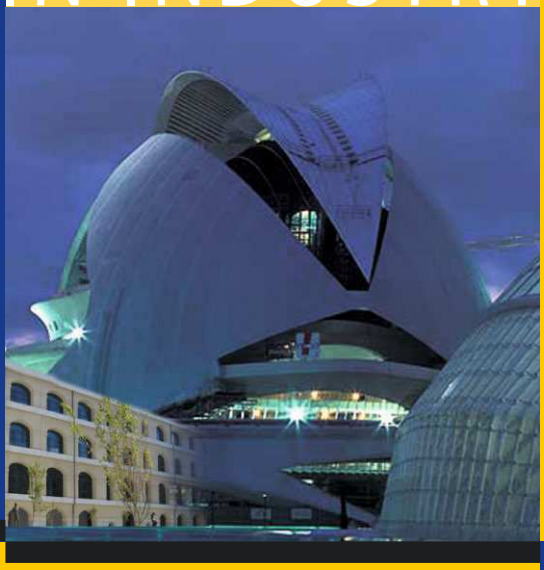
MATHEMATICS IN INDUSTRY 12

Luis L. Bonilla
Miguel Moscoso
Gloria Platero
Jose M. Vega
Editors



Progress in Industrial Mathematics at ECMI 2006

THE EUROPEAN CONSORTIUM FOR MATHEMATICS IN INDUSTRY

Springer

ECMI

MATHEMATICS IN INDUSTRY **12**

THE EUROPEAN CONSORTIUM
FOR MATHEMATICS IN INDUSTRY

SUBSERIES

Luis L. Bonilla
Miguel Moscoso
Gloria Platero
Jose M. Vega
Editors

Progress in Industrial Mathematics at ECMI 2006

With 361 Figures, 52 in color and 53 Tables

Springer

Editors

Luis L. Bonilla

Universidad Carlos III de Madrid
Avenida de la Universidad 30
28911 Leganes, Spain
Email: bonilla@ing.uc3m.es

Miguel Moscoso

Universidad Carlos III de Madrid
Avda. Universidad, 30
28911 Leganes (Madrid), Spain
Email: moscoso@math.uc3m.es

Gloria Platero

Instituto de Ciencia de Materiales
de Madrid, CSIC
28049 Madrid, Spain
Email: Gloria.platero@icmm.csic.es

José M. Vega

Universidad Politécnica
de Madrid
Plaza Cardenal Cisneros 3
28040 Madrid, Spain
Email: vega@fmetsia.upm.es

Library of Congress Control Number: 2007932402

Mathematics Subject Classification (2000): 00B20, 35-xx, 60-xx, 62-xx, 65-xx, 70-xx, 74-xx, 76-xx, 80-xx, 81-xx, 82-xx, 91-xx, 92-xx, 97-xx

ISBN 978-3-540-71991-5 Springer Berlin Heidelberg New York

Typeset by the editors and SPi using a Springer LaTeX macro-package

Cover design: *design & production* GmbH, Heidelberg
Printed on acid-free paper SPIN: 12049349 46/2244/YL - 5 4 3 2 1 0

To our beloved ones

Preface

Increasing computing power in the last decades has given mathematical modeling an ever greater impulse and made it a very important tool to solve problems coming from industry. The European Consortium for Mathematics in Industry (ECMI) was founded 20 years ago by mathematicians from ten European universities to foster the use of mathematics to help European industry and commerce to pose and solve their problems. The aims of ECMI are to (a) promote the use of mathematical models and mathematics in industry, (b) form applied mathematicians capable of working effectively in industry and (c) work for these goals at the European scale. Efficient problem solving often requires the use of results in different mathematical fields, yet no single applied mathematician may be able to cover the whole subject. By providing a European research network, ECMI can bring together experts from a wide geographical range.

Since 1986, ECMI has incorporated many more institutions and industries throughout Europe and it has been consolidated as a brand name for Industrial Mathematics. Twenty years later, the biannual ECMI conference was celebrated for the first time in Spain, at the Universidad Carlos III de Madrid. This is a young university created in 1989. Technological studies and departments are located at the Leganés campus where the conference was held. Moreover, University Carlos III participates in the Leganés Scientific and Technological Park, together with the Autonomous Region of Madrid and the city of Leganés. They contribute to place Madrid at the forefront of research and development in Spain.

The scientific program covered a wide variety of topics related to technological sectors (aerospace and automotive industry, materials and electronics, information and telecommunication technologies, energy and environment, biology, biotechnology, life sciences, imaging) and to finances and economics. The different origin of participants helped making the conference multidisciplinary. Active participation of industry was intended, with reasonable success. The present volume includes a part of the contributions to the conference, selected after a refereeing process. It is a pleasure to see that six

plenary speakers have submitted papers for this volume. Vincenzo Capasso in his "Alan Tayler" lecture, besides presenting his scientific work on statistical geometric measure applied to medicine and materials science, recalls some of the challenges for Mathematics in Industry listed in the first ECMI brochure produced by Alan Tayler and himself in 1994, relates them to the present situation of an enlarged Europe, and tells us how these challenges remain important and pressing for us today. Antonio Barrero (Seville), Alfredo Bermúdez (Santiago), Russel Caflisch (UCLA), Luis Campos (Lisbon) and Pierre Degond (Tolouse) illustrate with their contributions the breadth of applications and variety of techniques that are embraced by ECMI. ECMI's commitment to educating students in Industrial Mathematics is reflected in the fact that many papers were given by students. The Wacker Prize, offered for a Master's Level thesis on an industrial problem was awarded to Filippo Terragni, in line with the tradition of excellent work by previous winners. Many of the minisymposia and special sessions included the activities of ECMI Special Interest Groups. Of the 35 minisymposia organized for the conference, many are gathered in this book, usually preceded by a short explanation about their contents. A number of contributed papers complete the volume. I hope that these proceedings will contribute both to show interesting and relevant mathematical problems and methods, and to strengthen cooperation between academia and industry, the absence of which is a major weakness of the European Science-Technology system.

As President of ECMI and on behalf of the ECMI Council, I wish to thank all those who have contributed to the success of the Conference. Among them the participants, the speakers, the International Scientific Committee and the National and Local Organizing Committees. Organizing this meeting has been possible thanks to the efforts of many people both at the Spanish national and local level to whom we are very grateful. In particular all the members of the Modeling, Simulation and Industrial Mathematics Group at Universidad Carlos III worked hard to run a smooth and successful conference which would not have been possible without their help. The dedication of our university congress bureau, Congrega, was also essential for the conference success. Ms. Bárbara Tapiador's help was very important to process the manuscripts that are gathered in the present book. I am grateful to my co-editors, Gloria Platero, Miguel Moscoso and José Manuel Vega for their invaluable help.

Lastly, the support of our sponsors is gratefully acknowledged: Ministerio de Educación y Ciencia (grant MTM-2005-24569-E), Comunidad de Madrid (grant S-0505/ENE/0229), Universidad Carlos III de Madrid, Universidad Politécnica de Madrid, Consejo Superior de Investigaciones Científicas (CSIC), Instituto de Tecnológico de Química y Materiales "Álvaro Alonso Barba", Ayuntamiento de Leganés and Springer.

Madrid, May 2007 *Luis L. Bonilla, President of ECMI*

Contents

Part II Minisymposia

Part I

Plenary Lectures

On the Mean Geometric Densities of Random Closed Sets, and Their Estimation: Application to the Estimation of the Mean Density of Inhomogeneous Fibre Processes

Vincenzo Capasso and Alessandra Micheletti

Department of Mathematics, Università degli Studi di Milano, Via C. Saldini, 50, 20133 Milano, Italia
Vincenzo.Capasso@unimi.it, Alessandra.Micheletti@unimi.it

Dedicated to Alan Tayler

Preface [VC]

It has been a great honour for me to deliver the "Alan Tayler Lecture" in this ECMI Conference, to honour one of the leading founders and Presidents of ECMI. I have collaborated with Alan for many years, especially during my term as Chairman of the Educational Committee, and later during the first ECMI-HCM Project. While he was already very ill, he found the way to participate (even though only for a couple of days) in a workshop in Milan, opening ECMI to the Italian academic and industrial community, and highly supported the birth of MIRIAM (the Milan Research Centre for Industrial and Applied Mathematics).

I had a rewarding experience around the early 1990s producing, in a strict collaboration with Alan, the first ECMI Brochure [CT94] (see the ECMI web site) in order to advertise the specific role of ECMI within academia and industry in Europe.

It was clear to me that he had a vision of how to establish in Europe a co-operative action by the most active groups in the applications of mathematics to real world problems; I wish to remind the key issues stated in the brochure, since I may claim that these are still update.

"Realising the need of interaction between universities and research groups in industry, the European Consortium for Mathematics in Industry (ECMI) was founded in 1986 by mathematicians from ten European universities.

...

Mathematics, as the language of the sciences, has always played an important role in technology, and now is applied also to a variety of problems in commerce and the environment.

European industry is increasingly becoming dependent on high technology and the need for mathematical expertise in both research and development can only grow.

...

These new demands on mathematics have stimulated academic interest in Industrial Mathematics and many mathematical groups world-wide are committed to interaction with industry as part of their research activities.

In 1986 ten of these groups in Europe founded ECMI with the intention of offering their collective knowledge and expertise to European Industry.

The experience of ECMI members is that similar technical problems are encountered by different companies in different countries. It is also true that the same mathematical expertise may often be used in differing industrial applications.

If European industry is to compete in world markets it should take advantage of the competitive edge which may be gained from using European mathematical expertise.

No single European country is likely to have sufficient expertise of mathematical knowledge whereas ECMI can provide a comprehensive coverage of mathematical skills and their diverse applications." [CT94]

We are now facing the challenge of a larger European Union.

Alan had anticipated this by promoting an ECMI "patronage", financially supported by the EU, of those countries usually called "Central Europe", such as Čekia, Hungary, Poland, Romania, Slovakia.

I am sure that he would have liked to participate in the process of complete integration of all the new entries in the ECMI system.

Going back to the ECMI Brochure, a major scope of ECMI was identified as follows.

"C. TO OPERATE ON A EUROPEAN SCALE

Academic resources in Mathematics for Industry are also scarce and distributed across Europe; industrial needs are widely spread. Exchange and interactive programmes are necessary in training, research and industrial collaboration if there is to be an effective transfer of knowledge and skills. The EC is encouraging ECMI to involve relevant groups in Eastern Europe as Associate members."

As part of this encouragement, the EC provided funds to ECMI for organising a series of workshops in those countries, in collaboration with recognised colleagues at the local level. Thus anticipating the enlargement of the political Europe.

In my opinion, having the EC approved a significant enlargement of Europe towards East, listing soon 27 member states, ECMI, as an enlarged Consortium, should find new ways to exploit the best of the scientific resources of the old and the new member states together, to actively participate in the building up of a common competitive Europe. As far as scientific competence is concerned, there are excellencies in all regions of Europe, some of them well

identifiable also in the new member states; a genuine will to sustain competence of Europe should go through ways to exploit all of them, with the usual ECMI cooperative attitude.

Another anticipation envisaged by Alan has been the shift of meaning of the key word "Industry" in the ECMI system.

"This collaboration may also be extended to developing mathematical models for the environment, earth sciences, biology and finance." [CT94]

We have already achieved the inclusion of what we call **Economathematics**, and today we are facing a further shift of attention towards **Medicine and Biotechnology**.

All over the world leading experts of **Mathematics for/in Industry**, are participating actively in the development of **Mathematics for/in Medicine**, thus undertaking the further challenge of contributing to the development of innovative methods for diagnosis and treatment of relevant diseases, from cancer to infectious diseases.

My own presentation here is aimed to showing an example of how mathematics, originally developed for mining industry or more in general for material science and chemical industry, is now moving to deal with problems of interest in medicine.

At first this research was motivated by polymer industry in Europe, and constitutes one of the most important success stories of collaborative research within ECMI, that was supported within the first HCM Project coordinated by Alan Tayler. As a documentation of the cooperation between different research teams in Europe within the ECMI Special Interest Group on "Polymers", the volume "Mathematical Modelling for Polymer Processing. Polymerization, Crystallization, Manufacturing", edited by myself, was published as Volume 2 in the **ECMI Series on Mathematics in Industry by Springer-Verlag, Heidelberg 2002**, showing an additional success story of ECMI: the start of the **Springer Series on Mathematics in Industry**.

1 Introduction

Many processes of biomedical or material science interest may be modelled as birth-and-growth processes (germ–grain models), which are composed of two processes, birth (nucleation, branching, etc.) and subsequent growth of spatial structures (cells, vessel networks, etc.), which, in general, are both stochastic in time and space. These structures induce a random division of the relevant spatial region, known as random tessellation (see Fig. 1). A quantitative description of the spatial structure of a tessellation can be given, in terms of the mean densities of interfaces (n-facets).

In applications to material science a main industrial interest is controlling the quality of the relevant final product in terms of its mechanical properties; as shown, e.g. in [FC98], these are strictly related to the final morphology

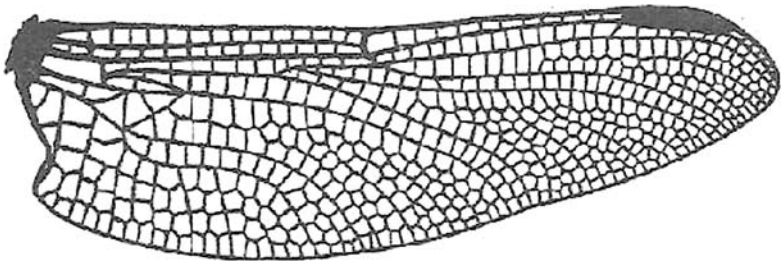


Fig. 1. The spatial tessellation generated by vessels in a dragonfly wing

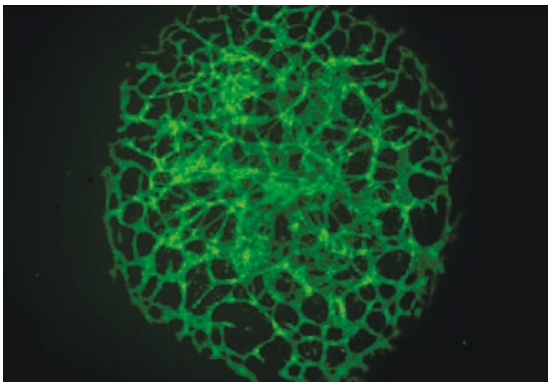


Fig. 2. Vascularization of an allantoid [Credit: Dejana et al. 2005]

of the solidified material, so that quality control in this case means optimal control of the final morphology.

In medicine, an important area of application of birth-and-growth processes and other models of stochastic geometry is tumour-induced angiogenesis. It can be modelled as a fibre process of Hausdorff dimension 1 in the relevant 2D or 3D space.

Tumour-induced angiogenesis is believed to occur when normal tissue vasculature is no longer able to support growth of an avascular tumour. At this stage the tumour cells, lacking nutrients and oxygen, become hypoxic. This is assumed to trigger cellular release of tumour angiogenic factors (TAFs) which start to diffuse into the surrounding tissue and approach endothelial cells (ECs) of nearby blood vessels. ECs subsequently respond to the TAF concentration gradients by forming sprouts, dividing, and migrating towards the tumour. A summary of these mechanisms can be found in the recent paper by Carmeliet [JK01] (see also Figs. 2–4 where examples of real or simulated vascular networks are depicted).

Initially, the sprouts arising from a parent vessel grow essentially parallel to each other. It is observed that once the finger-like capillary sprouts have

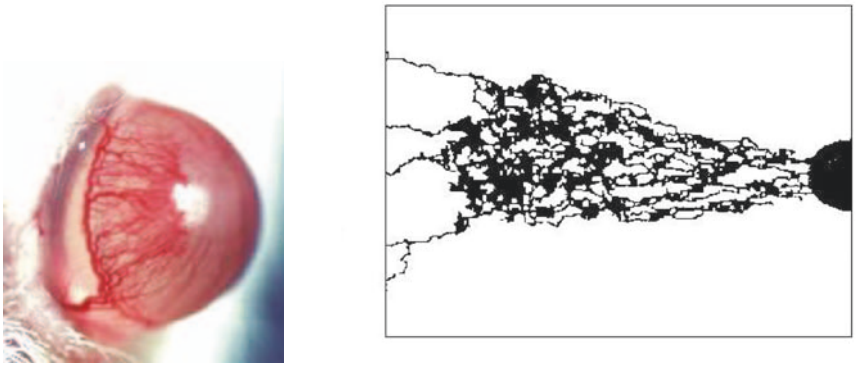


Fig. 3. *Left*: Angiogenesis on a rat cornea [Credit: Dejana et al. 2005]. The white spot is a pellet implanted in the cornea containing an angiogenetic substance, emulating the effect of a tumour. *Right*: A simulation of an angiogenesis due to a localized tumour mass (*black* region on the *right*) (from [CA99])

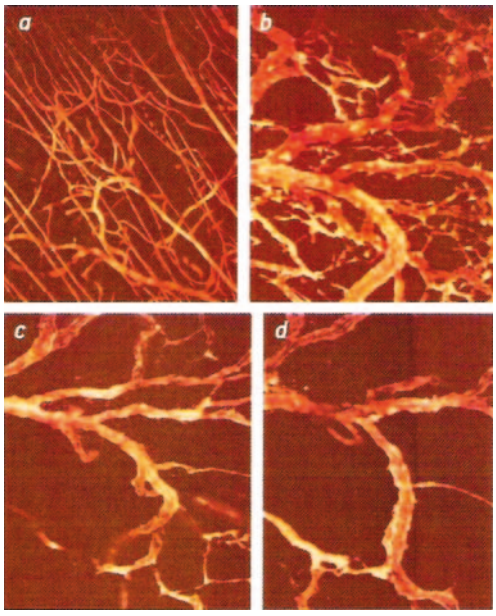


Fig. 4. Response of a vascular network to an antiangiogenic treatment (from [JK01])

reached a certain distance from the parent vessel, they tend to incline towards each other, leading to fusions called anastomoses. Such fusions lead to a network of vessels. On the other hand the sprout branching dramatically increases while approaching the tumour mass, eventually resulting in vascularization.

The coupling of the branching and growth process to the underlying chemical gradients is limited by the local density of the existing capillary network, thus leading to a mathematical strong coupling of this density and the kinetic parameters of the branching and growth process.

The study of angiogenesis has such potential for providing new therapies that it has received enthusiastic interest from the pharmaceutical and biotechnology industries. Indeed, dozens of companies are now pursuing angiogenesis-related therapies, and approximately 20 compounds that either induce or block vessel formation are being tested in humans. Although such drugs can potentially treat a broad range of disorders, many of the compounds now under investigation inhibit angiogenesis and target cancer. Intriguingly, animal tests show that inhibitors of vessel growth can boost the effectiveness of traditional cancer treatments (chemotherapy and radiation). Preliminary studies also hint that the agents might one day be delivered as a preventive measure to block malignancies from arising in the first place in people at risk for cancer.

In developing mathematical models of angiogenesis, the hope is to be able to provide a deeper insight into the underlying mechanisms which cause the process. It is therefore essential that predictive mathematical models are developed, capable of producing precise quantitative morphological features of developing blood vessels. Such models might be used for predicting the evolution of tumours (prognosis), and identifying optimal control strategies (medical treatment).

Unfortunately, a satisfactory modelling of angiogenesis requires a theory of stochastic fibre processes, evolving in time, and strongly coupled with underlying fields. In this case the theory of birth-and-growth processes (or branching-and-growth processes), developed for volume growth, cannot be applied to analyse realistic models, due to intrinsic mathematical difficulties, coming from the dependence of the kinetic parameters from the geometric spatial densities of the existing tumour, or capillary network itself [CM05, McDou06].

All these aspects induce stochastic time and space heterogeneities, thus motivating a more general analysis of the stochastic geometry of the process. The formulation of an exhaustive evolution model which relates all the relevant features of a real phenomenon dealing with different scales, and a stochastic domain decomposition at different Hausdorff dimensions, is a problem of high complexity, both analytical and computational.

Anyway statistical methods for the estimation of geometric densities may offer significant tools for diagnosis and dose/response analysis in medical treatments.

In the modelling of the above-mentioned systems it is of great importance to handle random closed sets of different (even though integer) Hausdorff dimensions. Following a standard approach in geometric measure theory, such sets may be described in terms of suitable measures. For a random closed set of lower dimension with respect to the environment space, the relevant measures induced by its realizations are singular with respect to the Lebesgue measure, and so their usual Radon–Nikodym derivatives are zero almost everywhere.

In Sect. 2 an original approach is reported, recently proposed by the research group of the authors, who have suggested to cope with these difficulties by introducing generalized densities (distributions) *á la Dirac–Schwartz*, for both the deterministic case and the stochastic case. In this last one, mean generalized densities are of interest.

These instruments may then help to formulate stochastic models (that is solving direct problems) for the over-mentioned applications; they also suggest methods for the solution of the related inverse problems, including methods of statistical analysis for the estimation of geometric densities of a stochastic fibre process that characterize the morphology of a real system. We apply such methods to real data, taken from the literature, and to simulated data, obtained by existing computational models of tumour-induced angiogenesis.

These methods can be used for validating computational models, and for monitoring the efficacy of possible medical treatment.

1.1 Nomenclature

We remind that a *random closed set (RACS)* Ξ in $\mathbb{R}^d$ is a measurable map

$$\Xi : (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow (\mathbb{F}, \sigma_{\mathbb{F}}),$$

where $\mathbb{F}$ denotes the class of the closed subsets in $\mathbb{R}^d$, and $\sigma_{\mathbb{F}}$ is the $\sigma-$algebra generated by the so-called *hit-or-miss topology* (see [Mat75]).

The theory of Choquet–Matheron shows that it is possible to assign a unique probability law associated with a *RACS* Ξ in $\mathbb{R}^d$ on the measurable space $(\mathbb{F}, \sigma_{\mathbb{F}})$ by assigning its *hitting functional* T_Ξ.

This is defined as

$$T_\Xi : K \in \mathcal{K} \longmapsto P(\Xi \cap K \neq \emptyset),$$

where $\mathcal{K}$ denotes the family of compact sets in $\mathbb{R}^d$.

Actually we may consider, equivalently, the restriction of T_Ξ to the family of closed balls $\{B_\varepsilon(x); x \in \mathbb{R}^d, \varepsilon \in \mathbb{R}_+ - \{0\}\}$.

In dependence of its regularity, a random closed set Θ_n with Hausdorff dimension n (i.e. $\dim_{\mathcal{H}} \Theta_n(\omega) = n$ for a.e. $\omega \in \Omega$), may induce a random Radon measure

$$\mu_{\Theta_n}(\cdot) := \mathcal{H}^n(\Theta_n \cap \cdot\,)$$

on $\mathbb{R}^d$ ($\mathcal{H}^n$ is the n-dimensional Hausdorff measure), and, as a consequence, an *expected measure*

$$\mathbb{E}[\mu_{\Theta_n}](\cdot) := \mathbb{E}[\mathcal{H}^n(\Theta_n \cap \cdot\,)]$$

(for a discussion about measurability of $\mathcal{H}^n(\Theta_n)$ we refer to [BM97, Z82]).

In several real applications, it is of interest to study the density (said *mean density*) of the measure $\mathbb{E}[\mu_{\Theta_n}]$ [BR04], and, in the dynamical case, its evolution in time [Mol92, Mol94]. Here we present a synthesis of a theory of

random distributions as generalized densities of random measures, and mean geometric densities as expected values of random generalized densities, as proposed in [CV06c]. In particular we introduce a *Delta formalism*, á la Dirac–Schwartz, for the description of random measures associated with random closed sets of lower dimensions, such that the well known usual Dirac delta at a point follows as a particular case (see, for instance, [Jones82,KF70,Vlad79]).

In dealing with mean densities, a concept of *absolutely continuous random closed set* arises in a natural way in terms of the expected measure; indeed, an interesting property of a random set in $\mathbb{R}^d$ is whether the expected measure induced by the random set is absolutely continuous or not with respect to the d-dimensional Lebesgue measure ν^d. Thus, it is of interest to distinguish between random closed sets which induce an absolutely continuous expected measure, and random closed sets which induce a singular one. To this aim we introduce definitions of *discrete*, *continuous*, and *absolutely continuous* random closed set, coherently with the classical 0-dimensional case, in order to propose an extension of the standard definition of discrete, continuous, and absolutely continuous random variable, respectively (see also [CV06a, CV06b]).

2 Generalized Densities

In the sequel we will refer to a class of sufficiently regular random closed sets in the Euclidean space $\mathbb{R}^d$, of integer dimension n.

Definition 1 (n-regular set). *Given an integer $n \in [0, d]$, we say that a closed subset S of $\mathbb{R}^d$ is* n-regular*, if it satisfies the following conditions:*

(*i*) $\mathcal{H}^n(S \cap B_R(0)) < \infty$ *for any* $R > 0$

(*ii*) $\lim_{r\to 0} \dfrac{\mathcal{H}^n(S \cap B_r(x))}{b_n r^n} = 1$ *for* $\mathcal{H}^n$*-a.e.* $x \in S$

Here b_n denotes the volume of the unit ball in $\mathbb{R}^n$.

Remark 1. Note that condition (ii) is related to a characterization of the $\mathcal{H}^n$-rectifiability of the set A ([Fal85], p. 256, 267, [AFP00], p. 83).

We may observe that if A_n is an n-regular closed set in $\mathbb{R}^d$, we have

$$\lim_{r\to 0} \frac{\mathcal{H}^n(A_n \cap B_r(x))}{b_n r^n} = \begin{cases} 1 & \mathcal{H}^n\text{-a.e. } x \in A_n, \\ 0 & \forall x \notin A_n; \end{cases}$$

as a consequence (by assuming $0 \cdot \infty = 0$), for $0 \le n < d$ we have

$$\begin{aligned} \lim_{r\to 0} \frac{\mathcal{H}^n(A_n \cap B_r(x))}{b_d r^d} &= \lim_{r\to 0} \frac{\mathcal{H}^n(A_n \cap B_r(x))}{b_n r^n} \frac{b_n r^n}{b_d r^d} \\ &= \begin{cases} \infty & \mathcal{H}^n\text{-a.e. } x \in A_n, \\ 0 & \forall x \notin A_n. \end{cases} \end{aligned}$$

It is well known that every positive Radon measure μ on $\mathbb{R}^d$ can be decomposed as

$$\mu = \mu_{\ll} + \mu_{\mathbb{P}\mathrm{erp}},$$

where $\mu_{\ll}$ and $\mu_{\mathbb{P}\mathrm{erp}}$ are the absolutely continuous, and the singular parts of μ, respectively, with respect to ν^d, the usual Lebesgue measure on $\mathbb{R}^d$.

It then follows that $\mu_{\ll}$ admits a (nontrivial) Radon–Nikodym derivative with respect to ν^d, which is known as its density; while the Radon–Nikodym derivative of $\mu_{\mathbb{P}\mathrm{erp}}$, with respect to ν^d, would be zero ν^d- a.e.

Anyhow in analogy with the usual Dirac delta function $\delta_{x_0}(x)$ associated with a point $x_0 \in \mathbb{R}^d$ (a 0-regular closed set), a density can be introduced also for $\mu_{\mathbb{P}\mathrm{erp}}$, in a generalized sense, according to Definition 2 [KF70].

