International Center for Mathematics

#31

Bulletin

January 2012

Editorial

Dear Colleagues,
CIM is already actively engaged in developing important and exciting activities for mathematicians with the support of the associated members.

To start, we are pleased to announce *The Summer School on Algebraic and Enumerative Combinatorics* that will be held in July, 2–13, 2012. The school will consist of four courses lead by mathematicians Francesco Brenti, Christian Krattenthaler, Marc Noy and Vic Reiner. While targeted to graduate students and post-doctorate researchers, the content will also prove beneficial to other level researchers, as well. In addition to the formal courses, there will also be short talks presented by summer school participants. The program will be hosted in the beautiful *Centro de Estudos Camilianos*, designed by the 1992 Laureate of the *Pritzker Architecture Prize* Álvaro Siza.

Another significant announcement is CIM's inaugural launch of the *Pedro Nunes Medal*. In 2011, the steering committees of CIM and SPM joined together to develop the proposal, which was enthusiastically supported. Later this year we will begin the process to convene an international panel to solicit and select the first recipients. Selection of medalists will be based on the meritorious research accomplishments with an emphasis on contributions to the development of mathematics in Portugal. Additional details will be provided in the near future.

The final activity I would like to bring to your attention is CIM's upcoming contribution to the international program *Mathematics of Planet Earth* (MPE 2013 — www.crm.umontreal.ca/Math2013/en/). To support this global effort, CIM is organizing two international conferences and corresponding advanced schools: *Planet Earth, Mathematics of Energy and Climate Change*, 25–27 March 2013, with the *Advanced School Planet Earth, Mathematics of Energy and Climate Change*, 18–23 March and 27–28 March 2013; and *Planet Earth, Dynamics, Games and Science*, 2–4 September 2013, with the *Advanced School Planet Earth, Dynamics, Games and Science*, 26–31 August and 5–7 September 2013.

Each international conference will have around 15 keynote speakers and 30 thematic sessions for a total of approximately 100 speakers. The invited keynote speakers are top researchers in mathematics and science whose discoveries have had a great impact on the progress of science. The *Portuguese Society of Mathematics* (SPM), the *Portuguese Society of Statistics* (SPE) and the *Portuguese Society of Operational Research* (APDIO) enthusiastically support the conferences and advanced schools we are organizing for MPE-2013. The *Advanced School Planet Earth, Mathematics of Energy and Climate Change* will be hosted at Faculdade de Ciências, Universidade Lisboa (FCUL). The *Advanced School Planet Earth, Dynamics, Games and Science* will be hosted at Escola Superior de Economia e Gestão, Universidade Técnica de Lisboa (ISEG-UTL). The two international conferences will be hosted in Calouste Gulbenkian Foundation.
In closing, on behalf of the CIM Board, I would like to thank you for your continued support and interest.

Alberto Adrego Pinto
President of CIM

Contents

Coming Events

4th Porto Meeting on Mathematics for Industry

Porto, Portugal
June 07-09, 2012
[http://cmup.fc.up.pt/cmup/mathindustry/2012/]

The purpose of this meeting is to focus the attention on the many and varied opportunities to promote applications of mathematics to industrial problems. Its major objectives are:

- Development and encouragement of industrial and academic collaboration, facilitating contacts between academic, industrial, business and finance users of mathematics.
- Through "bridging the industrial/academic barrier" these meetings will provide opportunities to present successful collaborations and to elaborate elements such as technology transfer, differing vocabularies and goals, nurturing of contacts and resolution of issues.
- To attract undergraduate students to distinctive and relevant formation profiles, motivate them during their study, and advance their personal training in Mathematics and its Applications to Industry, Finance, etc.

The meeting will be focused on short courses, of three one-hour lectures each, given by invited distinguished researchers, which are supplemented by contributed short talks by other participants and posters of case studies.

The meeting will be followed by the 86th European Study Group with Industry 2012 that will take place in ISEP—School of Engineering, Polytechnic of Porto, Portugal, between the 7th and the 12th of May 2012.

86th European Study Group with Industry 2012

ISEP—School of Engineering, Polytechnic of Porto, Portugal
May, 7-13, 2012

The purpose of these meetings is to strengthen the links between Mathematics and Industry by using Mathematics to tackle industrial problems, which are proposed by industrial partners. This meeting is part of the series of European Study Groups and will count with the participation of several European experts with a large experience in this type of events.

More information on Portuguese Study Groups is available at http://www.ciul.ul.pt/~freitas/esgip.html, while general information on study groups and related aspects is available at the International Study Groups website:

http://www.maths-in-industry.org,

the Smith Institute:

http://www.smithinst.ac.uk,

and the European Consortium for Mathematics in Industry:

http://www.ecmi-indmath.org/info/events.php.

Summer School on Algebraic and Enumerative Combinatorics,

Centro de Estudos Camilianos—S. Miguel de Seide
July 2-13, 2012

The Summer School on Algebraic and Enumerative Combinatorics will be held in July, 2-13, 2012, at the *Centro de estudos Camilianos*, in a building of Álvaro Siza, the 1992 Laureate of the *Pritzker Architecture Prize*. The *Centro de estudos Camilianos* is in S. Miguel de Seide, near to Guimarães, Portugal, where the participants are expected to be lodged. The school will focus on four courses, given by Francesco Brenti, Christian Krattenthaler, Marc Noy and Vic Reiner. The topics to be addressed by the speakers are, respectively, Combinatorics of Coxeter Groups, Map Enumeration, Asymptotic Enumeration of Topological Graphs and Reflection Group counting and q-counting, and the courses are mainly directed to graduate and post-graduate students, as well as researchers. There will also be time for some contributed short talks by participants.

More information available at:
http://www2.fc.up.pt/pessoas/agoliv/SC/default.htm

Summer School "Dynamic Models in Life Sciences"

The Summer School "Dynamic Models in Life Sciences" was coorganized by the Centro Internacional de Matemática (http://www.cim.pt), the European Society for Mathematical and Theoretical Biology (http://www.esmtb.org) and the European Mathematical Society (http://www.euro-math-soc.eu/) with financial support from the Fundação para a Ciência e a Tecnologia (http://www.fct.pt) and Centro de Matemática e Aplicações (Universidade Nova de Lisboa), Centro de Matemática e Aplicações Fundamentais (Universidade de Lisboa) and Centro de Investigação em Matemática e Aplicações (Universidade de Évora).

These summer schools are organized every year in a different European country. This was the first time Portugal hosted the event.

During one week in the warm weather of Evora in Summer (24th to 30th July 2011), 6 speakers presented the state-of-the art in their respective fields to more than 40 participants coming from 13 different countries. Participants mostly consisted of PhD students and post-docs in Mathematics, Biology and Physics. An important point was the Portuguese presence in the event, larger than expected.

Apart from the mini-courses (listed below), there was a poster session, where students could present their work and receive feedback from leading specialists in the field.

One afternoon was reserved to a visit to the Ducal Palace in Vila Viçosa followed by a traditional Alentejano dinner with local music.

Mini-courses: Dynamical models of Cancer (David Dingli, Mayo Clinic, USA); Adaptive dynamics and the evolution of pathogens (Eva Kisdi, University of Helsinki, Finland); Modelling Meso-evolution: adaptive dynamics and beyond (Hans Metz, Leiden University, The Netherlands); Stochastic and Deterministic Processes in Spatial Population Dynamics (Sergei Petrovskii, University of Leicester, UK); Mathematical Models in Hemodynamics. (Adelia Sequeira, Universidade Técnica de Lisboa, Portugal); Ecology and Eco-epidemiology. (Ezio Venturino, Universitá di Torino, Italy).

Organizers: Fernando Carapau (Évora), Fabio Chalub (Lisbon), Francisco Santos (Lisbon), Nico Stollenwerk (Lisbon).

An Interview

with Carla Gomes

by **JoãoGama** [Fac. de Economia da Univ. do Porto]
and **Márcia Oliveira** [LIAAD-INESC TEC, and FEP, Univ. do Porto]

Carla Gomes is a professor of computer science at Cornell University, with joint appointments in the computer science, information science, and Dyson School of applied economics and management departments. Her research has covered several themes in artificial intelligence and computer science, from the integration of constraint reasoning, operations research, and machine learning techniques for solving large-scale constraint reasoning and optimization problems, to the use of randomization techniques to improve the performance of exact search methods, algorithm portfolios, multi-agent systems, and game play.

Recently, Gomes has become immersed in the establishment of computational sustainability, a new interdisciplinary field that aims to develop computational methods to help balance environmental, economic, and societal needs to support a sustainable future. Gomes has started a number of research projects in biodiversity conservation, poverty mapping, the design of "smart" controls for electric cars, and pattern identification for material discovery (e.g., for fuel cell technology). Gomes obtained a PhD in computer science in the area of artificial intelligence and operations research from the University of Edinburgh. She also holds an MSc in applied mathematics from the Technical University of Lisbon. Gomes is the lead principal investigator on an award from the National Science Foundation's Expeditions in Computing program, the director of the newly established Institute for Computational Sustainability at Cornell, and a fellow of the Association for the Advancement of Artificial Intelligence. Gomes is currently a Fellow at the Radcliffe Advanced Study Institute at Harvard University.

Carla Gomes was an invited speaker at the 10th Intelligent Data Analysis Symposium, held in Porto from 29 to 31 October. The opportunity of having Carla Gomes in Porto motivated the present interview.

How and why did you start working in the field of Computational Sustainability?

Besides being a professor at the computer science and information science departments at Cornell University, I am also a professor at the Dyson School of applied economics and management, which is actually part of the college of agricultural and life sciences. So, at some point, some of my students from Dyson were working on problems concerning the wildlife corridors. Since they were designing corridors using very simplistic approaches, they asked my help in order to design them in a more rigorous way. Besides this, due to the land-grant missions of Cornell University and my previous intensive research on latent squares, I started working on the design of experiments involving fertilizers, using the so called spatially balanced latent squares. The scope of this work was, as I said, within the land-grant missions of Cornell. This means that, a long time ago, the State gave land to the Cornell University and so Cornell, as a counterpart, has to, at some extent, provide services to the community. One of the services is to advise farmers in how to use fertilizers. To do so, Cornell researchers run experiments on different fertilizers and tell farmers the amount of fertilizers they should use depending on the soil, on the weather, etc. Since my all career was based on working on latent squares, one day researchers from the crop and soil science asked my help in doing this. At that time, I thought that would be an easy problem to solve but I found out it was not, since we could only build spatially-balanced latent squares for squares of 6 by 6, which was too small, since they needed to experiment up to 30/35 fertilizers at the same time. And I got intrigued. After thinking deeply about the problem, we were able to make incredible progress and I started getting interested in learning about so many problems for which you really need serious computation. Due to my intersections with the college of agricultural and life sciences, pretty soon I realized that a lot of the problems in these fields involve management of resources, highly dynamic systems and huge volumes of data. For instance, you need to gather data to monitor the environment and then analyze and interpret it. But how can you extract and analyze patterns from such large volumes of data? In fact, I realized that in ecology, in biology and in environmental sciences, a lot of the problems have tremendous computational challenges and there are not that many computer scientists working in this area. That's why NSF (National Science Foundation) held a program called "Expedition in Computing", which aimed at providing grants to truly transformative research in computer science, that would set a new research direction for the field. Ideally, also with broad societal impact. This was a very competitive program and the entire USA submitted proposals (MIT, Stanford, Berkeley, Cornell, etc.). After several phases, they ended up selecting three proposals. One of them was our proposal. Since the topics of this program were really up to the candidates, I proposed to create this new field of Computational Sustainability and invest dramatically in terms of research in this area, especially because the computational challenges are so dramatic.

How Computational Sustainability will be transformative in terms of the impact in the future?

From NSF perspective, they were very much interested in programs that would be transformative in terms of computer science, but also with broad impact. From our perspective, we believe that Computational Sustainability can be transformative in the sense that the issues concerning sustainability are really deep research questions which often force us to look at aspects, such as dynamics, that we have not encountered when studying other problems. Since sustainability means planning today and thinking in terms of the future, you absolutely need to consider dynamics. So, basically, by looking at these problems, computer scientists are exposed to new issues that they haven't really worked with before and, therefore, we really need to have methodological advances to address them. This is the main reason why our own research is centered on these three topics: "dynamical systems", "constraint reasoning and optimization and statistics", "machine learning and data mining" (Figure 1). The most interesting thing in this is that the problems in each area per se really push the frontiers of the state-of-the-art today, and if we work at the synthesis of methodologies from different areas, this will lead to fundamental new methodological developments in the field. For example, the problem of material discovery that I've talked about in my presentation at IDA 2011, is a good example where data mining and machine learning per se are not going to solve the problem. Optimization per se is not going to solve the problem. You really need to develop methodologies that bring together ideas from different fields. And that it is really exciting.

People from optimization and from dynamical systems are model-based. In Machine Learning and Data Mining models are generated from data. There are different ways of using data and different mental attitudes.

My background is more in terms of optimization and reasoning but I realized that all these real-world models need to learn the parameters from older models. So, what you would say it is a machine learning problem, maybe I would say it is what we call an in voice optimization problem, in the sense that you know there's an optimization problem and now you want to know the parameters that minimize or maximize something. What's interesting is that all the models rely on data, so this artificial separation between optimization and machine learning does not make much sense. For instance, in the decision making process, one of the steps is "data acquisition". Here, you have sensors that collect data, for instance, about flight calls or signals, but what you want to know is species. So, after collecting data, you need to interpret it. To do so, it is common to perform model fitting. But, sometimes, you need to go back because you realize that your initial samples are very biased, and so you need to collect more data to increase the accuracy of the models. Another important step is related to policy optimization: I am going to make decision in terms of what areas to protect for the birds, which is based on the data I got. And, because of data, I may even have to go back to define my procedures in terms of collecting data, and even to reformulate my own objectives, because this project is not linear and gives you a lot of feedback.

This is a very iterative process, even though in terms of disciplines you have people doing machine learning, optimization, etc., but clearly there are interactions that have to be effective.

Your research topic – Computational Sustainability - involves very different areas.

Yes, we have a very interdisciplinary group, which involve mathematicians, biologists, sociologists, computer scientists, and that's exactly the point of our project.

You succeed in a very competitive society. What are your recommendations for those who aspire to become leaders on their research field?

Be passionate about your research. Always set high-standards, have the courage to ask and address very hard and risky questions, since those are the most rewarding when you succeed. Be driven and committed. Work hard, very hard. Focus on important questions and not trivial ones, and be obsessed about that. For women, you need to be really confident and positive, since they tend to think they are not good enough. Challenge yourself. In terms of research topics it is very important to know what the community is doing and to be aware of the hard topics, not only for you to follow the literature, but at the same time for you to have a chance to pose different questions, but related to those of the community. Sometimes you do not really know the answers to your problems, but you need to be able to deal with this kind of uncertainty. One thing that is very important is collaborations and networking. That it is how you make progress in research, how you bridge together areas that are completely different and that will generate fundamental new ways to solve the problems. For the young people, it is important to network, go to conferences, to try to interact with researchers, to actively look for collaborations, to get involved in research projects, to network a lot, to travel a lot, to be on program committees, etc. Basically, you need to learn how to "sell your work", which means that it is important to do great work, but it is also very important to be able to talk about your work in a way that's going to be easy to communicate with people. You need to learn how to give talks that are going to be appealing, that people find exciting, so they can get interested and follow up on your work. Write beautiful papers. Publish a lot, otherwise you may perish. It is often good to do a Pos-doc so that you can go and collaborate and get a lot of research going. Finally, make your own luck, your serendipity, and create opportunities by interacting, collaborating and doing a lot of things, since as Louis Pasteur said "chance favors the prepared minds".

Sympletic surface group representations and Higgs bundles

by **Peter Gothen***

1. Introduction

A surface group is the fundamental group of a surface. In this article we survey some results on representations of a surface group on a real vector space preserving a symplectic form. We emphasize in particular some results which have been obtained using holomorphic and algebraic geometry, through the use of Higgs bundles and a fundamental result known as the non-abelian Hodge Theorem. Though this theory itself is rather involved, the results on surface group representations can be explained without bringing it into play and this is one of our main aims.

This paper is organized as follows. After some preliminaries, we start by focusing on the case of representations in $\mathbb{R}^2$ with its standard symplectic form. Here we explain some seminal results of W. Goldman which are closely related to uniformization of surfaces by the hyperbolic plane.

We then move on to higher dimensional representations and explain some results which generalize those of Goldman and also point out some differences with the 2-dimensional situation.

Finally, we briefly outline how methods from holomorphic and algebraic geometry can be applied to the study of surface group representations through Higgs bundles and the non-abelian Hodge Theorem. This beautiful theory involves algebra, geometry, topology and analysis and has a long history. A few important milestones can be found in the work of Narasimhan-Seshadri [22], Atiyah-Bott [1], Donaldson [5], Hitchin [17], Corlette [4] and Simpson [24].

