complexity in the model. A cross sectional model with a single tube and a single valley in the corrugated surface was compared with a tri-dimensional thermo-hydrodynamical model for the first three sections of the collector, presenting a good agreement. A solar collector equipped with measuring devices was constructed. The results obtained for temperature and energy efficiency were compared with the model, obtaining a discrepancy below 7 %.

3 Conclusions

Collaboration between the research group mat+i and industry has been constant over time. These contacts have been developed through different formulas, from direct collaborations to periodic forums and the participation in Master activities.

Two examples were presented here. The first was a long-term project to model a metallurgical electrode. Numerical simulation was able to improve the understanding of the electrode's behaviour and helped to better operate it. The second one was the modelling of a prototype of a new design of solar collector. The adjustment of the simplified model was the main part of the work. At the same time, its simplicity of use allowed it to be included in the designing process.

Acknowledgements The author would like to thank Ana Álvarez, Alfredo Bermúdez, M. Carmen Muñiz, Peregrina Quintela and Pilar Salgado for allowing him to present their work in this presentation and for their help in the preparation of the talk.

References

1. Lery, T., Primicerio, M., Esteban, M.J., Fontes, M., Maday, Y., Mehrmann, V., Quadros, G., Schilders, W., Schuppert, A., Tewkesbury, H. (eds.): European Success Stories in Industrial Mathematics. Springer, Heidelberg (2012)
2. Technological Institute for Industrial Mathematics (itmati). http://itmati.com/ (2012)

3. The Spanish network for mathematics and industry (math-in.net). http://www.math-in.net/?q=en (2012)
4. Bermúdez, A., Bullón, J., Pena, F., Salgado, P.: A numerical method for transient simulation of metallurgical compound electrodes. Finite Elem. Anal. Des. **39**, 283–299 (2003)
5. Alonso, A., Valli, A.: Eddy Current Approximation of Maxwell Equations – Theory, Algorithms and Applications. Series: MS&A, vol. 4. Springer, London (2010)
6. Bermúdez, A., Rodríguez, R., Salgado, P.: A finite element method with lagrange multipliers for low-frequency harmonic Maxwell equations. SIAM J. Numer. Anal. **40**, 1823–1849 (2002)
7. Álvarez, A., Cabeza, O., Muñiz, M.C., Varela, L.M.: Experimental and numerical investigation of a flat-plate solar collector. Energy **35**, 3707–3716 (2010)

ECMI Master Programmes at the Faculty of Mathematics and Informatics, Sofia University

Stefka Dimova

Abstract The Faculty of Mathematics and Informatics, Sofia University, has been an ECMI member since 2011. The ECMI Educational Committee approved Sofia University as a provisional ECMI Teaching Centre at a meeting held on July 29th, 2011 in Milan, Italy. Here we present two of the Master programmes of the Faculty of Mathematics and Informatics, which correspond to the two branches—Techno-Mathematics and Econo-Mathematics—of the ECMI Model Master in Industrial Mathematics (ECMIMIM). These two programmes are “Computational Mathematics and Mathematical Modelling” and “Mathematical Modelling in Economics”. We show that they satisfy all the requirements of the ECMI Model Master in Industrial Mathematics.

1 Introduction

The ECMI Educational Committee (EC) approved Sofia University (SU) as a provisional ECMI Teaching Centre at a meeting held on July 29th, 2011 in Milan, Italy. Two Master programmes (MPs) at the Faculty of Mathematics and Informatics (FMI) are in a process of evaluation for relevance with respect to the ECMI Model Master in Industrial Mathematics (ECMIMIM). The MP “Computational Mathematics and Mathematical Modelling” (CMMM) is evaluated in the Techno-Mathematics branch, and the MP “Mathematical Modelling in Economics” (MME) is evaluated in the Econo-Mathematics branch. An inspection visit is expected to be the final step towards a definitive status of FMI, SU as an ECMI Teaching Centre.