Definition 2 (Generalized density). *We call* $\delta_{\mu_{\mathbb{P}\mathrm{erp}}}$, *the* generalized density (*or, briefly, the* density P) *of* $\mu_{\mathbb{P}\mathrm{erp}}$, *the quantity*

$$\delta_{\mu_{\mathbb{P}\mathrm{erp}}}(x) := \lim_{r\to 0} \frac{\mu_{\mathbb{P}\mathrm{erp}}(B_r(x))}{b_d r^d},$$

finite or not.

Clearly, if A_n is an n-regular closed set in $\mathbb{R}^d$ with $n < d$, then the measure

$$\mu_{A_n}(\cdot) := \mathcal{H}^n(A_n \cap \cdot)$$

is a singular measure with respect to ν^d. Based on Definition 1, the quantity

$$\delta_{A_n}(x) := \lim_{r\to 0} \frac{\mathcal{H}^n(A_n \cap B_r(x))}{b_d r^d},$$

(finite or not), can now be introduced as the (generalized) density associated with A_n.

With an abuse of notations, we may introduce the linear functional δ_{A_n} associated with the measure μ_{A_n}, as follows:

$$(\delta_{A_n}, f) := \int_{\mathbb{R}^d} f(x)\mu_{A_n}(\mathrm{d}x),$$

for any $f \in C_{\mathrm{c}}(\mathbb{R}^d, \mathbb{R})$, having denoted by $C_{\mathrm{c}}(\mathbb{R}^d, \mathbb{R})$ the space of all continuous functions from $\mathbb{R}^d$ to $\mathbb{R}$ with compact support. In accordance with the usual representation of distributions in the theory of generalized functions, we formally write

$$\int_{\mathbb{R}^d} f(x)\delta_{A_n}(x)\,\mathrm{d}x := (\delta_{A_n}, f).$$

Define the function

$$\delta^{(r)}_{A_n}(x) := \frac{\mathcal{H}^n(A_n \cap B_r(x))}{b_d r^d},$$

and correspondingly the associated measure

$$\mu_{A_n}^{(r)}(B) := \int_B \delta_{A_n}^{(r)}(x)\,\mathrm{d}x, \qquad B \in \mathcal{B}_{\mathbb{R}^d}.$$

As above, we may introduce the linear functional $\delta_{A_n}^{(r)}$ associated with the measure $\mu_{A_n}^{(r)}$, as follows:

$$\left(\delta_{A_n}^{(r)}, f\right) := \int_{\mathbb{R}^d} f(x)\mu_{A_n}^{(r)}(\mathrm{d}x),$$

It can be proven (see [CV06c]) that the sequence of measures $\mu_{A_n}^{(r)}$ weakly* converges to the measure μ_{A_n}; in other words, the sequence of linear functionals $\delta_{A_n}^{(r)}$ weakly* converges to the linear functional δ_{A_n}, i.e. $(\delta_{A_n}^{(r)}, f) \to (\delta_{A_n}, f)$ for any $f \in C_c(\mathbb{R}^d, \mathbb{R})$.

Consider now random closed sets.

Definition 3 (n-regular random set). *Given an integer n, with $0 \le n \le d$, we say that a random closed set Θ_n in $\mathbb{R}^d$ is* n-regular*, if it satisfies the following conditions:*

(i) For almost all $\omega \in \Omega$, $\Theta_n(\omega)$ is an n-regular set in $\mathbb{R}^d$
(ii) $\mathbb{E}[\mathcal{H}^n(\Theta_n \cap B_R(0))] < \infty$ for any $R > 0$

If Θ_n is a random closed set in $\mathbb{R}^d$, the measure

$$\mu_{\Theta_n}(\cdot) := \mathcal{H}^n(\Theta_n \cap \cdot)$$

is a random measure, and consequently δ_{Θ_n} is a *random linear functional* (i.e. (δ_{Θ_n}, f) is a real random variable for any test function f).

By extending the definition of expected value of a random operator à la Pettis (or Gelfand–Pettis) [AG80, Bosq00], we may define the *expected linear functional* $\mathbb{E}[\delta_{\Theta_n}]$ associated with δ_{Θ_n} as follows:

$$(\mathbb{E}[\delta_{\Theta_n}], f) := \mathbb{E}[(\delta_{\Theta_n}, f)] \tag{1}$$

and the *mean generalized density* $\mathbb{E}[\delta_{\Theta_n}](x)$ of $\mathbb{E}[\mu_{\Theta_n}]$ by the formal integral representation:

$$\int_A \mathbb{E}[\delta_{\Theta_n}](x)\,\mathrm{d}x := \mathbb{E}[\mathcal{H}^n(\Theta_n \cap A)],$$

with

$$\mathbb{E}[\delta_{\Theta_n}](x) := \lim_{r\to 0} \frac{\mathbb{E}[\mathcal{H}^n(\Theta_n \cap B_r(x))]}{b_d r^d}.$$

It can be shown [CV06c] that an equivalent definition of (1) can be given in terms of the expected measure $\mathbb{E}[\mu_{\Theta_n}]$ by

$$(\mathbb{E}[\delta_{\Theta_n}], f) := \int_{\mathbb{R}^d} f(x)\mathbb{E}[\mu_{\Theta_n}](\mathrm{d}x),$$

for any f such that the above integral makes sense.

By using the integral representation of (δ_{Θ_n}, f) and $(\mathbb{E}[\delta_{\Theta_n}], f)$, (1) becomes

$$\int_{\mathbb{R}^d} f(x)\mathbb{E}[\delta_{\Theta_n}](x)\,\mathrm{d}x = \mathbb{E}\left[\int_{\mathbb{R}^d} f(x)\delta_{\Theta_n}(x)\,\mathrm{d}x\right];$$

so that, formally, we may exchange integral and expectation.

Remark 2. When $n = d$, integral and expectation can be really exchanged by Fubini's theorem. Since in this case $\delta_{\Theta_d}(x) = \mathbf{1}_{\Theta_d}(x)$, ν^d-a.s., it follows that $\mathbb{E}[\delta_{\Theta_d}](x) = \mathbb{P}(x \in \Theta_d)$. In particular, in material science, the density $V_V(x) := \mathbb{P}(x \in \Theta_d)$ is known as the (*degree of*) *crystallinity.*

If $n = 0$ and $\Theta_0 = X_0$ is an absolutely continuous random point with p.d.f. p_{X_0}, then $\mathbb{E}[\mathcal{H}^0(X_0 \cap \cdot)] = \mathbb{P}(X_0 \in \cdot)$ is absolutely continuous, and its density $\mathbb{E}[\delta_{X_0}](x)$ is just the probability density function $p_{X_0}(x)$.

Thus, for any lower dimensional random closed set Θ_n in $\mathbb{R}^d$, while it is clear that $\mu_{\Theta_n(\omega)}$ is a singular measure, when we consider the expected measure $\mathbb{E}[\mu_{\Theta_n}]$, it may happen that it is absolutely continuous with respect to ν^d, thus having a classical Radon–Nikodym derivative, so that $\mathbb{E}[\delta_{\Theta_n}](x)$ is a classical real-valued integrable function on $\mathbb{R}^d$ (see [CV06c], and [CV06a]). It is then of interest to say whether or not a classical mean density can be introduced for sets of lower Hausdorff dimensions, with respect to the usual Lebesgue measure on $\mathbb{R}^d$. In order to respond to this further requirement, in [CV06a] we have proposed a concept of absolute continuity for random closed sets.

To avoid pathologies, as discussed in [ACaV06] (see also [CV06d]), we introduce now a class of random sets, which, in particular, include all random sets we are interested in the sequel.

Definition 4 ($\mathcal{R}$ class). *We say that a random closed set Θ in $\mathbb{R}^d$* belongs to the class $\mathcal{R}$ *if*

$$\dim_{\mathcal{H}}(\mathbb{P}\mathrm{artial}\Theta) < d \quad \textit{and} \quad \mathbb{P}(\mathcal{H}^{\dim_{\mathcal{H}}(\mathbb{P}\mathrm{artial}\Theta)}(\mathbb{P}\mathrm{artial}\Theta) > 0) = 1.$$

Definition 5 (Absolute continuity). *We say that a random closed set $\Theta \in \mathcal{R}$ is* (strongly) absolutely continuous *if*

$$\mathbb{E}[\mu_{\mathbb{P}\mathrm{artial}\Theta}] \ll \nu^d \tag{2}$$

on $\mathcal{B}_{\mathbb{R}^d}$.

Remark 3. Note that, if $\Theta \in \mathcal{R}$ with $\dim_{\mathcal{H}}(\Theta) = d$ is sufficiently regular so that $\dim_{\mathcal{H}}(\mathbb{P}\mathrm{artial}\Theta) = d - 1$, then it is absolutely continuous if

$$\mathbb{E}[\mathcal{H}^{d-1}(\mathbb{P}\mathrm{artial}\Theta \cap \cdot)] \ll \nu^d(\cdot).$$

Remark 4. In the particular case that $\Theta = X$ is a random variable, Definition 5 coincides with the usual definition of absolute continuity of a random variable.

In fact, $\dim_{\mathcal{H}} X = 0$, $\mathbb{P}\text{artial}X = X$, and $\mathbb{E}[\mathcal{H}^0(X)] = \mathbb{P}(X \in \mathbb{R}^d) = 1$, so $X \in \mathcal{R}$ and then Condition (2) is equivalent to

$$\mathbb{E}[\mathcal{H}^0(X \cap \cdot\,)] = \mathbb{P}(X \in \cdot\,) \ll \nu^d.$$

To conclude this section, we may then claim that, if Θ_n, with $0 < n < d$, is an absolutely continuous random closed set, then $\mathbb{E}[\mu_{\Theta_n}] \ll \nu^d$, so that its local mean density $\mathbb{E}[\delta_{\Theta_n}](x)$ is a classical real-valued integrable function on $\mathbb{R}^d$.

3 Approximation of Mean Densities

In many real applications, it is of interest the estimation of the local mean density $\mathbb{E}[\delta_{\Theta_n}]$ of an absolutely continuous lower dimensional random closed set such as a fibre process of dimension $n = 1$ in a space of dimension $d > 1$ (see, e.g. [BR04] and [SKM95]).

For facing the problem of the zero ν^2-measure for points or lines in $\mathbb{R}^2$ it is natural to make use of a 2-D box approximation of points or lines. As a matter of fact, a computer graphic representation of them is anyway provided in terms of pixels, which can only offer a 2-D box approximation of points in $\mathbb{R}^2$. This is the motivation of this and the following sections, which tend to suggest estimators for local mean densities of absolutely continuous random closed sets of lower dimensions in a given d-dimensional space [ACaV06].

Given a random closed set Θ_n with Hausdorff dimension n, we consider the enlarged set $\Theta_{n_{\oplus r}}$, which is now of dimension d, and hence of nontrivial measure ν^d. We observe that $\mathbb{P}(x \in \Theta_{n_{\oplus r}}) = T_{\Theta_n}(B_r(x))$.

Proposition 1. *[ACaV06] Let Θ_n be a random closed set with Hausdorff dimension n, and $A \in \mathcal{B}_{\mathbb{R}^d}$ such that $\mathbb{P}(\mathcal{H}^n(\Theta_n \cap \mathbb{P}\text{artial}A) > 0) = 0$. If*

$$\lim_{r\to 0} \frac{\mathbb{E}[\nu^d(\Theta_{n_{\oplus r}} \cap A)]}{b_{d-n} r^{d-n}} = \mathbb{E}[\mathcal{H}^n(\Theta_n \cap A)], \tag{3}$$

then

$$\mathbb{E}[\mathcal{H}^n(\Theta_n \cap A)] = \lim_{r\to 0} \int_A \frac{T_{\Theta_n}(B_r(x))}{b_{d-n} r^{d-n}}\, \mathrm{d}x.$$

Sufficient conditions for (3) have been given in [ACaV06].

As a consequence of Proposition 1, if we denote by $\mu^{\oplus r}$ the measure on $\mathcal{B}_{\mathbb{R}^d}$ defined by

$$\mu^{\oplus r}(A) := \int_A \frac{T_{\Theta_n}(B_r(x))}{b_{d-n} r^{d-n}}\, \mathrm{d}x,$$

then it follows that $\mu^{\oplus r}$ weakly* converges to $\mathbb{E}[\mu_{\Theta_n}]$.

For every fixed $r > 0$, the measure $\mu^{\oplus r}$ is absolutely continuous with respect to the d-dimensional Lebesgue measure with density

$$\delta_n^{\oplus r}(x) := \frac{T_{\Theta_n}(B_r(x))}{b_{d-n}r^{d-n}}.$$

Such a function defines a linear functional, say $\delta_n^{\oplus r}$, associated with the measure $\mu^{\oplus r}$ as follows

$$(\delta_n^{\oplus r}, f) := \int_{\mathbb{R}^d} f(x)\mu^{\oplus r}(\mathrm{d}x).$$

Note that many kinds of random closed sets satisfy the proposition above, like fibre processes, line and segment processes, Boolean models, etc. (see [ACaV06]). As a consequence, estimating the probability that the random set Θ_n intersects the ball $B_r(x)$ may suggest (global) estimators of $\mathbb{E}[\mu_{\Theta_n}]$, and possibly (local) estimators of the mean density $\mathbb{E}[\delta_{\Theta_n}]$ (see, e.g. [BR04]).

If Θ_n is absolutely continuous, then there exists an integrable function λ_{Θ_n} (the Radon–Nikodym derivative) such that, for all $A \in \mathcal{B}_{\mathbb{R}^d}$,

$$\mathbb{E}[\mathcal{H}^n(\Theta_n \cap A)] = \int_A \lambda_{\Theta_n}(x)\,\mathrm{d}x.$$

So, in this case, we have that

$$\lim_{r\to 0}\int_A \frac{T_{\Theta_n}(B_r(x))}{b_{d-n}r^{d-n}}\mathrm{d}x = \int_A \lambda_{\Theta_n}(x)\,\mathrm{d}x. \tag{4}$$

If Θ_n is a stationary random closed set, then $\delta_n^{\oplus r}(x)$ is independent of x and the expected measure $\mathbb{E}[\mu_{\Theta_n}]$ is motion invariant, i.e. it is absolutely continuous with density $\lambda_{\Theta_n}(x) = L \in \mathbb{R}_+$ for ν^d-a.e. $x \in \mathbb{R}^d$. It follows that

$$\lim_{r\to 0}\int_A \frac{T_{\Theta_n}(B_r(x))}{b_{d-n}r^{d-n}}\mathrm{d}x = \lim_{r\to 0}\frac{T_{\Theta_n}(B_r(0))}{b_{d-n}r^{d-n}}\nu^d(A),$$

and

$$\int_A \lambda(x)\,\mathrm{d}x = L\nu^d(A);$$

and so, by (4),

$$\lim_{r\to 0}\frac{T_{\Theta_n}(B_r(0))}{b_{d-n}r^{d-n}} = L.$$

Remark 5. When it is possible to exchange limit and integral in (4), by Proposition 1 we may claim that

$$\lim_{r\to 0}\frac{T_{\Theta_n}(B_r(x))}{b_{d-n}r^{d-n}} = \lambda_{\Theta_n}(x) \qquad \nu^d\text{-a.e. } x \in \mathbb{R}^d.$$

In the particular case $n = d$, we know that the measure $\mathbb{E}[\mu_{\Theta_d}]$ is always absolutely continuous with density $\lambda_{\Theta_d}(x) = \mathbb{P}(x \in \Theta_d)$. We may notice that $\delta_d^{\oplus r} = T_{\Theta_n}(B_r(x))$ and by Monotone Convergence Theorem we can exchange limit and integral, and so we have, as expected,

$$\lim_{r\to 0} T_{\Theta_d}(B_r(x)) = \mathbb{P}(x \in \Theta_d) = \lambda_{\Theta_d}(x).$$

Further, for $n = 0$, if $\Theta_0 = X$ is a random point in $\mathbb{R}^d$, we have $\mathbb{E}[\mathcal{H}^0(X \cap \cdot)] = \mathbb{P}(X \in \cdot)$. So, if X is absolutely continuous with probability density function f, we know that $\mathbb{E}[\mu_X] = \mathbb{P}_X$ is absolutely continuous with density f. In this case it can be shown that (3) holds, so that the sequence $\{\delta^{\oplus r}(x)\}$ converges to $f(x)$, as expected, which leads to the usual histogram estimation of $f(x)$ [ACaV06].

Example 1. As an additional example of applicability of the results above, let us consider the case in which Θ_n is given by a random union of absolutely continuous random closed sets of dimension $n < d$:

$$\Theta_n = \bigcup_{i=1}^{\Phi} E_i,$$

where Φ is a nonnegative discrete random variable with $\mathbb{E}[\Phi] < \infty$, and the E_i's are IID as E and independent of Φ. Then it follows that [ACaV06]

$$\lim_{r\to 0} \frac{T_{\Theta_n}(B_r(x))}{b_{d-n} r^{d-n}} = \mathbb{E}[\Phi] \lim_{r\to 0} \frac{T_E(B_r(x))}{b_{d-n} r^{d-n}},$$

provided that at least one of the two limits exists.

As a consequence, when it is possible to exchange limit and integral in (4), and so in particular when E is a stationary random closed set (which implies Θ_n stationary as well), we have

$$\lambda_{\Theta_n}(x) = \mathbb{E}[\Phi] \lim_{r\to 0} \frac{T_E(B_r(x))}{b_{d-n} r^{d-n}} = \mathbb{E}[\Phi]\lambda_E(x),$$

where λ_{Θ_n} and λ_E are the Radon–Nikodym derivatives of μ_{Θ_n} and μ_E, respectively. The above model may be used as a preliminary one for angiogenesis [CM05], but also for the earthworm burrow system in a soil [BR04, p.73].

4 Statistical Methods for Fibre Systems

We will here consider random fibre systems generated by Boolean models having a fibre as primary grain, that is a RACS Γ such that

$$\Gamma = \cup_{i\in\mathbb{N}} \Gamma_i \oplus x_i,$$

where

– $\{x_i\}_{i\in\mathbb{N}}$ is a spatial Poisson point process, possibly inhomogeneous, with intensity $\alpha(x), x \in \mathbb{R}^2$

– $\{\Gamma_i\}_{i\in\mathbb{N}}$ is a family of i.i.d. random fibres (i.e. random, a.s. bounded, 1-regular sets), passing a.s. through the origin

The resulting Boolean model is thus in general nonstationary and non-isotropic. The source of nonstationarity comes essentially from the nonstationarity of the germ process, i.e. from the *location* of the fibres, and not from intrinsic geometric irregularities of the fibres themselves. In fact the grains are assumed geometrically regular (1-regular) and with "good" statistical properties (i.i.d.). The main source of anisotropy instead comes from the distribution of fibres (grains) orientation, which may be nonuniform.

Note now that

$$T_{\Theta_n}(B_r(x)) = \mathbb{P}(x \in \Theta_{n_{\oplus r}}) = \mathbb{P}(\Theta_n \cap B_r(x) \neq \emptyset)$$

thus we may rewrite Equality (4) in the following way

$$\int_A \lambda_n(x)\mathrm{d}x = \lim_{r\to 0}\int_A \frac{\mathbb{P}(\Theta_n \cap B_r(x) \neq \emptyset)}{b_{d-n}r^{d-n}}\mathrm{d}x \tag{5}$$

$$= \lim_{r\to 0}\int_A \frac{T_{\Theta_n}(B_r(x))}{b_{d-n}r^{d-n}}\mathrm{d}x \tag{6}$$

$$= \lim_{r\to 0}\int_A \frac{\mathbb{P}(x \in \Theta_{n_{\oplus r}})}{b_{d-n}r^{d-n}}\mathrm{d}x. \tag{7}$$

Equalities (5)–(7) provide a way to introduce estimators of $\lambda_n(x)$ when Θ_n is a random fibre, or fibre system Γ, provided that the limit and the integrals in the right-hand terms of (5)–(7) can be exchanged, by estimating the quantities

$$\frac{T_\Gamma(B_r(x))}{b_{d-1}r^{d-1}} = \frac{\mathbb{P}(x \in \Gamma_{\oplus r})}{b_{d-1}r^{d-1}} = \frac{\mathbb{P}(\Gamma \cap B_r(y) \neq \emptyset)}{b_{d-1}r^{d-1}}.$$

We will call them *histogram-like* estimators, since the "enlargement" $\Gamma_{\oplus r}$ of the set Γ via the Minkowski addition of a d-dimensional ball, which approximates the fibre with a d-dimensional set, imitates the procedure used when we estimate the p.d.f. of a real random variable from an i.i.d. sample using moving histograms (see [Hard91, Pest98] for details), where we "enlarge" the Dirac-delta's measures concentrated on the sample points, approximating them with classical and sufficiently regular functions.

In the following we will provide two estimators for the mean geometric density of length, also called *intensity*, of the random fibre system Γ. The intensity can be used to characterize the mean geometric properties of the fibre system. Accordingly with the definitions introduced in the previous sections, the intensity of Γ is defined by

$$\lambda(x) \colon = \mathbb{E}[\delta_\Gamma](x) = \lim_{r\to 0}\frac{\mathbb{E}(\mathcal{H}^1(\Gamma \cap B_r(x)))}{r^d b_d}.$$

4.1 Basic Assumptions for the Estimation Procedure

Suppose to have one or more images of the random fibre system Γ under study and that the window $W \subseteq \mathbb{R}^d$ where Γ is observed can be divided in a partition of subwindows $\{A_k\}_{k=1,\dots,K}$ such that:

A1 $A_j \cap A_k = \emptyset, \quad \forall j \neq k$
A2 $\bigcup_{k=1}^{K} A_k = W$
A3 in each window A_k limit and integral in (4) can be exchanged when $\Theta_n = \Gamma$. This is the case for example if in A_k the fibre system is (*locally*) stationary
A4 the intensity $\lambda(x)$ is sufficiently "smooth" to be locally well approximated by piecewise constant functions, assuming different constant values in each window A_k

We will now introduce a (nonstationary!) example where the previous assumptions are satisfied. The example will be used in the following as a case study for the properties of our estimators.

4.2 An Example of Inhomogeneous Poisson Segment Process

Let $d = 2$ and consider the Boolean model Γ formed by:

- Germs: A spatial nonhomogeneous Poisson point process $\{x_i\}_{i\in\mathbb{N}}, x_i \in \mathbb{R}^2$ having intensity $\alpha(x) = \alpha(x_1, x_2) = cx_1^2$, and c is a constant.
- Grains: A family $\{S_i\}_{i\in\mathbb{N}}$ of (deterministic) closed sets all distributed like the segment $S = [0, l] \times \{0\}$ of fixed length l.

The resulting Boolean model is

$$\Gamma = \bigcup_{i\in\mathbb{N}} S_i \oplus x_i$$

(see Fig. 5 where a realization is depicted). Note that since the germ intensity $\alpha(x)$ is a function of class C^∞, Assumption A4 is trivially satisfied.

Let us assume that the following equality holds

$$\lambda(x) \colon = \lim_{r\to 0} \frac{\mathbb{E}\left(\nu^1(\Gamma \cap B_r(x))\right)}{2r} = \lim_{r\to 0} \frac{\mathbb{E}\left(\nu^1(\Gamma \cap Q_r(x))\right)}{2r}, 0 \tag{8}$$

where $Q_r(x)$ is a square centred at x with side $2r$. This assumption is reasonable, since both cubes and spheres form a system of generators of the Borel σ-algebra in $\mathbb{R}^d$. Then Assumption A3 is satisfied thanks to the following.

Proposition 2. *Let Γ be the random segment system described above, $\lambda(x)$ be the mean intensity of length of the system, and suppose that Assumption (8) is satisfied, i.e.*

$$\lambda(x) = \lim_{r\to 0} \frac{E\left(\nu^1(\Gamma \cap Q_r(x))\right)}{2r}.$$

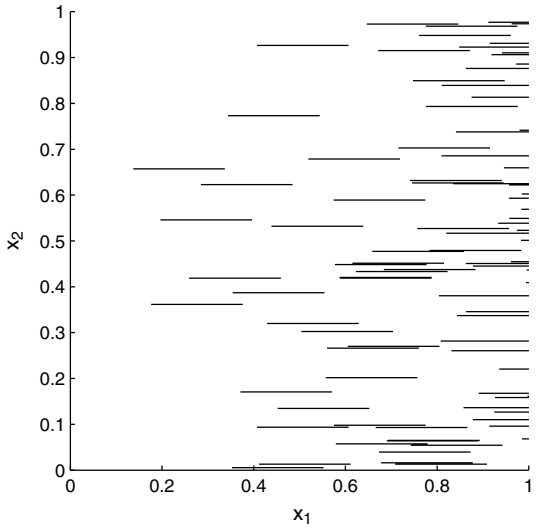


Fig. 5. A realization of the Boolean model in the example. The point process $\{x_i\}$ of germs is formed by the left-hand extremes of the segments

Then, the quantity

$$\bar{\lambda}(x) \colon = \lim_{r \to 0} \frac{\mathbb{P}(\Gamma \cap Q_r(x) \neq \emptyset)}{2r}. \tag{9}$$

exists and is finite and we have, a.s.

$$\bar{\lambda}(x) = \lambda(x)$$

for ν^2- almost all $x \in \mathbb{R}^2$.

For the proof of this proposition see [CM06]. Let us remark that, in the proof, the particular functional form of $\bar{\lambda}(x)$ is not relevant.

5 Estimators of the Intensity

In the assumptions stated in Sect. 4.1, for all $x \in A_k$, let us denote by

$$\begin{aligned} \lambda_k \colon &= \lim_{r \to 0} \frac{\mathbb{E}(\nu^1(\Gamma \cap B_r(x)))}{b_d r^d} \\ &= \lim_{r \to 0} \frac{\mathbb{P}(x \in \Gamma_{\oplus r} \cap A_k)}{2r} \\ &= \lim_{r \to 0} \frac{T_\Gamma(B_r(x))}{2r} \end{aligned}$$

the (constant) intensity of the random fibre system in the subwindow A_k. We have explicited all the previous equalities since we will obtain different estimators, based on the estimate of the quantities:

1. $\mathbb{P}(x \in \Gamma_{\oplus r} \cap A_k)$
2. $T_\Gamma(B_r(x))$,

respectively.

Let us build first an estimator based on the estimate of $\mathbb{P}(x \in \Gamma_{\oplus r} \cap A_k)$. Let us overlap to A_k a grid of points $z_1, \ldots, z_p \in A_k$ and build the set $\Gamma_{\oplus r} \cap A_k$. Then a first estimator of λ_k is

$$\hat{\lambda}^1_{k,r,p} := \frac{1}{2rp} \sum_{i=1}^{p} \mathbf{1}_{z_i \in \Gamma_{\oplus r} \cap A_k}, \tag{10}$$

where $\mathbf{1}_{z_i \in \Gamma_{\oplus r} \cap A_k}$ are i.i.d. Bernoulli random variables assuming value one with probability $\mathbb{P}(x \in \Gamma_{\oplus r} \cap A_k)$ which is independent of $x \in A_k$ in our assumptions. Since estimator (10) is the arithmetic mean of these variables, by applying the strong law of large numbers (SLLN) and Slutsky Theorem (see, e.g. [Pest98]) we obtain

$$\mathbb{E}(\hat{\lambda}^1_{k,r,p}) = \frac{\mathbb{P}(x \in \Gamma_{\oplus r} \cap A_k)}{2r} \longrightarrow \lambda_k, \text{ for } r \to 0 \tag{11}$$

$$\mathrm{Var}(\hat{\lambda}^1_{k,r,p}) = \frac{(\mathbb{P}(x \in \Gamma_{\oplus r} \cap A_k))(1 - \mathbb{P}(x \in \Gamma_{\oplus r} \cap A_k))}{4r^2 p} \longrightarrow 0, \tag{12}$$
$$\text{for } r \to 0, p \to \infty, rp \to \infty$$

that is the asymptotic unbiasedness and weak consistency of the estimator, when $r \to 0, p \to \infty$ with $rp \to \infty$.