We have left out many important and fascinating aspects of surface group representations. To finish this introduction we mention a few places where the interested reader may find further information and references and also other points of view. Nice surveys are provided in Goldman [14] (emphasizing the point of view of geometric structures on surfaces) and Burger-Iozzi-Wienhard [3] (emphasizing methods from bounded cohomology). For an application of Higgs bundle theory to representations in isometry groups of hermitian symmetric spaces of the non-compact type, see the survey [2].

* CMUP, Faculdade de Ciências da Universidade do Porto

2. Surface group representations and character varieties

Let Σ be a compact oriented surface without boundary of genus g. The fundamental group of Σ has the standard presentation

$$\pi_1\Sigma = \langle a_1, b_1, \dots, a_g, b_g \mid \prod_{1\le i\le g}[a_i, b_i] = 1\rangle$$

in terms of generators and relations.

Let G be a connected semisimple Lie group. (2.1) In this paper we are mainly interested in the case when $G = \mathrm{Sp}(2n, \mathbb{R})$ is the real symplectic group but we shall also have occasion to consider the cases $G = \mathrm{GL}(n, \mathbb{C})$ and $G = \mathrm{PSL}(2, \mathbb{R}) := \mathrm{SL}(2, \mathbb{R})/\{\pm I\}$. Since all of these groups are defined via a linear action on a vector space, the motivation for the following definition is clear

Definition 2.1.—A *representation* of $\pi_1\Sigma$ in G is a homomorphism

$$\rho : \pi_1\Sigma \to G.$$

In view of (2.1) a representation ρ is uniquely prescribed by a $2g$-tuple $(A_1, B_1, \dots, A_g, B_g)$ of matrices in G satisfying the relation $\prod[A_i, B_i] = 1$. Thus, if we denote the set of all representations by

$$\mathrm{Hom}(\pi_1\Sigma, G) = \{\rho : \pi_1\Sigma \to G\}.$$

we get an identification

$$\mathrm{Hom}(\pi_1\Sigma, G) \cong \left\{(A_1, B_1, \dots, A_g, B_g) \mid \prod_{1\le i\le g}[A_i, B_i] = 1\right\} \subset G^{2g}.$$

with a subspace of the set of $2g$-tuples of matrices in G. (2.2)

3. Fuchsian representations

Consider the upper half plane model of the hyperbolic plane

$$\mathbb{H}^2 = \{z = x + iy \mid y > 0\}.$$

The metric is $ds^2 = (dx^2 + dy^2)/y^2$ which has constant curvature -1. The group of orientation preserving isometries of $\mathbb{H}^2$ can be identified with $\mathrm{PSL}(2, \mathbb{R})$, acting on $\mathbb{H}^2$ via

Möbius transformations:

$$z \mapsto \frac{az+b}{cz+d}.$$

for a ×2-matrix

$$A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}(2,\mathbb{R}).$$

A subgroup $\Gamma \subset \mathrm{PSL}(2,\mathbb{R})$ is *Fuchsian* if it is discrete. In this case the orbit space $\mathbb{H}^2/\Gamma$ is a surface of constant negative curvature. If $\mathbb{H}^2/\Gamma$ is compact, it must have genus at least 2, as follows from the Gauss-Bonnet Theorem.

Conversely, assume that Σ is a compact oriented surface without boundary of genus $g \geq 2$. Then Σ admits a hyperbolic metric and is therefore locally isometric to $\mathbb{H}^2$. The local isometries patch together to give the globally defined *developing map*

$$\tilde{\Sigma} \to \mathbb{H}^2,$$

where $\tilde{\Sigma} \to \Sigma$ is the universal cover. This map is a homeomorphism and therefore we obtain a Fuchsian representation

$$\rho : \pi_1\Sigma \to \mathrm{PSL}(2,\mathbb{R})$$

an a corresponding isometry

$$\Sigma \cong \mathbb{H}^2/\pi_1\Sigma.$$

4. Reductive representations and the character variety

Suppose that G is a linear group with a defining fundamental representation V, such as all of the previously mentioned groups (with the exception of $\mathrm{PSL}(2,\mathbb{R})$). It is then clear what we should mean by a *reductive* (or *semisimple*) representation. Namely, it should be one for which the fundamental representation V is semisimple, i.e., such that each invariant subspace has an invariant complement.[1] We denote by

$$\mathrm{Hom}^+(\pi_1\Sigma, G) \subset \mathrm{Hom}(\pi_1\Sigma, G)$$

the subspace of reductive representations.

Of course we should consider representations equivalent if they correspond under some change of basis in the fundamental representation V. Therefore we make the following definition.

Definição 4.1.—Representations ρ_1 and ρ_2 are *isomorphic* if there exists a $g \in G$ such that

$$\rho_1(\gamma) = g\rho_2(\gamma)g^{-1} \qquad \text{for all } \gamma \in \pi_1\Sigma.$$

We wish to consider the set of all isomorphism classes of representations. For technical reasons, which we shall explain below, we restrict attention to reductive representations.

Definição 4.2.—The *character variety* for representations of $\pi_1\Sigma$ in G is the orbit space

$$\mathscr{R}(\pi_1\Sigma, G) = \mathrm{Hom}^+(\pi_1\Sigma, G)/G,$$

where G acts by overall conjugation:

$$g \cdot \rho(\gamma) = g\rho(\gamma)g^{-1} \qquad \text{for } \gamma \in \pi_1\Sigma. \tag{4.1}$$

Since by (2.2), the space $\mathrm{Hom}^+(\pi_1\Sigma, G)$ is contained in G^{2g} it has a natural topology and we give $\mathscr{R}(\Sigma, G)$ the quotient topology. The restriction to reductive representations makes it possible to show that in this topology the character variety is Hausdorff.

There is a very natural notion of deformation equivalence of representations $\rho : \pi_1\Sigma \to G$ which can be conveniently encoded in the language of character varieties. Two representations ρ_0 and ρ_1 are said to be *deformation equivalent* if there is a continuous family of representations $\rho_t : \pi_1\Sigma \to G$, $t \in [0,1]$ connecting them. Since G is connected we have the following result.

Proposition 4.3.—Two representations ρ_0 and ρ_1 are deformation equivalent if and only if the points they represent in the character variety $\mathscr{R}(\pi_1\Sigma, G)$ belong to the same connected component.

Thus, if we wish to classify representations of $\pi_1\Sigma$ up to deformation equivalence, we are actually looking to determine the set of path connected components

$$\pi_0(\mathscr{R}(\pi_1\Sigma, G)).$$

5. Invariants of representations

Let $\rho : \pi_1\Sigma \to G$ be a representation. We shall associate an invariant $c(\rho) \in \pi_1 G$ as follows. Let $\widetilde{G}$ be the universal covering group of G. Then we have an exact sequence

$$1 \to \pi_1 G \to \widetilde{G} \xrightarrow{p} G \to 1.$$

Take elements $\tilde{A}_i, \tilde{B}_i \in \widetilde{G}$ such that

$$p(\tilde{A}_i) = \rho(a_i) \quad \text{and} \quad p(\tilde{B}_i) = \rho(b_i).$$

The invariant is then defined as

$$c(\rho) = \prod_{i=1}^{g} [\tilde{A}_i, \tilde{B}_i] \in \pi_1 G. \tag{5.1}$$

Let $H \subseteq G$ be a maximal compact subgroup. Then G retracts onto H and hence the invariant takes values in $\pi_1 H \cong \pi_1 G$. For example, if $G = \mathrm{SL}(2,\mathbb{R})$, we have defined an integer invariant

$$c(\rho) \in \pi_1 \mathrm{SO}(2) \cong \mathbb{Z}.$$

In this case the invariant $c(\rho)$ is known as the *Toledo invariant*.

Remark 5.1.—An equivalent definition of the invariant can be given as follows. Given a representation $\rho : \pi_1\Sigma \to G$, let $E_\rho = \tilde{\Sigma} \times_{\pi_1\Sigma} G$ be the corresponding flat principal G-bundle. The invariant is defined to be the characteristic class $c(\rho) \in H^2(\Sigma, \pi_1 G) \cong \pi_1 G$ which classifies topological G-bundles.

It is clear that isomorphic representations have the same Toledo invariant. Hence we can define the subspace of the character variety consisting of representations of Toledo invariant $d \in \mathbb{Z}$ to be

$$\mathscr{R}_d(\pi_1\Sigma, (2,)) = \{[\rho] \mid c(\rho) = d\}.$$

6. The Milnor-Wood inequality and Goldman's Theorem

For the remainder of this article, we shall assume that $g \geq 2$.

A classical theorem of Milnor [20] states that not all integers are possible values for the Toledo invariant $c(\rho)$ of a representation $\rho : \pi_1\Sigma \to \mathrm{SL}(2,\mathbb{R})$. To be precise, the following, usually known as the *Milnor-Wood inequality*, holds

$$|c(\rho)| \leq g - 1. \tag{6.1}$$

It is natural to ask whether there is any connection between the Toledo invariant of a representation and its geometric properties, such as being Fuchsian. The following theorem of Goldman answers this affirmatively.

Theorem 6.1 [Goldman [11,12]].—A representation

$$\rho : \pi_1\Sigma \to \mathrm{SL}(2,\mathbb{R})$$

is Fuchsian if and only if $c(\rho) = g - 1$.

Remark 6.2.—One might ask what is the significance of the sign of the Toledo invariant. Define the matrix

$$T = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \in \mathrm{GL}(2,\mathbb{R}).$$

Note that $|T| = -1$. One can check that conjugation takes representations with Toledo invariant d to representations with Toledo invariant $-d$, in other words, $c(T^{-1}) = -c(\rho)$. Hence there is an identification

$$\mathscr{R}_d(\pi_1\Sigma, \mathrm{SL}(2,\mathbb{R})) \cong \mathscr{R}_{-d}(\pi_1\Sigma, \mathrm{SL}(2,\mathbb{R})) \tag{6.2}$$

and, whenever convenient, we can restrict attention to $d \geq 0$.

Representations $\rho : \pi_1\Sigma \to \mathrm{SL}(2,\mathbb{R})$ which satisfy $|c(\rho)| = g - 1$ are called *maximal*.

The question of deformation equivalence of representations into $\mathrm{SL}(2,\mathbb{R})$ was also answered by Goldman.

Theorem 6.3 [Goldman [13]].—For any d with $|d| < g - 1$, the space $\mathscr{R}_d(\pi_1\Sigma, \mathrm{SL}(2,\mathbb{R}))$ is connected. If $|d| = g - 1$, the space $\mathscr{R}_d(\pi_1\Sigma, \mathrm{SL}(2,\mathbb{R}))$ has 2^{2g} connected components.

Goldman also proved that each of the 2^{2g} connected components of $\mathscr{R}_{g-1}(\pi_1\Sigma, \mathrm{SL}(2,\mathbb{R}))$ project isomorfically onto a unique connected component of $\mathscr{R}(\pi_1\Sigma, \mathrm{PSL}(2,\mathbb{R}))$ under the natural map

$$\mathscr{R}(\pi_1\Sigma, \mathrm{SL}(2,\mathbb{R})) \to \mathscr{R}(\pi_1\Sigma, \mathrm{PSL}(2,\mathbb{R})).$$

Recall that the *Teichmüller space* $\mathscr{T}$ of Σ may be viewed as the space of hyperbolic structures.[2] Thus, in the light of the discussion in Section 3,

$$\mathscr{T} = \Big\{\rho : \pi_1\Sigma \to \mathrm{PSL}(2,\mathbb{R}) \mid \rho \text{ is Fuchsian}\Big\}/\mathrm{PSL}(2,\mathbb{R}),$$

where $\mathrm{PSL}(2,\mathbb{R})$ acts by overall conjugation as in (4.1). Hence it follows from Goldman's Theorem that each of the components of maximal representations can be identified with Teichmüller space.

7. Representations in the symplectic group

Let $(x_1, y_1, \ldots, x_n, y_n)$ be coordinates on $\mathbb{R}^{2n}$. The *real symplectic group* $\mathrm{Sp}(2n,\mathbb{R})$ is the group of linear transformations of $\mathbb{R}^{2n}$ which preserve the standard symplectic form

$$\omega = dx_1 \wedge dy_1 \wedge \cdots \wedge dx_n \wedge dy_n.$$

In particular, $\mathrm{Sp}(2,\mathbb{R}) \cong \mathrm{SL}(2,\mathbb{R})$. It turns out that certain key properties of representations $\pi_1\Sigma \to \mathrm{SL}(2,\mathbb{R})$ generalize to representations in $\mathrm{Sp}(2n,\mathbb{R})$. However there are also some important differences.

Note that the maximal compact subgroup of $\mathrm{Sp}(2n,\mathbb{R})$ is the unitary group $\mathrm{U}(n)$. Hence the topological invariant of representations $\pi_1\Sigma \to \mathrm{Sp}(2n,\mathbb{R})$ takes values in

$$\pi_1 \mathrm{U}(n) \cong \mathbb{Z}.$$

There is also a Milnor-Wood type inequality for representations in the symplectic group, which states that

$$|c(\rho)| \leq n(g-1) \tag{7.1}$$

for any representation $\rho : \pi_1\Sigma \to \mathrm{Sp}(2n,\mathbb{R})$. This inequality—as well as other generalizations—is the result of the work of many people, we mention the general results of Dupont [6] and the result of Turaev [25] which gives (7.1) in its sharp form.

Just as for the case $n = 2$, representations of the form $\pi_1\Sigma \to \mathrm{Sp}(2n,\mathbb{R})$ with $|c(\rho)| = n(g-1)$ are called *maximal*.

The question of deformation equivalence of representations in $\mathrm{Sp}(2n,\mathbb{R})$ for general n so far only has a complete answer for maximal representations. We have the following results.

[1] In general one may define a representation ρ to be semisimple if the linear representation obtained by composing ρ with the adjoint representation $\mathrm{Ad} : G \to \mathrm{Aut}(\mathfrak{g})$ of G on its Lie algebra $\mathfrak{g}$ is semisimple.

THEOREM 7.1 [[15]].—The character variety

$$\mathscr{R}_{2(g-1)}(\pi_1\Sigma, \mathrm{Sp}(4,\mathbb{R}))$$

has $3\cdot 2^{2g}+2g-4$ connected components.

THEOREM 7.2 [[10]].—Assume that $n\geq 3$. Then the character variety

$$\mathscr{R}_{n(g-1)}(\pi_1\Sigma, \mathrm{Sp}(2n,\mathbb{R}))$$

has $3\cdot 2^{2g}$ connected components.

One might expect $\mathscr{R}_d(\pi_1\Sigma, \mathrm{Sp}(2n,\mathbb{R}))$to be connected for $|d|\leq n(g-1)$. However, so far this has only been proved for $n=2$, by García-Prada and Mundet [9].

Some of the components of maximal representations are natural generalizations of Teichmüller space which, as we have seen, appears as the components of maximal representations for $G=\mathrm{SL}(2,\mathbb{R})$. These are known as *Hitchin components* and were first studied by Hitchin [18]. To explain this, write $\mathbb{V}=\mathbb{R}^2$ for the standard 2-dimensional representation of $\mathrm{SL}(2,\mathbb{R})$. The m-fold symmetric power

$$S^m\mathbb{V}\subset \mathbb{V}^{\otimes m}$$

is an irreducible representation of $\mathrm{SL}(2,\mathbb{R})$ of dimension $m+1$. The standard symplectic form $dx_1\wedge dx_2$ on $\mathbb{R}^2$ induces a non-degenerate bilinear form ω on the symmetric power $S^m\mathbb{V}$ which is antisymmetric when m is odd (and symmetric when m is even). Hence, for $m=2n-1$, ω is a symplectic form on $S^m\mathbb{V}\cong^{2n}$ and we have a natural embedding

$$r:\mathrm{SL}(2,\mathbb{R})\hookrightarrow \mathrm{Sp}(2n,\mathbb{R}).$$

DEFINITION 7.3.—A representation $\rho:\pi_1\Sigma\to\mathrm{Sp}(2n,\mathbb{R})$ is called a *Hitchin representation* if it is deformation equivalent to a representation of the form $r\circ\rho_0$, where $\rho_0:\pi_1\Sigma\to\mathrm{SL}(2,\mathbb{R})$ is Fuchsian.

Hitchin [18] proved that there are exactly 2^{2g} connected components of $\mathscr{R}_{n(g-1)}(\pi_1\Sigma,\mathrm{Sp}(2n,\mathbb{R}))$ consisting of Hitchin representations. In complete analogy with the case $n=2$, these components are all homeomorphic to a euclidean space $\mathbb{R}^N$ and projectively equivalent to a unique connected component of representations in $\mathrm{PSp}(2n,\mathbb{R})$. These components are the Hitchin components referred to above. However, in contrast to the case of $n=1$, non-Hitchin components exist for $n\geq 2$, as follows from the result of Hitchin just mentioned and Theorems 7.1 and 7.2.