S. Dimova (✉)
Faculty of Mathematics and Informatics, Sofia University “St. Kliment Ohridski”, 5 James Bourchier Blvd., 1164 Sofia, Bulgaria
e-mail: dimova@fmi.uni-sofia.bg

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*, Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3_62,

The first specialization on Computational mathematics at FMI was created back in the 1959/1960 academic year, inside the Department of Higher Analysis. It comprised *Numerical methods, Linear programming, Computers and programming, Theory of information* as compulsory disciplines. Among the diploma works, given to the first students, graduating from this specialization, were "Modelling the harmonization of 8-bar melodies" and "Modelling of the belote game". So from its very beginning the computational mathematics was closely connected with the mathematical modelling, what is more, it has been considered as a tool for mathematical modelling. A three-stage profiled education—Block A (3.5–4 years, Bachelor), Block B (1.5–2 years, Master), Block C (3 years, PhD)—was established at FMI in the 1970/1971 academic year (let us note, 29 years before the Bologna process!!). Three of the branches of Block B were Computational mathematics, Mathematical modelling and Operations research. During the years the names have been changed slightly, but the "mathematical modelling" has remained in the heart of all of the applied branches of education. This tradition has been kept till now and has determined the names of our Master programmes.

2 The Bachelor Programme in Applied Mathematics at FMI

The duration of the Bachelor programme (BP) in Applied mathematics at FMI (and in other Bulgarian universities) is 4 years, with 240 ECTS credits (for information see http://www.fmi.uni-sofia.bg/). The requirements for admission to the ECMIMIM are "180 ECTS of undergraduate study at university level (Bachelor degree)". So the students, graduating with a Bachelor degree from Bulgarian universities, have additional 60 ECTS credits. The compulsory courses alone in the Bachelor programme give all the prerequisites from Block A of the ECMIMIM, all courses in Block B for Techno-Mathematics and almost all courses in Block B for Econo-Mathematics. The elective courses for 67.5 ECTS credits enable the students to choose subjects in the field of the desired Master programme. The students coming from other Universities for the CMMM are required to get up to 38 additional ECTS credits in order to reach the minimal basic level.

All stated above supports the opinion (accepted by the Review subcommittee of the ECMI Educational Committee) that regardless the obligatory state regulation for only 90 ECTS credits and 3-semester education, the FMI Master programmes fulfill the requirements of ECMIMIM. The 15 ECTS credits for a Master thesis, given in accordance with the Bulgarian state regulations, were the main difference between our Master programmes and ECMIMIM. Currently, additional 15 ECTS are being given for a Diploma project (the preparatory part of the Master thesis), thus the total number of ECTS credits for the Master thesis is 30.

Numerical analysis, Equations of mathematical physics, Numerical methods for differential equations, Mechanics of continua, Probability theory and mathematical *statistics, Applied statistics* are among the compulsory courses in our BP providing the required basis for further education at the MPs. In addition, there are three more specific courses, which help the students to make their further choice of MP.

The aim of the course *Mathematical modelling* is not only to present the general ideas and schemes of mathematical modelling but also to illustrate them with various examples by the Classical mechanics, Biology and Medicine, Physics, Chemistry, etc., and thus—to demonstrate how some phenomena and processes apparently different by nature turn out to be similar from the point of view of the mathematical model.

The course *Mathematical introduction to economics* gives first knowledge on the theory of the firms, profit maximization, cost minimization, consumer theory, Paretto optimality, competitive behavior and monopoles.

Macroeconomics 1 is an introduction to fundamental concepts and models in macroeconomics. The course aims to introduce the basic concepts and methods of analysis in macroeconomics. IS-LM analysis constitutes an important part of the course, dynamic models are also studied. A diverse set of mathematical techniques is employed to study economic phenomena.

3 ECMI Master Programmes at FMI

3.1 MP Computational Mathematics and Mathematical Modelling

The educational goal of this MP is to provide the students with solid theoretical knowledge and practical skills in one of the following areas:

- development and analysis of mathematical models of processes in Physics, Chemistry, Biology, Ecology and Engineering;
- development and studying of effective numerical methods and algorithms for solving the mathematical problems, obtained through the modelling;
- identifying and using the most relevant of the available software for scientific computations.

The professional goal of CMMM Master programme is to prepare the students to work in interdisciplinary research teams, to create mathematical models of real processes in at least one domain of science and engineering and to solve them by using contemporary numerical methods and high-performance computing.