Note that this estimator is not much affected by edge effects, if the "enlargement" of Γ is performed correctly. If the fibres go across the whole window or have extremes internal to the window but far from the window border, edge effects are not present. For fibres having extremes close to the window border, edge effects can be reduced by reducing also the width r of the enlargement (see Fig. 6).

Let us now introduce an estimator based on the estimate of $T_\Gamma(B_r(x)), x \in A_k$. Let us again consider a grid of points $z_1, \ldots, z_p$ overlapped on the window A_k, such that $B_r(z_i) \subseteq A_k$ for all $i = 1, \ldots, p$ (this assumption has again the aim of reducing the edge effects). We then define

$$\hat{\lambda}^2_{k,r,p} := \frac{1}{2rp} \sum_{i=1}^{p} \mathbf{1}_{\Gamma \cap B_r(z_i) \neq \emptyset}. \tag{13}$$

where again $\mathbf{1}_{\Gamma \cap B_r(z_i) \neq \emptyset}$ is a Bernoulli random variable assuming value 1 with probability

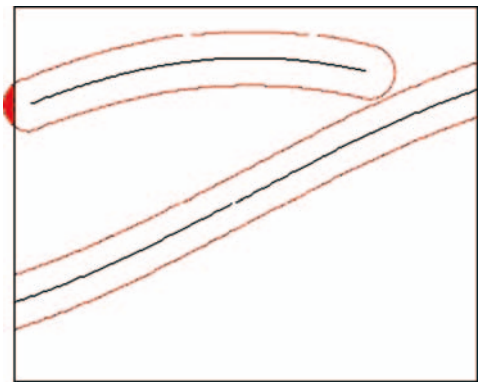


Fig. 6. Examples of edge effects: If the extreme of the fibre is internal and too close to the window border, a piece of enlargement is not considered in the estimator $\hat{\lambda}^1_{k,r}$

$$\mathbb{P}(\Gamma \cap B_r(z_i) \neq \emptyset) = T_\Gamma(B_r(z_i)) = T_\Gamma(B_r(x)),$$

$\forall x \in A_k$. Thus again by using the SLLN and Slutsky Theorem we obtain the asymptotic unbiasedness and weak consistency of this estimator, in fact, for any $x \in A_k$,

$$\mathbb{E}(\hat{\lambda}^2_{k,r,p}) = \frac{T_\Gamma(B_r(x))}{2r} = \frac{\mathbb{P}(x \in \Gamma_{\oplus r} \cap A_k)}{2r} \longrightarrow \lambda_k, \text{ for } r \to 0 \tag{14}$$

$$\mathrm{Var}(\hat{\lambda}^2_{k,r,p}) = \frac{(T_\Gamma(B_r(x)))(1 - T_\Gamma(B_r(x)))}{4r^2 p} \tag{15}$$

$$= \frac{(\mathbb{P}(x \in \Gamma_{\oplus r} \cap A_k))(1 - \mathbb{P}(x \in \Gamma_{\oplus r} \cap A_k))}{4r^2 p} \to 0, \tag{16}$$

$$\text{for } r \to 0, p \to \infty, rp \to \infty.$$

6 Application of the Estimators to the Simulated Inhomogeneous Poisson Segment Process

In this section we will apply the estimators $\hat{\lambda}^1_{k,r,p}$ and $\hat{\lambda}^2_{k,r,p}$ introduced in Sect. 5 to the working example introduced in Sect. 4.2 and we will also derive the rate of convergence to 0 of the variance, in order to assess a method for choosing an "optimal bandwidth" of enlargement r, depending on p. Since the true intensity of this process is known, we use first this example to test empirically the properties of our estimators. In Sect. 7 we will apply the estimators to real or simulated processes where the true intensity is unknown.

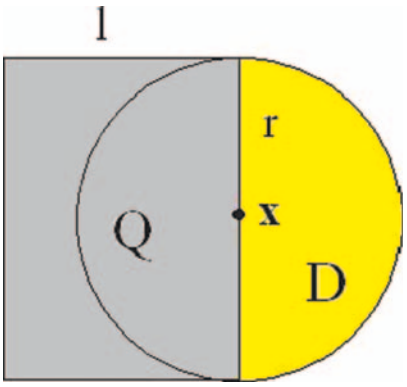


Fig. 7. The regions where a germ must appear in order that a segment hits $B_r(x)$

Assume that $r < l$, where l is the length of the segments forming the Boolean model, consider the quantities which have been estimated in Sect. 5 in the specific case of our example, where $x = (x_1, x_2) \in \mathbb{R}^2$ and refer to Fig. 7 for the definition of the regions Q and D.

$$\begin{aligned}
\mathbb{P}(x \in \Gamma_{\oplus r}) &= \mathbb{P}(B_r(x) \cap \Gamma \neq \emptyset) \\
&= 1 - \mathbb{P}(B_r(x) \cap \Gamma = \emptyset) \\
&= 1 - \mathbb{P}(\text{no germs fall in } Q \cup D) \\
&= 1 - \exp\left[- \int_Q \alpha(x)\mathrm{d}x - \int_D \alpha(x)\mathrm{d}x \right] \\
&= 1 - \exp\left[-2r \int_{x_1 - l}^{x_1} c\bar{x}_1^2 \mathrm{d}\bar{x}_1 - \int_{x_1}^{x_1 + r} \int_{x_2 - \sqrt{r^2 - (\bar{x}_1 - x_1)^2}}^{x_2 + \sqrt{r^2 - (\bar{x}_1 - x_1)^2}} c\bar{x}_1^2 \mathrm{d}\bar{x}_1 \mathrm{d}\bar{x}_2 \right] \\
&= 1 - \exp\Big[-\frac{2}{3} rc \left(x_1^3 - (x_1 - l)^3 \right) \qquad (17) \\
&\quad - \int_{x_1}^{x_1 + r} 2c\bar{x}_1^2 \sqrt{r^2 - (\bar{x}_1 - x_1)^2} \mathrm{d}\bar{x}_1 \Big]. \qquad (18)
\end{aligned}$$

Now by computing a Taylor series expansion of (17)–(18) in a right neighborhood of $r = 0$, we obtain

$$\mathbb{P}(x \in \Gamma_{\oplus r}) = 2r\lambda(x) - 2r^2(\lambda(x))^2 + o(r^2).$$

By substituting this expansion in the expressions of the expected value and variance of the estimators $\hat{\lambda}^1_{k,r,p}$ and $\hat{\lambda}^2_{k,r,p}$, which are the same, given by (11), (12), and (14), (16), we get, for all $x \in A_k$ and $i = 1, 2$

$$\mathbb{E}(\hat{\lambda}^i_{k,r,p}) = \lambda(x) - r\lambda^2(x) + o(r) \tag{19}$$

$$\mathrm{Var}(\hat{\lambda}^i_{k,r,p}) = \frac{\lambda(x)}{2rp} - \frac{3\lambda^2(x)}{2p} + o\left(\frac{1}{p}\right). \tag{20}$$

The optimal enlargement or bandwidth r can then be computed by minimizing the mean square error, which (by neglecting infinitesimal terms of higher order) is given by

$$\begin{aligned}\mathrm{MSE}(\hat{\lambda}^i_{k,r,p}) &= \mathrm{Var}(\hat{\lambda}^i_{k,r,p}) + \mathrm{Bias}^2(\hat{\lambda}^i_{k,r,p}) \\ &= \frac{\lambda(x)}{2rp} - \frac{3\lambda^2(x)}{2p} + r^2\lambda^4(x).\end{aligned}$$

By minimization one obtains

$$r_{\mathrm{optimal}} = \arg\min_r \mathrm{MSE}(\hat{\lambda}^i_{k,r,p}) = [4p\lambda(x)]^{-1/3}.$$

Note that the optimal bandwidth can be computed only if the true intensity $\lambda(x)$ is known, which is obviously not the case in general. The problem can be overcome in various ways, for example assuming that λ belongs to a given family of functions depending on parameters which can be estimated from the data, or with iterative methods, via the use of an initial guess for $\lambda(x)$ or for r_{optimal}. A discussion for the case of kernel density estimators of the p.d.f. of real-valued random variables can be found in [Hard91, Chap. 4].

7 Experimental Results

We applied estimators $\hat{\lambda}^1_{k,r,p}$ and $\hat{\lambda}^2_{k,r,p}$ to simulated data coming from the model described in Sect. 4.2. The simulation has been performed in the window $[0,1] \times [0,1]$; the constant c appearing in the intensity of germs has been assumed $c = 400$, and the length of segments was fixed to $l = 0.2$. The window $[0,1] \times [0,1]$ was divided into ten vertical stripes of equal width. The two estimators have been computed on each subwindow both by using a deterministic grid of p points z_i, coinciding with the grid of pixels of the image, and by overlapping a random grid of p uniformly distributed points $z_i, i = 1, \ldots, p$. The second method is less affected by correlation problems which may arise from points which have a spatially close location, but has higher computational costs. The optimal bandwidth r has been computed via the true value of $\lambda(x)$ in the centroid of each subwindow. The results are reported in Fig. 8. Since the estimators are biased, with first order bias given in (19), we corrected the estimators by subtracting $-r\lambda^2(x)$. The corrected estimators are reported in Fig. 9, and show a good agreement with the true value of the intensity of the process.

We also computed confidence bands for the estimators, both corrected and uncorrected for bias, by simulating 100 processes with the same intensity, performing on each simulated pattern the estimation procedure and taking the

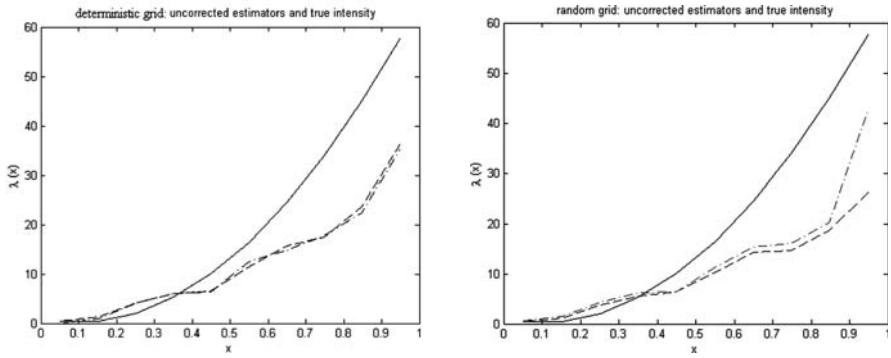


Fig. 8. *Left*: estimate with a deterministic grid; *right*: estimate with a random grid. *Dashed line* $= \hat{\lambda}^1_{k,r,p}$, *dotted-dashed line* $= \hat{\lambda}^2_{k,r,p}$, *continuous line* = true value of $\lambda(x)$. The random grid used for the right-hand picture was formed by $p = 2,000$ uniformly distributed points. The number of pixels in each subwindow, used for the deterministic grid, is $11,628$

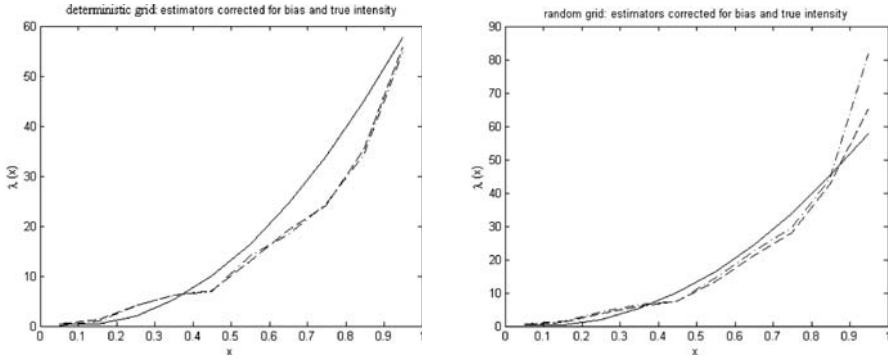


Fig. 9. *Left*: estimate with a deterministic grid; *right*: estimate with a random grid. *dashed line* $= \hat{\lambda}^1_{k,r,p}$, *dotted-dashed line* $= \hat{\lambda}^2_{k,r,p}$, *continuous line* = true value of $\lambda(x)$. The estimators have been corrected for bias using the true value of the intensity. The random grid used for the bottom picture was formed by $p = 2,000$ uniformly distributed points. The number of pixels in each subwindow, used for the deterministic grid, is $11,628$

minimum and maximum values of the estimated intensity in each subwindow. The results are reported in Figs. 10 and 11. From the experimental results the estimators obtained by overlapping to the subwindows a deterministic equally spaced grid seem not to be equivalent to the ones obtained by overlapping a random grid of uniformly distributed points. The deterministic ones seem to have a larger variance than the random ones, and the random ones still show some negative bias, even after the correction, probably due to the terms of higher order which we neglected. The difference in the variance is due to the fact that in the derivation of the expected value and variance of the two esti-

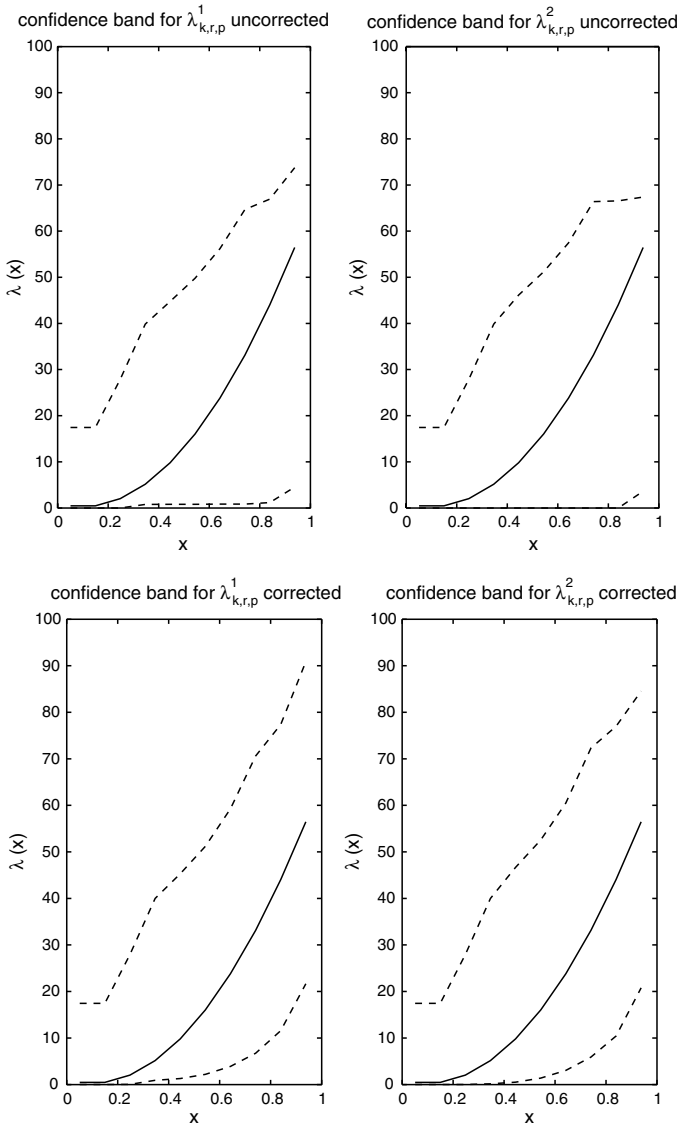


Fig. 10. *Dashed line* min–max confidence band computed by estimating the intensity over 100 simulations of the process; *continuous line* true value of $\lambda(x)$. For the estimation we used a *deterministic* equally spaced grid of points (coinciding with the pixels of the image) overlapped to each subwindow

mators, we assumed that the indicator functions appearing in their definition were i.i.d. Unfortunately the indicators are not independent if the points z_i are located on a regular grid, of width dx comparable with the length l of the segments of the Boolean model or with the "enlargement bandwidth" r. Note

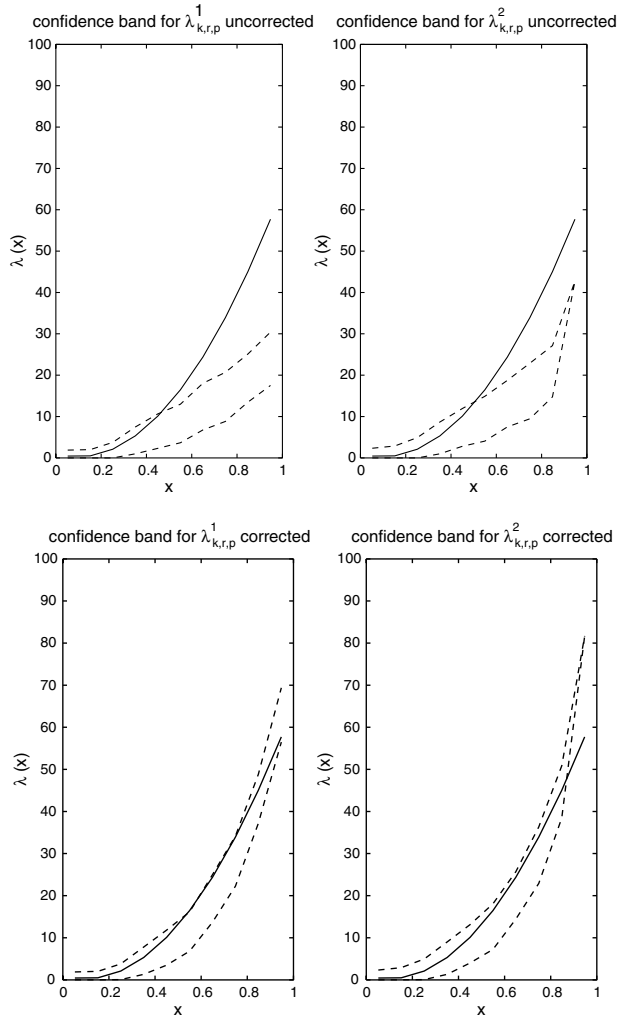


Fig. 11. *Dashed line* min–max confidence band computed by estimating the intensity over 100 simulations of the process; *continuous line* true value of $\lambda(x)$. For the estimation we used a *random* grid of $2,000$ uniformly distributed points overlapped to each subwindow

that the results obtained using a random grid could be improved by augmenting the number of random points of the grid, with a consequent increase of the computational costs.

Since in real applications the true intensity of the fibre process is unknown, we also applied an iterative method to compute the intensity. The method starts by enlarging of the same quantity r_{start} (initial guess) the fibres in all the subwindows; then the estimate procedure is applied and the estimate of

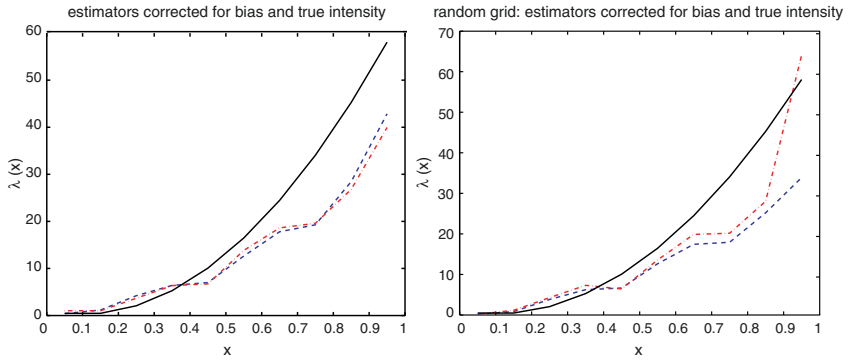


Fig. 12. Estimates with an iterative method. *Left figure*: estimate with a deterministic grid. The number of pixels in each subwindow, used for the deterministic grid, is $11,628$. *Right figure*: Estimate with a random grid. The random grid used for the estimate was formed by $p = 2,000$ uniformly distributed points. The estimators have been corrected for bias; *dashed line* $= \hat{\lambda}^1_{k,r,p}$, *dotted-dashed line* $= \hat{\lambda}^2_{k,r,p}$, *continuous line* = true value of $\lambda(x)$

$\lambda(x)$ is computed in each subwindow. The estimated intensity is then used to compute the optimal enlargement bandwidth r in the next iteration and an update of $\hat{\lambda}_{k,r,p}$ is computed. A given tolerance constant tol is fixed and the procedure is iterated up to when

$$\sup_x |\hat{\lambda}^i(x)_{m+1} - \hat{\lambda}^i(x)_m| < \text{tol},$$

where $\hat{\lambda}^i(x)_m$ is the intensity function estimated at iteration m ($i = 1, 3$ for the two considered estimators). The study of the termination of the iterative procedure is left to subsequent papers.

The results are reported in Fig. 12. Also in this case the estimators have been corrected for bias, using the estimated intensity for the correction instead of the true value of $\lambda(x)$. From the experimental results the termination and the results of the algorithm does not seem to depend strongly on the initial guess. The convergence looks faster for $\hat{\lambda}^1_{k,r,p}$ if we use a deterministic grid, and for $\hat{\lambda}^2_{k,r,p}$ if we use a random grid.

Min–max confidence bands have been computed over 100 simulations of the process also with the iterative method, using both a deterministic and a random grid; the results are reported in Figs. 13 and 14.

Remark 6. In this case estimator $\hat{\lambda}^1_{k,r,p}$ computed with a random grid seems to behave badly with respect to the others, in particular when the true intensity is high. Nevertheless this estimator has many computational advantages when applied to subwindows which have not a rectangular shape, since overlapping a random grid of points to a window having any shape and counting what points are falling inside the enlarged fibres, is much easier than selecting random points which have a spherical neighbourhood of fixed width r entirely

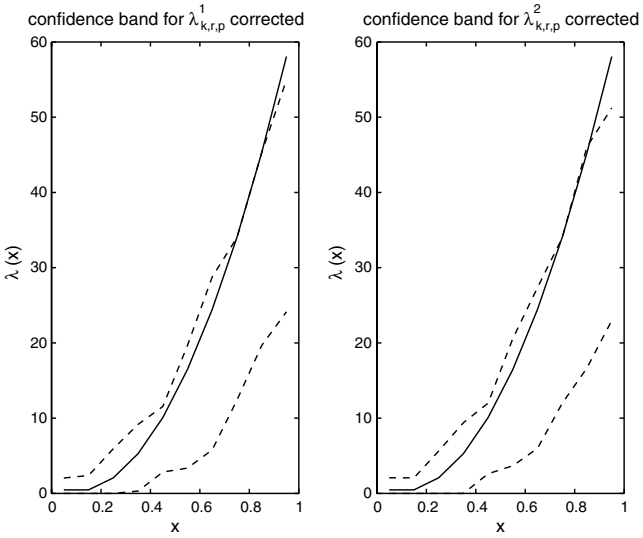


Fig. 13. Confidence bands for $\lambda(x)$ using $\hat{\lambda}^1_{k,r,p}$ (*left figure*) and $\hat{\lambda}^2_{k,r,p}$ (*right figure*). *Dashed line* = min–max confidence band computed by estimating the intensity over 100 simulations of the process; *continuous line* = true value of $\lambda(x)$. For the estimation we used a *deterministic* equally spaced grid of points (coinciding with the pixels of the image) overlapped to each subwindow

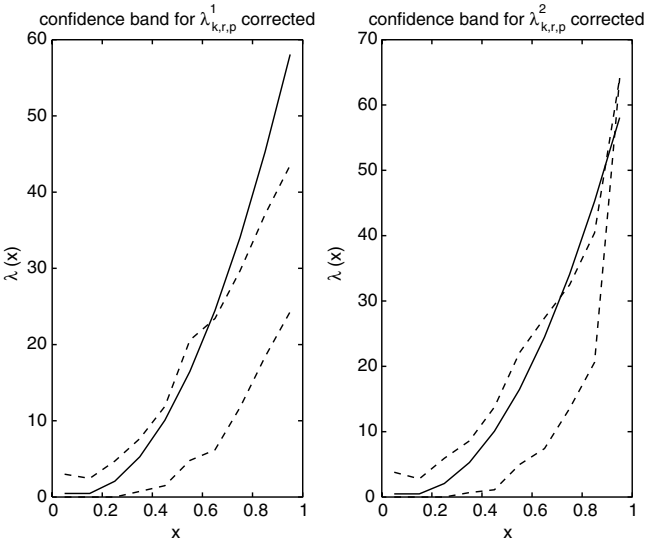


Fig. 14. Confidence bands for $\lambda(x)$ using $\hat{\lambda}^1_{k,r,p}$ (*left figure*) and $\hat{\lambda}^2_{k,r,p}$ (*right figure*). *Dashed line* = min–max confidence band computed by estimating the intensity over 100 simulations of the process; *continuous line* = true value of $\lambda(x)$. For the estimate we used a random grid formed by $p = 2,000$ uniformly distributed points

contained in the subwindow. Thus estimator $\hat{\lambda}^1_{k,r,p}$ will be more often used in the real applications which need a nonrectangular division in subwindows for a good analysis.