There are other ways in which maximal representations in $\mathrm{Sp}(2n,\mathbb{R})$ share properties with representations in $\mathrm{SL}(2,\mathbb{R})$. Recall that the mapping class group of Σ acts properly discontinuously on Teichmüller space. Generalizing this fact, it was proved by Labourie [19] and Wienhard [26], that the mapping class acts properly discontinuously on the whole space $\mathscr{R}_{n(g-1)}(\pi_1\Sigma,G)$ of maximal representations.

8. HIGGS BUNDLES

In this final section we briefly outline how methods from holomorphic and algebraic geometry provide insights leading to some of the above mentioned results on surface group representations.

The first step is to equip the surface Σ with a complex structure, i.e. local coordinate systems taking values in $\mathbb{C}$ with biholomorphic coordinate changes. This makes Σ into a Riemann surface which we shall denote by X.

We shall assume that the reader is familiar with the basic language of complex manifolds and holomorphic bundles (see, e.g., Miranda [21] or Griffiths-Harris [7]). However, we briefly recall a couple of central notions.

Let $E\to X$ be a rank n *holomorphic vector bundle*. Roughly speaking, this is a holomorphic family of complex vector spaces E_x parametrized by $x\in E$ which locally looks like the trivial product family $X\times\mathbb{C}^n$. The *rank of* E, denoted by $\mathrm{rk}(E)$ is the dimension of the vector spaces E_x. A holomorphic vector bundle of rank one is called a *line bundle*.

The *determinant bundle* $\det(E)$ of a rank n vector bundle $E\to X$ is a holomorphic line bundle naturally associated to E. It has the property that there is a canonical identification of fibres $\det(E)_x\cong\Lambda^n E_x$, where the latter denotes the top exterior power of the vector space E_x.

A *section* of a holomorphic vector bundle $E\to X$ is a holomorphic map $s:X\to E$ such that $s(x)\in E_x$ for all $x\in X$. We denote by $H^0(X,E)$ the space of sections of $E\to X$.

The *canonical bundle* $K\to X$ is by definition the holomorphic cotangent bundle of X. It is a holomorphic line bundle. A section of K is nothing but a holomorphic one-form on X.

DEFINITION 8.1.—A *Higgs bundle* on X is a pair (E,Φ), where $E\to X$ is a holomorphic vector bundle and

$$\Phi\in H^0(K\otimes\mathrm{End}(E))$$

is a holomorphic 1-form on X with values in the bundle $\mathrm{End}(E)$ of endomorphisms of E.

We can view the *Higgs field* Φ as a holomorphic bundle map $\Phi:E\to E\otimes K$. Higgs bundles (E_1,Φ_1) and (E_2,Φ_2) are *isomorphic* if there is an isomorphism $E_1\cong E_2$ intertwining the Higgs fields Φ_1 and Φ_2.

There is an integer invariant, called *the degree of* E and denoted by $\deg(E)$ which topologically classifies the vector bundle. It can be identified with the total number of zeros and poles of any meromorphic section of the line bundle $\det(E)$, taking into account multiplicities. The degree has the following useful properties. If

$$0\to E_1\to E\to E_2\to 0$$

is a short exact sequence of vector bundles, then $\deg(E)=\deg(E_1)+\deg(E_2)$. Moreover, if L and M are line bundles, then $\deg(L\otimes M)=\deg(L)+\deg(M)$.

The notion of degree of a vector bundle is required for defining the following notion of polystability of a Higgs bundle, which is central for the link with representations of surface groups.

DEFINITION 8.2.—A Higgs bundle (E,Φ) with E of degree zero is *polystable* if every holomorphic subbundle $F\subset E$ such that $\Phi(F)\subset F\otimes K$ satisfies $\deg(F)\leq 0$ and, moreover, if such an F satisfies $\deg(F)=0$, then there is a another holomorphic subbundle $F^\perp\subset E$ such that $E=F\oplus F^\perp$ and $\Phi(F^\perp)\subset F^\perp\otimes K$.

The fundamental result linking surface group representations with Higgs bundles is the following, known as *non-abelian Hodge Theorem*. It was proved by Hitchin [17] and Donaldson [5] and (for more general bundles and also higher dimensional base varieties) by Corlette [4] and Simpson [23].

THEOREM 8.3.—There is a bijective correspondence between isomorphism classes of reductive representations of π_1X in $\mathrm{GL}(n,\mathbb{C})$ and isomorphism classes of polystable Higgs bundles of rank n and degree 0.

In order to apply these ideas to representations of π_1X in Lie groups G beyond the case of $G=\mathrm{GL}(n,\mathbb{C})$, a more elaborate theory of *G-Higgs bundles* is required, as was already realized by Hitchin [17,18]. We shall not go into the full details of this theory here (the interested reader may consult, for example, [2,8].) In the case of representations of π_1X in the symplectic group, the relevant notion is the following.

DEFINITION 8.4.—An $\mathrm{Sp}(2n,\mathbb{R})$-*Higgs bundle* on X is a triple (V,β,γ), where $V\to X$ is a rank n holomorphic vector bundle,

$$\beta\in H^0(X,K\otimes S^2V) \qquad \text{and}$$
$$\gamma\in H^0(X,K\otimes S^2V^*).$$

There is an obvious notion of isomorphism of $\mathrm{Sp}(2n,\mathbb{R})$-Higgs bundles. Note also that we can view β and γ as holomorphic bundle maps

$$\beta: V^*\to V\otimes K,$$
$$\gamma: V\to V^*\otimes K,$$

which are symmetric. Hence we can associate in a natural way a Higgs vector bundle (i.e. a Higgs bundle in the sense of Definition 8.2) of rank $2n$ and degree 0 by letting

$$E=V\oplus V^*,\qquad \Phi=\begin{pmatrix}0&\beta\\ \gamma&0\end{pmatrix}. \tag{8.1}$$

The non-abelian Hodge Theorem takes the following form in the case of representations in $\mathrm{Sp}(2n,\mathbb{R})$.

THEOREM 8.5.—There is a bijective correspondence between isomorphism classes of reductive representations of π_1X in $\mathrm{Sp}(2n,\mathbb{R})$ of Toledo invariant $d\in\mathbb{Z}$ and isomorphism classes of polystable $\mathrm{Sp}(2n,\mathbb{R})$-Higgs bundles with $\mathrm{rk}(V)=n$ and $\deg(V)=d$.

We now illustrate the power of Higgs bundle theory by outlining a simple proof of the Milnor-Wood inequality (7.1). Let (V,β,γ) be a polystable $\mathrm{Sp}(2n,\mathbb{R})$-Higgs bundle with $\deg(V)=d>0$. By polystability of the Higgs bundle (E,Φ) defined in (8.1), the map $\gamma:V\to K\otimes V^*$ must be non-zero. Let $N\subset V$ and $\tilde{I}\subset K\otimes V^*$ be the subbundles associated to the kernel and image of γ respectively. Let $I=\tilde{I}\otimes K^{-1}\subset V^*$. Then γ induces a non-zero holomorphic section $\tilde{\gamma}$ of the line bundle

$$\det((V/N)^*\otimes I\otimes K).$$

which therefore has positive degree:

$$\deg(N)-\deg(V)+\deg(I)+\mathrm{rk}(I)(2g-2)\geq 0. \tag{8.2}$$

Moreover, the subbundles $N\subset E$ and $V\oplus I\subset E$ are both preserved by Φ and hence polystability gives

$$\deg(N)\leq 0, \tag{8.3}$$
$$\deg(V)+\deg(I)\leq 0. \tag{8.4}$$

Combining (8.2), (8.3) and (8.4) we obtain

$$\deg(V)\leq \mathrm{rk}(I)(g-1).$$

From this the Milnor-Wood inequality (7.1) is immediate. But a further important consequence can be drawn: if equality holds in (7.1) we must have $\mathrm{rank}(I)=n$ and equality in (8.2). It follows that we have an isomorphism

$$\gamma: V\xrightarrow{\cong} V^*\otimes K.$$

In other words, γ induces a non-degenerate K-valued quadratic form on V! This can be used to induce a structure of orthogonal bundle on $V\otimes K^{-1/2}$ (for any choice of square root of K). This gives us *new invariants* of representations of surface groups in $\mathrm{Sp}(2n,\mathbb{R})$, namely the Stiefel-Whitney classes of the orthogonal bundle.[3] This explains the appearance of more connected components

[2] This identification is a consequence of Riemann's uniformization Theorem.

[3] A different point of view on these invariants was provided by Guichard-Wienhard [16]

of $\mathscr{R}_{n(g-1)}(\pi_1\Sigma, \mathrm{Sp}(2n,\mathbb{R}))$ in Theorems 7.1 and 7.2. For more information, in particular on the rather delicate issue of the exact count of the connected components, we refer to [15,10].

References

[1] M. F. Atiyah and R. Bott, *The Yang-Mills equations over Riemann surfaces*, Philos. Trans. Roy. Soc. London Ser. A **308** (1982), 523–615.

[2] S. B. Bradlow, O. García-Prada, and P. B. Gothen, *Maximal surface group representations in isometry groups of classical hermitian symmetric spaces*, Geometriae Dedicata **122** (2006), 185–213.

[3] M. Burger, A. Iozzi, and A. Wienhard, *Higher Teichmüller spaces: from* SL(2, ℝ) *to other Lie groups*, Handbook of Teichmüller Theory III, IRMA Lectures in Mathematics and Theoretical Physics, European Math. Soc., 2010, to appear.

[4] K. Corlette, *Flat G-bundles with canonical metrics*, J. Differential Geom. **28** (1988), 361–382.

[5] S. K. Donaldson, *Twisted harmonic maps and the self-duality equations*, Proc. London Math. Soc. (3) **55** (1987), 127–131.

[6] J. L. Dupont, *Bounds for characteristic numbers of flat bundles*, Springer LNM 763, 1978, pp. 109–119.

[7] W. Fulton and J. Harris, *Representation theory: a first course*, Springer-Verlag, 1991.

[8] O. García-Prada, P. B. Gothen, and I. Mundet i Riera, *The Hitchin-Kobayashi correspondence, Higgs pairs and surface group representations*, preprint, 2009, arXiv:0909.4487 [math.AG].

[9] O. García-Prada and I. Mundet i Riera, *Representations of the fundamental group of a closed oriented surface in* Sp(4, ℝ), Topology **43** (2004), 831–855.

[10] Oscar García-Prada, Peter B. Gothen, and Ignasi Mundet i Riera, *Representations of surface groups in the real symplectic group*, preprint, 2009, arXiv:0809.0576 [math.AG].

[11] W. M. Goldman, *Discontinuous groups and the Euler class*, Ph.D. thesis, University of California, Berkeley, 1980.

[12] _____, *Representations of fundamental groups of surfaces*, Springer LNM 1167, 1985, pp. 95–117.

[13] _____, *Topological components of spaces of representations*, Invent. Math. **93** (1988), 557–607.

[14] _____, *Locally homogeneous geometric manifolds*, Proceedings of the International Congress of Mathematicians 2010 (R. Bhatia, A. Pal, G. Rangarajan, V. Srinivas, and M. Vanninathan, eds.), World Scientific, 2011.

[15] P. B. Gothen, *Components of spaces of representations and stable triples*, Topology **40** (2001), 823–850.

[16] O. Guichard and A. Wienhard, *Topological invariants of Anosov representations*, J. Topol. **3** (2010), no.3, 578–642.

[17] N. J. Hitchin, *The self-duality equations on a Riemann surface*, Proc. London Math. Soc. (3) **55** (1987), 59–126.

[18] _____, *Lie groups and Teichmüller space*, Topology **31** (1992), 449–473.

[19] F. Labourie, *Cross ratios, Anosov representations and the energy functional on Teichmüller space*, Ann. Sci. Éc. Norm. Supér. (4) **41** (2008), no. 3, 437–469.

[20] J. W. Milnor, *On the existence of a connection with curvature zero*, Comment. Math. Helv. **32** (1958), 216–223.

[21] R. Miranda, *Algebraic curves and Riemann surfaces*, Graduate Studies in Mathematics, vol. 5, American Mathematical Society, Providence, RI, 1995.

[22] M. S. Narasimhan and C. S. Seshadri, *Stable and unitary vector bundles on a compact Riemann surface*, Ann. Math. **82** (1965), 540–567.

[23] C. T. Simpson, *Constructing variations of Hodge structure using Yang-Mills theory and applications to uniformization*, J. Amer. Math. Soc. **1** (1988), 867–918.

[24] _____, *Higgs bundles and local systems*, Inst. Hautes Études Sci. Publ. Math. **75** (1992), 5–95.

[25] V. G. Turaev, *A cocycle of the symplectic first Chern class and the Maslov index*, Funct. Anal. Appl. **18** (1984), 35–39.

[26] A. Wienhard, *The action of the mapping class group on maximal representations*, Geom. Dedicata **120** (2006), 179–191.

Photo by Pedro de Mendonça

An Interview

with Robert MacKay

by **João Lopes Dias** [CEMAPRE and ISEG, Universidade Técnica de Lisboa]

Robert MacKay is currently Professor of Mathematics, Director of Mathematical Interdisciplinary Research and Director of the Centre for Complexity Science at the University of Warwick, United Kingdom. After completing his undergraduate studies at the University of Cambridge, he did his Ph.D. at the University of Princeton, United States. Since then he has held positions at several prestigious universities and research centres, including a professorship at the University of Cambridge. His research interests include Dynamical Systems and its applications to Physics, Engineering, Chemistry, Biology and Economics. He is a Fellow of the Royal Society of London, of the Institute of Physics and of the Institute of Mathematics and its Applications.

Professor MacKay visited Lisbon in June and participated at the Jornadas LxDS-CIM-SPM, at the Colóquio de Matemática DM-ISEG, and at the Doctoral Program in Complexity Sciences ISCTE-IUL/FCUL.

When did you realize your interest for mathematics?
My mother remembers worrying that I was late home from primary school one day aged 4 or 5 and she found me in the garden counting crocuses! I used to do a lot of calculating.

You did Part III Maths at Cambridge, the oldest and most famous mathematics examination in the world. What did you think about the Cambridge experience, in particular your mathematical education?
I received a good education in Cambridge. Mathematics in Cambridge includes a lot of physics, which suited me well, as I'd originally intended to study Physics. So I learnt a lot about mechanics, waves, electromagnetism, fluid dynamics, and quantum mechanics, as well as an initiation into analysis, linear algebra, groups, probability, ODEs, PDEs, numerical analysis and so on.

Was there anyone who you recall as being a major influence on your future choices and views?
I was particularly influenced by James Lighthill and Michael McIntyre, notably on the theory of waves, and wrote essays on solitons and on waves in stratified atmospheres. My director of studies John Hinch pointed me in good directions, like to read Hirsch and Smale on dynamical systems. During my final year Nigel Weiss and Mike Proctor welcomed me into their Astrophysics research seminars, which I appreciated as an opportunity to see what research is like.

After Cambridge you went to Princeton for a PhD. Why did you choose the Plasma Physics Lab?
I wanted to work on a problem of potential social value that was nevertheless mathematically challenging. So I chose plasma physics, with a view to realising controlled nuclear fusion energy. I wanted also to see something different from Cambridge: wonderful as it had been I was sure the world had other good things to offer. Nigel Weiss and Mike Proctor recommended I should go to Princeton Plasma Physics Laboratory.

Who influenced you most at Princeton?
The main influences on me at Princeton were my PhD supervisor John Greene who gave me good problems to work on; John Mather whose course I attended for two and a half years non-stop; and fellow students like Rafael de la Llave with whom we met regularly to go through papers and books and to bring talks and conferences to each other's notice.

How did you develop an interest for dynamical systems?
My father copied Robert May's 1976 Nature paper on chaos in population dynamics for me while I was an undergraduate, my director of studies recommended Hirsch and Smale's book as summer reading, and Alistair Mees offered a Pt III project on "Period three implies chaos". All these struck me as fun but not serious enough mathematics, so I did a Pt III project on wave propagation in inhomogeneous atmospheres instead. But in Princeton the plasma physics programme included an introduction to dynamical systems theory and I got together with a bunch of students mainly from the Physics department to read papers and books and educate each other on the topic. We started going to John Mather's course, who treated various topics in dynamical systems theory, culminating in what is now called Aubry-Mather theory. We made day trips to a conference on Nonlinear Dynamics in New York in 1979 and I think that is when I decided nonlinear dynamics was what I wanted to do. When John Greene gave a seminar three months later about his 1979 J Math Phys paper I asked if he could suggest anything similar to do and he put me on to numerical investigation of period doubling in area-preserving maps and I was hooked.

You are very much interested in the interactions between dynamical systems theory and concrete problems arising in several different areas of knowledge. How do you manage to talk to people outside mathematics?
It takes a lot of time to understand differences in use of language, the unstated assumptions and world-view, and the often huge literature, and then to formulate worthwhile mathematical versions of their problems. I do not feel particularly good at it.