Four groups of courses make these goals achievable:

- **Mathematical Modelling**: *Mathematical models and computational experiment, Mathematical modelling in Physics, Mathematical modelling in Biology, Hydrodynamics, Mechanics of continua, Non-linear mathematical models*;

- **Contemporary Numerical Methods**: *Numerical methods for differential equations, Finite elements method—algorithmic foundations, Numerical methods for system with sparse matrices, Parallel algorithms, Numerical integration*;
- **Theory and Analysis of the Numerical Methods and the Continuous Models**: *Theory of the finite difference schemes, Theory of the finite element methods, Applied functional analysis, Sobolev spaces and applications in PDE, Chaotic dynamical systems*;
- **Other Tools**: *Spline-functions and applications, Wavelets and applications, Fractals, Fourier transform, Wavelets and signal processing, Computer graphics, Software for scientific computations*.

The appropriate choice of courses from each group makes the CMMM master students capable to implement the full cycle of the **Computational Experiment as a tool for investigating a real-life problem: physical model, mathematical model, analytical investigation (as far as possible), discrete model, algorithm, computer programme, numerical experiments, parametric investigation**.

3.2 MP Mathematical Modelling in Economics, with Two Specializations: Economics; Mathematical Finance and Actuarial Science

The aim of the MME MP is to develop the student's mathematical and computational skills to handle problems in business and finance providing them with:

- theoretical and practical knowledge applicable in Economics, Finance, Insurance, Company Management;
- ability to handle large amounts of data by numerical and statistical methods;
- skills in identifying and using the most relevant of the available software.

The MME MP goal is to stimulate the students to model, study and optimize particular events and processes in Economics. Profound knowledge in micro- and macroeconomics, financial tools and markets, risk evaluation, insurance, data processing, combined with computer skills, are acquired by the students within a number of courses, e.g. *Numerical methods and their applications in economics, Variational calculus with application to economics, Microeconomics, Macroeconomics 2, Open economy macroeconomics, Econometrics, Financial mathematics, Time series, Probability models, Mathematical risk theory, Stochastic analysis and applications, Multicriteria optimization, Optimal control, Nonlinear control systems, Credit risk, Life insurance, European practices in insurance*. Cooperation with insurance companies provides additional opportunities to analyze real models and data.

3.3 Something More for the Two Master Programmes

The CMMM and MME MPs are well provided with computer laboratories, libraries and software products. Some of the courses are taught in English, if there are foreign students following the programmes.

Mathematical modelling seminar, common for the two MPs and compulsory for the students, is going on for 3 years now. Several Bulgarian firms and institutions have been involved: R and D Bulgaria, ProSystLabs, SAP, Rila Solutions, Sirma Group, Institute of Information and Communication Technologies, Institute of Metal Science, BAS. Some of the Master students, working in firms, have presented the current problems they were dealing with.

Three students at CMMM programme and four students at MME made their theses in English. The topics were on Techno- and Econo-Mathematics, e.g. "Mathematical modelling of electrochemical processes in Li-ion batteries", "Adaptive algebraic multigrid for finite element elliptic equations with random coefficients", "SPEA 2 for Mean-VaR portfolio selection under real constraints", "Long horizon risk estimation using ARMA-GARCH processes". Three former CMMM Master students, graduated 2010/2011 and 2011/2012, are now PhD students at the Fraunhofer ITWM, Kaiserslautern.

Two students, following the MPs CMMM and MME, took part at ESSIM'2011 in Milan. It was with great satisfaction that we learned about their excellent research performance. We are sending now four Master students to ESSIM'2012 in Dresden.

Bilateral Erasmus agreements between FMI, SU and two ECMI universities—Kaiserslautern University of Technology, Germany and Johannes Kepler University of Linz, Austria—are signed for the academic years 2012/2014 (they are among the 33 Erasmus agreements between FMI, SU and European universities). Two CMMM Master students will study at Kaiserslautern during 2012/2013 academic year, a student from Linz is coming to FMI. We hope our connections with the ECMI Universities will grow.

In conclusion, the adheration of FMI to ECMI has to be considered as a promising attempt for development, encouragement and promotion of mathematical research in Bulgarian and European industry and economy.

Index

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3,

Authors

M. Fontes et al. (eds.), *Progress in Industrial Mathematics at ECMI 2012*,
Mathematics in Industry 19, DOI 10.1007/978-3-319-05365-3,