The estimators have then been applied to some simulations of real fibre processes, where the true intensity is not known. In Fig. 15a simulation of the generation and branching of vessels driven by a chemotactic field generated by a tumour is reported. The tumour is located on the right-hand side of the window and the vessels start growing and branching from the left-hand side of the window in the right direction. The chemotactic field has a gradient in the x direction and influences both the speed of growth and the branching of the vessels. The intensity has been estimated both with a deterministic and a random grid, by dividing the observation window into ten vertical stripes of the same width. The estimators have been corrected for bias. The results are reported in Fig. 15. In Fig. 16 two simulations are reported where the

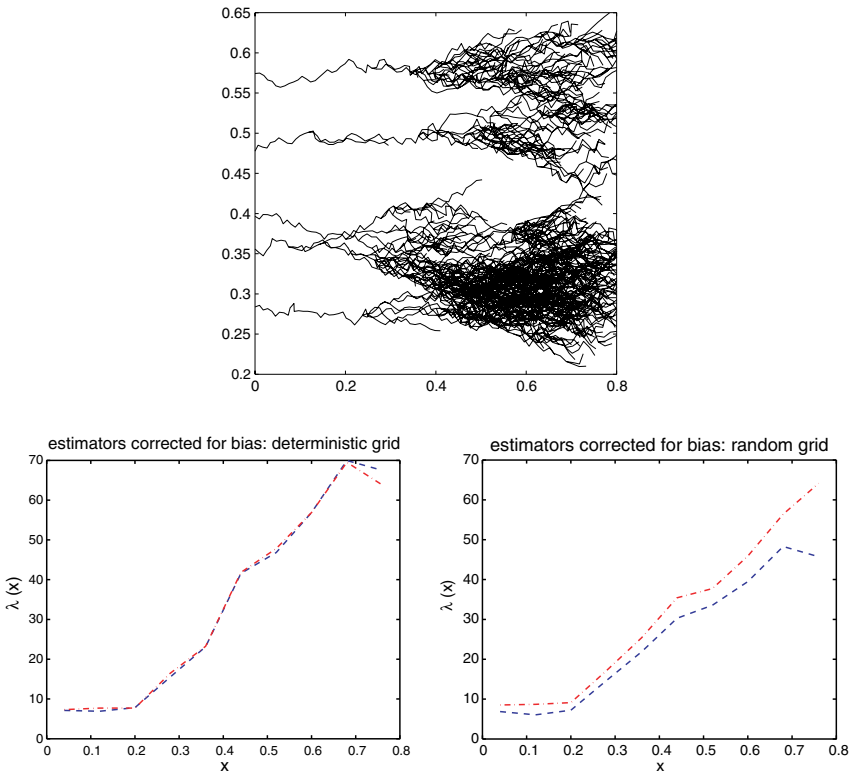


Fig. 15. Estimate of the fibre intensity of an angiogenetic process with a chemotactic field having a gradient in the x direction. *Bottom left*: estimate with a deterministic grid; *bottom right*: estimate with a random grid of 2,000 points. The estimators have been corrected for bias. *Dashed line* $= \hat{\lambda}^1_{k,r,p}$, *dotted-dashed line* $= \hat{\lambda}^2_{k,r,p}$

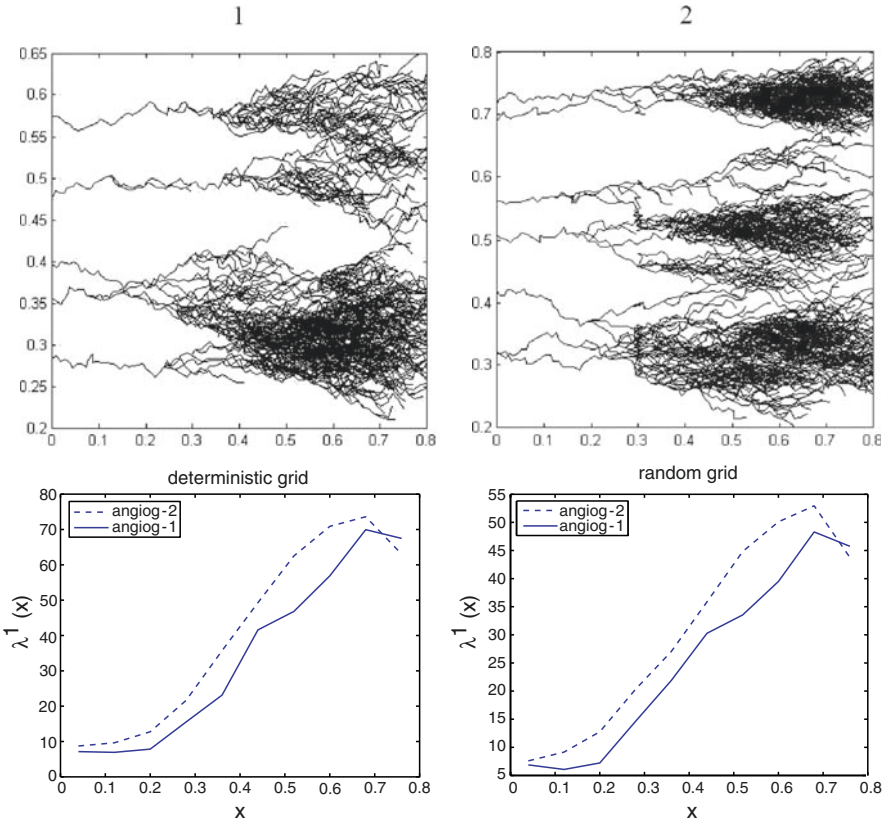


Fig. 16. Comparison between the two simulated angiogenetic processes depicted in the *top line*. *Bottom left*: comparisons of $\hat{\lambda}^1_{k,r,p}$ for the two processes estimated with a deterministic grid; *bottom right*: comparisons of $\hat{\lambda}^1_{k,r,p}$ for the two processes estimated with a random grid. In both cases the first process reveals an intensity lower than the second, and this was really the case in the performed simulation

intensities of branching where different. The difference is not much evident by simply looking at the patterns, but the estimate of the intensity reveals that the pattern on the left has a lower intensity than the pattern on the right for any value of x, and this was really the case, since the frequency of branching and speed of growth was settled higher in the right-hand pattern. This is thus an example where quantitative analysis is essential for the characterization and differentiation of the geometry.

In Fig. 17 an analogous process but driven by a chemotactic field with a spherical symmetry around a point-shaped tumour is reported. Because of the observed symmetry, in this case the window of observation has been divided

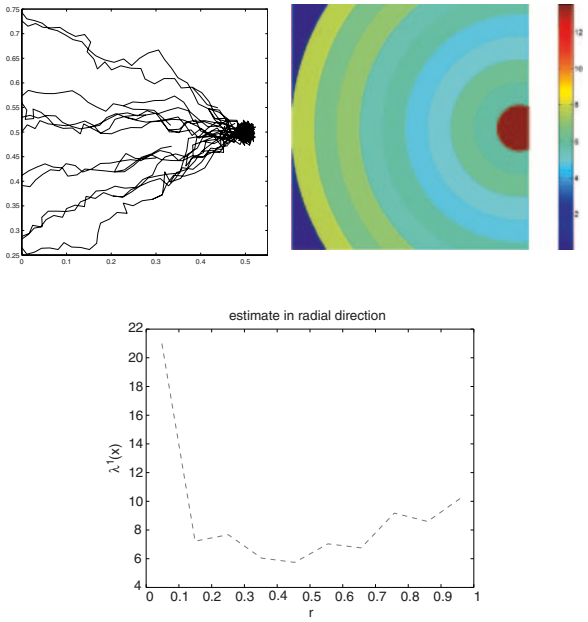


Fig. 17. Estimate of the fibre intensity of an angiogenetic process driven by a chemotactic field with a spherical symmetry. *Top line*: the fibre process and an estimate of $\hat{\lambda}^1_{k,r,p}$ using a random grid and dividing the window into ten spherical shells centred at the tumour; *bottom*: plot of $\hat{\lambda}^1_{k,r,p}$ with respect to the radial coordinate, centred at the tumour

into 10 spherical shells centred at the tumour location. Both the estimated values in each subregion in a 2D visualization and the plot of the estimated intensity with respect to the radial coordinate are reported. In this case, since the subwindows are not rectangular, only estimator $\hat{\lambda}^1_{k,r,p}$ has been computed (see Remark 6).

In Fig. 18 an estimator $\hat{\lambda}^1_{k,r,p}$ has been computed on three images of a vascular networks generated in allantoids (see [CM05] for a discussion of the relevance of these studies in tumour treatment). Two of the three allantoids have been treated with two different doses of an antiangiogenic substance, which should inhibit the formation of vessels. The figure on the left refers to an untreated control allantoid. Because of the spherical symmetry of the images, also in this case the observation window has been divided into spherical shells centred at the centroid of the allantoid. The results of the estimate reveal, in a quantitative way, that the increase of the dose of the substance results in a less widespread network and in a lower intensity of length of the vessels.

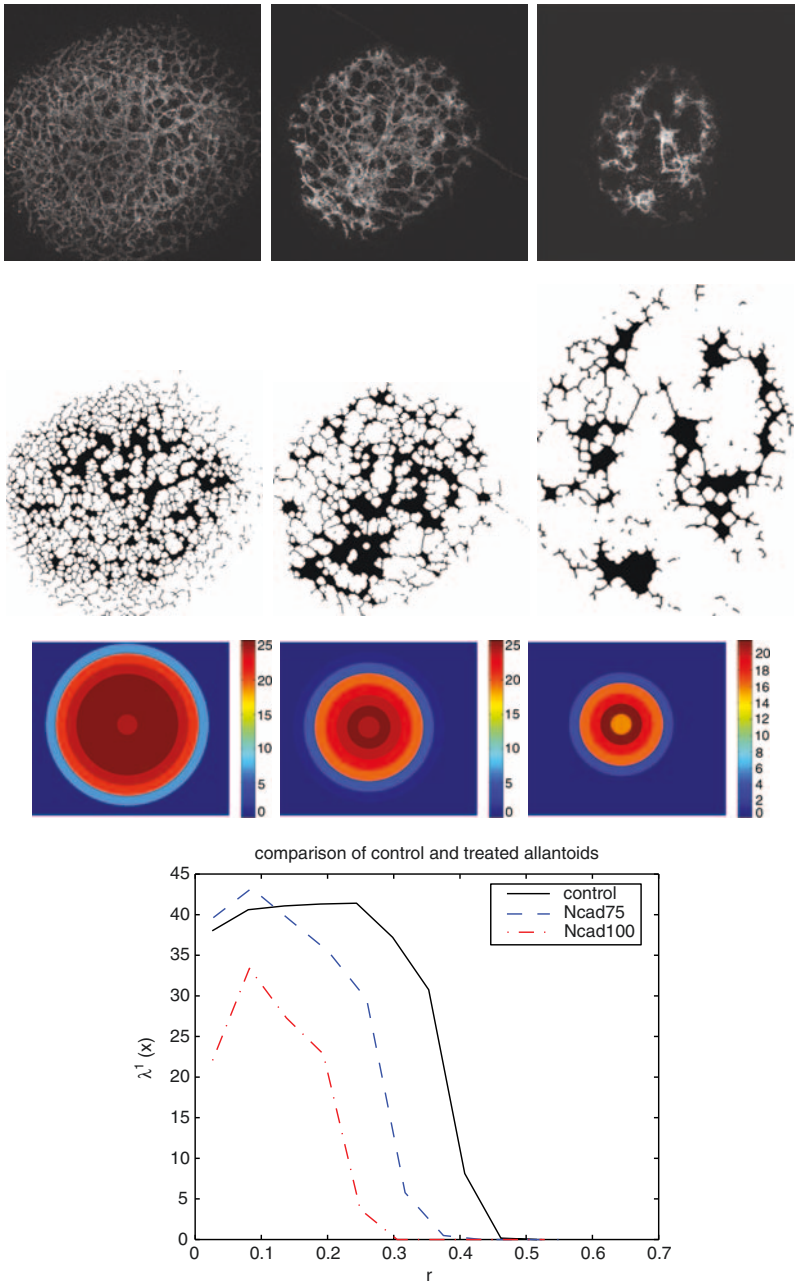


Fig. 18. Vascularization in allantoids. *First line*, from *left to right*: control experiment (untreated), treated with 0.75 mg of antiangiogenetic substance, treated with 1 mg of antiangiogenic substance. *Second line*: scheletonization of the upper images. *Third line*: 2D representation of the intensity estimate of the fibres in the skeletons; the space has been divided into ten spherical concentric shells. *Bottom line*: comparison of the radial estimates of the intensities of the three allantoids

Acknowledgements

It is a pleasure to acknowledge useful discussions with many Colleagues, at different Universities and Research Centres. Particular thanks are due to Professor Luigi Ambrosio (Scuola Normale Superiore, Pisa), Professor Elisabetta Dejana (IFOM, Milan), and Dr. Elena Villa (Department of Mathematics, University of Milan).

References

[ACaV06] Ambrosio, L., Capasso, V., Villa, E.: On the approximation of geometric densities of random closed sets. RICAM Report N. 2006–14, 2006.

[ACoV06] Ambrosio, L., Colesanti, A., Villa, E.: First order Steiner formulas for some classes of closed sets. An application to stochastic geometry. In preparation, 2006.

[AFP00] Ambrosio, L., Fusco, N., Pallara, D.: Functions of Bounded Variation and Free Discontinuity Problems. Clarendon Press, Oxford (2000)

[AG80] Araujo, A., Giné, E.: The Central Limit Theorem for Real and Banach Valued Random Variables. John Wiley & Sons, New York (1980)

[BM97] Baddeley, A.J., Molchanov, I.S.: On the expected measure of a random set. In: Proceedings of the International Symposium on Advances in Theory and Applications of Random Sets (Fontainebleau, 1996). World Sci. Publishing, River Edge, NJ, **1**, 3–20 (1997)

[BR04] Benes, V., Rataj, J.: Stochastic Geometry: Selected Topics. Kluwer Academic Publishers, Norwell (2004).

[Bosq00] Bosq, D.: Linear Processes in Function Spaces. Theory and Applications. Lecture Notes in Statistics 149. Springer-Verlag, New York (2000).

[Ca03] Capasso, V. (ed): Mathematical Modelling for Polymer Processing. Polymerization, Crystallization, Manufacturing. Mathematics in Industry Vol 2. Springer Verlag, Heidelberg (2003)

[CM00a] Capasso, V., Micheletti, A.: The local mean volume and surface densities for inhomogeneous random sets. Rend Circ. Mat. Palermo Suppl., **65**, 49–66 (2000).

[CM00b] Capasso, V., Micheletti, A.: Local spherical contact distribution function and local mean densities for inhomogeneous random sets. Stochastics and Stoch. Rep., **71**, 51–67, (2000).

[CM05] Capasso, V., Micheletti, A.: Stochastic geometry and related statistical problems in biomedicine. In: A. Quarteroni et al. (Eds), Complex Systems in Biomedicine. Springer, Milano (2005).

[CM06] Capasso, V., Micheletti, A., Kernel-like estimators of the intensity of inhomogeneous fibre processes. Preprint (2006).

[CT94] Capasso, V., Tayler, A. (Eds): ECMI Brochure, Bari (1994)

[CV06a] Capasso, V., Villa, E.: On the continuity and absolute continuity of random closed sets. Stoch. An. Appl. **24**, 381–397, (2006)

[CV06b] Capasso, V., Villa, E.: Some remarks on the continuity of random closed sets. In: Lechnerová R., Saxl I., Beneš V. (eds), Proceedings of the Intenational Conference in Stereology, Spatial Statistics and Stochastic Geometry, UCMP, Prague, 69–74, (2006)

[CV06c] Capasso, V., Villa, E.: On the geometric densities of random closed sets. RICAM Report 13/2006, Linz, (2006).

[CV06d] Capasso, V., Villa, E.: On mean densities of inhomogeneous geometric processes arising in material science and medicine. Preprint (2006)

[CA99] Chaplain, M.A.J., Anderson, A.R.A.: Modelling the growth and form of capillary networks. In: Chaplain, M.A.J. et al. (eds) On Growth and Form. Spatio-temporal Pattern Formation in Biology. John Wiley & Sons, Chichester (1999).

[Fal85] Falconer, K.J.: The Geometry of Fractal Sets. Cambridge University Press, Cambridge (1985).

[Fed96] Federer, H.: Geometric Measure Theory. Springer, Berlin (1996)

[FC98] Friedman, L.H., Chrzan, D.G.: Scaling theory of the Hall-Petch relation for multilayers. Phys. Rev. Letters, **81**, 2715-2718 (1998).

[Hard91] Hardle, W.: Smoothing techniques. With Implementation in S, Springer-Verlag, New York (1991).

[KF70] Kolmogorov, A.N., Fomin S.V.: Introductory Real Analysis. Prentice-Hall, Englewood Cliffs (N.J.), (1970).

[JK01] Jain, R.K., Carmeliet, P.F.: Vessels of Death or Life. Scientific American **285**, 38-45 (2001).

[Jones82] Jones, D.S.: The Theory of Generalised Functions. Cambridge University Press, Cambridge (1982)

[Mat75] Matheron, G.: Random sets and integral geometry. John Wiley & Sons, New York (1975)

[McDou06] McDougall, S.R., Anderson, A.R.A., Chaplain, M.A.J.: Mathematical modelling of dynamic tumour-induced angiogenesis: Clinical implications and therapeutic targeting strategies. J. Theor. Biology, **241**, 564–589 (2006)

[Mol92] Møller, J.: Random Johnson-Mehl tessellations. Adv. Appl. Prob., **24**, 814–844 (1992)

[Mol94] Møller, J.: Lectures on Random Voronoi Tessellations. Lecture Notes in Statistics 87. Springer-Verlag, New York, Berlin, Heidelberg (1994)

[Pest98] Pestman, W.R.: Mathematical Statistics. An Introduction. Walter de Gruyter, Berlin (1998).

[SKM95] Stoyan, D., Kendall, W.S., Mecke, J.: Stochastic Geometry and its Application, John Wiley & Sons, New York (1995)

[Vlad79] Vladimirov, V.S.: Generalized Functions in Mathematical Physics. Mir Publishers, Moscow (1979).

[Z82] Zähle, M.: Random processes of Hausdorff rectifiable closed sets. Math. Nachr. **108**, 49–72 (1982)

Synthesis of Micro and Nanoparticles from Coaxial Electrified Jets

A. Barrero[1] and I.G. Loscertales[2]

[1] Escuela Técnica Superior de Ingenieros, Universidad de Sevilla, 41092 Sevilla, Spain
[2] Escuela Técnica Superior de Ingenieros Industriales, Universidad de Málaga, 29013 Málaga, Spain

Summary. The use of electrohydrodynamic (EHD) forces to generate highly charged coaxial jets of immiscible fluids, with diameters in the micro and nanoregime, has unravel itself as a quite interesting choice for producing complex nanostructures from a vast variety of precursors, provided they can solidify, polymerize or gel, in times comparable or shorter than the living time of the coaxial nanojet. For time ratios larger than one, the result of the process are micro or nanocapsules, while for time ratios smaller than one coaxial nanofibres are produced. We show examples of both situations, with organic and inorganic precursors. On the other hand, realization of the process in a liquid bath opens the door to production of controlled micro and nanosized complex emulsions.

1 Introduction

It is well known that the physical properties of a piece of a given substance (thermal and electrical conductivity, strength, toughness, etc.) depend not only on the substance itself but also on its characteristic size. In effect, let us consider an ideal experiment consisting of a material piece whose characteristic length L can be shortened in a controlled way by an external observer. The observer would find out that the values of the physical properties of the material piece undergone a dramatic change when L reaches values sufficiently small. The explanation for such an anomalous behaviour lays on the fact that the surface of a piece of matter decreases with L much more slowly than its volume does and, contrarily to what happens in our familiar macroscopic world, the atomic and molecular interactions of the surface becomes dominant compared to those in the volume once the nanoscopic limit is reached. The length at which the change of properties takes place is, roughly speaking, of the order of 100 nm so this length may be thought as the boundary below which nanotechnology and nanoscience apply. Therefore, its application domain ranges from isolated atoms/molecules to bulk materials, where length

and timescales of the phenomena become comparable to those of the structure. Nanotechnology implies the ability to generate and to use structures, components, and devices with a size range from about 0.1 nm (atomic and molecular scale) to about 100 nm (or larger in some situations) by control at atomic, molecular, and macromolecular levels. Nanotechnology is a major breakthrough that will yield new tools for fundamental discoveries with broad impact on technology, materials, biomedical, energy, and environment. Moreover, their interdisciplinary character allows for unparalleled synergy between previously unrelated fields and therefore their applications are extremely diverse. Some few examples of potential applications that are being actively investigated are: advanced drug delivery via nanoparticles in medicine and pharmaceutics fields; chemical and biodetectors for security and other civilian uses; nanostructured catalysts in chemical and fuel industries; metallic and ceramic nanostructured materials with engineered properties, molecular manipulation of polymeric macromolecules, and nanostructured coatings, among others, in material science; nanofabrication of electronic products in electronics, etc. Commercially viable technologies are already available for some ceramic, metallic, and polymeric nanoparticles, nanostructured alloys, colorants and cosmetics, tissue engineering, electronic components such as those for media recording, and hard-disk reading, to name a few. In biomedicine, tissue engineering, for example, applies to regeneration of bones, arteries, and other organs by using biocompatible polymers: polycaprolactone (PCL) and polylactide-co-glycolide acid (PLGA). Basically, it is based on the fact that cells get together and rearrange faster around fibres with smaller diameters (500 nm) than the cells. These scaffolds made of woven fibres, which have proved to be a very efficient growing environment, are being used as biocompatible films to cover prostheses to avoid rejection. This stimulating tissue growth also applies in the cicatrization of wounds and burns. Another example of synergy between nanotechnology and medicine is the use of nanoparticles in drug delivery. The technique involves binding a therapeutic compound to a nanoparticle, or encapsulating it within a nanoshell. A key advantage of nanoshells is that they can be targeted to specific cell populations through conjugation with a monoclonal antibody. When the nanoshells reach the target site, their therapeutic contents are released by breaking them using a low intensity light source such as a laser; shells with controlled porous wall could be also used for the appropriate outflow of the drug. Drug delivery using nanoparticles provides high target specificity, with high potential for treatment of localized neurological disorders and cancer with therapeutic compounds which have side effects in the rest of the body. An alternative, noninvasive procedure for tumour ablation, which has been tested in mice, consists in the intravenous injection of nanoparticles with a dielectric core coated by a thin gold shell, Loo (2005). Based on the relative dimensions of the shell thickness and core radius (typical diameter of the shell is in 100 nm range), nanoshells may be designed to scatter and/or absorb light over a broad spectral range including the near-infrared (NIR), a wavelength region that provides maximal penetration

of light through tissue. Immunotargeted nanoshells are engineered to absorb light, allowing selective destruction of targeted carcinoma cells through photothermal therapy. Production of micrometer- or even nanometer-sized particles and fibres can be tackled from two different approaches: bottom-up and top-down methods. Bottom-up refers to methods where materials and devices are built from molecular components which assemble themselves chemically using principles of molecular recognition. Bottom-up should broadly speaking be able to produce devices in parallel and much cheaper than top-down methods, but getting control over the methods is difficult when nanostructures become larger and more bulky than what is normally made by chemical synthesis. On the contrary, in top down methods, nanoobjects are obtained from the appropriate splitting of much larger physical systems without atomic level control along the process.

Top-down methods to produce micro- and nanoparticles require the division of a macroscopic (i.e. millimetric) piece of matter, generally a liquid, into tiny offsprings of micro- or nanometric size. Surface tension strongly opposes the huge increase of area inherent to this dividing process. Thus, to produce such small particles, energy must be properly supplied to the interface. This energy is the result of a mechanical work done on the interface by any external force field, i.e. hydrodynamic forces, electrical forces, etc. Two kinds of approaches can be distinguished, depending on how the energy is supplied. In one approach, such as in the mechanical emulsification techniques, the force fields (extensional and shear flows) employed to break up the interface between two immiscible fluids are so inhomogeneous that, in general, the offspring droplets present a very broad size distribution. Nevertheless, a good degree of monodispersity might be achieved for a particular combination of the emulsification parameters (shear rate, rotation speeds, temperature, etc.) and a given combination of substances. However, such a desirable condition might not exist if one of the substances is changed, if a new one is added, or if a different size is desired. The same occurs if capsules must be formed. Furthermore, in many instances, the formation of the structure depends on chemical interactions, usually preventing the process from being applicable to a broad combination of substances.

In the other approach, which has the advantage of being based on purely physical mechanisms, the force field stretches, steadily and smoothly, the fluid interface without breaking it until at least one of its radii of curvature reaches a well-defined micro or nanoscopic dimension d; at this point, the spontaneous break up of the stretched interface by capillary instabilities yields monodisperse particles with a size of the order of d, Barrero and Loscertales (2007). These types of flows are known as capillary flows due to the paramount role of the surface tension. For example, the formation and control of single and coaxial jets with diameters in the micrometer/nanometer range, and their eventual varicose breakup, lead to particles without structure (single jets) or compound droplets (coaxial jets), with the outer liquid encapsulating the inner one. On the other hand, if the liquid solidifies before the jet breaks, one obtains fibres

(single jet) or coaxial nanofibres or hollow nanofibres (coaxial jets). The mean size of the particles obtained with these methods ranges from hundreds of micrometers to several nanometers, although the nanometric range is generally reached when electric fields are employed. The particles obtained using this approach are, in general, nearly monodisperse and its employment enables, in the case of capsules, a precise tailoring of both the capsule size and the shell thickness.

2 Capillary Flows Driven by Electrical Forces

2.1 Electrospray

The interaction of an intense electrical field with the interface between a conducting liquid and a dielectric medium has been known to exist since William Gilbert (1600) reported the formation of a conical meniscus when an electrified piece of amber was brought close enough to a water drop. The deformation of the interface is caused by the force that the electrical field exerts on the net surface charge induced by the field itself. Experiments show that the interface reaches a motionless shape if the field strength is below a critical value, whereas for stronger fields the interface becomes conical, issuing mass and charge from the cone tip in the form of a thin jet of diameter d. In the latter case, the jet becomes steady if the mass and charge it emits are supplied to the meniscus at the same rate. Taylor (1964) explained the conical shape of the meniscus as a balance between electrostatic and surface tension stresses; since then the conical meniscus has been referred to as the Taylor cone. The thin jet eventually breaks up into a stream of highly charged droplets with a diameter of the order of d. This electrohydrodynamic (EHD) steady-state process is the so-called steady cone-jet electrospray after Cloupeau and Prunet-Foch (1989), or just electrospray, see Fig. 1, Pantano et al. (1994).

The electrospray has been applied for bioanalysis (Fenn et al. 1989), fine coatings (Siefert 1984), synthesis of powders (Rullison and Flagan 1994), and electrical propulsion (Martinez-Sanchez et al. 1999), among other technological applications. Recently, the electrosprays in cone–jet mode were also stabilized inside dielectric liquid baths, Barrero et al. 2004; hence, the technique could be applied to the production of simple and double emulsions of the type water in oil, oil in water, and oil–water–oil.