Is it too hard for a mathematician to read their literature?
For some topics there are good reviews or collections of papers setting out the subject. That is the easiest way in. There are also some good books, but they tend to be too much one author's view or to miss the state of the art.

Do you share the view that there is not a clear distinction between pure and applied mathematics, just good or bad mathematics?
The usage of the terminology is unhelpful. What is called "applied mathematics" is often not applied to anything, and some "pure mathematics" is applied to many areas. The distinction is sometimes more between attention to rigour which for the purposes of applications can limit one's analysis so much that the result is irrelevant for the original problem versus making approximations and plausible assumptions in order to get at least some form of relevant answer. Both approaches have their place and indeed a good analysis of a problem may involve moving between the two extremes in an iterative process that builds an answer that is both rigorous and relevant. The important thing is to be clear about what one is claiming. The other distinction is one of motivation: is your mathematical work driven by scientific problems or pure mathematical curiosity? Again there is a place for both.

What different cultures do you find within mathematics?
Apart from the pure v. applied culture difference, there is the algebra v. geometry difference. Some prefer symbol manipulation, others pictures. I'm more on the geometry side but I like explicit formulae when they are available.

Some of your research has been strongly motivated by scientific problems from physics, biology and social sciences. Do you find any fundamental relation between problems in those areas?
I tend to think laterally, which can be fruitful though I recognise that it is also limited, as it won't provide major paradigm shifts. Thus, for example, I hit on the idea in 1994 that the way the cochlea frequency-analyses sound may be mode-conversion rather than critical layer absorption. Mode conversion seemed to be unknown to the physiologists and the fluid mechanics working in this area, though in retrospect it is what Andrew Huxley was proposing in 1969; but I knew about it from my training in plasma physics. I think it is the right explanation, though have not persuaded a suitable journal to publish my paper yet. I'm currently in the process, with colleague Nick Chater in the Business School, of trying to formulate a thermodynamics of economics, aided by the abstract framework of Lieb and Yngvason, but there is a long history of such attempts and it may be a mistake to force economics into a physics mould.

Could you describe your work trajectory, from renormalisation of area-preserving maps to complexity science and emergence phenomena? How did your choice of problems evolve from the previous ones?
I went to Princeton to do research in plasma physics, but found that basic problems like the magnetic field line flow in a tokamak were not understood, except in the axisymmetric case, for which there is a foliation by invariant tori. To study the question of invariant tori for non-symmetric perturbations, I considered area-preserving maps. Following numerical observations of Kadanoff and Shenker I formulated a renormalisation theory for the breakup of invariant circles and verified it numerically. It was subsequently proved with computer-assisted estimates by Koch, following a direction that I proposed in 1994 going back to the continuous-time problem. Anyway, that led me into understanding the transport through the gaps of broken tori, where I interpreted Mather's action difference as a flux across a cantorus. I also developed a sufficient condition for non-existence of invariant tori that is easy to implement and with enough work is exhaustive. At IHES, Charles Tresser invited me to join a project on the boundary of chaos for circle maps, in which we proved that the boundary of complicated dynamics is itself complicated. At Warwick I pursued a number of further themes in Hamiltonian and non-Hamiltonian dynamics. One was stimulated by numerics presented at a conference by Philip Saffman on stability of periodic water waves, where using Hamiltonian theory I was able to explain his results; I followed that theme to also explain the diagram for instability of Karman's vortex street. Another came as a by-product from a visit of Philip Boyland in which he introduced me to the topological behaviour of dynamics on surfaces: this led me to some nice results for example about the rotation set and toroidal chaos for homeomorphisms of the torus. Another was stimulated by breakfast with Serge Aubry at a workshop in Minnesota, when he explained his anti-integrable limit to me and I realised I could use the idea to prove all sorts of results about area-preserving maps, and also improve his results for some quantum-mechanical models of solid-state physics. Also at this time I realised I could extend my renormalisation theory for area-preserving maps to the statistical mechanics of some classical models of solid-state physics. Perhaps my first attempts to tackle complex systems came when I took over from David Rand management of a grant with Dave Broomhead on Dynamics of large-scale networks. We didn't really achieve much on the subject, but it laid the seeds. Instead we developed an approach to extract topological information from time series; this was a precursor of what is now the very popular domain of computational topology. An important event was on a visit to Aubry he returned from a conference very excited about discrete breathers: spatially localised time-periodic solutions of networks of oscillators that physicists saw in numerics. He asked if we could prove their existence using the anti-integrable limit and I said yes and did. This initiated a series of results on their stability and interaction. In parallel I pursued a number of ideas in

topological dynamics, the best of which was prompted by happening to read Thurston and Weeks' Scientific American article on Three-manifolds. I was struck by the example they gave of a two-manifold, namely the configuration space of a triple linkage, which they showed has genus 3. I asked myself whether the free dynamics of the linkage might be Anosov and following numerics by a PhD student Tim Hunt in Cambridge, managed to prove this in a certain parameter regime. In Cambridge I was also invited to help steer a project on spatially extended dynamics. I tried out an idea I had for responding to Sinai and Bunimovich's challenge to make a coupled map lattice with non-unique phase on the group. One of the postdocs Guy Gielis explained to me some interesting stochastic systems that showed similar effects and I realised we could simulate them using coupled map lattices. This was probably my real entry into complex systems. Using the understanding gained, I proposed a mathematical formulation of the trendy concept of "emergence". Actually I did this first in response to a new PhD student David Sanders in 2000 when I'd just returned to Warwick, who wanted a project on emergence. More recently I went through Dobrushin's proof of ergodicity for weakly dependent probabilistic cellular automata and realised it could be expressed more nicely in terms of a metric on spaces of multivariate probabilities, which I have found useful in talking about the amount of emergence and the dependence on parameters. This is just a sample of things I've worked on and how I got into them. It is mostly serendipitous: just happening to pick up something where I could see I could do something, putting together things I'd already understood, interacting with interesting people.

Is stochastic dynamics closer to real systems than deterministic dynamics? Do you think that that is a fruitful direction for future work?

My view of stochastic dynamics is that the random terms represent aspects of the system that we choose not to attempt to model more accurately. In the absence of further knowledge or analytical ability this can be a sensible approach. Nevertheless, there are examples where the effect of some deterministic dynamics is rigorously equivalent to some noise process, the randomness being with respect to initial conditions, and then it makes sense to use the stochastic model. For example, a Langevin equation is widely used for the dynamics of slow degrees of freedom in a Hamiltonian system whose fast degrees of freedom are mixing. Anosov an I have scketched a derivation of this.

In the last few decades the number of active researchers and the quality of the mathematical work produced in Portugal has grown considerably. In your professional life have you ever had this perception?

It has been my privilege to interact with the dynamical systems group from Porto for at least 20 years and to supervise three PhD students from Portugal. And I've just taken on another one.

In contrast to older times, today mathematics is very much a collaborative effort. Do you have any preference between working alone or in teams? Is the challenge different?

There is still plenty of room in mathematics for single author research. But there are advantages to collaborations: broader perspective, shared work, wider dissemination.

Who is your favourite mathematician? Why?

I have many heroes, for example Moser, Arnol'd, Anosov and Sinai. I like what they have written and I like them as people (though unfortunately Arnol'd and Moser are no longer alive). Moser made many important advances in Hamiltonian dynamics; he was particularly nice to me, accepting me early in my career even though my approach was very non-standard mathematically and suggesting fruitful lines of research. Arnol'd was brilliant in a wide range of directions; he could be famously caustic but he was always nice to me and willing to answer my questions in considerable detail. Anosov I feel is a greatly under-rated mathematician: the insights he had in the 1960s about the Holder continuity of the foliations of hyperbolic dynamical systems and its implications for their measure theory are profound; I enjoyed making his acquaintance and showing him my mechanical Anosov system. Sinai I feel is the main architect of the theory of how deterministic dynamical systems can behave stochastically: he showed that the Markov partitions that had been constructed for special systems are a general feature of hyperbolic dynamical systems and that they give a correspondence of the dynamics to a generalisation of Markov processes called Gibbsian processes (which allow infinite-range but decaying memory). He has a very warm character and has been very supportive of my work. Going further back in time, I'd say Poincaré is my biggest hero: he developed so much interesting mathematics and presented it in such a readable way. And before him there was Newton, who was so creative, but apparently an awful character.

Detailed mathematical models in neurobiology—Storing information in membrane conductances dynamics

by **Eduardo Conde-Sousa*** and **Paulo Aguiar****

1. Introduction

Neurons are Nature's solution to the problem of information processing and information storage. Nervous systems have been engineered by evolution to sense information from the environment, process this information and store experiences for the purpose of improving future decisions. Ubiquitous in all these stages is the necessity of information buffers. In the case of mammals, there are different mechanisms providing storage in a wide range of time scales: from the ephemeral facilitation of a synapse to the life-long memories of childhood. As expected, neuronal dynamics are an extremely rich subject from a mathematically point of view. In this paper we focus on a model for a short-term memory mechanism called *working memory*. Regions of the mammal brain engaged in providing this functional resource are capable of retaining neuronal spatial patterns of activity for the duration of a few seconds. Basically, working memory provides a temporary buffer where information is held for short-time, while it is being actively used in cognitive tasks; this information can then be passed on to longer-term storage mechanisms or be simply discarded and forgotten. We humans use our working memory system when we temporarily retain a phone number or a name, when we mentally perform an arithmetic calculation, or when our wives tell us by phone the grocery list.

Our goal in this article is to give a glimpse into some of the methodologies used in theoretical neuroscience targeting a particular problem: to describe a mathematical model, closely fitted into the biophysical constrains of the nervous system, that helps understanding how working memory can be produced in a network of neurons. Our approach is different from other working memory models [1] in the sense that it does not rely on synaptic plasticity[1] nor connectivity structure to store information. In our model we store information in the dynamical states of the neuron's membrane conductances. An important feature in working memory systems is that it is possible to retain complex activity patterns after a single exposure to the stimuli. This constrain is better supported by the time scales found in conductances dynamics, than by synaptic plasticity temporal properties, even if we take into account short-term plasticity mechanisms.

In a population of N interconnected neurons engaged in working memory, we define as information content the particular subset of neurons that are co-activated

{1} Synapses are the structures that mediate most of the communication and transfer of signals between neurons. A strong synapse produces a large signal in the target neuron while a weak synapse will produce a small response. By modulating the synaptic strengths it is possible to both store information and to change the computational/functional capabilities of populations of neurons. The present dogma in neuroscience is that information is stored in the efficacy, or strength, of synapses.

* Faculdade de Ciências da Universidade do Porto
** Centro de Matemática da Universidade do Porto

at a given time. We assume that different sensory/perceptual configurations produce in this population different patterns of active units. In the neuronal population, the units not belonging to the memory pattern have low frequency stochastic activity (0.5Hz) while the units belonging to the pattern have a higher frequency activity (20Hz). The sizes of these patterns, i.e. the number of active units in any pattern, is considered to be roughly the same, which is in accordance with the notion of activity level control in neuronal circuits. The population can act as a working memory system if, after a short-period of a few hundred milliseconds where a subset of neurons is consistently co-activated (the duration of a "one-shot" stimuli, such as hearing a number a single time), this spatial pattern of activity is auto-preserved after the stimuli has been removed, for a duration of several seconds. In addition to this core property, our working memory model has to satisfy the following conditions:

- stochastic activations of spurious neurons should not be stored nor should affect the stored pattern
- the retained activity pattern should be stable for several seconds
- an inhibitory input within physiological values should be able to clear the memory pattern and restore the network to its basal, low frequency, stochastic firing
- deactivation of isolated neurons in the pattern should not compromise the rest of the pattern's integrity

This article is organized in the following way. First we describe in detail the mathematical model used to set the dynamics for each neuron individually, which parameters are used and how the numerical simulations are performed. The following section describes how the neuronal population is assembled, what network architecture is used and the properties of the synaptic connectivity. The following two sections describe results of the model: first we present results regarding the single neuron model, after that we present the results regarding the collective behavior of the neuronal population as a working memory system. We conclude with some final remarks.

2. Single Neuron Model

The neuronal dynamics are set using the Hodgkin-Huxley (HH) formalism in order to produce biophysically accurate descriptions of all neuronal conductances. Neurons are modeled as a single cylindrical compartment of length $L = 20\mu m$ and diameter $\text{diam} = 20\mu m$. That is, given the context and the questions being addressed, there are no *a priori* reasons to assume a special role to be taken by the spatial properties of the neuron. Therefore, the complex neuronal tree topology is collapsed and pointwise neurons are considered—hence the main variable is time only. Together with the leakage current and synaptic currents, a set of four ionic currents are considered in the membrane potential model. In addition to the canonical delayed rectifier potassium current and transient sodium current present in the HH equation, our model adds two more currents: a calcium current, $I_{Ca,L}$, which produces an influx of calcium whenever the membrane potential is becoming depolarized, and a nonspecific cationic current which is dependent on the intracellular calcium concentration, I_{CAN}. Together, these two currents can act synergetically to prolong depolarization in the membrane potential. These are currents that are known to exist in many neuronal types in the nervous system [2]. All ionic current dynamics are taken from Senselab [3] database. The time evolution of the membrane potential is described by the equation:

$$C_m \dot{V} = -I_L - I_{Na} - I_{Kdr} - I_{Ca,L} - I_{CAN} - I_{syn}$$

where $C_m = 1\mu F/cm^2$ is the membrane capacitance; V represents the membrane potential in mV; $I_L = g_L \times (V - E_L)$ is the leakage current, where $g_L = 4.2 \times 10^{-5} S/cm^2$ is the leakage conductance and $E_L = -65\text{mV}$ is the leakage reversal potential; I_{Na} and I_{Kdr} are the transient Na^+ and K^+ currents responsible for action potentials; $I_{Ca,L}$ is the high-threshold Ca^{2+} current and I_{CAN} is the intracellular calcium concentration nonspecific cation current mentioned earlier; finally, I_{syn} is the sum of all synaptic currents impinging on the neuron.

All numerical analysis/simulations were preformed in the simulation environment NEURON [4].

2.1. Sodium and Potassium currents (I_{Na} and I_{Kdr})

The fast Na^+ and K^+ currents are modeled according to the canonical Hodgkin-Huxley kinetics [5] with small modifications proposed by Traub and Miles to model hippocampal pyramidal cells [6]. The key parameters are: maximal sodium conductance $\bar{g}_{Na} = 80\text{mS/cm}^2$, maximal potassium conductance $\bar{g}_K = 20\text{mS/cm}^2$, sodium reversal potential $E_{Na} = 50\text{mV}$ and potassium reversal potential $E_K = -70\text{mV}$.

The model for these currents consists of:

$$I_{Na} = \bar{g}_{Na} \times m^3 \times h \times (V - E_{Na})$$
$$I_{Kdr} = \bar{g}_K \times n^4 \times (V - E_R).$$

The kinetic equation for the gating variables is

$$\dot{y} = -\frac{y - y_\infty(V)}{\tau_y(V)}.$$

where

$$y_\infty = \frac{\alpha_y}{\alpha_y + \beta_y},$$

$$\tau_y = \frac{1}{\alpha_y + \beta_y}.$$

and $y \in \{m, h, n\}$.

The activation and inactivation gate functions are:

$$\alpha_m = \frac{0.32 \times (73 - V)}{\exp\left(\frac{73-V}{4}\right) - 1}.$$

$$\beta_m = \frac{0.28 \times (V - 100)}{\exp\left(\frac{V-100}{5}\right) - 1}$$

$$\alpha_h = 0.128 \times \exp\left(\frac{72 - V}{18}\right)$$

$$\beta_h = \frac{4}{1 + \exp\left(\frac{95-V}{5}\right)}$$

$$\alpha_n = \frac{0.032 \times (75 - V)}{\exp\left(\frac{75-V}{5}\right) - 1}$$

$$\beta_n = 0.5 \times \exp\left(\frac{70 - V}{40}\right).$$

2.2. High-threshold Ca^{2+} current($I_{Ca,L}$)

The high-threshold $I_{Ca,L}$ current is modeled according to the equation [7]:

$$I_{Ca,L} = \bar{p}_{Ca,L} \times m^2 \times \text{GHK}(V, cai, cao).$$

were $\bar{p}_{Ca,L} = 0.03\text{cm/s}$ is the Ca^{2+} membrane permeability, cai and cao are respectively the intracellular and extracellular calcium concentration, and GHK is the Goldman-Hodgkin-Katz equation.

The kinetic equation for the activation variable is

$$\dot{m} = -\frac{m - m_\infty(V)}{\tau_m(V)}$$

where

$$m_\infty = \frac{1}{1 + exp(\frac{V+10}{-10})}$$

$$\tau_m = \frac{1}{\alpha_m + \beta_m}$$

$$\alpha_m = \frac{1.6}{1 + exp(-0.072 \times (V - 5))}$$

$$\beta_m = 0.02 \times \frac{1.31 - V}{1 - exp\left(\frac{V-1.31}{5.36}\right)}.$$

2.3 Intracellular calcium dynamics

The dynamics of the intracellular calcium concentration, denoted as cai, are modeled by a fast removal process due to an active pump, and by calcium entry which is due to the current $I_{Ca,L}$, as described in [8].