Although the equations (Navier–Stokes and Maxwell equations) and boundary conditions governing the electrospray are known, the numerical simulation of the electrospray is quite complex due to (a) the disparity of length scales between the diameter of the jet and the needle diameter, or aperture, through which liquid is being injected, which can vary more than three orders of magnitude, (b) the existence of one (or more) free surface that must be consistently determined as part of the solution of the problem, and (c) the fact that the region where the interface breaks is time dependent in

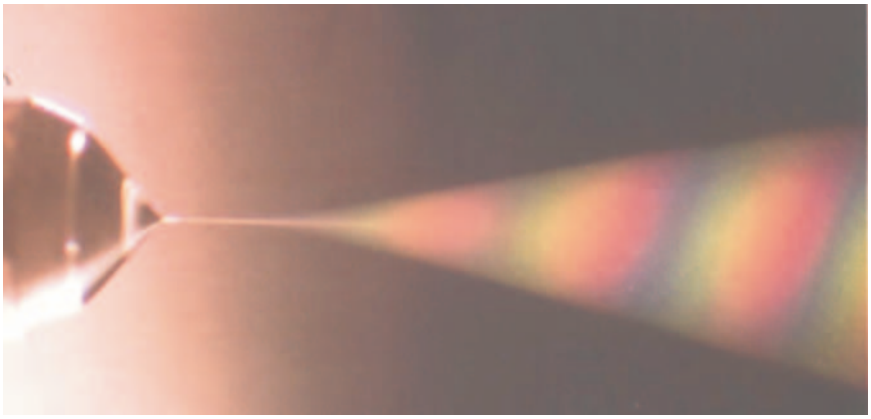


Fig. 1. Cone, jet, and spray in an electrospray; the electrosprayed liquid was methanol. The size of the charged droplets ranged between 380 and 720 nm, which are the wavelength of the blue and red radiation. As shown in the picture, droplets scatter the blue component avoiding its pass throughout the spray while the other components of the white light pass through the droplet cloud

spite of the steady character of the flow upstream of the breaking zone. For these reasons most works on electrospray have focused on experiments, which under the guide of the dimensional analysis have provided the widely accepted relationship between the current I and the flow rate q transported through the jet, Fernández de la Mora and Loscertales (1994),

$$\frac{I}{I_0} = g(\beta)\left(\frac{q}{q_0}\right)^{1/2} \quad \text{with} \quad I_0 = \left(\frac{\epsilon_0}{\rho}\right)^{1/2} \quad q_0 = \frac{\gamma\epsilon_0}{\rho K}, \tag{1}$$

where γ is the surface tension, ρ, K, and $\beta\epsilon_o$ are density, electrical conductivity and permittivity of the liquid respectively, ϵ_o is the vacuum permittivity, and $g(\beta) \sim \beta^{-1/4}$ is a dimensionless function that has been experimentally determined (Gañán-Calvo et al. 1997). However, the scaling law for the jet diameter d is still controversial because experimental errors in the reported measurements of the mean droplet diameter do not allow one to distinguish between the different proposed size laws. The scaling size laws that appear most frequently in the literature can be cast in the form

$$\frac{d}{d_0} = f(\beta)\left(\frac{q}{q_0}\right)^{n} \quad \text{with} \quad d_0 = \frac{\gamma\epsilon_0^2}{\rho K^2}, \tag{2}$$

where $f(\beta)$ is a dimensionless function of order of unity and exponent n takes the values 1/3, 1/2, and 2/3 depending on the authors. For electrosprays, the minimum flow rate at which it can operate in steady-state conditions is approximately given by $q_{\min} \sim q_0$, which for liquids with electrical conductivities of the order of $1\,\mathrm{S\,m^{-1}}$, the minimum jet diameter becomes of the order of a few nanometers.

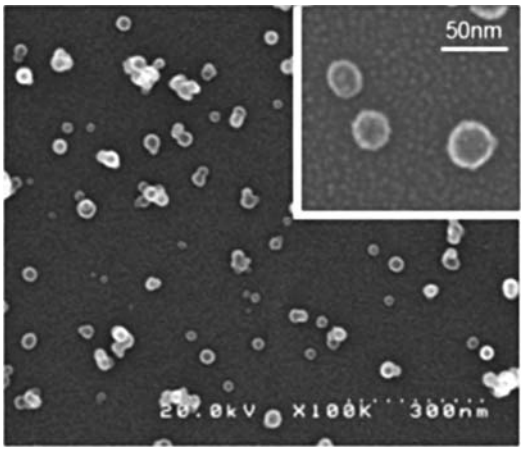


Fig. 2. Nonagglomerated spherical titanium oxide nanoparticles were prepared using an electrospray assisted chemical vapor deposition (ES-CVD) process. From Nakaso et al. (2003)

Numerical simulation of the cone jet electrospray has been considered in a recent paper, Higuera (2003); details of the equations and boundary conditions can be found there. To avoid the numerical difficulty of dealing with two highly disparate length scales, which appears in the case of liquids of relatively high electrical conductivity. Higuera did not consider the full problem from the needle to the final jet region (before breakup) but the cone-to-jet transition region and used the cone and the jet as asymptotic boundary conditions. The numerical analysis included the effect of the liquid viscosity, which had been neglected in prior experiments, and he approximately recovered the $I \sim Q^{1/2}$ law. An excellent review on the physics of electrosprays may be found in Fernandez de la Mora (2007).

The electrospray technique has proved its ability for the production of single nanoparticles; the ones shown in Fig. 2 are an example.

2.2 Electrospinning

The EHD flow described above can be also used to obtain very thin fibres if the jet solidifies before breaking into charged droplets. This process, known as electrospinning, occurs when the working fluid is a complex fluid, such as the melt of polymers of high molecular weight dissolved in volatile solvent, Doshi and Reneker (1995), Fridrikh et al. (2003). The rheological properties of these melts, sometimes enhanced by the solvent evaporation from the jet, slowdown, and even prevent the growth of varicose instabilities. As is well known, large values of liquid viscosity delay the jet breakup by reducing the growth rate of axisymmetric perturbations, so longer jets may be obtained. However, nonsymmetric perturbation modes can grow due to the net charge

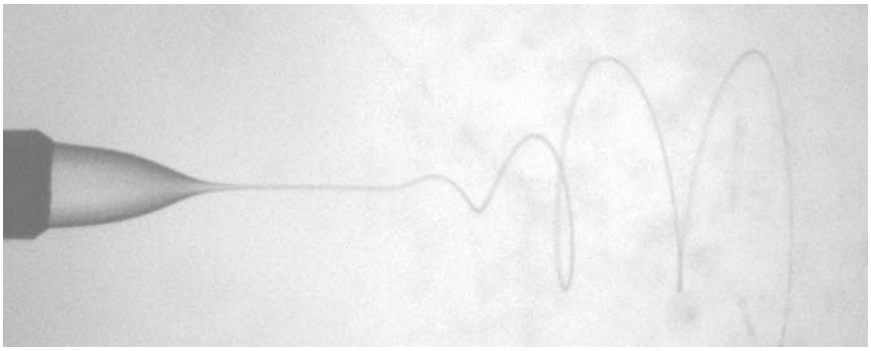


Fig. 3. Whipping instability in an electrified jet of glycerine in a bath of hexane. Courtesy of Mr. A. Gomez-Marín

carried by the jet. Indeed, if a small portion of the charged jet moves slightly off axis, the charge distributed along the rest of the jet will push that portion farther away from the axis, thus leading to a lateral instability known as whipping or bending instability. A picture capturing the development of the whipping instability in a jet of glycerine in a hexane bath is shown in Fig. 3.

The chaotic movement of the jet under this instability gives rise to very large tensile stresses, which lead to a dramatic jet thinning. The solidification process, and thus the production of micro- or nanofibres, is enhanced by the spectacular increase of the solvent evaporation rate due to the thinning process. This technique is very competitive to produce nanofibres as compared with other existing ones (i.e. phase separation, self-assembly, and template synthesis, among others), and it is therefore the subject of intense research.

2.3 Electrified Coaxial Jets

A new technique, which also uses EHD forces to generate coaxial jets of immiscible liquids, with diameters in the nanometer range, has been recently reported, Loscertales et al. 2002. The method is being used to synthesize nanoparticles with core-shell structure. Basically, the technique consists of the injection at appropriate flow rates of two immiscible liquids through two concentrically located needles. The inner diameter of the inner needle ranges from the order of 1 mm to tens of micrometers, whereas its outer diameter sets limits to the cross-section of the outer needle.

The outer needle is connected to an electrical potential of several kilovolts relative to a ground electrode. The inner needle is kept to an electrical potential that, depending on the conductivity of the outer liquid, can be varied from that of the outer needle to that of the extractor. For a certain range of values of the electrical potential and flow rates, a structured Taylor cone

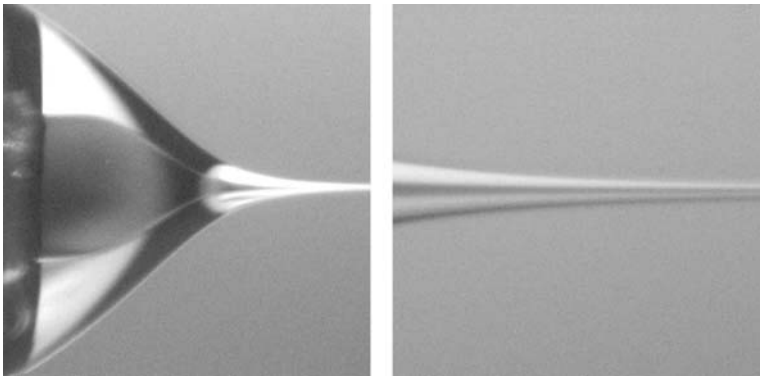


Fig. 4. Picture on the *left* shows a structured liquid Taylor cone; a downstream detail of the two coaxial jets emitted from the vertexes of the two menisci is given in picture on the *right*

is formed at the exit of the needles with an inner meniscus surrounded the inner one, see picture on the left in Fig. 4. A liquid thread is issued from the vertex of each one of the two menisci, giving rise to a compound jet of two coflowing liquids see picture on the right (Fig. 4). At the minimum jet section, the two-layered jet has an outer diameter of 4 μm.

To obtain this compound Taylor cone, at least one of the two liquids must be sufficiently conductive. Similarly to simple electrosprays, the electrical field pulls the induced net electric charge located at the interface between the conducting liquid and a dielectric medium and sets this interface into motion; because this interface drags the bulk fluids, it may be called the driving interface. The driving interface may be either the outermost or the innermost one; the latter happens when the outer liquid is a dielectric. When the driving interface is the outermost, it induces a motion in the outer liquid that drags the liquid–liquid interface. When the drag overcomes the liquid–liquid interfacial tension, a steady-state coaxial jet may be formed. On the other hand, when the driving interface is the innermost, its motion is simultaneously diffused to both liquids by viscosity, setting both in motion to form the coaxial jet. Scaling laws showing the effect of the flow rates of both liquids on the current transported by these coaxial jets and on the size of the compound droplets were recently investigated (Lopez-Herrera et al. 2003).

3 Core-Shell Nanoparticles

3.1 Nanocapsules and Hollow Nanospheres

The last technique has been applied, upon coaxial jet breakup, to microencapsulate aqueous solutions. An outer jet of Somos 6120, a Du Pont photopolymer

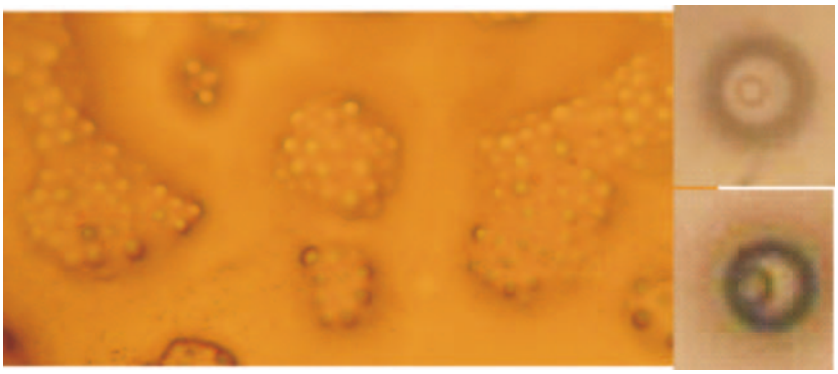


Fig. 5. Collection of near monodisperse capsules. Magnified views of two capsules formed under different parametrical conditions are also given in the two pictures on the *right*. In the *upper* one picture, the outer diameter is 10 µm, whereas the diameter of the capsule shown in the *lower* one is 8 µm

and a coflowing water inner jet were generated as described before. Compound droplets of water coated by Somos resulted from the jet breakup, so that a spray of compound droplets was formed and collected on a plate damped with water. In this case, the outer shell of the droplets was hardened with an ultraviolet light reactor. Before the hardening process, the charged aerosol was neutralized by corona discharge, so that losses were minimized. The liquid flow rates in this experiment were selected to obtain capsules in the micrometer range, because capsules in this range can be optically recorded to allow for visual observation, Fig. 5. Capsules of olive oil surrounding water of 150 nm of mean diameter have been also obtained with this technique. Some examples of applications of this approach to produce capsules include the encapsulation of water-based flavours within oil-based substances, and the opposite (oil-based flavours within water-based polymers) for food enrichment applications, Bocanegra et al. (2005).

Also, combination with sol–gel chemistry has proven fruitful, Larsen et al. (2003). In this case, the outer liquid was a sol–gel formulation, while the inner one was a regular nonstructured or "regular" liquid (like oil, water, glycerine, etc.). By adjusting the sol properties and the operating parameters, we have been able of producing hollow spheres, with mean diameters ranging from 10 µm down to 0.4 µm, and with shell thickness between 1 µm and less than 50 nm. Some of these results are shown in Fig. 6. Although the capsules were initially filled with the "regular" liquid, since the polymerization or gel transition forms porous solids, the inner liquid was easily solvent-extracted, so that after solvent evaporation a void cavity was left. In any of the above cases, the time of flight of the liquid capsules (that is the time from their formation up to their collection on a collector) was controlled to allow for either phase transition or polymerization (gelation). This can be easily done by reducing

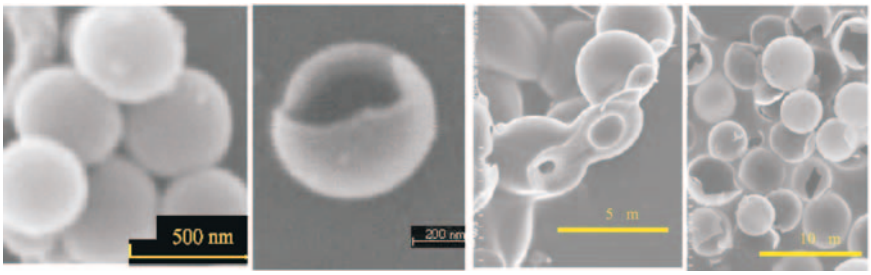


Fig. 6. Hollow spheres of SiO_2. Both diameter and shell thickness can be controlled by adjusting the flow rates

the charge level on the freshly formed capsules; this was accomplished in our lab by setting up a corona discharge of opposite polarity in the surrounding atmosphere.

3.2 Hollow Nanofibres and Coaxial Nanofibres

Another recent application of this technique relates to the production of nanofibres, compound nanofibres and hollow nanotubes. There are many procedures to build nanotubes of different materials, other than the popular carbon nanotubes. In general, the vast majority of these procedures resort to templates, Cepak and Martín (1999). A solid nanotemplate (i.e. a nanofibre or a pore membrane) is formed, around which nanotubes are grown. This growth usually happens in liquid phase, and it resorts to self-assembly of the proper molecules onto the surface of the template. The first complexity is due to this self-assembly process, which unfortunately appears to be very chemistry dependant. Usually, the recipe that works for one particular precursor does not work for another, even for very similar molecules. Once the shell is built around the template, still the template itself must be removed. This is typically done by degrading or decomposing the template thermically, or chemically, etc. This necessarily requires the shell to be more "resistant" than the template. In brief, the procedure is a multistep process, apart of the restriction imposed by the chemistry. One of the advantages of using compounds nanojets to produce nanotubes resorts on the fact that self-assembly is not a limiting step since the shape of the jet itself already constrain the material to the proper cylindrical shape. But during the same process, the inner liquid, which is also stretched to a cylindrical shape, plays the role of the template, thus limiting the inner surface of the nanotube. Furthermore, the template is not solid, but liquid, so that removing the template is much easier and much less energy consuming. Therefore, if solidification (or polymerization, or gelation) of the outer liquid occurs prior to the jet break up, then the nanotube is form in just one step, Loscertales et al. (2004).

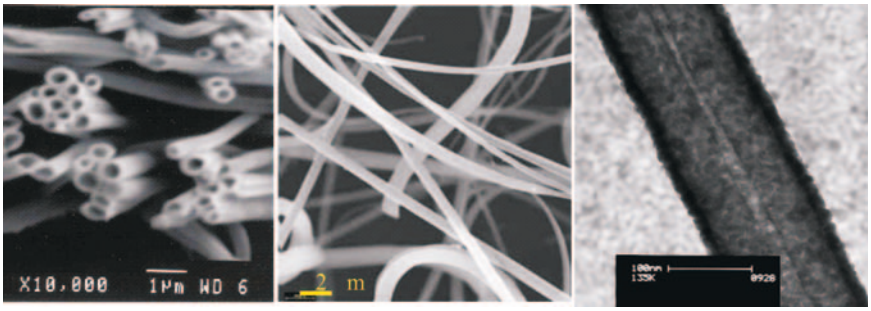


Fig. 7. Examples of hollow nanotubes and compound nanofibres

Figure 7 shows some examples of hollow nanotubes and compound nanofibres produced from electrified coaxial jets. Pictures have been taken with a scanning electron microscope, except the right one, which was taken with transmission electron microscope. The picture on the left shows nanotubes of SiO_2, with diameters of the order of 500 nm, and shell thickness of the order of 70 nm. The one in the middle shows ZnO_2 nanotubes, with diameters from 1 μm down to 400 nm; the wall thickness was of the order of 80 nm. Finally, the right one shows a coaxial character compound nanofibre of poly-ethyleneoxide (PEO) on the outside, and stained PEO in the inside. The outer and inner diameters are of 100 and 15 nm, respectively.

3.3 Simple and Double Emulsions

Finally, another extension of the EHD atomization is that when the surrounding atmosphere is not a gas nor vacuum, but a liquid insulator, Barrero et al. (2004). The same atomization process is possible within a liquid, which opens up the possibility of producing monodisperse micro- and nanoemulsions, Marín et al. (2007). Although work is still on its way, we have investigated the scaling laws for both the current and the size of the droplets. In this new situation, the role of surfactants, emulsifiers, and polymers in solution may be essential in order to stabilize such nanoemulsions. On top of that, the process may be executed with a compound Taylor cone instead (Fig. 8), so that double emulsions of nanometric size can be directly formed, still with a well controlled mean size and small size dispersion. Finally, the charged nature of the dispersed phase can be an advantage to control their trajectories and to select where to deposit them; this could be used to generate well controlled layers of nanoparticles on top of macroscopic objects to emulate colloidosomes, see for instance (Dinsmore et al. 2002).

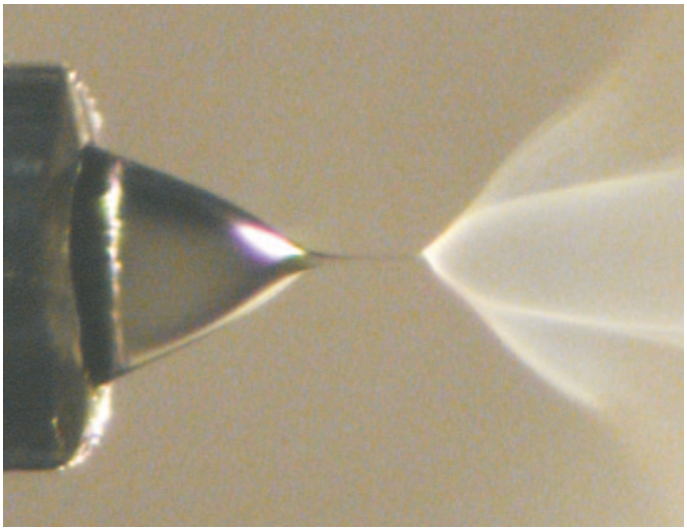


Fig. 8. Taylor cone of glycerol in a bath of hexane. The needle OD is 0.8 mm. The hydrosol in this case is formed by droplets of two different sizes: the main droplets, of 2 µm in diameter, and the satellite droplets, of about 0.8 µm in diameter

4 Conclusions

Some topdown methods to produce micro- and nanoparticles require one to divide a macroscopic (i.e. millimetric) piece of liquid into tiny offsprings of micro- or nanometric size. One of them uses electrical forces to generate coaxial jets with a diameter in the micro- and nanometric size ranges. Micro- or nanocapsules are formed upon jet breakup, whereas if the jet solidifies coaxial nanofibres or hollow nanofibres are obtained. This method can produce micro- and nanoparticles with or without inner structure. Generally, it enables both a precise control of the particle size and a narrow size distribution, which makes it attractive and competitive with other existing techniques. A noticeable feature of the method lies in the fact that the core-shell particles may be obtained in just one step; this is a clear advantage over multistep processes such as the emulsification techniques. However, the throughput of this EHD method is usually too small for many industrial purposes, restricting their use to some analytical applications. Increasing the production rate requires the operation of parallel devices. The main problems when trying to operate in parallel come from the shielding effect of the space charge created by the highly charged aerosol and from the electric crosstalk between neighbouring devices. Accordingly, the design of efficient approaches for operating parallel devices will probably become a very active area of research in the near future.

Acknowledgements

This work has been founded by the Spanish Ministry of Science and Technology under contract BFM2001-3860-C02-01 and by Yflow SL. The help provided by M. Lallave (Yflow), J.E. Díaz (Yflow) and D. Galán (Yflow), and A. Gómez (University of Seville) is also acknowledged.

References

1. Barrero A. and Loscertales I.G. *Annual Rev. Fluid Mech.* **39**, 89-106, 2007.
2. Barrero A., López-Herrera J., D. Boucard A. and Loscertales I.G. *J. Colloid Interf. Sci.* **272**, 104-108, 2004.
3. Bocanegra R., Gaonkar A.G., Barrero A., Loscertales I.G., Pechack D., Marquez M., *J. Food Sci.* **70** 492-497, 2005.
4. Cepak V.M. and Martín, C.R. *Chem. Mater.* **11**, 1363, 1999.
5. Cloupeau M. and Prunet-Foch B. *J. Electrost.* **22** 135-59, 1989.
6. Dinsmore A.D., Hsu, M.F., Nikolaides M.G., Márquez M., Bausch A.R., and Weitz D.A. *Science* **298**, 1006-1009, (2002).
7. Doshi J. and Reneker D.R. *J. Electrost.* **35**, 151-60, 1995.
8. Fenn J.B., Mann M., Meng C.K., and Wong S.F. *Science* **246**, 64-71 (1989).
9. Fernández de la Mora J. and Loscertales I.G. *J. Fluid Mech.* **260** 155-84, 1994. Fernández de la Mora J. *Annual Review of Fluid Mech.*, **39**, 217-244, 2007.
10. Fridrikh S.V., Yu J.H., Brenner M.P., Rutledge G.C. *Phys. Rev. Lett. 90*, 144502, 2003.
11. Gañán-Calvo A.M., Dávila J., Barrero A. *J. Aerosol Sci.* **28**, 249-75 1997.
12. Gilbert W. De Magnete (1600). Transl. P.F. Mottelay. Dover, UK. (1958)
13. Higuera F.J. *J. Fluid Mech. 484*, 303-327, 2003.
14. Larsen G., Velarde-Ortiz R., Minchow K., Barrero A., Loscertales I.G. *J. Am. Chem. Soc.* **125**, 1154-55, 2003.
15. Loo C., Lowery A., Halas N., West J., and Drezek R. *Nanoletters* **5**, 4, 709-711, (2005).
16. López-Herrera J., Barrero A., López A., Loscertales I.G., Márquez M. *J. Aerosol Sci.* **34**, 535-552, 2003.
17. Loscertales I.G., Barrero A., Guerrero I., Cortijo R., Márquez M. *Science* **295**, 1695-98, 2002.
18. Loscertales I.G., Barrero A , Márquez M., Spretz R., Velarde-Ortiz R., Larsen G. *J. Am. Chem. Soc.* **126**, 5376-77, 2004.
19. Marín A.G., Loscertales I.G., Márquez, M., Barrero A. *Phys. Rev. Lett.* **98**, 014502, 2007.
20. Martínez-Sanchez M., Fernandez de la Mora J., Hruby V., Gamero-Castano M., Khayms, V. *Proc. 26th Int. Electr. Propuls. Conf.*, Kitakyushu, Japan, pp. 93-100. Electr. Rocket Propuls. Soc. 1999.
21. K. Nakaso, B. Han, K.H. Ahn, M. Choi, and K. Okuyama. *J. Aerosol Sci.* **34**, 869-881, 2003.
22. Pantano C., Gañán-Calvo A.M. and Barrero A. *J. Aerosol Sci.* **25**, 1065-77, 1994.
23. Rulison A.J. and Flagan R.C. *J. Am. Ceramic Soc.* **77**, 3244-50, 1994.
24. Siefert W. *Thin Solid Films* **120**, 267-74, 1984.
25. Taylor G.I., *Proc. Royal Soc London* **A280**, 383-397, 1964.

Numerical Simulation of Induction Furnaces for Silicon Purification

A. Bermúdez[1], D. Gómez[1], M.C. Muñiz[1], P. Salgado[2], and R. Vázquez[1]

[1] Departamento de Matemática Aplicada, Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Spain
{mabermud, malola, mcarmen, marafa}@usc.es
[2] Departamento de Matemática Aplicada, EPS Lugo, Universidade de Santiago de Compostela, 27002 Lugo, Spain
mpilar@usc.es

Summary. This paper deals with mathematical modelling and numerical simulation of induction heating furnaces for axisymmetric geometries. The mathematical model presented consists in a coupled thermo-magneto-hydrodynamic problem with phase change. We propose a finite element method and an iterative algorithm to solve the equations. Some numerical results for an industrial furnace used for silicon purification are shown.

1 Introduction

Silicon (Si) is the second most abundant element in the earth crust after oxygen. In natural form, it can be found mainly as silicon dioxide (Silica, SiO_2) and silicates. In particular, quartz and sand are two of the most common forms. Silicon is produced industrially by reduction of silicon dioxide, as quartz or quartzite, with carbon by a reaction which can be written in a simple way as follows:

$$\mathrm{Si\ O_2 + 2C = Si + 2CO.} \tag{1}$$

Silicon has a wide variety of applications depending on its purity. Indeed, silicon is referred to by the approximate percentage of silicon contained in the material and the maximum amount of trace impurities present. Thus, *silicon metal* (or metallurgical grade silicon) refers to the silicon which contains about 1% of other elements. Its main application is as alloying of other metals like aluminum to produce cast parts, mainly for automotive industry. It is also a basic material in chemical industry for silicones. *Ferrosilicon* can contain more than 2% of other materials and represents the largest application of silicon. Almost all ferrosilicon products are consumed by the iron and steel industries.

Pure elementary silicon when doped with traces of elements such as boron and phosphorus is one of the best semiconductors. These substances have a myriad of applications in modern technology, because they are the core of any

analog or digital electronic circuit. The use of silicon in semiconductor devices demands a much greater purity than afforded by metallurgical grade silicon. In fact, it is the purest silicon used in industry; it is known as the *9-nines* silicon (99,9999999% of purity).

With the growing of the photovoltaic industry, there is a great request of *solar silicon*, name given to the silicon suitable for use in photovoltaic applications, such as solar cells. Solar silicon must be extremely pure, even if the specifications of purity are less strict than for semiconductor silicon.

Induction heating techniques have been widely applied in the last years in the metallurgical and semiconductor industry for the purification of silicon ingots. Figure 1 illustrates the basic components of an induction heating system: a power supply, an induction coil and a workpiece, which is the piece to be heated. The power supply sends alternating current through the coil that circulates around the coil generating a magnetic field. When the workpiece is placed in the coil, the magnetic field induces eddy currents in it that, by the Joule effect, produce heat. It is this heat which warms up the workpiece.

Based on the induction heating technique, various kinds of induction furnaces are employed for different purposes, such as metal smelting ([CETAL, CRST93]), metal hardening ([CSL04, WKN94]) or crystal growing ([MR97, KP03]). In this work we consider an induction melting furnace as the one represented in Fig. 2. It consists of a cylindrical vessel (usually called the crucible)

Fig. 1. Induction system

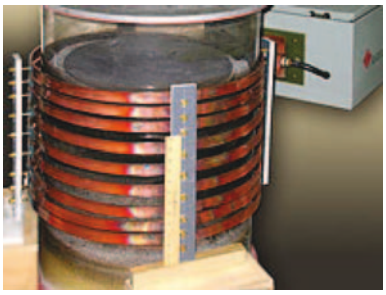


Fig. 2. Induction furnace

made from a material such as graphite which is surrounded by an inductor coil made of a very conductive material (copper, for instance). Silicon is placed inside the crucible and the coil is supplied with an alternating current. The goal is to melt the silicon that initially is introduced in solid state.