The used parameters are:

$depth = 0.1\mu\text{m}$

$cai_\tau = 1\text{ms}$

$cai_\infty = 5 \times 10^{-5}\text{mM}$

2.4. Ca^{2+} dependent nonspecific cation current (I_{CAN})

The adopted model for Ca^{2+}-dependent nonspecific cation current [9] is described as:

$$I_{CAN} = \bar{g}_{CAN} \times m^2 \times (V - E_{CAN})$$

with parameters $\bar{g}_{CAN} = 0.01\text{mS/cm}^2$ and $E_{CAN} = 0\text{mV}$. For our version of the model we modified the middle point of the activation function to 0.5×10^{-3} (before was 1.0×10^{-3}). This change allowed a small increase in the sensibility of this current to lower concentrations of intracellular calcium.

2.5 Synaptic Current

The synaptic current is modeled by the sum:

$$I_{syn} = \Sigma_i g_{syn,i} \times \left(V - E_{syn,i}\right)$$

where i runs over the set of pre-synaptic neurons. In other words, the term I_{syn} aggregates the currents from all synapses established with a particular neuron. All synapses in the model are excitatory, and their dynamics are modeled according to the biological NMDA synapse type. The core synaptic conductance profile is modeled by a dual exponential function:

$$g_{syn}(t) = \bar{g}_{syn} \cdot a\left[e^{-\frac{t}{\tau_{decay}}} - e^{-\frac{t}{\tau_{rise}}}\right]$$

where $g_{syn}(t)$ represents the synaptic conductance after t milliseconds of the synaptic activation and a is chosen so that the maximum value of g_{syn} matches $\bar{g}_{syn}$, the maximum synaptic conductance. The values for the rising time constant τ_{rise}, the decay time constant τ_{decay} and the synaptic reversal potential E_{syn} are: $\tau_{rise} = 5.0\,[\text{ms}]$, $\tau_{decay} = 70.0\,[\text{ms}]$ and $E_{syn} = 0.0\,[\text{mV}]$.

The NMDA synaptic current has the property of depending on the post-synaptic membrane potential: independently of the synaptic activation by the pre-synaptic neuron, an effective synaptic current will only be elicited if the post-synaptic membrane is sufficiently depolarized. In other words, this type of synapses act as an "AND" operator and has strong functional implications in the dynamics of neuronal networks. We follow a well established model and represent the NMDA synaptic conductance multiplying the dual exponential conductance profile by a factor representing the magnesium

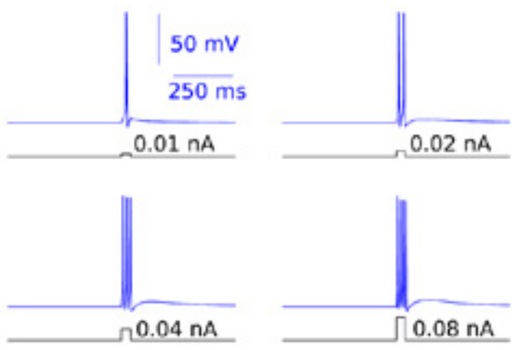


Figure 1.—Membrane potential response (in blue) to 40ms constant current stimuli of different amplitudes (in black).

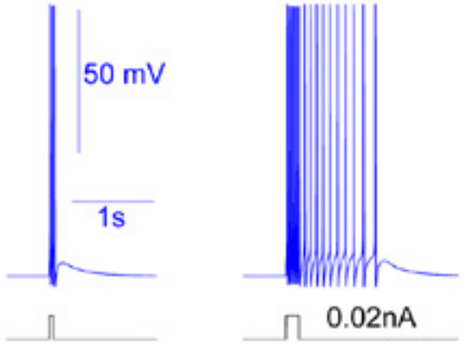


Figure 2.—Two electrode currents are injected in the neuron with amplitude of 0.02nA. The stimulus have a duration of 50ms (left) and 150ms (right). With a 150ms current injection the neuron sustain activity for a period of 900ms, after the end of the current injection. When the current injection is shorter, this period of sustained activity is non existing.

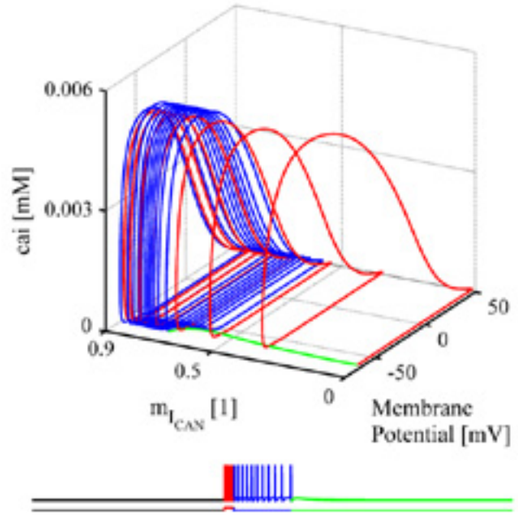


Figure 3.—An electrode current is injected in the neuron with duration 150ms and amplitude 0.02nA. The upper panel contains the phase space with variables $m_{I_{CAN}}$ (activation variable of CAN current), cai (intracellular calcium concentration [mM]) and membrane potential [mV]. The middle panel contains the membrane potential time evolution and the lower trace corresponds to the electrode current injected. For every new action potential, the activation variable $m_{I_{CAN}}$ increases (red trace). After the end of the current injection, the I_{CAN} activation variable slowly decreases to $m_{I_{CAN}} \approx 0.7$ (blue trace) and during a considerable amount of time, acting synergetically with the $I_{Ca,L}$ current, it is sufficient to sustain activity in the neuron. After this period, the depolarization induced by I_{CAN} is not sufficient to generate new action potentials and the membrane potential converges to the -65mV equilibrium potential (green trace).

block which characterize the post-synaptic dependence [10]. The NMDA model also accounts for the ratio of calcium current to total current [11] flowing through these channels as they introduce a relevant contribution to the increase of the intracellular calcium concentration.

3. Network topology

Working memory is a emergent property of the collective behavior of specific populations of neurons. The communication between neurons is determined by the connectivity matrix and in this model we use random connections to set the network architecture. Given two neurons, i and j, the probability of a synapse from pre-synaptic neuron i to post-synaptic neuron j is $P(i \to j) = ConnRate$, where $ConnRate$ represents the predefined connectivity rate of the model. The highest values in the mammalian brain are close to 30%, in regions with dense recurrent connections such as area CA3 in the hippocampus. As a central goal in this model is to store new information without involving synaptic changes, all peak conductances are taken from a common distribution and are then fixed for all numerical simulations.

While not plastic, the absolute values of the synaptic conductances are crucial in setting the activity level of the network. Two constraints are used to quantify the synaptic peak conductances, and therefore constraint all connections in the network:

1. one neuron must be able to fire stochastically, without entering a state of persistent activity due to interactions with active neurons;
2. when consistently excited, one neuron must be able to sustain activity for a period of tens of seconds as a result of the interactions with other active neurons.

By "consistently excited" we mean a series of coherent excitations in a small time window. The size of time window has to be balanced between small enough to be compatible with the notion of "one-shot learning" and big enough to make the probability of stochastic activations producing such an excitation profile virtually zero. A length of 200ms is chosen for the stimulation time window.

4. Results

For clarity purposes the results are separated into single neuron dynamics and population's collective behavior, where the emergence of a working memory system is analyzed.

4.1 Single neuron firing properties

In the absence of stimulation currents and stochastic noise, the neuron's membrane potential rests in the stable equilibrium point of about –65mV. Conversely, when a current of amplitude 0.01nA is injected for a period of 40ms to an isolated neuron an action potential (AP) is produced. Higher current amplitudes naturally lead to more APs in the same time period (see Fig. 1).

Each AP, or simply *spike*, results from the fast, but transient, Na^+ current and from the delayed rectifier K^+ current. Every time a spike is generated, the high-threshold calcium current activates and the intracellular calcium concentration rises. However, for short stimulation intervals like this one of 40ms, the slowly adapting I_{CAN} activation variable (long time constant) suffers little or no variation, leading to a negligible change in the I_{CAN} current. Thus, the dynamics of the neuron's membrane potential can be seen mainly as a result of the well know and well studied interaction between I_{Na} and I_{Kdr} currents. This is the situation where the neuron fires due to stochastic network activations or to very short, isolated and non-consistent external stimulation.

On the other hand, longer, consistent activations leading to several spikes with short latency give rise to a different behavior in the neuron's currents dynamics. The consecutive activation of the high-threshold calcium current generates a progressive increase in the I_{CAN} activation variable. The long time constant of the I_{CAN} activation variable enables a coarse integration leading to values close to 1.0. For a stimulation period of 150ms, the I_{CAN} activation variable reaches values in $m \in [0.85; 0.9]$ which are enough to sustain the activity for a period of several hundred milliseconds (see Fig. 2). A consistent stimulation in a time window of 100 – 200ms is compatible with the "one-shot learning" paradigm.

The synergy between I_{CAN} and $I_{Ca,L}$ can be better appreciated in a phase graph (Fig. 3), where the axis are the neuron's membrane potential, the I_{CAN}'s activation variable m and the calcium's intracellular concentration which is heavily modulated by $I_{Ca,L}$.

It is important to emphasize that in order to consistently excite one neuron, the variable $m_{I_{CAN}}$ must rise

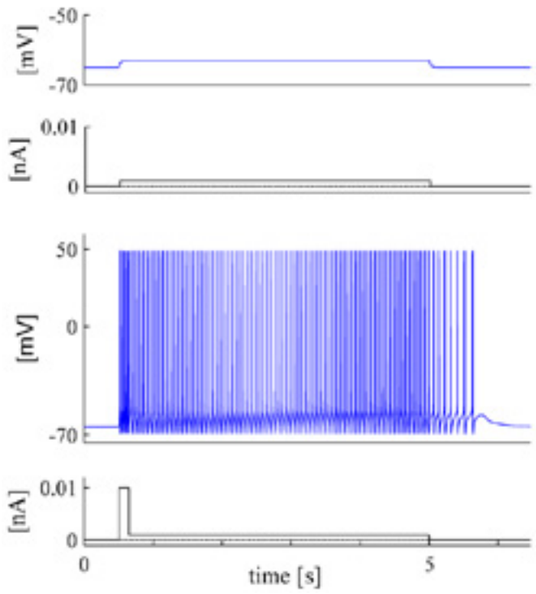


Figure 4.— A residual current injection of 0.001nA only produces a small depolarization in the membranes potential and is incapable of eliciting APs (upper traces). On the other hand, if this residual current follows a 150ms, 0.01nA amplitude core stimulus, the other way vanishing activity now becomes persistent (lower traces).

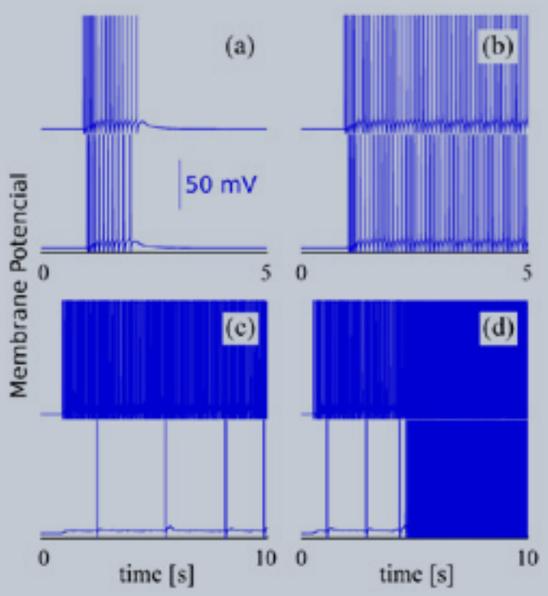


Figure 5.—Parametrization of the NMDA peak conductances w. Two neurons connected through a single NMDA synapse are used to assess the proper synaptic conductance values. A synaptic value of $w=4.0\times10^{-5}\mu S$ is insufficient to provide enough recurrent excitation to sustain activity after the short period of stimulation on a small fraction of the population; the membrane potential traces for the two neurons connected through such w are shown in panel **(a)**. Above $w=5.0\times10^{-5}\mu S$, the recurrent connections are already sufficient to preserve the activity pattern; again the membrane potential traces for the neurons connected through the new w are shown in panel **(b)**. This proper functional behavior is maintained up to $w=20.0\times10^{-5}\mu S$; the activity in the neuron belonging to the memory pattern does not propagate to other neuron even if they are subject to stochastic activations—panel **(c)**. Higher synaptic conductance values, such as $w=22.0\times10^{-5}\mu S$, start to invoke spurious activations and instability on the activity pattern; the activity in the neuron belonging to the memory pattern is now capable of recruiting additional neurons, thus corrupting the working memory—panel **(d)**.

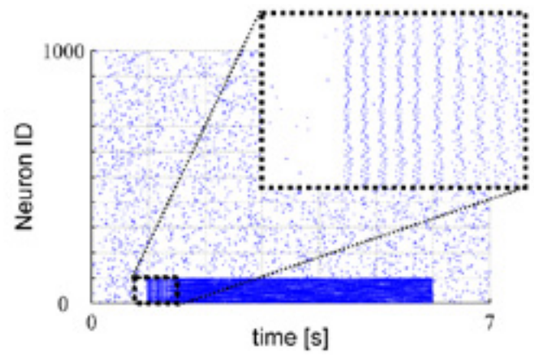


Figure 6.—Working memory in action. A sub-population of 100 neurons is activated and is capable of sustaining its activity for a long period of time without becoming corrupted (by loosing or adding elements).

enough (above 0.7). This *CAN* current is only activated by rising the internal concentration of calcium, which in turn depends on the existence of APs. Thus, the stimulus current must be sufficient to trigger a sequence of APs during a period between 100 and 200ms, depending on the achieved firing rate. Lower firing rates require longer periods of consistent stimulation. For example, with a 150ms and 0.02nA amplitude current injection, the neuron fires five times over a period of approximately 120ms (corresponding to an average firing rate of 30Hz) which is sufficient to sustain the activity for a few hundreds of milliseconds after the stimulation finishes. However, if a small residual current is provided after the stimulus ends, the neuron can retain its activity for much longer periods of time (Fig. 4). This property is of considerable importance as it sets the conditions in which a working memory system can work.

4.2 Network behavior

Unless otherwise stated, all presented network results refer to simulations with a population of 1000 neurons, with a recurrent connectivity rate of 25%. The stimulation protocol consists of exciting 100 neurons (10% of the population), creating the so called memory pattern of activity. The stimulation lasts for 200ms and generates 5 spikes on each neuron (individual firing rate of 20Hz). All induced spikes are not completely synchronized: a uniform random jitter of ±10ms is introduced to emphasize the robustness of the system to small amounts of noise. This robustness comes mainly from the long-lasting NMDA conductance profiles and from the fact that the passive properties of neuron membrane act as a low-pass filter with a time constant, in the case of this model's parameters, of ≈ 24ms (obtained from C_m/g_L). These two mechanisms significantly enlarge the integration time scale for the synaptic inputs.

In addition to the spike jitter in the memory pattern activation, noise is constantly provided to all neurons, belonging or not to the memory pattern. Noise is introduced as stochastic activations following a Poisson process with an average interspike interval of 2000ms (0.5Hz), in agreement with cortical neurons experimental data.

In the connected population, the residual stimulation required to sustain the memory pattern is provided by the recurrent connections. The NMDA's synaptic peak conductances are therefore of noteworthy importance. The calculated range of values which satisfy the two constraints mentioned in section 3, given all neuronal model parameters, is $[5\times10^{-5}, 20\times10^{-5}]\mu S$ (see Fig. 5). These values correspond to the *total* NMDA synaptic conductance required to drive the target neuron in this safe zone and are therefore independent of the population size; in other words, they represent the target value for the sum of all synaptic conductances and are obtained considering a complete synchrony in the synaptic activations. This is a strong assumption but, again, the long-lasting NMDA conductance profiles and the long membrane's time constant produces a synchronicity time window in the order of a few tenths of milliseconds. This interval encapsulates the variability in the memory pattern activations and renders irrelevant the need for complete, sub-millisecond, synaptic synchronization. The calculated values for the total synaptic conductance under the synchrony assumption are then safely used as estimators for the total synaptic conductance under less stringent synchronicity constraints.

While the required total synaptic conductance is a function of the neuron's parameters and independent of the population size, the individual NMDA synaptic conductances depend on the number of synaptic inputs each neuron receives. This number follows a binomial distribution with parameters N, the population size, and *ConnRate*, the connectivity rate.

The working memory behavior of the system can be visualized in a raster plot, where the spikes of all neurons are represented as dots (Fig. 6). The consistent but short activation of a constellation of neurons in the population forms a memory activity pattern which sustains for several seconds due to the synergy between the $I_{Ca,L}$ and I_{CAN} currents, and the stabilizing current provided by the recurrent connections. Two relevant points worthwhile mentioning is that both the activity pattern firing frequency (in the range of 15–20Hz) as well as the magnitude of the inhibitory conductance necessary to reset the memory pattern (in the range of 0.1μS for a duration of 200ms—compatible with $GABA_B$) are in accordance with neurobiological data [2].

4.3 Model scaling

The simulation results shown use a population of 1000 neurons. It is interesting to notice that as the number of neurons rises, our working memory model becomes more robust to variability in the NMDA's peak conductances.

Given the synaptic constraints (see Fig. 5), we can conclude that a neuron belonging to the activity pattern must receive a total synaptic conductance of, at least, $5\times10^{-5}\mu S$ from the other neurons belonging to the pattern. Therefore, if m is the minimum number of connec-

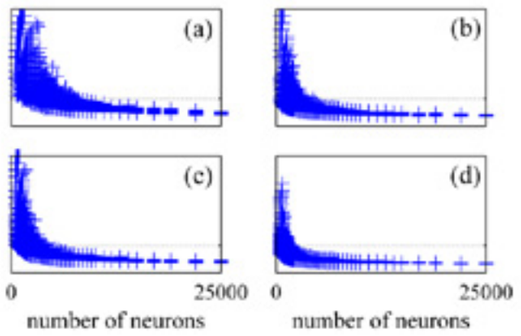


Figure 7.—Larger population sizes produce more robust working memory systems. Each graph represents the value of M/m obtained for randomly generated connectivity matrices as a function of the total number of neurons in the population (between 500 and 25000), the percentage neurons belonging the activity pattern (5% in **(a)** and **(b)**, and 10% in **(c)** and **(d)**), and the connectivity rate (10% in **(a)** and **(c)**, and 20% in **(b)** and **(d)**). For each population size, 500 samples are drawn. The dashed line marks the value M/m=4 below which the variability in the synaptic conductances becomes better contained within the calculated bounds.

tions each neuron in the pattern receives from other neurons in the pattern, the average synaptic efficacy must be $\geq 5 \times 10^{-5}/m$. On the other hand, a neuron outside the activity pattern must not be activated by the neurons in the pattern; i.e. it must not receive more than a total synaptic conductance of $20 \times 10^{-5}\mu S$. If M is the maximum number of connections each exterior neuron receives from neurons belonging the pattern, than the average synaptic efficacy must be $\leq 20 \times 10^{-5}/M$. Thus, the average synaptic efficacy must be between $5 \times 10^{-5}/m$ and $20 \times 10^{-5}/M$ which is only possible if $5 \times 10^{-5}/m \leq 20 \times 10^{-5}/M$. This means that $M/m \leq 4$.

As the population size grows, the fluctuations in the number of input synapses each neuron receives becomes less relevant (scales with $1/\sqrt{N}$) and the excitation reaching neurons inside the memory pattern, and outside, becomes more homogeneous. Less variability in the total synaptic conductances means that corruption of the memory activity pattern becomes less probable. A comparison of the population sizes required to obtain highly robust working memory systems, as a function of the pattern size and *ConnRate*, is shown in Fig. 7.

5. Final remarks

We have shown how detailed biophysical models and their numerical analysis can be used to shed light to complex problems in neurobiology. These type of models are not simply a mathematical challenge: their proximity to biology makes them ideal to construct new hypothesis, produce predictions, catalyze new experiments and ultimately improve our understanding of how our brains can process and store information.

Acknowledgments

Research funded by the European Regional Development Fund through the programme COMPETE and by the Portuguese Government through the FCT—Fundação para a Ciência e a Tecnologia under the project PEst-C/MAT/UI0144/2011.
Eduardo Conde-Sousa was supported by the grant SFRH/BD/65633/2009 from FCT.

References

[1] D. Durstewitz, J. K. Seamans, and T. J. Sejnowski. Neurocomputational models of working memory. *Nat Neurosci*, 3 Suppl:1184–1191, Nov 2000.

[2] E. R. Kandel, J.H. Schwartz, and T. M. Jessell. *Principles of Neural Science*. Mc Graw Hill, 2000.

[3] Chiquito J Crasto, Luis N Marenco, Nian Liu, Thomas M Morse, Kei-Hoi Cheung, Peter C Lai, Gautam Bahl, Peter Masiar, Hugo Y K Lam, Ernest Lim, Huajin Chen, Prakash Nad- karni, Michele Migliore, Perry L Miller, and Gordon M Shepherd. Senselab: new developments in disseminating neuroscience information. *Brief Bioinform*, 8(3):150–162, May 2007.

[4] M. L. Hines and N. T. Carnevale. The neuron simulation environment. *Neural Comput*, 9(6):1179–1209, Aug 1997.

[5] A. L. Hodgkin and A. F. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J Physiol*, 117(4):500–544, Aug 1952.

[6] R. D. Traub and R. Miles. Multiple modes of neuronal population activity emerge after modifying specific synapses in a model of the ca3 region of the hippocampus. *Ann N Y Acad Sci*, 627:277–290, 1991.

[7] D. A. McCormick and J. R. Huguenard. A model of the electrophysiological properties of thalamocortical relay neurons. *J Neurophysiol*, 68(4):1384–1400, Oct 1992.

[8] A. Destexhe, A. Babloyantz, and T. J. Sejnowski. Ionic mechanisms for intrinsic slow oscillations in thalamic relay neurons. *Biophys J*, 65(4):1538–1552, Oct 1993.

[9] L. D. Partridge and D. Swandulla. Calcium-activated non-specific cation channels. *Trends Neurosci*, 11(2):69–72, Feb 1988.

[10] C. E. Jahr and C. F. Stevens. A quantitative description of nmda receptor-channel kinetic behavior. *J Neurosci*, 10(6):1830–1837, Jun 1990.

[11] N. Burnashev, Z. Zhou, E. Neher, and B. Sakmann. Fractional calcium currents through recombinant glur channels of the nmda, ampa and kainate receptor subtypes. *J Physiol*, 485 (Pt 2):403–418, Jun 1995.

Modeling and simulation of the human cardiovascular system

by **Alexandra Bugalho de Moura*** and **Adélia Sequeira****

Abstract.—The use of mathematical modeling and numerical simulation to study blood circulation and related pathologies is an active interdiciplinary field of research. It has a great social and economical impact mainly due to cardiovascular diseases, that represent one of the leading causes of death and morbidity in industrialized countries.

Due to the complexity of the human cardiovascular system, the use of computational models to study blood flow in healthy and pathological situations is a challenge to mathematicians and engineers. Nevertheless, it constitutes nowadays a reliable tool which is increasingly used in clinical applications, such as the placement of stents in arteries with atherosclerotic plaques, or the understanding of aneuerysm growth and rupture.

In this article some of the fundamental aspects of mathematical modeling and numerical simulation of blood circulation will be described, highlighting in particular the pathological case of cerebral aneurysms.

1. Simulating blood circulation: a challenge to mathematicians

Over the last years, the development and application of mathematical models, seconded by the use of efficient and accurate numerical algorithms, has allowed for impressive progresses in the understanding of the human cardiovascular system, in both healthy and pathological situations [5,12,9]. The developments in scientific computation techniques and computers capacity have also contributed to patient-specific studies, providing valuable clinical information in the perspective of diagnosis, treatment or surgical planning [5,15,12,9,13,14]. Indeed, the increasing demand from the medical community for scientifically rigorous investigations of cardiovascular diseases has been a major impulse to the progress in this field. However, modeling and simulating the human circulation still remains a very difficult and challenging task. The geometrical structure of the vascular tree and the heterogeneous composition of blood, the mechanical and biochemical interactions between blood and the vessel walls, the pulsatile nature of blood flow, together with auto-regulation processes and the link between global and local circulation, are extremely complex physiological phenomena. Therefore, it is impossible to construct a three-dimensional (3D) mathematical model of the circulatory system including all those characteristics, and therefore simplifications are mandatory. On the other hand, it is recognized that cardiovascular pathologies, like atherosclerosis or aneurysms, are closely related with local hemodynamics, such as areas of flow

Keywords.—Mathematical modeling, numerical approximation, computer simulations, computational fluid dynamics (CFD), blood flow, cardiovascular system, fluid-structure interaction, geometrical multiscale modeling, aneurysm.

* Department of Mathematics and CEMAT, Instituto Superior Técnico, Technical University of Lisbon.
E-MAIL: xiluva@gmail.com

** Department of Mathematics and CEMAT, Instituto Superior Técnico, Technical University of Lisbon.
E-MAIL: adelia.sequeira@math.ist.utl.pt

reversal or low and oscillatory wall shear stress [5,7,2,12]. The progress in the power of modern computers along with the progress in imaging, visualization and geometry reconstruction techniques, as well as the improvement of sophisticated numerical algorithms, allow for the development and analysis of highly complex models. The final goal is to set up patient-specific models and simulations incorporating data and measurements taken from each single patient, that will be able to predict the results of medical diagnosis and therapeutic planning with reasonable accuracy and using non-invasive means. This is a highly multidisciplinary field of research, requiring the collaboration between mathematicians, bio-engineers and medical doctors.

2. Mathematical models for the cardiovascular system

It is known that cardiovascular diseases are associated to local hemodynamics [7,5,2], that is, to local blood flow dynamics in specific regions of the cardiovascular tree. Strictly speaking, blood is not a fluid, but a suspension of particles in a fluid named plasma [8]. However, in medium to large sized vessels, blood can be considered as an incompressible continuum fluid described by the incompressible Navier-Stokes equations, accounting for the conservation of momentum and mass (1).

2.1. The fluid equations

Given $\Omega \subset \mathbb{R}^3$ an open and bounded domain of interest, usually a portion of a vessel, and $I =]0, T]$ the time interval, the continuity and momentum equations for incompressible and isothermal fluids are given by:

$$\begin{cases} \rho\left(\dfrac{\partial \mathbf{u}}{\partial t} + \mathbf{u}\cdot\nabla\mathbf{u}\right) - \operatorname{div}\boldsymbol{\sigma}(\mathbf{u}, P) = 0, & \text{in } \Omega, \forall t \in I, \\ \operatorname{div}\mathbf{u} = 0, & \text{in } \Omega, \forall t \in I, \end{cases} \tag{1}$$

where ρ is the density of blood, assumed constant since the fluid is considered incompressible, and $\mathbf{u}$ and P are the unknown velocity and pressure fields, respectively. The fluid flow is interily known if the velocity vector and the pressure at each spacial point and instant of time are known. $\boldsymbol{\sigma}(\mathbf{u}, P)$ is the so called Cauchy stress tensor, defining the internal forces of the fluid, hence its rheology [8]. Blood is often considered to be a Newtonian fluid in large to medium sized vessels, meaning that it flows like water: the internal tangential forces are proportional to the velocity gradient, with the constant of proportionality being the fluid viscosity,

$$\boldsymbol{\sigma}(\mathbf{u}, P) = -P\mathbf{I} + 2\mu\mathbf{D}(\mathbf{u}),$$

where μ is the constant viscosity, and $\mathbf{D}$ is the strain rate tensor given by

$$\mathbf{D}(\mathbf{u}) = \frac{1}{2}\left(\nabla\mathbf{u} + \nabla\mathbf{u}^T\right). \tag{2}$$

However, blood exhibits non-Newtonian properties, mainly due to the mechanical characteristics of red blood cells [8,9]. The shear-thinning behavior of blood is one of its main non-Newtonian properties, characterized by the decrease of the apparent viscosity with increasing shear rate. In this case, the viscosity is not constant and depends on the shear rate:

$$\dot{\gamma} = \sqrt{\frac{1}{2}\mathbf{D}(\mathbf{u}) : \mathbf{D}(\mathbf{u})}. \tag{3}$$

To account for this property of blood, a generalized Newtonian rheological model can be considered [8,9,13,14] with the Cauchy stress tensor given by:

$$\boldsymbol{\sigma}(\mathbf{u}, P) = -P\mathbf{I} + 2\mu(\dot{\gamma})\mathbf{D}(\mathbf{u}).$$

Different viscosity functions $\mu(\dot{\gamma})$ define different generalized Newtonian models that can be of shear-thinning, shear-thickening, or yield stress type, according to the behavior of the apparent viscosity with respect to the shear rate. One of the most used shear-thinning generalized Newtonian models for blood is the Carreau model, for which the viscosity function is given by:

$$\mu(\dot{\gamma}) = \mu_\infty + \left(\mu_0 - \mu_\infty\right)\left(1 + (\lambda\dot{\gamma})^2\right)^{\frac{n-1}{2}}, \tag{4}$$

where $\lambda > 0$, and $n \in \mathbb{R}$ are constants, and the coefficients μ_0 and μ_∞ are the asymptotic viscosity values at low and high shear rates, respectively. In this case, since the blood is shear-thinning, we have $\mu_0 > \mu_\infty > 0$. All these parameters should be obtained from curve fitting to experimental data. In particular, in several works [9,13,14] the parameter values of the viscosity function were estimated from experimental viscosity data obtained for normal human blood: $\mu_0 = 0.456$ Poi, $\mu_\infty = 0.032$ Poi, $\lambda = 10.03$ s, and $n = 0.344$. Different experimental data will give rise to different parameter values.

2.2 Initial and boundary conditions

In order to be well-posed, i.e., to have a unique solution that depends continuously on the data, equations (1) and (3) must be endowed with initial and boundary conditions. The initial condition is given by $\mathbf{u} = \mathbf{u}_0$, for $t = 0$, in Ω. Due to the lack of in vivo data usually $\mathbf{u}_0 = \mathbf{0}$. This means that the simulation starts with a zero solution and it is necessary to compute the solution for several time instants in order to have clinically relevant solutions.

Regarding the boundary conditions, two types of boundaries should be considered (see Fig. 1): the physical artery wall, and the artificial boundaries resulting from the truncation of the domain. Indeed, due to the geometrical complexity of the cardiovascular system, the computational cost of 3D simulations, and the fact that 3D detailed information is usually needed only in specific regions of interest, the portion of the artery at study should be truncated.

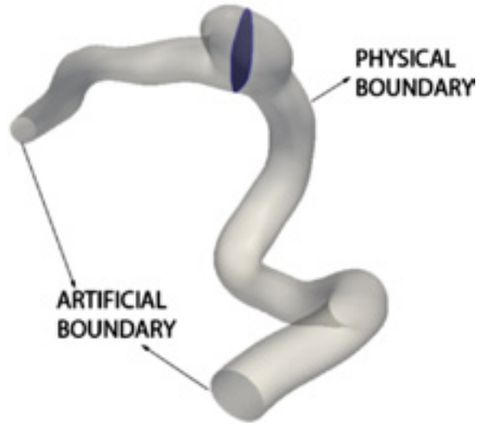


Figure 1.—The 3D computational domain of interest (blood vessel with an aneurysm), showing the physical boundary formed by the artery wall and the inflow and outflow artificial sections due to the truncation of the domain.

On the physical boundary, that we denote by Γ_w, boundary conditions are prescribed using physical arguments. If the movements of the vessel wall due to the blood flow load are not considered, i.e., if the artery wall is assumed to be rigid, then at that boundary the velocity is zero $\mathbf{u} = \mathbf{0}$, describing the total adherence of the fluid to the wall (no-slip condition). This simplifying hypothesis is assumed very often [2,9,12,13,14,15].

2.3 Compliance of the artery wall: fluid-structure interaction (FSI)

If the compliance of the wall is taken into account, the velocity of the fluid on the wall should be the same as the velocity of the moving wall: $\mathbf{u} = \mathbf{g}$, where $\mathbf{g}$ is the wall velocity given by a mathematical model that describes its motion [10,4]. The vascular wall is a very complex soft tissue, composed of several different layers, and it is very difficult to devise appropriate and accurate models describing their dynamical behavior. This is still a subject of active research and, for that reason, the simplest 3D linear hyperelastic model is often applied (see [10,4,11] and references therein):

$$\rho_w \frac{\partial^2 \boldsymbol{\eta}}{\partial t^2} - \operatorname{div}_0(\mathbf{P}) = \mathbf{0}, \qquad \text{on } \Omega_s^0, \tag{5}$$

where Ω_s^0 is the computational domain of the structure artery wall in the reference configuration[1], $\boldsymbol{\eta}$ is the displacement vector with respect to the reference configuration Γ_w^0, ρ_w is the wall density, div_0 stands for the divergence operator with respect to the Lagrangian coordinates and $\mathbf{P} = \mathbf{P}(\boldsymbol{\eta}) = \mathbf{FS}$ is the first Piola-Kirchhoff tensor (see [10,4]), with $\mathbf{S} = \mathbf{S}(\boldsymbol{\eta})$ the second Piola-Kirchhoff tensor and $\mathbf{F} = \mathbf{F}(\boldsymbol{\eta}) = \mathbf{I} + \nabla_0\boldsymbol{\eta}$ the deformation gradient tensor.