The idea for the purification process is based on the fact that if silicon is melted and resolidified, the last parts of the mass to solidify contain most of the impurities. Thus, in zone melting, the first silicon purification method to be widely used industrially, rods of metallurgical grade silicon are heated to melt at one end. Then, the heater is slowly moved down the length of the rod, keeping a small length of the rod molten as the silicon cools and resolidifies behind it. Since most impurities tend to remain in the molten region, when the process is complete most of the impurities in the rod will have been moved into the end that was the last to be melted. This end is then cut off and discarded, and the process repeated if a still higher purity is needed. Usually, these methods are combined with chemical ones which involve the injection of gasses into (or onto) a molten silicon bath and are chosen to remove undesirable elements through formation of solid of gaseous reaction products.

An important advantage of induction heating is that the melt is very well stirred, since the Lorentz forces generated by the induced fields cause a movement in the liquid material.

The inductive system can be designed to maintain the silicon in a liquid state, control the shape of its free surface and to provide a strong electromagnetic stirring, ensuring a rapid transfer of pollutants from the bulk liquid to its surface. This stirring also aids in melting the charge since the moving fluid transfers heat from the crucible wall to the solid. The numerical simulation is used to control the design of the induction system, discussing, for instance, the effect of the power and the frequency on the process. One of the important items is the crucible.

From the mathematical point of view, the overall process is rather complex, involving thermal, electromagnetic, hydrodynamic and mechanical phenomena. In order to perform a numerical simulation of the furnace, the physical process is expressed as a coupled nonlinear system of partial differential equations arising from the thermo-magneto-hydrodynamic problem. In the last years several papers have been published which deal with the thermo-electromagnetic problem ([BGMS1, BGMS2, CETAL, CRST93, KP03]), with the magneto-hydrodynamic problem ([HSSH93, NEK99]) or with the thermo-magneto-hydrodynamic problem, but not fully coupled ([HO94, KHT96]). The authors have already dealt with the thermo-electromagnetic problem with phase change, using a finite element method [BGMS1]. The present work starts from the problem and the algorithms proposed in [BGMS1] and introduces the hydrodynamic problem and the convective heat transfer in the heat equation.

The outline of this chapter is as follows. In Sect. 2 we present the coupled mathematical model, assuming cylindrical symmetry. The equations of the electromagnetic model are expressed in terms of the magnetic vector potential. Moreover, the heat equations are written in terms of the enthalpy,

to take into account the phase change. The hydrodynamic model is described by the incompressible Reynolds-averaged Navier–Stokes equations, to handle the effects of turbulence. In Sect. 3 we propose an iterative algorithm to solve the coupled problem. Finally, in Sect. 4 we present some numerical results for an industrial furnace devoted to the purification of silicon.

2 Statement of the Problem: Mathematical Modelling

We consider an induction furnace consisting of an induction coil surrounding a workpiece as the one sketched in Fig. 3. The goal is to compute the distribution of heat in the workpiece caused by the eddy currents, considering phase change and convective heat transfer.

Let Ω_0 be the radial section of the workpiece, and $\Omega_1, \Omega_2, \ldots, \Omega_m$ the radial sections of the windings of the coil. In fact, to be able to consider the problem in an axisymmetric setting, the induction coil is replaced by m rings with toroidal geometry. Moreover, Ω_a will denote the air around the conductors, so that $\Omega = \Omega_a \cup \Omega_0 \cup \Omega_1 \cup \cdots \cup \Omega_m$ will denote the two dimensional domain of the model (see Fig. 4). In principle, Ω is a half-plane and we should impose "boundary conditions" at infinity. For the sake of simplicity, we cut the domain far from the conductors and impose boundary conditions on the artificial boundary (see [BGMS2] for a BEM–FEM method to deal with the unbounded domain).

2.1 The Electromagnetic Model

Since we are considering alternating currents, all of the fields have the form:

$$\boldsymbol{\mathcal{F}}(\mathbf{x}, t) = \mathrm{Re}\,[\mathrm{e}^{\mathrm{i}\omega t}\,\mathbf{F}(\mathbf{x})], \tag{2}$$

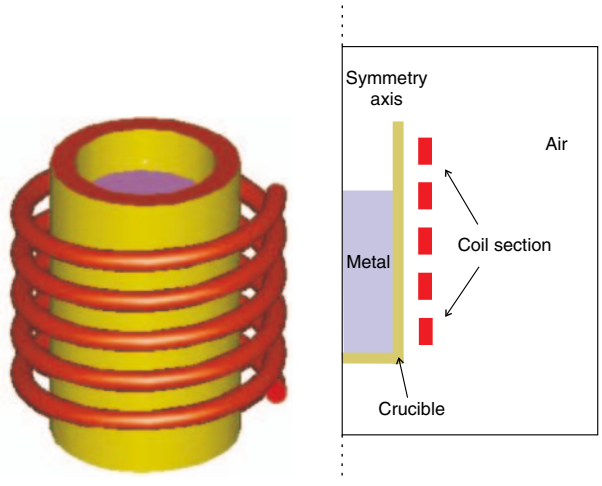


Fig. 3. Sketch of the induction furnace and diametral section

where t is time, $\mathbf{x} \in \mathbb{R}^3$ is the space position, ω is the angular frequency, i the imaginary unit and $\mathbf{F}(\mathbf{x})$ is the complex amplitude of the field. Moreover, as the induction furnace we are interested in works in a low-frequency regime, the Maxwell's equations can be reduced to the so-called *eddy current* model:

$$\mathbf{curl}\,\mathbf{H} = \mathbf{J}, \tag{3}$$

$$\mathrm{i}\omega\mathbf{B} + \mathbf{curl}\,\mathbf{E} = \mathbf{0}, \tag{4}$$

$$\mathrm{div}\,\mathbf{B} = 0, \tag{5}$$

$$\mathrm{div}\,\mathbf{D} = \varrho, \tag{6}$$

to which we have to add the equation imposing the intensity current, I, flowing along the coil. In (3)–(6) $\mathbf{H}$, $\mathbf{J}$, $\mathbf{B}$, $\mathbf{E}$ and $\mathbf{D}$ are the complex amplitudes associated with the magnetic field, the current density, the magnetic induction, the electric field and the electric displacement, respectively, while ϱ denotes the charge density.

The system (3)–(6) needs to be completed by the constitutive relations

$$\mathbf{B} = \mu\mathbf{H}, \tag{7}$$

$$\mathbf{D} = \varepsilon\mathbf{E}, \tag{8}$$

where μ is the magnetic permeability and ε is the electric permittivity. We also need the Ohm's law

$$\mathbf{J} = \begin{cases} \sigma\mathbf{E} & \text{inside conductors,} \\ 0 & \text{in air,} \end{cases} \tag{9}$$

where σ is the electric conductivity.

Remark 1. In fact, the current density in the conductors is given by

$$\mathbf{J} = \sigma(\mathbf{E} + \mathbf{u} \times \mathbf{B}), \tag{10}$$

where $\mathbf{u}$ is the velocity field. In our problem the second term is only important when the furnace works at low frequencies and very high intensities, so we are neglecting it, for the sake of simplicity.

Due to the symmetry of the problem, we are interested in using a cylindrical coordinate system (r, θ, z), with the z-axis coinciding with the symmetry axis of the domain. Hereafter we denote $\mathbf{e}_r$, $\mathbf{e}_\theta$ and $\mathbf{e}_z$ the local orthonormal basis associated with this system of coordinates. Now we assume cylindrical symmetry, which means that no field depends on the angular variable θ. We further assume that the current density field has nonzero component only in the tangential direction $\mathbf{e}_\theta$, namely

$$\mathbf{J}(r, \theta, z) = J_\theta(r, z)\mathbf{e}_\theta.$$

A well-known result allows us to conclude from (5) that $\mathbf{B}$ is the curl of a magnetic vector potential, denoted by $\mathbf{A}$:

$$\mathbf{B} = \mathbf{curl}\,\mathbf{A}. \tag{11}$$

For the sake of uniqueness we take $\mathbf{A}$ to be divergence-free (Coulomb gauge), and we can also conclude that $\mathbf{A}$ is of the form

$$\mathbf{A}(r,\theta,z) = A_\theta(r,z)\mathbf{e}_\theta. \tag{12}$$

From (3), (4), (9) and (11) we deduce that there exist constants $C_k \in \mathbb{C}$, $k = 0,\ldots,m$, such that

$$\mathrm{i}\omega A_\theta + \sigma^{-1} J_\theta = \frac{C_k}{r} \quad \text{in } \Omega_k, \tag{13}$$

recalling that Ω_k, $k = 1,\ldots,m$ denotes each connected component of the conductor, and that Ω_0 is the workpiece (see [BGMS1] or [BGMS2] for details).

The expression of the curl of a vector field in cylindrical coordinates and equations (3), (7), (12) and (13) combined together yield

$$-\left(\frac{\partial}{\partial r}\left(\frac{1}{\mu r}\frac{\partial(rA_\theta)}{\partial r}\right) + \frac{\partial}{\partial z}\left(\frac{1}{\mu}\frac{\partial A_\theta}{\partial z}\right)\right) + \mathrm{i}\omega\sigma A_\theta = \frac{\sigma}{r}C_k, \tag{14}$$

in any connected component of the conducting domain, and

$$-\left(\frac{\partial}{\partial r}\left(\frac{1}{\mu r}\frac{\partial(rA_\theta)}{\partial r}\right) + \frac{\partial}{\partial z}\left(\frac{1}{\mu}\frac{\partial A_\theta}{\partial z}\right)\right) = 0 \tag{15}$$

in the air.

To be able to solve equations (14)–(15) we assume that the current intensities flowing in each ring are given data. Thus we add to the model the following equations

$$\int_{\Omega_k} J_\theta \,\mathrm{d}r\mathrm{d}z = \mathrm{I}_k, \quad k = 1,\ldots,m,$$

I_k being the intensity traversing Ω_k. For a further discussion about the model one can see [BGMS1], [BGMS2] or [CETAL]. An explanation about the physical meaning of the constants C_k can be seen in [CETAL] or [RS96]. An important result is that C_k must be zero in any simply connected region, in particular $C_0 = 0$ in the workpiece. From the mathematical point of view, these constants can be considered as Lagrange multipliers associated with the intensity constraints above.

Electromagnetic Boundary Conditions

As we have already said, the unbounded domain is cut far from the conductors to have a bounded domain. We shall denote by Γ^A the boundary of this

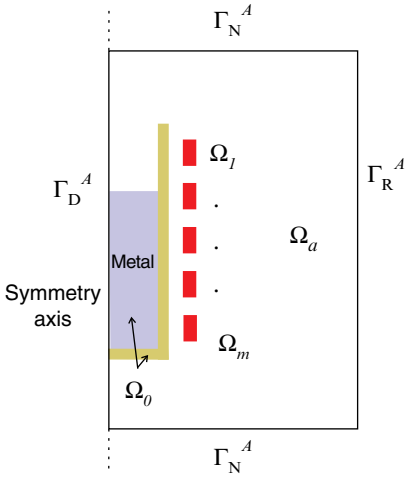


Fig. 4. Computational domain for the electromagnetic problem

computational domain and set $\Gamma^A = \Gamma_{\rm R}^A \cup \Gamma_{\rm N}^A \cup \Gamma_{\rm D}^A$ (see Fig. 4). Following [CETAL] the boundary conditions we impose are

$$\frac{\partial(rA_\theta)}{\partial r} + A_\theta = 0 \quad \text{on } \Gamma_{\rm R}^A, \tag{16}$$

$$\frac{\partial(rA_\theta)}{\partial z} = 0 \quad \text{on } \Gamma_{\rm N}^A, \tag{17}$$

$$A_\theta = 0 \quad \text{on } \Gamma_{\rm D}^A. \tag{18}$$

2.2 The Thermal Model

The above model must be coupled with the heat equation to study the thermal effects of the electromagnetic fields in the workpiece. As the furnace is designed to reach temperatures higher than the melting point of the metal we shall use the heat transfer equation in transient state with change of phase. Furthermore, since the molten metal is subject to electromagnetic and buoyancy forces, we also need to consider convective heat transfer. Let us suppose that we already know the velocity field $\mathbf{u}$ which is null in the solid part of the workpiece, then the equation for energy conservation is

$$\left(\frac{\partial e}{\partial t} + \mathbf{u} \cdot \mathbf{grad}\, e\right) - \mathrm{div}(k_{\rm eff}(\mathbf{x}, T)\ \mathbf{grad}\, T) = \frac{|\mathbf{J}|^2}{2\sigma} \ \text{ in } \Omega_0, \tag{19}$$

where e is the enthalpy, T is the temperature and $k_{\rm eff}$ is the effective thermal conductivity, which is the sum of the turbulent and molecular conductivities,

$k_{\text{eff}} = k + k_{\text{t}}$. The turbulent thermal conductivity is computed by using the formula

$$k_{\text{t}} = \frac{\eta_{\text{t}}}{\sigma_{\text{t}}}, \tag{20}$$

where η_{t} is the turbulent dynamic viscosity given by (39) below, and σ_{t} is the turbulent Prandtl's number, which is taken to be equal to 0.9.

We remark that the thermal conductivity k depends on temperature. We also assume that other material properties as the electric conductivity σ, the magnetic permeability μ and the dynamic viscosity η may depend on temperature.

The coupling between the thermal and the electromagnetic submodels is made by the heat released in the workpiece due to the Joule effect. This heat is represented in (19) by the term on the right-hand side, involving $\mathbf{J}$ which is obtained from (13). In fact, since the electromagnetic equations are expressed in the frequency domain, the heat source is determined by taking the mean value in a cycle (see [BGMS1]).

In (19) the terms between parenthesis on the left-hand side can be rewritten as the material time derivative of enthalpy, which we shall denote by $\dot{e}$. Moreover, assuming cylindrical symmetry and the fact that T does not depend on the angular coordinate θ, the heat equation becomes

$$\dot{e} - \frac{1}{r}\frac{\partial}{\partial r}\left(rk(r,z,T)\frac{\partial T}{\partial r}\right) - \frac{\partial}{\partial z}\left(k(r,z,T)\frac{\partial T}{\partial z}\right) = \frac{|J_\theta|^2}{2\sigma}. \tag{21}$$

Notice that, from (13), we obtain

$$J_\theta = -\mathrm{i}\omega\sigma A_\theta \quad \text{in } \Omega_0, \tag{22}$$

because $C_0 = 0$ in Ω_0.

Thermal Boundary Conditions

The computational domain for the thermal problem is the workpiece, i.e. Ω_0. We shall denote its symmetry axis by Γ_{S}, and by $\Gamma_{\text{R}}^{\text{T}}$ the part of the boundary that is not on the symmetry axis (see Fig. 5). Then, (21) is completed with the following radiation–convection condition on the boundary $\Gamma_{\text{R}}^{\text{T}}$:

$$k(\mathbf{x},T)\frac{\partial T}{\partial \mathbf{n}} = \alpha(T_{\text{c}} - T) + \gamma(T_{\text{r}}^4 - T^4), \tag{23}$$

where α is the coefficient of convective heat transfer, T_{c} and T_{r} are the external convection and radiation absolute temperatures, respectively, the coefficient γ is the product of emissivity by Stefan–Boltzmann constant, and $\mathbf{n}$ is the outward unit normal vector to the boundary. Besides, on the axis Γ_{S} we set

$$k(\mathbf{x},T)\frac{\partial T}{\partial \mathbf{n}} = 0.$$

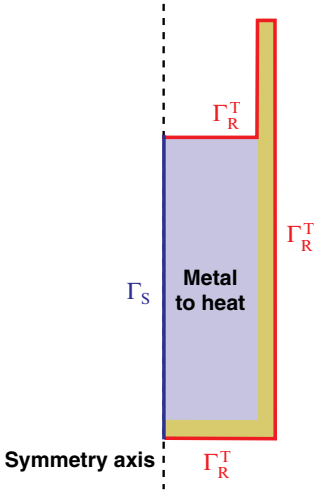


Fig. 5. Computational domain for the thermal problem

2.3 The Hydrodynamic Model

Let $\Omega_l(t)$ be the radial section of the molten metal, and $\Gamma_x(t)$, $\Gamma_d(t)$ and $\Gamma_n(t)$ the different parts of the boundary at time t (depicted in Fig. 6). We assume that the fluid motion is governed by the incompressible Navier–Stokes equations:

$$\rho(\mathbf{x},T)\left(\frac{\partial \mathbf{u}}{\partial t}+\mathbf{u}\cdot\nabla\mathbf{u}\right)-\operatorname{div}(\eta(\mathbf{x},T)D(\mathbf{u}))+\nabla p=\mathbf{f}\quad\text{in }\Omega_l(t), \tag{24}$$

$$\operatorname{div}\mathbf{u}=0\quad\text{in }\Omega_l(t), \tag{25}$$

where ρ denotes the density, $\mathbf{u}$ is the velocity field, η is the dynamic viscosity, p is the pressure and $D(\mathbf{u})$ denotes the symmetric part of $\operatorname{grad}\mathbf{u}$, namely

$$D=\frac{\operatorname{grad}\mathbf{u}+\operatorname{grad}\mathbf{u}^t}{2}.$$

We remark that the hydrodynamic domain is the molten region of the metal, which varies as the metal melts or solidifies, so it depends on time. Moreover, both density and viscosity are material properties which depend on temperature, so for the solution of the thermal problem is essential to solve the hydrodynamic problem.

The right-hand side term $\mathbf{f}$ contains the forces supported by the fluid due to natural convection (buoyancy forces) and those due to the electromagnetic field (Lorentz force):

$$\mathbf{f}=\rho(\mathbf{x},T)\mathbf{g}+\mathbf{J}\times\mathbf{B}, \tag{26}$$

where $\mathbf{g}$ is the acceleration of gravity.

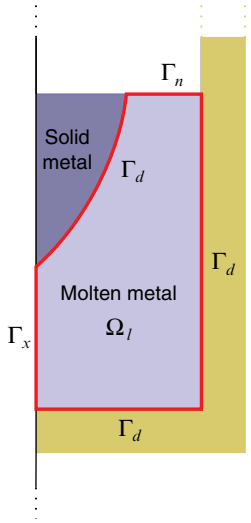


Fig. 6. Computational domain for the hydrodynamic problem

The term representing the Lorentz force is obtained from the solution of the electromagnetic problem. Since in the electromagnetic model we work in the frequency domain, the Lorentz force is determined by taking the mean value in a cycle, namely

$$\frac{\omega}{2\pi}\int_0^{2\pi/\omega} \boldsymbol{\mathcal{J}}(\mathbf{x},t)\times\boldsymbol{\mathcal{B}}(\mathbf{x},t)\,\mathrm{d}t, \tag{27}$$

where $\boldsymbol{\mathcal{J}}$ and $\boldsymbol{\mathcal{B}}$ denote the current density and the magnetic induction, respectively, and ω is the angular frequency.

In (24) we can rewrite the terms into the parenthesis as the material time derivative of the velocity. If we do so, and use cylindrical coordinates, we obtain the equations we will use in our model:

$$\rho\dot{u}_r - \frac{1}{r}\left[\frac{\partial}{\partial r}\left(\eta r\frac{\partial u_r}{\partial r}\right)+\frac{r}{2}\frac{\partial}{\partial z}\left(\eta\left(\frac{\partial u_r}{\partial z}+\frac{\partial u_z}{\partial r}\right)\right)-\frac{u_r}{r^2}\right]+\frac{\partial p}{\partial r}=\mathbf{f}_r, \tag{28}$$

$$\rho\dot{u}_z - \frac{1}{r}\left[\frac{\partial}{\partial r}\left(\eta\frac{r}{2}\left(\frac{\partial u_r}{\partial z}+\frac{\partial u_z}{\partial r}\right)\right)+r\frac{\partial}{\partial z}\left(\eta\frac{\partial u_z}{\partial z}\right)\right]+\frac{\partial p}{\partial z}=\mathbf{f}_\mathbf{z}, \tag{29}$$

$$\frac{1}{r}\frac{\partial}{\partial r}(ru_r)+\frac{\partial u_z}{\partial z}=0, \tag{30}$$

where we recall that $\rho=\rho(r,z,T)$ and $\eta=\eta(r,z,T)$.

Initial and Boundary Conditions for the Hydrodynamic Model

Equations (28)–(30) are completed with the following initial and boundary conditions

$$\mathbf{u} = \mathbf{0} \quad \text{on } \Gamma_d(t)\,, \tag{31}$$

$$S\mathbf{n} = 0 \quad \text{on } \Gamma_n(t)\,, \tag{32}$$

$$S\mathbf{n} = 0 \quad \text{on } \Gamma_x(t)\,, \tag{33}$$

$$\mathbf{u} = \mathbf{0} \quad \text{in } \Omega_{\rm l}(0)\,, \tag{34}$$

where S denotes the Cauchy stress tensor, $S = 2\eta D(\mathbf{u}) - pI$, and $\mathbf{n}$ is the outward unit normal vector to the boundary.

An Algebraic Turbulence Model: Smagorinsky's Model

We recall that the Reynolds number is a dimensionless quantity which gives the ratio of inertial forces to viscosity forces. It is given by

$$Re = \frac{\rho \mathbf{V} L}{\mu}.$$

When this number goes beyond a threshold the flow becomes turbulent, and it makes practically impossible to model its behaviour using the Navier–Stokes equations, due to the extremely fine required computational mesh. For numerical simulation purposes the Navier–Stokes equations are replaced with the so-called Reynolds-averaged Navier–Stokes equations (see [MP94]):

$$\rho(\mathbf{x},T)\left(\frac{\partial \bar{\mathbf{u}}}{\partial t} + \bar{\mathbf{u}}\cdot\nabla\bar{\mathbf{u}}\right) - \operatorname{div}(\eta(\mathbf{x},T)D(\bar{\mathbf{u}})) - \operatorname{div} R + \nabla\bar{p} = \mathbf{f} \quad \text{in } \Omega_{\rm l}(t), \tag{35}$$

$$\operatorname{div}\bar{\mathbf{u}} = 0 \quad \text{in } \Omega_{\rm l}(t), \tag{36}$$

where $\bar{\mathbf{u}}$ denotes the mean velocity and $\bar{p}$ the mean pressure. The tensor R is called the Reynolds stress tensor, and it represents the contribution of the turbulent part to the mean flow.

The Boussinesq assumption consists in taking the Reynolds tensor as

$$R = -\frac{1}{3}tr(R)I + 2\eta_{\rm t} D(\bar{\mathbf{u}}), \tag{37}$$

where I is the identity tensor and $\eta_{\rm t}$ is the turbulent viscosity. Using this assumption we can now rewrite equation (35) as

$$\rho(\mathbf{x},T)\left(\frac{\partial \bar{\mathbf{u}}}{\partial t} + \bar{\mathbf{u}}\cdot\nabla\bar{\mathbf{u}}\right) - \operatorname{div}(\eta_{\rm eff}(\mathbf{x},T)D(\bar{\mathbf{u}})) + \nabla\bar{p^*} = \mathbf{f} \quad \text{in } \Omega_{\rm l}(t), \tag{38}$$

where $p^* = p - \frac{1}{3}\mathrm{tr}(R)$ and $\eta_{\rm eff}$ is the effective viscosity, which is given by $\eta_{\rm eff} = \eta + \eta_{\rm t}$. Different models are obtained depending on the way in which the turbulent viscosity $\eta_{\rm t}$ is computed. A very simple and easy to implement model is the one proposed by Smagorinsky (see [MP94]), which consists in taking

$$\eta_{\rm t} = \rho c h^2 |D(\bar{\mathbf{u}})|, \quad c \cong 0.01 \tag{39}$$

where $h(x)$ is the mesh size of the numerical method around point x.

3 Numerical Approximation

To obtain a suitable discretization of the material time derivative in (21) and (28) we have used the characteristics method (see [PIR82]).

Electromagnetic and thermal problems have been spatially discretized by a piecewise linear finite elements associated with a triangular mesh. The electromagnetic problem is solved in the workpiece, the inductors and the air, while the heat transfer equation is only solved in the workpiece.

The hydrodynamic problem has been spatially discretized by the finite element couple P_1-bubble/P_1, which is known to satisfy the *inf–sup* condition (see [BF91]). We remark that the hydrodynamic problem is only solved in the liquid domain Ω_l, which must be determined at each time step.

We also notice that, at each time step, the three problems form a coupled nonlinear system. Indeed, in the thermal problem the heat source depends on the solution of the electromagnetic problem, while the convective heat transfer needs from the hydrodynamic problem. Moreover, the Lorentz force in the hydrodynamic problem needs the solution of the electromagnetic problem. On the other hand, parameters k, σ, μ, ρ and η depend on temperature, and so does enthalpy. Furthermore, the radiation–convection boundary condition in the thermal problem depends on T^4. To handle the coupling between the three problems we propose a fixed point algorithm which is schematized in Fig. 7 below.

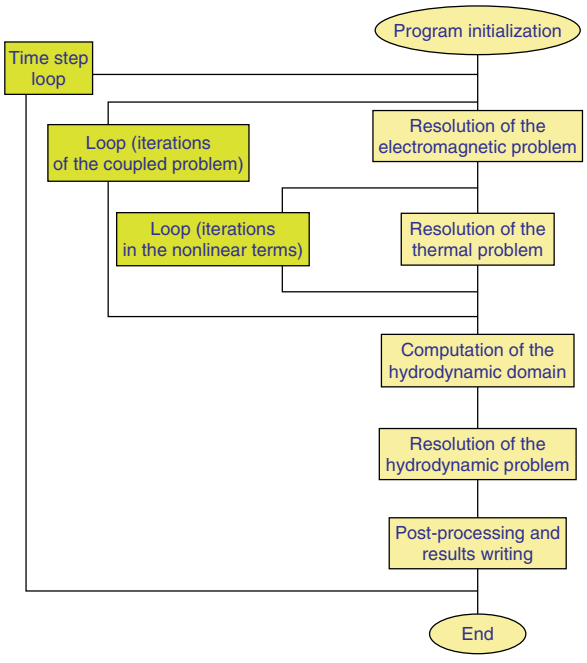


Fig. 7. Scheme of the algorithm

Remark 2. As we have seen in Remark 1 the velocity term in the Ohm's law is not considered, so the electromagnetic problem does not need the solution of the hydrodynamic problem. Moreover, in the thermal problem, the velocity field comes from the solution at the previous time step. Thus, we are allowed to solve the hydrodynamic problem segregated from the two other problems, which saves much computational time.

4 Numerical Results

In this section we present some numerical results obtained by using the algorithm introduced above, which has been implemented in a computer `Fortran` program. More precisely we have applied the algorithm to simulate an industrial furnace used for silicon purification.

We consider a workpiece consisting of a graphite crucible surrounded by an alumina layer and containing silicon. Since solid silicon is not very conductive, a graphite susceptor is required to heat the silicon charge; heating the silicon is then done by conduction and radiation from the graphite until the silicon melts and it conducts electric current. All of materials are initially at 30°C. The induction coil is made of water-cooled copper. The geometrical data of this furnace are summarized in Fig. 8 and Table 1. A detail of the computational

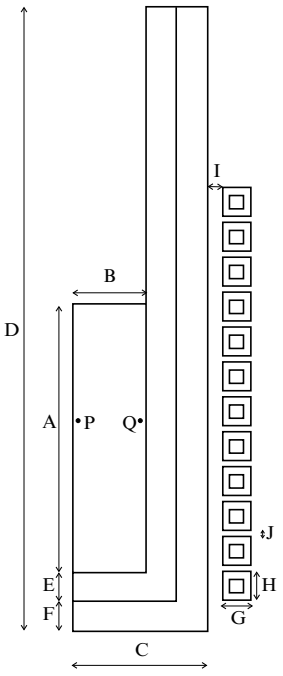


Fig. 8. Sketch of the geometry

Table 1. Geometrical data

A – Height of silicon:	0.45 m
B – Inner radius of crucible:	0.125 m
C – Outer radius of crucible:	0.225 m
D – Crucible height:	1.05 m
E – Crucible width:	0.05 m
F – Alumina layer width:	0.05 m
G – Turn diameter:	0.05 m
H – Turn height:	0.05 m
I – Distance between coil and crucible:	0.025 m
J – Distance between the turns:	0.01 m
Number of coil turns:	12
P, Q – Measure points	

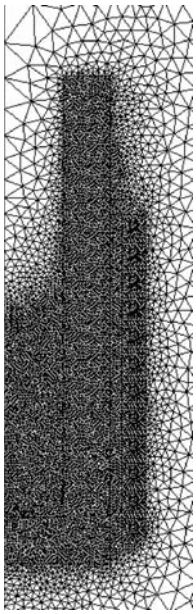


Fig. 9. Detail of the mesh

mesh can be seen in Fig. 9. The physical properties of the three materials in the workpiece depend on temperature and have been obtained from literature. Since we are not considering the thermal model in the coil, the electromagnetic properties of copper are supposed to be constant. Several simulations have been carried out, considering values of 100 Hz for the frequency and 5,500 Å for the intensity.

Figure 10 shows the temperature field in the workpiece, for 30 min and 180 min, respectively. In Fig. 11 we represent the temperature in the silicon

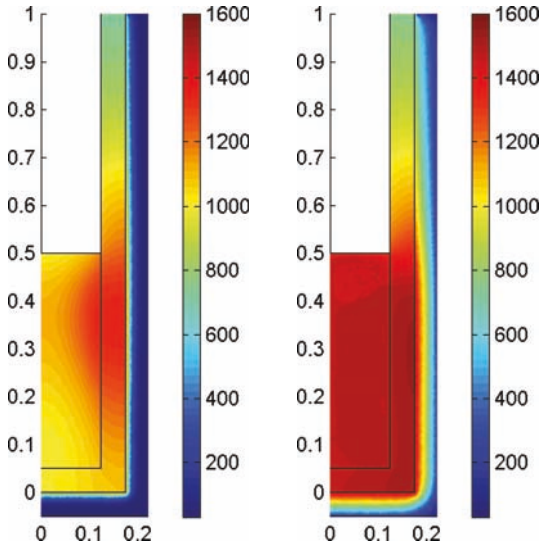


Fig. 10. Temperature field for $t = 30\,\text{min}$ (*left*) and $t = 180\,\text{min}$ (*right*)

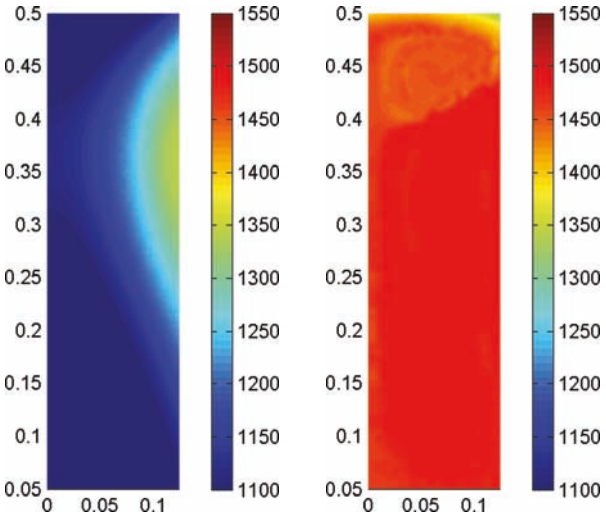


Fig. 11. Silicon temperature for $t = 30\,\text{min}$ (*left*) and $t = 180\,\text{min}$ (*right*)

for the same times. During the first 30 min, the temperature of the workpiece increases and the silicon begins to melt (the melting point is 1,412 °C) and after 180 min the silicon is completely liquid. Figure 12 shows the modulus of current density also for 30 min and 180 min, respectively. Notice that, since solid silicon is not an electric conductor, the induced current density concentrates in the graphite. As silicon temperature increases, so does its electrical conductivity and the induced current density on its surface.

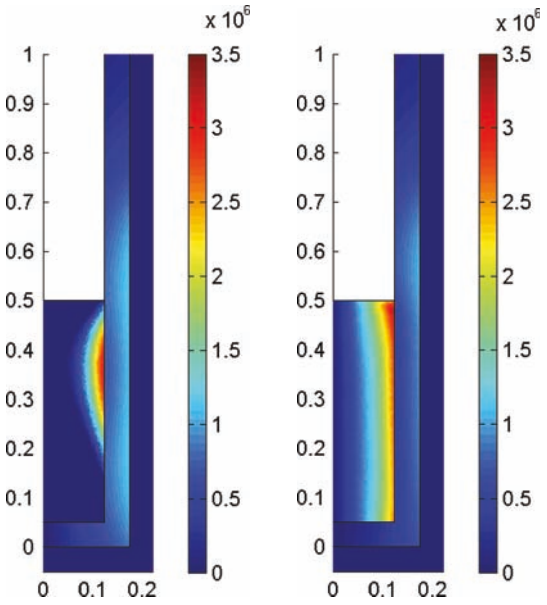


Fig. 12. Modulus of current density for $t = 30\,\text{min}$ (*left*) and $t = 180\,\text{min}$ (*right*)

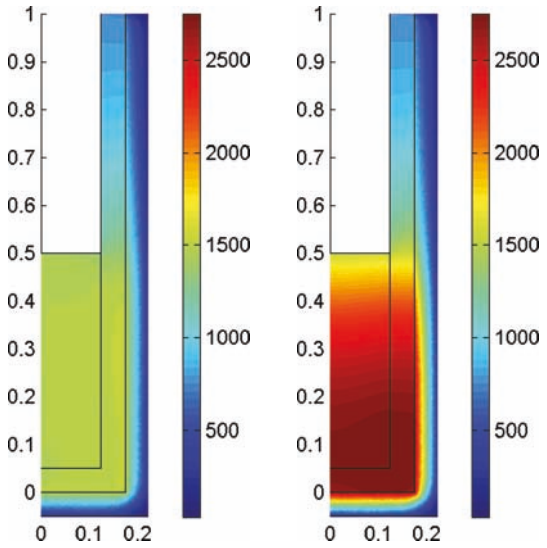


Fig. 13. Temperature with and without convection term ($t = 180\,\text{min}$)

Figures 13 and 14 illustrate the importance of considering convective heat transfer when computing the temperature field. In Fig. 13 one can check how neglecting the convection term in the heat equation could cause the materials to reach very high and unrealistic temperatures that, in particular, would

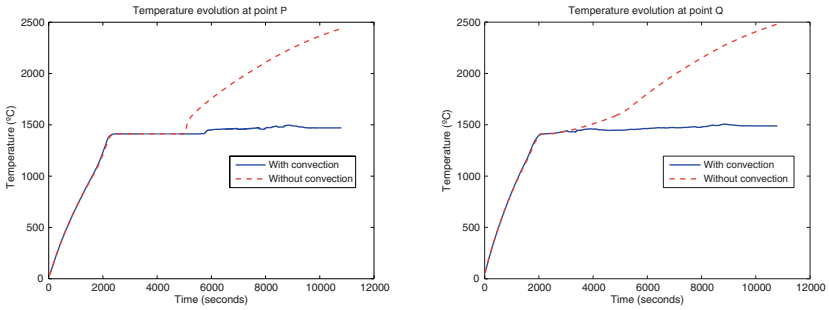


Fig. 14. Evolution of temperature at points P and Q

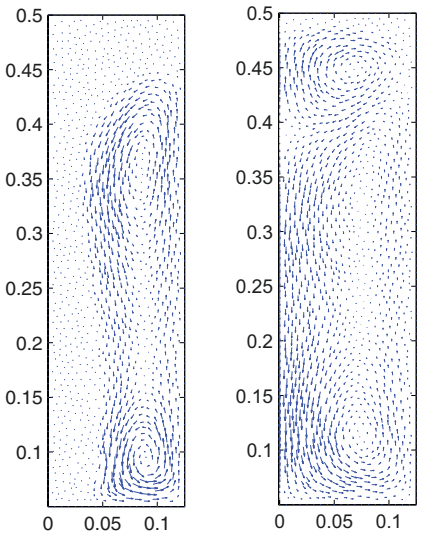


Fig. 15. Velocity field $t = 90\,\text{min}$ (*left*) and $t = 180\,\text{min}$ (*right*)

cause the crucible to melt. The same conclusions can be obtained from Fig. 14, that shows the evolution in time of the temperature of two different points in the silicon: a point P close to the symmetry axis and another point Q close to the graphite crucible, considering or not the convection term.

We complete these results by representing, in Fig. 15, the velocity field for times $t = 90$ and $t = 180\,\text{min}$, respectively. We can appreciate the swirls due to Lorentz forces.

Acknowledgements

This work has been partly supported by MEC-FEDER (Spain) through research project DPI2003-01316 and Ferroatlántica I+D company under contract. The last author has received financial support by a predoctoral grant

FPI from the Spanish Ministry of Education and Science, cofinanced by the European Social Fund.

References

[BGMS1] Bermúdez, A., Gómez, D., Muñiz, M.C., Salgado, P: Transient numerical simulation of a thermoelectrical problem in cylindrical induction heating furnaces. Adv. Comput. Math., To appear

[BGMS2] Bermúdez, A., Gómez, D., Muñiz, M.C., Salgado, P: A FEM/BEM for axisymmetric electromagnetic and thermal modelling of induction furnaces, Syst. Int. J. Numer. Meth. Engng, To Appear.

[BF91] Brezzi, F., Fortin, M.: Mixed and hybrid finite element methods. Springer Verlag, New York, (1991)

[CSL04] Cajner, F., Smoljan, B., Landek, D.: Computer simulation of induction hardening, J. Mater. Process. Technol. **157-158**, 55–60 (2004)

[CETAL] Chaboudez, C., Clain, S., Glardon, R., Mari, D., Rappaz, J., Swierkosz, M.: Numerical Modeling in Induction Heating for Axisymmetric Geometries, IEEE Trans. Magn., **33**, 739–745 (1997)

[MR97] Chen, Q.S., Gao, P., Hu, W.R.: Effects of induction heating on temperature distribution and growth rate in large-size SiC growth system, J. Cryst. Growth, **266**, 320–326 (2004)

[CRST93] Clain, S., Rappaz, J., Swierkosz, M., Touzani, R.: Numerical modelling of induction heating for two-dimensional geometries, Math. Models Appl. Sci., **3**, 805–822 (1993)

[HO94] Henneberger, G., Obrecht, R.: Numerical calculation of the temperature distribution in the melt of industrial crucible furnaces In: Second International Conference on Computation in Electromagnetics. (1994)

[HSSH93] Henneberger, G., Sattler, Ph. K., Shen, D., Hadrys, W.: Coupling of magnetic and fluid flow problems and its application in induction melting apparatus, IEEE Trans. Magn., **29**, 1589–1594 (1993)

[KHT96] Katsumura, Y., Hashizume, H., Toda, S.: Numerical Analysis of fluid flow with free surface and phase change under electromagnetic force, IEEE Trans. Magn., **32**, 1002–1005 (1996)

[KP03] Klein, O., Philip, P.: Transient numerical investigation of induction heating during sublimation growth of silicon carbide single crystals, J. Cryst. Growth, **247**, 219–235 (2003)

[MP94] Mohammadi, B., Pironneau, O.: Analysis of the k-epsilon turbulence model. Wiley/Masson, New York (1994)

[NEK99] Natarajan, T.T., El-Kaddah, N.: A methodology for two-dimensional finite element analysis of electromagnetically driven flow in induction stirring systems, IEEE Trans. Magn., **35**, n.3, 1773–1776 (1999)

[PIR82] Pironneau, O.: On the transport-diffusion algorithm and its applications to the Navier-Stokes equations. Numer. Math., **38**, n.3, 309–332 (1982)

[RS96] Rappaz, J., Swierkosz, M.: Mathematical modelling and numerical simulation of induction heating processes. Appl. Math. Comput. Sci., **6**, n.2, 207–221 (1996)

[WKN94] Wanser, S., Krähenbühl, L., Nicolas, A.: Computation of 3D induction hardening problems by combined finite and boundary elements methods. IEEE Trans. Magn., **30**, 3320–3323 (1994)

Growth and Pattern Formation for Thin Films

Russel E. Caflisch

Department of Mathematics and Department of Material Science & Engineering, University of California, 405 Hilgard Avenue, Los Angeles, CA 90095-1555, USA
caflisch@math.ucla.edu

Summary. Epitaxy is the growth of a thin film by attachment to an existing substrate in which the crystalline properties of the film are determined by those of the substrate. In heteroepitaxy, the substrate and film are of different materials, and the resulting mismatch between lattice constants can introduce stress into the system. We have developed an island dynamics model for epitaxial growth that is solved using a level set method. This model uses both atomistic and continuum scaling, since it includes island boundaries that are of atomistic height, but describes these boundaries as smooth curves. The strain in the system is computed using an atomistic strain model that is solved using an algebraic multigrid method and an artificial boundary condition. Using the growth model together with the strain model, we simulate pattern formation on an epitaxial surface.

1 Introduction

Epitaxy is the growth of a thin film on a substrate in which the crystal properties of the film are inherited from those of the substrate. Since an epitaxial film can (at least in principle) grow as a single crystal without grain boundaries or other defects, this method produces crystals of the highest quality.

The geometry of an epitaxial surface consists of step edges and island boundaries, across which the height of the surface increases by one crystal layer, and adatoms which are weakly bound to the surface. Epitaxial growth involves deposition, diffusion, and attachment of adatoms on the surface. Deposition is from an external source, such as a molecular beam. The principal dimensionless parameter (for growth at low temperature) is the ratio $D/(a^4F)$, in which a is the lattice constant and D and F are the adatom diffusion coefficient and deposition flux. It is conventional to refer to this parameter as D/F, with the understanding that the lattice constant serves as the unit of length. Typical values for D/F are in the range of 10^4–10^8.

2 Island Dynamics

Burton, Cabrera, and Frank [2] developed the first detailed theoretical description for epitaxial growth. In this "BCF" model, the adatom density solves a diffusion equation with an equilibrium boundary condition ($\rho = \rho_{\rm eq}$), and step edges (or island boundaries) move at a velocity determined from the diffusive flux to the boundary. Modifications of this theory were made, for example in [11], to include line tension, edge diffusion, and nonequilibrium effects. These are "island dynamics" models, since they describe an epitaxial surface by the location and evolution of the island boundaries and step edges. They employ a mixture of coarse graining and atomistic discreteness, since island boundaries are represented as smooth curves that signify an atomistic change in crystal height.

Adatom diffusion on the epitaxial surface is described by a diffusion equation of the form

$$\partial_{\rm t}\rho - D\nabla^2\rho = F - 2{\rm d}N_{\rm nuc}/{\rm d}t \tag{1}$$

in which the last term represents loss of adatoms due to nucleation, and desorption from the epitaxial surface has been neglected. Attachment of adatoms to the step edges and the resulting motion of the step edges are described by boundary conditions at an island boundary (or step edge) Γ for the diffusion equation and a formula for the step-edge velocity v. The simplest of these is

$$\begin{aligned} &\rho = \rho_* \\ &v = D[\partial\rho/\partial n] \end{aligned} \tag{2}$$

in which the brackets indicate the difference between the value on the upper side of the boundary and the lower side. Two choices for ρ_* are $\rho_* = 0$, which corresponds to irreversible aggregation in which all adatoms that hit the boundary stick to it irreversibly, and $\rho_* = \rho_{\rm eq}$ for reversible aggregation. For the latter case, $\rho_{\rm eq}$ is the adatom density for which there is local equilibrium between the step and the terrace [2]. Numerical details on implementation of the level set method for thin film growth are provided in [5].

2.1 Nucleation

For the case of irreversible aggregation, a dimer (consisting of two atoms) is the smallest stable island, and the nucleation rate is

$$\frac{{\rm d}N_{\rm nuc}}{{\rm d}t} = D\sigma_1\langle\rho^2\rangle, \tag{3}$$

where $\langle\cdot\rangle$ denotes the spatial average of $\rho(\mathbf{x},t)^2$ and

$$\sigma_1 = \frac{4\pi}{\ln[(1/\alpha)\langle\rho\rangle D/F]} \tag{4}$$

is the adatom capture number as derived in [1]. The parameter α reflects the island shape, and $\alpha \simeq 1$ for compact islands. Expression (3) for the nucleation rate implies that the time of a nucleation event is chosen deterministically. Whenever $N_{\rm nuc}L^2$ passes the next integer value (L is the system size), a new island is nucleated. Numerically, this is realized by raising the level set function to the next level at a number of grid points chosen to represent a dimer.

The choice of the location of the new island is determined by probabilistic choice with spatial density proportional to the nucleation rate ρ^2. This probabilistic choice constitutes an atomistic fluctuation that must be retained in the level set model for faithful simulation of the epitaxial morphology. For growth with compact islands, computational tests have shown additional atomistic fluctuations can be omitted [16].

Additions to the basic level set method, such as finite lattice constant effects and edge diffusion, are easily included [17]. The level set method with these corrections is in excellent agreement with the results of kinetic Monte Carlo (KMC) simulations.

2.2 The Level Set Method

Within the level set approach, the union of all boundaries of islands of height $k+1$, can be represented by the level set $\varphi = k$, for each k. For example, the boundaries of islands in the submonolayer regime then correspond to the set of curves $\varphi = 0$. The function ϕ is the level set function that evolves according to

$$\frac{\partial \phi}{\partial t} + v|\nabla \phi| = 0. \tag{5}$$

All the physical information is in the normal component v of the velocity function. Islands grow because atoms diffuse toward and attach to island boundaries, and shrink because they can detach from an island boundary.

3 Discrete Elasticity

In heteroepitaxy, strain is introduced into the epitaxial system due to the lattice mismatch between the two constituents of the material. Because of the strain, atoms are displaced by a vector $\mathbf{u}$ from their lattice position. The following discussion of atomistic strain and stress follows that in [19].

To describe the strain energy at each atom, $\mathbf{i} = (i, j, k)$, introduce the translation operators, $T_k^{\pm}$, and the discrete difference operators, $D_k^{\pm}$, D_k^0, defined as follows:

$$T_k^{\pm} f(\mathbf{i}) = f(\mathbf{i} \pm \mathbf{e_k}),$$

$$D_k^{+} f(\mathbf{i}) = \frac{(T_k^{+} - 1) f(\mathbf{i})}{h},$$

$$D_k^- f(\mathbf{i}) = \frac{(1 - T_k^-) f(\mathbf{i})}{h},$$

$$D_k^0 f(\mathbf{i}) = \frac{(T_k^+ - T_k^-) f(\mathbf{i})}{2h},$$

where h is the lattice constant and $\mathbf{e}_k$ is the vector in the kth direction for $k = 1, 2, 3$ with $\|\mathbf{e}_k\| = h$. Throughout this paper, we assume the lattice constant $h = 1$ for simplicity. We use i for the depth-like index, with $-\infty < i \leq n$. Here n is the maximum height of the material. An ABC is sought at $i = 0$, assuming that there is no force for $i < 0$.

Let $\mathbf{u}(\mathbf{i}) = (u_k(\mathbf{i}))_{k=1,\ldots,d}$ be the displacement at the discrete point $\mathbf{i}$ relative to an equilibrium lattice. The discrete strain components defined below ((6) and (7)) can be used to describe the discrete elastic energy. For $k, \ell = 1, 2, 3$ and $p, q = \pm$,

$$S_{k\ell}^{\pm}(\mathbf{u}(\mathbf{i})) = D_\ell^{\pm} u_k(\mathbf{i}), \tag{6}$$

$$S_{k\ell}^{pq}(\mathbf{u}(\mathbf{i})) = \frac{1}{2}(D_\ell^q u_k(\mathbf{i}) + D_k^p u_\ell(\mathbf{i})). \tag{7}$$

The discrete energy density at a point $\mathbf{i}$ is then given by

$$E(\mathbf{i})(\mathbf{u}, \mathbf{u}) = \sum_{k,p} \alpha_k^p (S_{kk}^p(\mathbf{u}))^2 + \sum_{k \neq \ell, p, q} \left\{ 2\beta_{k\ell}^{pq} (S_{k\ell}^{pq}(\mathbf{u}))^2 + \gamma_{k\ell}^{pq} S_{kk}^p(\mathbf{u}) S_{\ell\ell}^q(\mathbf{u}) \right\}.$$

The total energy is the sum

$$\mathcal{E} = \sum_{\mathbf{i}} E(\mathbf{i}). \tag{8}$$

The atomistic strain is determined by minimizing this energy with respect to variations in $\mathbf{u}$.

An effective numerical method for solving the atomistic strain equations using an algebraic multigrid method was developed in [4]. Moreover an artificial boundary condition can be imposed in the substrate close to the interface with the film, to greatly accelerate the computation [10].

4 Directed Self-Assembly

Regular patterns of nanoscale features, such as quantum dots [6, 7, 12], on an epitaxial surface are of considerable interest for possible applications, ranging from memory and logical devices to lasers. Features of this size are difficult to obtain by standard "top-down" approaches, such as lithography. The spontaneous growth of quantum dot arrays is a promising "bottom-up" approach, but it has proved difficult to control the size and spacing of quantum dots obtained in this way. Directed self-assembly is an intermediate approach, in

which formation of the desired patterns is guided by prepatterning of the epitaxial system. For example, subsurface dislocation arrays have been suggested as a prepatterning method [8,18]. These buried dislocations introduce a long-range strain field, which alters the potential energy surface (PES) of the system. Similarly, islands that are capped by a buffer layer of a different material introduce a long-range strain field. It has been shown by density-functional theory (DFT) calculations for metal systems [15] and semiconductor systems [14] that both the adsorption energy $E_{\rm ad}$ and the transition energy $E_{\rm trans}$ of the PES change upon strain.

We model epitaxial growth on a surface with a spatially varying, anisotropic PES, using the following modification of the adatom diffusion equation (1)

$$\frac{\partial \rho}{\partial t} = F + \nabla \cdot (\mathbf{D} \nabla \rho) - 2\frac{\mathrm{d}N}{\mathrm{d}t} + \nabla \cdot \left(\frac{\rho}{k_B T} \mathbf{D} (\nabla E_{\rm ad}) \right). \tag{9}$$

In (9), $\mathbf{D}$ is a diffusion tensor where the diagonal entries are labeled $D_i(\mathbf{x})$ and $D_j(\mathbf{x})$, and correspond to diffusion along the two directions i and j. For simplicity no other direction for diffusion is included (but could easily be incorporated). The last term is the thermodynamic drift, where $k_{\rm B}$ is the Boltzmann constant, and T is the temperature. We enforce a boundary condition $\rho(\mathbf{x}) = \rho_{\rm eq}(D_{\rm det}(\mathbf{x}), \mathbf{x})$, where $D_{\rm det}(\mathbf{x})$ is a (spatially varying) detachment rate [3].

We assume a simple sinusoidal variation of $E_{\rm ad}$ and $E_{\rm trans}$. Figure 1 shows the resulting patterns for PES with spatial variation that is one dimensional (left) and two dimensional (right). These simulation results bear a striking resemblance to the quantum dot patterns obtained in the experimental results of [8].

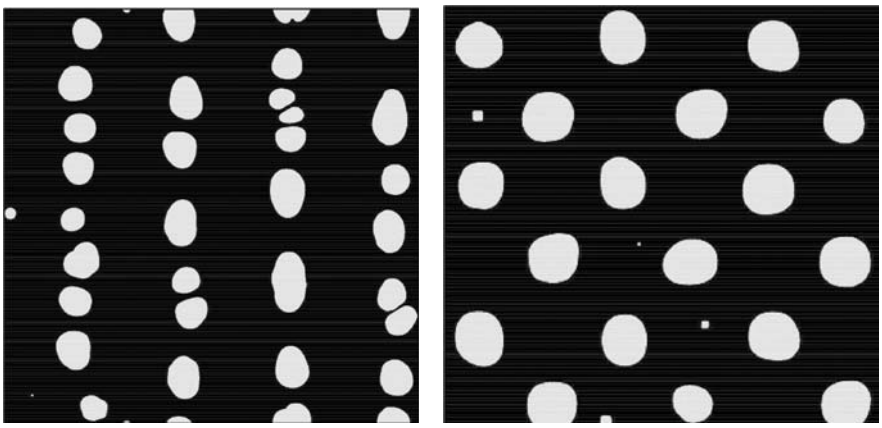


Fig. 1. Pattern formation for monolayer height islands due to a spatially varying PES, with sinusoidal variation in 1D (*left*) and 2D (*right*)

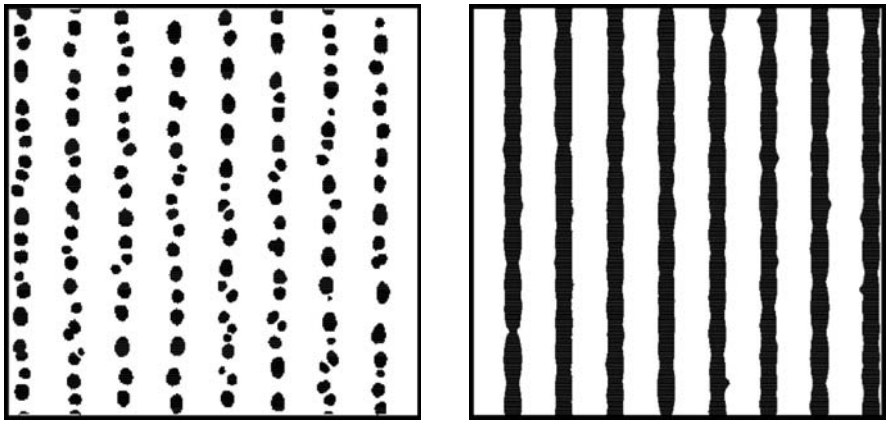


Fig. 2. Morphologies at coverages $\Theta = 0.1$ ML (*left*) and $\Theta = 0.3$ ML (*right*) obtained with a PES that has a much narrower variation

The morphologies shown so far were all obtained at a submonolayer precoalescence coverage of $\Theta = 0.2$ ML and with a PES that varies sinusoidally. Figure 2 shows the patterns that are obtained by a function that has sharper peaks that those of a sine function. The resulting islands at coverage $\Theta = 0.1$ monolayer (ML) are highly aligned. Moreover, at $\Theta = 0.3$ ML, all the islands that are aligned along the j-direction have coalesced in this direction, forming monolayer height "wires." For more details on these computations, see [13].