To have a description of the bood-vessel interaction problem, the fluid equations (1) and (3) are coupled with the structural equations (5). That is achieved by imposing the following matching conditions on Γ_w^t,[2] for all $t \in I$:

$$\begin{cases} \mathbf{u} = \dot{\boldsymbol{\eta}}, \\ -(\boldsymbol{\sigma}(\mathbf{u}, p) + p_{ext}\mathbf{I})\cdot\mathbf{n} = \boldsymbol{\Phi}\cdot\mathbf{n} \end{cases} \tag{6}$$

where p_{ext} is a given external pressure which, without loss of generality, is considered to be zero, $\boldsymbol{\Phi}$ is the stress exerted by the structure on the fluid and $\mathbf{n}$ is the outward unit vector to Γ_w^t. In (6) the first equality is the no-slip condition that guarantees the total adherence of the fluid to the structure ($\dot{\boldsymbol{\eta}}$ is the wall movement velocity), while the second equality establishes the continuity of the normal stresses.

{1} As it is customary in solid mechanics, the structure equations are written in the reference configuration (Lagrangian frame), while the fluid equations are set up in the current configuration (Eulerian frame), see for instance [10].

{2} Notice that now the fluid domain changes in time, due to the wall motion: $\Omega = \Omega^t$, and $\Gamma_w = \Gamma_w^t$.

It is necessary to provide appropriate initial conditions to the structure, compatible with the FSI problem. The dependence of the fluid domain on the structure equations solution makes it very difficult to guarantee the well-posedness of the FSI problem, which is still an open problem [10,4].

2.4 Artificial boundaries: the geometrical multiscale approach

The prescription of boundary conditions on the artificial sections constitutes a great challenge, since they cannot be deduced from physical arguments, and in vivo data on the flow rate or pressure are very difficult to obtain. The artificial sections can be divided into two types, the inflow sections, closer to the heart and also called upstream sections (usually the computational domain only has one inflow section), and the outflow sections, closer to the systemic circulation and also called downstream sections (it is common to have more than one in the computational domain). Very often standard boundary conditions, such as Neumann homogeneous conditions at outflow sections, are imposed: $\sigma(\mathbf{u}, P) \cdot \mathbf{n} = \mathbf{0}$. However, these conditions do not account for the remaining parts of the cardiovascular system. The computational solution is highly dependent on the choice of the boundary conditions on the artificial sections, so that such solutions can become not reliable and their use in clinical applications are compromised. Indeed, the cardiovascular system is closed, and the local hemodynamics greatly depends on the systemic circulation (see e. g. [12]). For that reason, the global behavior of blood flow should be taken into account in local 3D simulations. In order to do that, models of different geometrical scales, with different levels of accuracy and computational cost are considered, according to the level of detail required. This approach leads to the so called *Geometrical Multiscale Modeling* of the cardiovascular system [6,10].

In regions where detailed information is necessary, 3D models are applied. These are the most computationally costly and can only be applied to small regions of the vasculature.

If the purpose is to simulate large arterial trees, 1D simplified models should be used [3,6,10,4,11,1]. These models are obtained from the 3D FSI problem by averaging and assuming cylindrical geometry of the vessels. The 1D models are less accurate, providing only average quantities such as flow rate and mean pressure, yet they

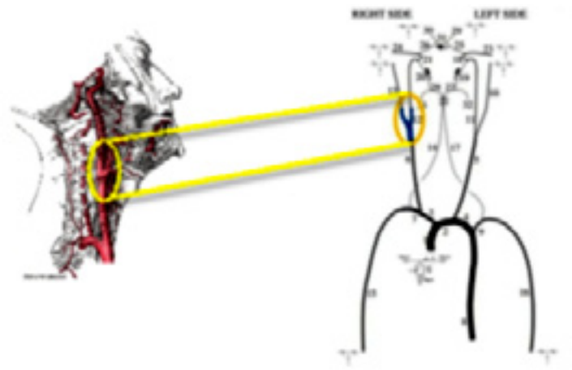


Figure 2.—Schematic of the coupling of a 3D model of the carotid bifurcation with a 1D arterial network, and with 0D models at its extremities to take into account the capillaries resistance.

are much less expensive from the computational view point and describe very well the wave propagation nature of blood flow in arteries [10,4,11]. Indeed, the 1D model for blood flowing in arteries is given by an hyperbolic system of equations, that in physiological situations is under a sub-critical flow regime.

Simpler lumped parameter models can be derived from the 1D models by further averaging in space, resulting in a system of ordinary differential equations (ODEs) [3,6,12]. Since lumped parameter models do not depend on space, they are also named 0D models. They describe the variation in time of the averaged pressure and flow rate in a specific region of the circulatory system, such as the venous bed, the pulmonary circulation, or the heart. There is a strong analogy between lumped parameter models and electric networks. Indeed the flow rate can be seen as the electric current and the mean pressure as the voltage. Furthermore, the lumped parameters are precisely the resistance, related with the blood viscosity, the inductance, related with the blood inertia, and the capacitance, related with the wall compliance.

Coupling together models of the three different levels gives rise to the geometrical multiscale modeling of the cardiovascular system. The couplings are achieved by imposing the continuity of mean pressure and flow rate [6,10,4]. In this manner, reduced 1D or 0D models can be coupled to the artificial sections of the 3D model in order to provide proper boundary conditions, accounting for the remaining parts of the cardiovascular system [10,4,11,12]. This procedure allows to perform reliable computational simulations of local blood flow with clinical impact. Fig. 2 illustrates the coupling of all the three hierarchical models. The region of interest is the carotid bifurcation, which often undergoes atherosclerotic plaques.

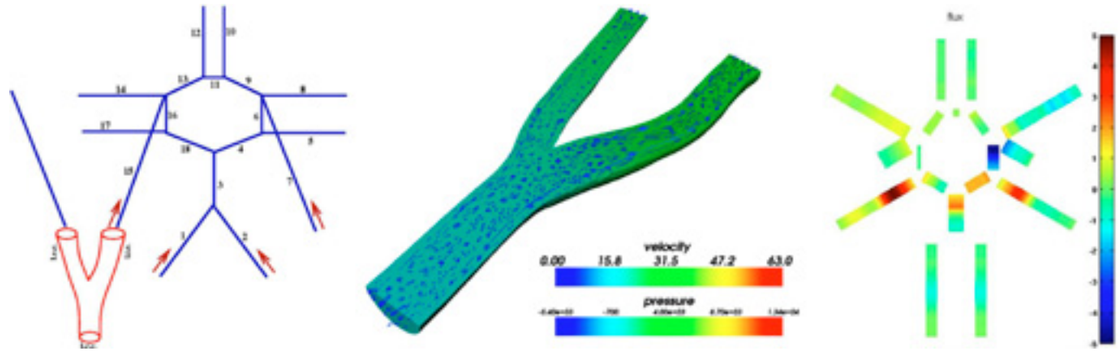


Figure 3.—Realistic 3D carotid bifurcation coupled to the circle of Willis. Left: scheme of the coupling between the 3D carotid bifurcation and 1D reduced models. The internal carotid downstream section is coupled to a 1D network of the circle of Willis, while the external carotid downstream section is coupled to a single 1D tube. Center: pressure [dyn/cm^2] and velocity [cm/s] solution in the 3D carotid bifurcation. Right: the values of the flow rate [cm^3/s] in the circle of Willis [10,11].

In Fig. 3, the numerical solution of the coupling of a 3D FSI model of the carotid bifurcation with a 1D model of the circle of Willis is represented (taken from [10,11]). The 1D description of the circle of Willis properly accounts for the absortion and propagation of pressure waves, so that the 3D simulation on the carotid bifurcation is reliable.

Although they are less detailed, the reduced 1D and 0D models provide very useful simulations at very low computational cost, and are often applied as stand alone models, not necessarily coupled with 3D models. For instance in [1], 1D models are used to study anatomic variations of the circle of Willis, and in [3], 1D models and a 0D model for the heart are applied to study the circulation effects of amputating one leg.

In [13,14] the sensitivity of the numerical fluid solution in cerebral aneurysms to changes on the outflow conditions is studied, including the use of reduced models. In [12] the geometrical multiscale approach is used to obtain reliable results with clinical applications in heart paediatric surgery.

3. Numerical simulation of hemodynamics in cerebral aneurysms

Cerebral aneurysms are pathological dilations of the cerebral vascular wall, which induce modifications in the mechanical properties of the artery wall, including its weakning that may lead to rupture. The rupture of cerebral aneurysms causes sudden death in 50% of the patients, and provokes permanent disabilities in a great number of the remaining cases. It is a silent pathology, without any symptomatology until rupture, except for a very small number of cases. It is therefore a devastating disease that is believed to affect approximately 5% of the population. The causes for initiation, growth and rupture of cerebral aneurysms are still unknown, although it is accepted that there is a correlation between aneurysm progression or genetic and hemodynamic factors [2,5,15]. In what concerns the hemodynamic factors, the numerical simulations play a very important and unique role in the comprehension of this disease, allowing in particular to obtain patient-specific reliable results and their visualization in a non-invasive way [9,13,14,15]. Through computational simulations it is also possible to easily compute hemodynamic indicators, such as the wall shear stress (WSS), that are very difficult to measure in vivo or in vitro. Precisely, the WSS and other related quantities, such as the WSS gradient, are known to be determinant in the initiation, growth and rupture of aneurysms [2,15]. Thus, numerical simulations have

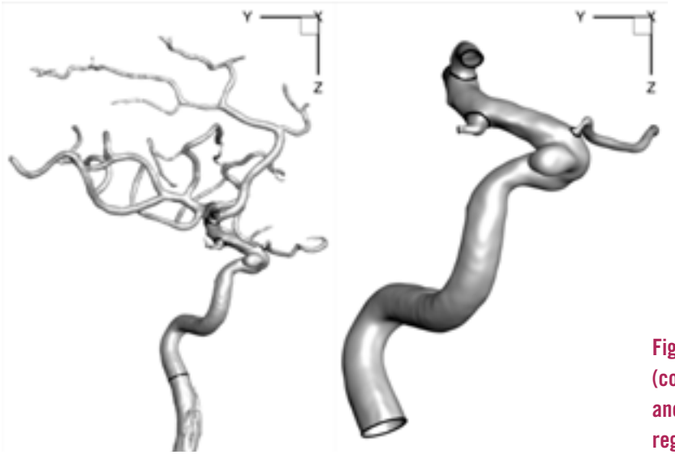


Figure 4.—Reconstructed geometry from CTA (computed tomography angiogram) of a cerebral aneurysm (left), and definition of the computational region of interest (right) [13].

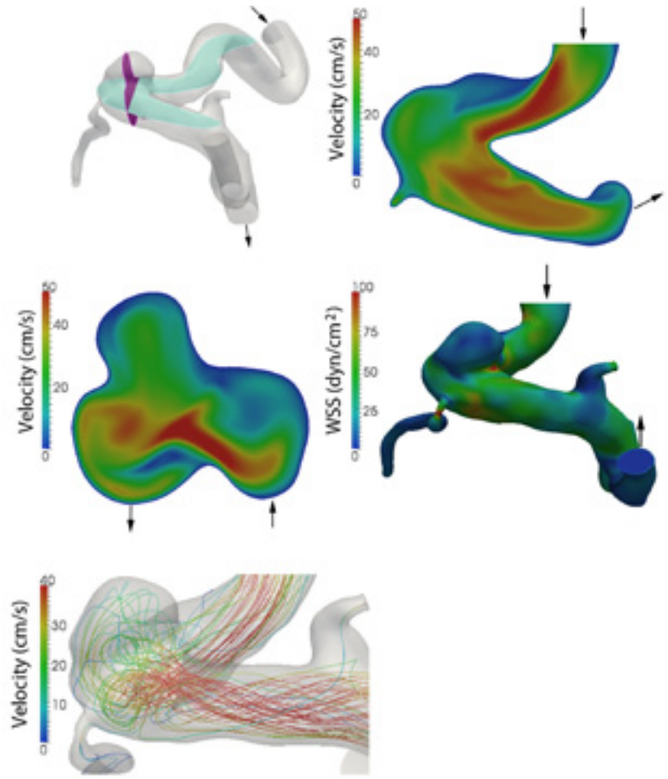


Figure 5.—Illustration of the chosen cross-sections (top-left), velocity magnitude (cm/s) in both cross-sections (top-right and middle-left), WSS magnitude (dyn/cm^2) (middle-right), and velocity pathlines coloured by the velocity magnitude (bottom). Note that the velocity cross-section that includes the aneurysm is such that the upstream flow is on the right and downstream flow on the left [13].

nowadays an increasing impact in the clinical practice of patients with cerebral aneurysms. They lead to a better understanding of the disease and try to predict its natural progression, namely its rupture and consequent potential letal bleeding, contributing also to its treatment. Therefore, computer simulations constitute a tool to support medical and clinical decision, both in the analysis and diagnosis of anatomic and physiological results, as well as in the prediction of surgical outcomes and post surgical complications. The results may also contribute to improve treatment and surgical techniques, such as endovascular surgery.

As already mentioned, the hemodynamics highly depends on the morphology of the blood vessels, that is, on its geometry, being thus specific of each patient. In particular, for the study of cerebral aneurysms, reliable simulations depend not only on the choice of appropriate mathematical models and accurate numerical algorithms, but also on their application in patient-specific computational geometries, obtained from medical acquisition, as for instance computational tomography (CT). In order to have patient-specific computational domains, it is necessary to reconstruct the medical images, which consists essentially in three steps [9,13,14]:

(1) Segmentation: identification of the region of interest in the grey scale medical image. In this case it is important to distinguish between lumen and artery wall;

(2) Surface reconstruction: mathematical definition of the 3D surface, usually performed by means of the marching tetrahedra algorithm (see [9,13,14] and references therein);

(3) Smoothing: the 3D reconstructed surface has non physiological irregularities related with the medical image quality, which has noise due to its acquision, that should be eliminated through a smoothing process usually carried out by a bi-Laplacian algorithm [9,13,14].

Once one has the reconstructed medical image (see Fig. 4, left) [13], it is necessary to define the computational region of interest, where to perform the 3D numerical simulations (see Fig. 4, right) [9,13,14]. Afterwards, it is necessary to define a computational 3D mesh, by decomposition into simpler geometrical figures, usually tetrahedra, in which the numerical algorithms are applied. In order to attain accurate solutions in patient-specific simulations, it is essential to have a large number of very small elements, usually in the order of millions.

From the simulation results of the velocity and pressure fields, it is possible to compute the hemodynamic indicators associated to aneurysm risk of growth and rupture, such as the WSS and its variations. In Fig. 5 are depicted the numerical results of the simulation performed in the geometry of Fig. 4, taken from [13]. It is possible to observe that the region of higher WSS is the neck of the aneurysm, precisely where the flux of the main vessel occurs, as it can be seen by the pathlines, while the lower values of WSS are found within the aneurysm sac.

References

[1] J. Alastruey, K.H. Parker, J. Peiró, S.M. Burd, S.J. Sherwin. Modelling the circle of Willis to assess the effects of anatomical variations and occlusions on cerebral flows. Journal of Biomechanics, 40(8):1794–1805, 2007.

[2] J.R. Cebral, F. Mut, D. Sforza, C.M. Putman. Association of hemodynamic characteristics and cerebral aneurysm rupture. AJNR American Journal of Neuroradiology, 2011, in press.

[3] L. Formaggia, D. Lamponi, M. Tuveri, and A. Veneziani. Numerical modeling of 1D arterial networks coupled with a lumped parameters description of the heart. Computer

Methods in Biomechanics & Biomedical Engineering, 9(5):273–288, 2006.

[4] L. Formaggia, A. Moura and F. Nobile. On the stability of the coupling of 3D and 1D fuid-structure interaction models for blood flow simulations. ESAIM: Mathematical Modelling and Numerical Analysis (M2AN), 41(4):743–769, 2007.

[5] L. Formaggia, A. Quarteroni, A. Veneziani, Eds. Cardiovascular Mathematics: modeling and simulation of the circulatory system. Modeling, Simulation and Applications Series, Springer-Verlag, Italia, Milano, 2009.

[6] L. Formaggia, A. Veneziani. Reduced and multiscale models for the human cardiovascular system. Lecture Notes VKI Lecture Series, 7, 2003.

[7] K. Perktold and G. Rappitsch. Mathematical modeling of arterial blood flow and correlation to artherosclerosis. Technology and Health Care, 3:139–151, 1995.

[8] G.P. Galdi, R. Rannacher, A.M. Roberston, S. Turek. Hemodynamical Flows: Modeling, Analysis and Simulation. Birkhäuser, 2008.

[9] A. Gambaruto, J. Janela, A. Moura, A. Sequeira. Sensitivity of hemodynamics in patient-specific cerebral aneurysms to vascular geometry and blood rheology. Mathematical Biosciences and Engineering, 8(2):411–425, 2011.