5 Conclusions

The island dynamics/level set method is capable of simulating epitaxial growth with processes such as adatom detachment from islands that would slow down other approaches. It can also be effectively combined with an atomistic strain code to simulate heteroepitaxial growth. The combined method can be used to study pattern formation due to strain in self-assembly and directed self-assembly.

Acknowledgements

This research was supported in part by the MARCO Center on Functional Engineered NanoArchitectonics (FENA) and by the NSF through grant DMS-0402276.

References

1. G.S. Bales and D.C. Chrzan. Dynamics of irreversible island growth during submonolayer epitaxy. *Phys. Rev. B*, **50**, 6057–6067, 1994.
2. W.K. Burton, N. Cabrera and F.C. Frank. The growth of crystals and the equilibrium structure of their surfaces. *Phil. Trans. Roy. Soc. London Ser. A*, **243**, 299–358, 1951.
3. R.E. Caflisch, W.E., M.F. Gyure, B. Merriman, and C. Ratsch, Phys. Rev. E. **59**, 6879 (1999).
4. R.E. Caflisch, Y.-J. Lee, S. Shu, Y. Xiao and Jinchao Xu. An application of multigrid methods for a discrete elastic model for epitaxial systems JCP, 219 (2006) 697–714.
5. S. Chen, M. Kang, B. Merriman, R.E. Caflisch, C. Ratsch, R. Fedkiw, M.F. Gyure and S. Osher. Level set method for thin film epitaxial growth. *Journ. Comp. Phys.*, **167**, 475–500, 2001.
6. D.J. Eaglesham and M. Cerullo, Phys. Rev. Lett. **64**, 1943 (1990).
7. S. Guha, A. Madhukar, and K.C. Rajkumar, Appl. Phys. Lett. **57**, 2110 (1990).
8. H.J. Kim, Z.M. Zhao, and Y.H. Xie, Phys. Rev. B **68**, 205312 (2003).
9. L.D. Landau and E.M. Lifshitz, *Theory of Elasticity*, Butterworth-Heinemann, Oxford, UK, 1986.
10. R.E. Sunmi Lee, Caflisch and Y.-J. Lee Artificial Boundary Conditions for Discrete Elasticity SIAM J. Applied Math. 66 (2006) 1749–1775.
11. B. Li and R.E. Caflisch. Analysis of island dynamics in epitaxial growth. *Multiscale Model. Sim.*, **1**, 150–171, 2002.
12. Y.-W. Mo, D.E. Savage, B.S. Swartzentruber and M.G. Lagally, Phys. Rev. Lett. **65**, 1020 (1990).
13. X. Niu, R. Vardavas, R.E. Caflisch and C. Ratsch. A Level Set Simulation of Directed Self-Assembly during Epitaxial Growth. Phys. Rev. B (2006) to appear.
14. E. Penev, P. Kratzer and M. Scheffler, Phys. Rev. B **64**, 085401 (2001).
15. C. Ratsch, A.P. Seitsonen and M. Scheffler, Phys. Rev. B **55**, 6750 (1997).
16. C. Ratsch, M.F. Gyure, S. Chen, M. Kang and D.D. Vvedensky. Fluctuations and scaling in aggregation phenomena. *Phys. Rev. B*, **61**, 10598–10601, 2000.
17. C. Ratsch, M.F. Gyure, R.E. Caflisch, F. Gibou, M. Petersen, M. Kang, J. Garcia and D.D. Vvedensky. Level-set method for island dynamics in epitaxial growth. *Phys. Rev. B*, **65**, #195403, U697–U709, 2002.
18. A.E. Romanov, P.M. Petroff and J.S. Speck, Appl. Phys. Lett. **74**, 2280 (1999).
19. A.C. Schindler, M.F. Gyure, D.D. Vvedensky, R.E. Caflisch, C. Connell and G. D. Simms. Theory of Strain Relaxation in Heteroepitaxial Systems Phys. Rev. B 67 (2003): art. no. 075316

On Waves in Fluids: Some Mathematical, Physical and Engineering Aspects

L.M.B.C. Campos

Centro de Ciências e Tecnologias Aeronáuticas e Espaciais (CCTAE)
and Secção de Mecânica Aeroespacial (SMA), Instituto Superior Técnico (IST),
1049-001 Lisboa, Portugal
lmbcampos.aero@mail.ist.utl.pt

Summary. The subject of waves in fluids is addressed from three complementary points-of-view: (Sect. 2) 60 mathematical forms of the acoustic wave equation in fluids, applying to linear and non-linear, non-dissipative and dissipative, sound waves in homogeneous or inhomogeneous, steady or unsteady media, at rest or in motion, e.g. potential and vortical flows; (Sect. 3) the physical interactions between (i) sound waves due to pressure fluctuations in a compressible fluid, with (ii) magnetic waves in an ionized fluid under external magnetic fields, (iii) internal waves in a stratified fluid under gravity and (iv) inertial waves due to Coriolis forces on a rotating fluid, viz. magneto-acoustic-gravity-inertial waves; (Sect. 4) some engineering problems in the area of aerocoustics, which has applications to aircraft, helicopters, rockets and other aerospace vehicles, including acoustic fatigue, sonic boom, interior noise and airport noise, concentrating on the last aspect.

1 Introduction

The classical wave equation describes the propagation of (i) linear (ii) non-dissipative sound waves in a (iii) homogeneous and steady medium (iv) at rest. There are many practical situations in which one or more of the assumptions (i)–(iv) do not hold, hence the importance to extend the acoustic wave equation to (i) inhomogeneous and unsteady media, for which mean state properties (such as mass density and sound speed), may depend, respectively, on position and time; (ii) moving media, e.g. potential mean flows, or vortical mean flows, such as shear flows or swirling flows; (iii) dissipation by thermal conduction and bulk and shear viscosity and (iv) non-linear effects, either weak or strong, depending on whether only second-order or also higher-order non-linearities are included.

Acoustic waves occur in the low atmosphere and in the ocean, are important in speech, hearing, music and high-fidelity sound reproduction and have applications in ultrasonics (e.g. crack detection), as well as unwanted effects (noise and acoustic fatigue). They are one (i) of the four types of waves in

fluids, viz.: (ii) internal waves, in a stratified fluid under gravity occur in the ocean and in the atmosphere of the earth and other planets; (iii) inertial waves associated with the Coriolis force on rotating fluids, affect weather and climate on the earth and occur on other rotating celestial bodies like planets and stars and (iv) magnetic waves in an ionized fluid under an external magnetic field occur in fusion reactors and magnetohydrodynamic generators, in the earth's molten core and high atmosphere (ionosphere) and in the plasma which constitutes stars and permeates the interstellar medium. Their coupling leads to magneto-acoustic-gravity-inertial waves.

Aeroacoustics is a major area of application of acoustics, since it is relevant to many problems of aeronautics and astronauts e.g. (i) the noise of jet and propeller engines at take-off and climb is a major contributor to airport noise; (ii) at approach to land, with the engines at idle, the aerodynamic noise may by comparable; (iii) the sonic boom of supersonic aircraft has so far restricted commercial flight to subsonic speeds overland; (iv) the noise level of rockets is high enough to cause acoustic fatigue of launcher structures and satellite payloads and (v) the helicopter, due to the rotor and gearbox mechanisms it uses, poses noise and vibration problems which limit the exploitation of its ability to hover and fly low and slow near populated areas.

2 Sixty Acoustic Wave Equations

There are at least 60 forms of the acoustic wave equation in fluids (thus excluding solids), which may be grouped in nine classes. The derivation of the most general wave equation in each class can be made by elimination among the equations of fluid mechanics; in some cases variational and other methods can be used as alternatives. Thus, together with overlaps between different classes, there may be several derivations of the same wave equation and multiple cross-checks. In the present account one wave equation in each class is indicated, often but not always the most general [1, 2]. The acoustic wave equation has the same form for all acoustic variables (e.g. potential, gas pressure, mass density and velocity perturbations) for linear non-dissipative sound in an homogeneous steady medium, e.g. for the classical wave equation in a medium at rest or convected wave equation in a uniform flow. In more general conditions this is not usually the case and different acoustic variables satisfy different wave equations, so it is reasonable to aim for the simplest. Note also that non-linear waves are those with steep waveforms, viz. large amplitude waves are non-linear, but small amplitude waves with steep wavefronts ('ripples') are also non-linear.

2.1 Nine Classes of Acoustic Wave Equations

The classical wave equation

$$c_0^{-2}\ddot{\phi} - \nabla^2\phi = 0, \tag{1}$$

where ϕ is the acoustic potential, c_0 the sound speed and dot denotes time derivative $\ddot{\phi} = \partial^2\phi/\partial t^2$ assumes (i) an homogeneous and steady medium; (ii) medium at rest; (iii) linear perturbations and (iv) no dissipation. Next will be presented nine classes of acoustic wave equations, which generalize the classical wave equation.

2.2 Class I: Linear, Non-dissipative Sound in a Potential Mean Flow

The medium in assumed to be a potential flow of velocity $\mathbf{v}_0$, gas pressure p_0, mass density ρ_0 and sound speed c_0 which may depend on position (inhomogeneous medium) and/or on time (unsteady medium). Note that a potential flow is homentropic; in this case there is an acoustic potential ϕ. The wave equation can be deduced from equations of fluid mechanics [3, 4] or a variational method [5, 6]. The variational method uses the acoustic velocity and pressure perturbations:

$$\mathbf{v} = \nabla\phi, \tag{2a}$$

$$p = -\rho_0 \mathrm{d}\phi/\mathrm{d}t, \tag{2b}$$

where $\mathrm{d}/\mathrm{d}t$ is the material derivative for the mean flow:

$$\mathrm{d}/\mathrm{d}t = \partial/\partial t + \mathbf{v}_0 \cdot \nabla. \tag{3}$$

The difference of the kinetic energy per unit volume (4a) and compression energy (4b) in the quadratic approximation:

$$E_v = \rho_0 v^2 = \frac{1}{2}\rho_0(\nabla\phi)^2, \tag{4a}$$

$$E_p = \frac{p^2}{2\rho_0 c_0^2} = \frac{1}{2}\rho_0 c_0^{-2}(\mathrm{d}\phi/\mathrm{d}t)^2, \tag{4b}$$

specifies the acoustic Lagrangian:

$$\pounds(\phi, \dot{\phi}, \nabla\phi; \mathbf{x}, t) = \frac{1}{2}\left[(\nabla\phi)^2 - c_0^2(\dot{\phi} + \mathbf{v}_0 \cdot \nabla\phi)^2\right], \tag{5}$$

which satisfies the principle of stationary action:

$$0 = \delta \int \mathrm{d}^3\mathbf{x} \int \mathrm{d}t \; \pounds(\phi, \dot{\phi}, \nabla\phi; \mathbf{x}, t), \tag{6}$$

leading to the Euler–Lagrange equation:

$$\frac{\partial}{\partial t}\frac{\partial \pounds}{\partial \dot{\phi}} + \nabla \cdot \left[\frac{\partial \pounds}{\partial(\nabla\phi)}\right] = 0. \tag{7}$$

The substitution of (5) in the latter (7) specifies the wave equation in a potential flow (W1–W9 – there are nine particular cases):

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{c_0^2}\frac{\mathrm{d}\phi}{\mathrm{d}t}\right)-\frac{1}{\rho_0}\nabla\cdot(\rho_0\nabla\phi)=0. \tag{8}$$

In the case of an homogeneous, steady uniform flow it reduces to the convected wave equation:

$$c_0^{-2}\mathrm{d}^2\phi/\mathrm{d}t^2-\nabla^2\phi=0 \tag{9}$$

and in the general case it has ten terms

$$\begin{aligned}&\ddot{\phi}-c_0^2\nabla^2\phi-c_0^2\nabla\phi\cdot\nabla(\log\rho_0)-2\dot{\phi}c_0^{-1}\dot{c}_0+2(\mathbf{v}_0\cdot\nabla\dot{\phi})\\&\quad+(\dot{\mathbf{v}}_0\cdot\nabla\phi)-2\dot{\phi}\mathbf{v}_0\cdot\nabla(\log c_0)-2(\dot{\mathbf{v}}_0\cdot\nabla\phi)c_0^{-1}\dot{c}_0\\&\quad+(\mathbf{v}_0\cdot\nabla)(\mathbf{v}_0\cdot\nabla\phi)-2(\mathbf{v}_0\cdot\nabla\phi)\mathbf{v}_0\cdot\nabla(\log c_0)=0\end{aligned} \tag{10}$$

as follows (i) the first two terms form the classical wave equation (1); (ii) the third and fourth terms correspond to an inhomogeneous, unsteady medium at rest; (iii) the fifth term accounts for uniform low Mach number convection; (iv) the sixth to eighth terms includes inhomogeneous, unsteady low Mach number mean flow; (v) the ninth term represents uniform high Mach number convection and (vi) the tenth term includes non-uniform high Mach number mean flow.

2.3 Class II: Non-linear, Non-dissipative Sound in a Potential Mean Flow

The starting point is the exact continuity equation:

$$\nabla^2\Phi=\nabla\cdot\mathbf{V}=\frac{1}{\Gamma}\frac{\mathbf{D}\Gamma}{\mathrm{d}t}=\frac{1}{\Gamma C^2}\frac{\mathrm{D}P}{\mathrm{d}t}, \tag{11}$$

where is Φ total potential, $\mathbf{V}$ the total velocity, P the total pressure, Γ the total mass density and C the total sound speed:

$$C^2=c_*^2-(\gamma-1)\left[\dot{\Phi}+\frac{1}{2}(\nabla\Phi)^2\right], \tag{12}$$

where c_* denotes the stagnation sound speed. For homentropic flow with enthalpy H:

$$\frac{1}{\Gamma}\frac{\mathrm{D}P}{\mathrm{d}t}=\frac{\mathrm{D}H}{\mathrm{d}t}=\dot{H}+\mathbf{V}\cdot\nabla H=\dot{H}+\nabla\Phi\cdot\nabla H, \tag{13}$$

the Bernoulli equation

$$H+\dot{\Phi}+\frac{1}{2}(\nabla\Phi)^2=\text{const}, \tag{14}$$

leads to the exact potential equation

$$\ddot{\Phi}-\left\{c_*^2-(\gamma-1)\left[\dot{\Phi}+(\nabla\Phi)^2/2\right]\right\}\nabla^2\Phi\quad+2\Phi\cdot\nabla\dot{\Phi}+\nabla\Phi\cdot[(\nabla\Phi\cdot\nabla)\nabla\Phi]=0, \tag{15}$$

which includes non-linear terms up to the fourth-order. Assuming that the mean flow is non-uniform but steady (it is not possible to distinguish non-linear waves from a non-uniform, unsteady mean state), the exact wave equation is:

$$\frac{\mathrm{D}}{\mathrm{d}t}\left(\frac{\delta\phi}{\delta t}\right) - \left[c_*^2 - (\gamma-1)\frac{\delta\phi}{\delta t}\right]\nabla^2\phi - c_*^2\nabla\phi\cdot\nabla(\log\rho_0) = 0, \tag{16}$$

where the exact (17a) and self-convected (17b) material derivatives are used:

$$\frac{\mathrm{D}}{\mathrm{d}t} = \frac{\partial}{\partial t} + \mathbf{v}_0\cdot\nabla + \nabla\phi\cdot\nabla, \tag{17a}$$

$$\frac{\delta}{\delta t} = \frac{\partial}{\partial t} + \mathbf{v}_0\cdot\nabla + \frac{1}{2}\nabla\phi\cdot\nabla. \tag{17b}$$

There are six particular cases (W10–W15) of the wave equation (16); it has 15 terms:

$$\begin{aligned} 0 = {} & \ddot{\phi} - c_0^2\nabla\phi - c_0^2\nabla\phi\cdot\nabla(\log\rho_0) + 2(\mathbf{v}_0\cdot\nabla\dot{\phi}) - 2\dot{\phi}\mathbf{v}_0\cdot\nabla(\log c_0) \\ & + (\mathbf{v}_0\cdot\nabla)(\mathbf{v}_0\cdot\nabla\phi) - 2(\mathbf{v}_0\cdot\nabla\dot{\phi})\mathbf{v}_0\cdot\nabla(\log c_0) + (\gamma-1)\nabla^2\phi + 2\nabla\dot{\phi}\cdot\nabla\phi \\ & + \nabla\phi\left[(\mathbf{v}_0\cdot\nabla)\nabla\phi\right] + \mathbf{v}_0\left[(\nabla\phi\cdot\nabla)\nabla\phi\right] + \nabla\phi\left[(\nabla\phi\cdot\nabla)\mathbf{v}_0\right] \\ & - (\nabla\phi)^2\mathbf{v}_0\cdot\nabla(\log c_0) + \frac{\gamma-1}{2}(\nabla\phi)^2\nabla^2\phi + \nabla\phi\cdot\left[(\nabla\phi\cdot\nabla)\nabla\phi\right] \end{aligned} \tag{18}$$

(i) the first three coincide with the classical wave equation for linear waves in a steady inhomogeneous medium at rest; (ii) the fourth to seventh terms apply to a linear waves, in a moving medium (10); (iii) the eighth to eleventh terms account for quadratic non-linearities in a homogeneous medium; (iv) the twelfth and thirteenth terms include quadratic non-linearities in an inhomogeneous medium and (v) the 14th and 15th terms show that the highest-order non-linearities are cubic [7, 8].

2.4 Class III: Linear, Non-Dissipative Sound in a Quasi-one-Dimensional Duct

Consider (Fig. 1) a straight duct with longitudinal coordinate x, non-uniform cross-section $A(x)$, steady shape (no coupling to elastic walls), containing a one-dimensional mean flow (which is always potential).

The Lagrangian per unit length:

$$\pounds^* = A\pounds = \frac{1}{2}\rho_0 A\left[\phi'^2 - c_0^{-2}(\dot{\phi} + v_0\phi')^2\right], \tag{19}$$

where prime denotes derivative with regard to x, viz. $\phi' \equiv \partial\phi/\partial x$, leads to the high-speed wave equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{c_0^2}\frac{\mathrm{d}\phi}{\mathrm{d}t}\right) - \frac{1}{\rho_0 A}(\rho_0 A\phi')' = 0, \tag{20}$$

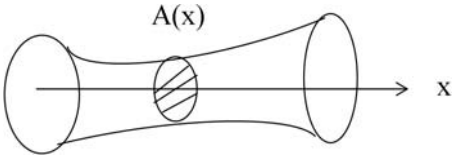


Fig. 1. Quasi-one-dimensional propagation in duct of varying cross-section

which is similar to a one-dimensional form of Class I, replacing in (10) the mass density ρ_0 per unit volume by the mass of fluid per unit length $\rho_0 A$; it has nine particular cases (W16–W24). It consists of 11 terms:

$$\begin{aligned} 0 = \ddot{\phi} - c_0^2 \left[\phi'' + \phi'(A'/A + \rho_0'/\rho_0)\right] - 2\dot{\phi}\dot{c}_0/c_0 + 2v_0\phi' + \dot{v}_0\phi' \\ -2\dot{\phi}v_0 c_0'/c_0 - 2v_0\phi' c_0'/c_0 + v_0(v_0\phi') - 2v_0^2\phi' c_0'/c_0, \end{aligned} \tag{21}$$

namely (i) the first five apply to a horn [9–12] i.e. a duct of non-uniform cross-section without flow and (ii) the last six to a nozzle [13–17], i.e. a duct of non-uniform cross-section with mean flow.

2.5 Class IV: Non-Linear, Non-Dissipative Sound in a Quasi-One-Dimensional Duct

The combination of non-linearity (Class II) with a duct of non-uniform cross-section (Class III), leads to (Class IV) which has six particular cases (W25–W30). The most general is non-linear high-speed nozzle wave equation:

$$\frac{\mathrm{D}}{\mathrm{d}t}\left(\frac{\delta\phi}{\delta t}\right) - \left[c_0^2 - (\gamma-1)\frac{\delta\phi}{\delta t}\right]\left(\phi'' + \phi'\frac{A'}{A}\right) - c_0^2\phi'\frac{\rho_0'}{\rho}, \tag{22}$$

where (i) the first term involves the non-linear (17a) and self-convected (17b) material derivatives; (ii) the second term has as a factor (12) the non-linear sound speed; (iii) the remaining factor in the second term is the Laplacian replaced by duct wave operator [9–12] and (iv) the last term applies to an inhomogeneous medium. The most general wave equation (W30) of Class IV is

$$\begin{aligned} 0 = \ddot{\phi} - c_0^2\phi'' - c_0^2\phi' A'/A - c_0^2\phi'\rho_0'/\rho_0 - 2\dot{\phi}v_0 c_0'/c_0 + v_0(v_0\phi')' \\ -2v_0^2\phi' c_0'/c_0 + 2\dot{\phi}\dot{\phi}' + (\gamma-1)\phi'\phi'' + (\gamma-1)\dot{\phi}\phi' A'/A \\ +\phi'(v_0\phi')'v_0 + \phi'\phi'' + (\gamma-1)v_0\phi'^2 A'/A - \phi'^2 v_0 c_0'/c_0 \\ +\frac{\gamma-1}{2}\dot{\phi}\phi'^2 + \frac{\gamma-1}{2}\phi'^2\phi'' + \frac{\gamma-1}{2}\phi'^3\frac{A'}{A}, \end{aligned} \tag{23}$$

has 17 terms (i) the first four apply to linear waves in an inhomogeneous horn; (ii) the terms five to seven concern linear waves in an inhomogeneous nozzle; (iii) the terms eight to ten specify quadratic non-linearities in a horn; (iv) the terms 11–14 represent quadratic non-linearities in a nozzle and (v) the terms 15–17 represent cubic non-linearities. Note that all all cases of potential flows have been covered as shown in Table 1.

Table 1. Acoustics of potential flows

Waves	Free-space	1-D ducts
Linear	Class I	Class III
Non-linear	Class II	Class IV

2.6 Class V: Acoustic Waves in a Unidirectional Shear Flow

The acoustics of vortical flows is considered next, in the particular cases of (Sect. 2.5) shear flows [18–24]; (Sect. 2.6) rotating flows [25–27]. In both cases, since the mean flow is vortical, there is no acoustic potential; the scalar wave equation is obtained for the acoustic pressure perturbation. In a potential mean flow the are two acoustic modes plus decoupled vorticity (by Kelvin's theorem); in a vortical mean flow the sound couples to vorticity leading to a third-order wave equation. The simplest shear flow is unidirectional (24a) and leads to a material derivative (24b):

$$\mathbf{v}_0 = U(y,z)\mathbf{e}_x, \tag{24a}$$

$$\mathrm{d}/\mathrm{d}t = \partial/\partial t + U(y,z)\partial/\partial x. \tag{24b}$$

The acoustic wave equation, for acoustic pressure in unidirectional shear flow, has four (W31–W34) particular cases and consist of four terms:

$$\begin{aligned} 0 = \frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{1}{c_0^2}\frac{\mathrm{d}^2 p}{\mathrm{d}t^2} - \nabla^2 p\right] + \frac{\mathrm{d}}{\mathrm{d}t}\left[\nabla p \cdot \nabla(\log \rho_0)\right] \\ +2\rho_0\left(\frac{\partial U}{\partial y}\frac{\partial^2 p}{\partial x \partial y} + \frac{\partial U}{\partial z}\frac{\partial^2 p}{\partial x \partial y}\right) \end{aligned} \tag{25}$$

as follows (i) the first two coincide with the convected wave equation (9) for homentropic flow without shear; (ii) the third term applies to isentropic, non-homentropic flow [compare with (10)] without shear and (iii) the fourth term shows that the presence of shear in the mean flow leads to a third-order wave equation.

2.7 Class VI: Acoustics of Sheared and Swirling Axisymmetric Mean Flow

For a rotating fluid, assuming an axisymmetric mean flow and using cylindrical coordinates, the mean flow velocity:

$$\mathbf{v}_0(r) = U(r)\mathbf{e}_z + r\Omega(r)\mathbf{e}_\theta, \tag{26}$$

consists of an axial shear and azimuthal rotation. There are 12 forms of the acoustic wave equation (W36–W60); the most general (W50) reduces in the low Mach number swirl and shear approximation:

$$(\Omega + \mathrm{d}U/\mathrm{d}r)^2 \ll r^2[c_0(r)]^2 \tag{27}$$

to the form:

$$0 = \frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{1}{c_0^2}\frac{\mathrm{d}^2 p}{\mathrm{d}t^2} - \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial p}{\partial r}\right) - \frac{1}{r^2}\frac{\partial^2 p}{\partial \theta^2} - \frac{\partial^2 p}{\partial z^2}\right] + 2\frac{\mathrm{d}U}{\mathrm{d}r}\frac{\partial^2 p}{\partial z \partial r} + 2\frac{\mathrm{d}\Omega}{\mathrm{d}r}\left(\frac{\partial^2 p}{\partial \theta \partial r} - \frac{1}{r}\frac{\partial p}{\partial \theta}\right) + \frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{\partial p}{\partial r}\frac{\partial}{\partial r}(\log \rho_0)\right] + \frac{2\Omega}{r}\frac{\partial p}{\partial \theta}\frac{\partial p}{\partial r}\frac{\partial}{\partial r}(\log \rho_0) \tag{28}$$

consisting of five terms (i) the first term corresponds to the convected wave operator (9) in cylindrical coordinates; (ii) the second term accounts for sheared mean flow [compare with (25)]; (iii) the third term corresponds to swirling mean flow and (iv) the fourth and fifth terms represent isentropic, non-homentropic mean flow.

2.8 Class VII: Viscous and Resistive Dissipation of Linear Sound

The magnitude of the viscous dissipation of sound is comparable to that for thermal conduction so both must be considered. The vorticity decouples and satisfies diffusion equation

$$\mathbf{\Omega} \equiv \nabla \times \mathbf{v}, \tag{29a}$$

$$\dot{\mathbf{\Omega}} = \nu \nabla^2 \mathbf{\Omega}, \tag{29b}$$

showing that it is dissipated only by shear viscosity ν; thus it is sufficient to consider a wave equation only for the dilatation $\Psi = \nabla \cdot \mathbf{v}$; the latter is dissipated by shear viscosity ν, bulk viscosity β and thermal conductive diffusivity α. There are two particular cases (W51 and W52) of the linear dissipative acoustic wave equation in an homogeneous medium at rest; it consists of five terms:

$$\dddot{\Psi} - c_0^2 \nabla \dot{\Psi} = (4\nu/3 + \beta + \alpha)\nabla^2 \ddot{\Psi} - c_0^2(\alpha/\gamma)\nabla^4 \Psi - \alpha(4\mu/3 + \beta)\nabla^4 \dot{\Psi} \tag{30}$$

(i) the first two correspond to the classical wave equation (1) differentiated to the third-order in time; (ii) the third term corresponds to small diffusities and is of the second-order in space and time; (iii) the fourth term is of the fourth-order in space and involves the adiabatic exponent γ and (iv) the last term involves the product of diffusivities, so it applies to large diffusivities and is of the fourth-order in space and first-order in time.

2.9 Class VIII: One-dimensional Viscous Non-Linear Waves in a Quasi-One-Dimensional Duct of Variable Area

The quadratic non-linearities are sufficient to lead to wave front steeping and shock formation; the linear dissipation opposes this, leading to shock widening