[10] A. Moura. The Geometrical Multiscale Modeling of the Cardiovascular System: Coupling 3D and 1D Fluid-Structure Interaction Models. PhD Thesis, Politecnico di Milano, 2007.

[11] A. Moura. Coupling multiscale fluid-structure interaction models for blood flow simulations. In: Vascular Wall and Endothelium, Lisboa, 2007, Actas de Bioquímica 9 (J. Martins e Silva and Carlota Saldanha, eds.), pp. 137–141, 2008.

[12] G. Pennati, G. Dubini, F. Migliavacca, C. Corsini, L. Formaggia, A. Quarteroni, and A. Veneziani. Multiscale Modelling with Application to Paediatric Cardiac Surgery. MOX Report 43/2010, 2010.

[13] S. Ramalho, A. Moura, A. Gambaruto, and A. Sequeira. Sensitivity to outflow boundary conditions and level of geometry description for a cerebral aneurysm. International Journal for Numerical Methods in Biomedical Engineering, 2011, in press.

[14] S. Ramalho, A. Moura, A.M. Gambaruto, and A. Sequeira. Influence of Blood Rheology and Outflow Boundary Conditions in Numerical Simulations of Cerebral Aneurysms. In: Mathematical Models and Methods in Biomedicine. American Institute of Mathematical Sciences (AIMS), pp. 143–168, 2011.

[15] D.M. Sforza, C.M. Putman, J.R. Cebral. Hemodynamics of cerebral aneurysms. Annual Review on Fluid Mechanics, 41:91–107, 2009.

Ficha Técnica

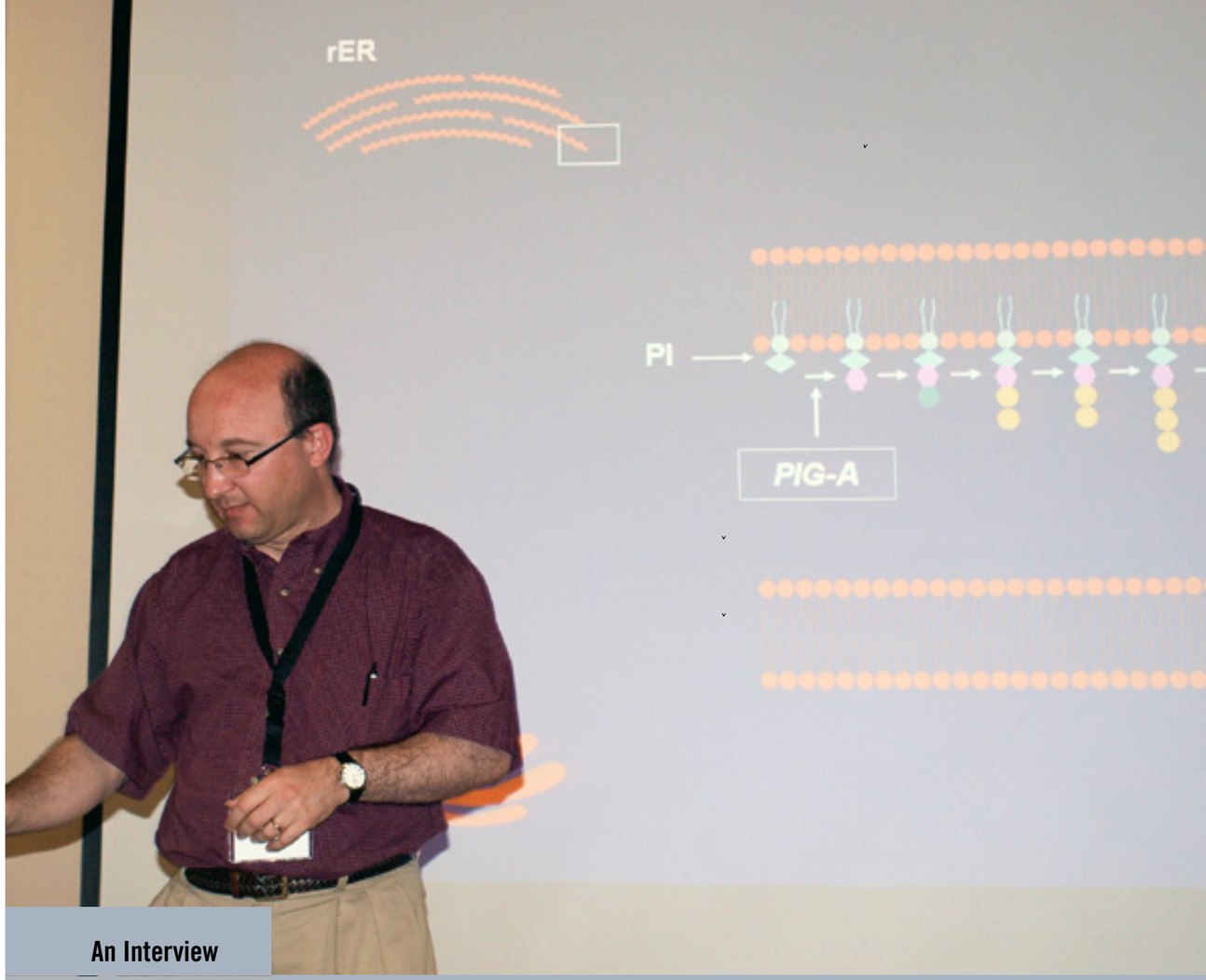
Editors
Adérito Araújo (alma@mat.uc.pt)
António Fernandes (amfern@math.ist.utl.pt)
Sílvio Gama (smgama@fc.up.pt)

Address
IIIUL-Universidade de Lisboa
Av. Prof. Gama Pinto, 2
1649-003 Lisboa

The CIM Bulletin is published twice a year
Material intended for publication should be sent to one of the editors.

The Bulletin is available at www.cim.pt

The CIM acknowledges the financial support of FCT–Fundação para a Ciência e a Tecnologia



An Interview

with David Dingli

by **Francisco Santos** and **Fabio Chalub** [Universidade Nova de Lisboa]

David Dingli was in Evora for the Summer School "Dynamic Models in Life Sciences", where he presented a set of lectures called "Hematopoietic Stem Cells and Hematopoiesis". In these lectures, he introduced the audience to the state-of-the-art in the dynamics of stem cells, and their relation to blood disorders. After the conference he gave this interview to Francisco Santos and Fabio Chalub, summer school co-organizers.

Please summarize your academic/professional trajectory (just a short bio).

I am a hematologist and treat patients with various types of blood malignancies. My research initially focused on the generation of trackable, replication competent viruses to treat cancer. It became clear very early on that the interactions between oncolytic viruses, the tumor cell population and the immune system are quite complex with various outcomes. Understanding these dynamics required mathematics and as a result, while in graduate school, I enrolled in various classes to learn more mathematics. I got hooked and decided to spend more time in mathematical biology after finishing my training in hematology. I was fortunate enough to spend 2 years at the Program for Evolutionary Dynamics [PED] at Harvard University working with Professor Martin Nowak and his group. There, I established strong collaborations that continue to this day. As a result, now I devote a considerable amount of my time on mathematical modeling of various hematologic disorders. However, my laboratory still continues to work on the use of viruses to treat cancer and a considerable part of my modeling is still centered on tumor virotherapy.

How do you assess the importance of mathematics to your research?

It is not possible to understand dynamic systems without mathematics. Whenever we are dealing with a process that changes in time, we have to use mathematics for a meaningful understanding of the process. For example, with tumor virotherapy we need to know how the virus spreads in the tumor, the kinetics of the process, the rate of virus generation, cell killing etc and then try to design viruses with "optimal properties" for cancer therapy. Mathematical models are a great asset also by providing an in silico testing ground for innumerable therapeutic scenarios that can be explored rapidly and cheaply. The in vivo experiments that are time consuming and expensive can be used to test the most interesting scenarios predicted by such modeling.

How do you assess the importance of mathematics to medical practice?

The physician of the future must have a good basis in mathematics. Advances in technology mean that nowadays, acquisition of data is not the limiting factor. One can see what has happened with the 'omics' revolution. However, we're still far away from being able to understand the data being generated. Such an understanding will require new theory that can only come from mathematics, just as physics moved forward when calculus was discovered (or invented). Similarly, one can obtain multiparameter data in real time on patients in the intensive care unit. One can imagine scenarios where modeling of such data will enable understanding of the trajectory of the illness and plan therapeutic interventions of the right magnitude and at the right time to move the patient away from the ultimate stable equilibrium (death) and back to a state of health.

In your research, you work with physicists, mathematicians, etc. Was it easy to start this collaboration? Did you have to start by building a "common language"?

One of the most enriching aspects of my research has been this interaction with physicists, mathematicians and computer scientists. I was fortunate that the PED is a melting pot for scientists from different disciplines to meet and discuss science. We all come from different backgrounds and training of a physicist is quite different from that of a physician. However, it was not difficult to find common ground and start collaborating. Such interactions are mutually rewarding in the sense that if I had to explain the detailed molecular biology of a process to my colleagues, I had to understand it well myself and then strip it down to the bare bones. This is an essential exercise that helped me identify gaps in my knowledge of the subject but also enabled me to ask relevant questions for the field that ultimately translated into many joint publications. This exercise serves to establish the "common language" that you mention. However, the main issue is one of "synchronization of thinking"—a physicist looks at a problem differently from a physician. For them, cells and balls are very similar, and tumor growth is similar to nucleation of a crystal etc.

You have long-term scientific collaborations with researchers in Portugal. How did it start?

I met Professor Jorge M. Pacheco at the PED in the summer of 2005. We started almost simultaneously there and not only did we come from different backgrounds, but we also went to PED for different purposes—Jorge was working on evolutionary game theory while I wanted to study tumor virotherapy. One afternoon, we went for a walk along the Charles River and started talking about blood disorders, stem cell and bone marrow transplantation. A few incisive questions from Jorge on that fateful day established that collaboration that has been going ever since and resulted in various trips to Portugal and Jorge also visited me at Mayo Clinic. Since those initial days, the collaboration has expanded to include Dr Francisco Santos and Professor Fabio Chalub where we have applied principles from EGT to cancer.

You said once that the reward from the clinical practice is essentially immediate, while the one from research takes a long time. How do you compare the pleasure of these two facets of your work?

I enjoy meeting patients, trying to understand their illness, how it affects them and then personalize therapy for them. Often it is possible to help patients quickly with pain control, improved quality of life and then long term therapy, usually for their hematopoietic tumor. I come to know not only the patient but their family, their hobbies, interests, travel etc. In this way, each patient is unique and there is no redundancy. The frustrating part is when I reach the limit of therapy that is available…one always wants safer, more effective therapies and ultimately the cure. Research provides a different form of gratification—the intellectual musings, hypothesis generation, the critical experiments, writing code (and debugging), running simulations, etc. all take time. What I enjoy the most is the creativity, the imagination, and in some way, even the "shortcuts" that we sometimes take in our modeling efforts, all with the aim of getting some results, the first glimpse of the output.

Do you believe Summer Schools like the one in Évora are a good starting point for students wishing to work in math-biology?

Years ago, I attended a two week summer school on "mathematics and computers in medicine" organized by the late Professor Lee Segel at the Santa Fe Institute. During those two weeks I learned not only mathematics but more importantly how to apply mathematics to various medical problems—from virology to cancer, the immune response etc. The summer school in Evora was fantastic—the depth and breadth of topics explored was immense and the students had an opportunity to see how the immense power of mathematics can be used to address problems in ecology, cancer, evolution, imaging, population dynamics etc. Once students understand how to apply their skills to biological problems, they are only limited by their imagination. Such schools are essential for training of tomorrow's scientists.

Finally, do you have any practical advice for students willing to start an interdisciplinary work?

The best way to learn about interdisciplinary research is to find a biological problem of interest and read about it as much as you can. One cannot model what one does not understand—once the student is well aware of what is known about the problem, then it often becomes clear what questions can be posed in a mathematical framework for the problem at hand. Finding a mentor with a track record of research and publications in the field and who has a string of prior graduate students that have been with her/him, will increase the probability of success in the field.

Well, David, thanks a lot for this interview.
We are looking forward to seeing you again in Portugal soon!

Summer School on Algebraic and Enumerative Combinatorics

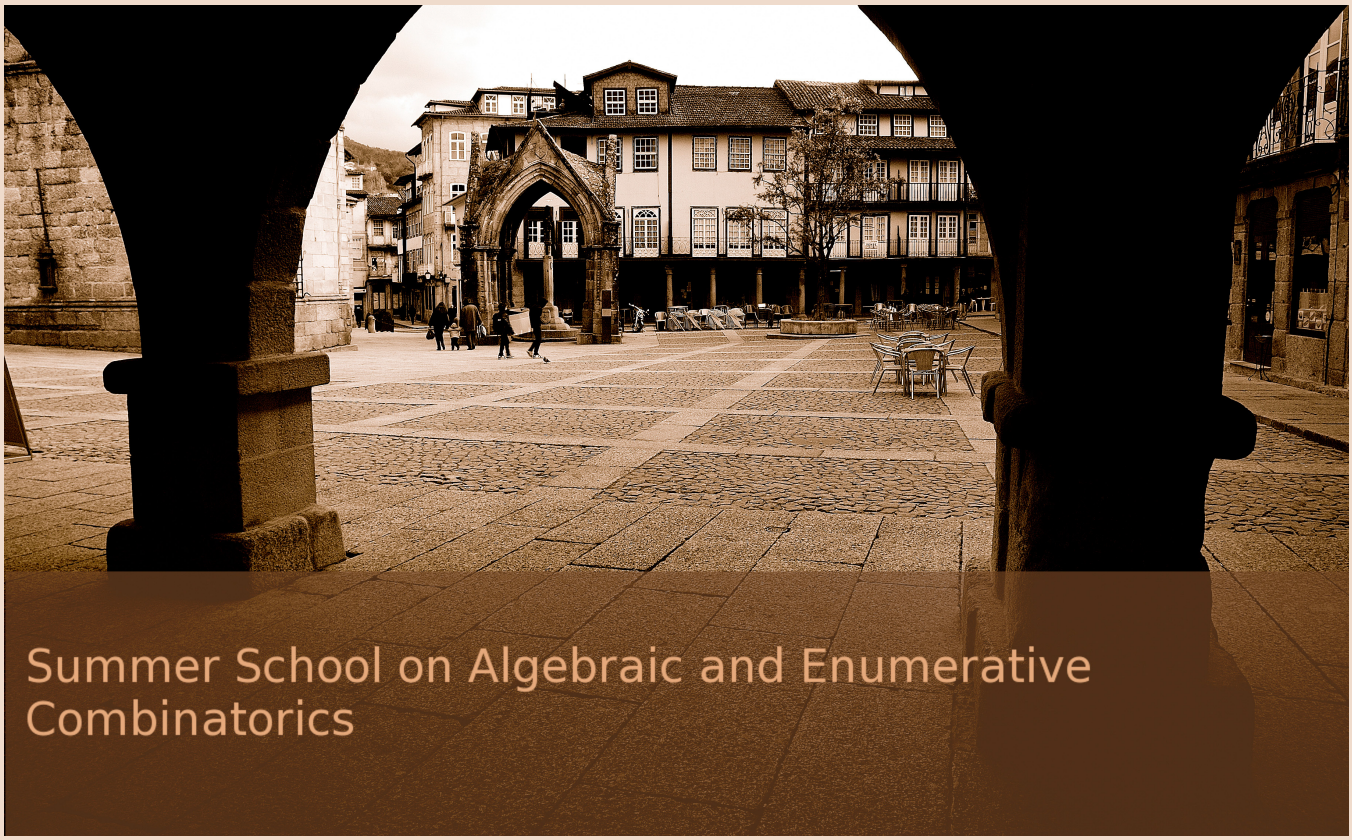
S. Miguel de Seide, Portugal

Dates: July 2-13, 2012

Courses: Casa de Camilo, S. Miguel de Seide, Famalicão, Portugal
Lodging: Guimarães, Portugal

Speakers

- **Francesco Brenti** Combinatorics of Coxeter Groups
- **Christian Krattenthaler** Map Enumeration
- **Marc Noy** Asymptotic enumeration of topological graphs
- **Vic Reiner** Reflection group counting and q-counting



Based on a photography by Paulo Gandra

Organized by

- Carlos André, Universidade de Lisboa
- Olga Azenhas, Universidade de Coimbra
- Alessandro Conflitti, Universidade de Coimbra
- António Guedes de Oliveira, Universidade do Porto
- Ricardo Mamede, Universidade de Coimbra
- Maria Leonor Moreira, Universidade do Porto

for

Sponsors:

Câmara Municipal de Vila Nova de Famalicão

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA EDUCAÇÃO E CIÊNCIA

CMUC Centro de Matemática Universidade de Coimbra

CMUP Centro de Matemática Universidade do Porto

Centro de Estruturas Lineares e Combinatórias

Please visit http://www2.fc.up.pt/pessoas/agoliv/SC/default.htm